འབྲུག་རྒྱལ་འཛིན་གཙུག་ལག་སློབ་དཔེ།།

ཤེས་རབ་རྩེ་མཐོ་རིམ་སློབ་གྲྭ།

ROYAL UNIVERSITY OF BHUTAN

SHERUBTSE COLLEGE

Definitive Programme Document for

Bachelor of Science in Data Science

Royal University of Bhutan

April, 2023

Table of Contents

1 Programme Specification

1.1 Basic Information on the Programme

Name of the home base college:	Sherubtse College
Title of the award or awards:	Bachelor of Science in Data Science
The duration and mode of study:	4 years, full time
Award granting Body:	The Royal University of Bhutan
Date of Initial Approval:	28-29 April 2021, 17th USRC (validated)

1.2 Aims and Learning Outcomes of the Programme

1.2.1 Aims of the Programme

The BSc in Data Science programme aims to equip students with sound theoretical and practical knowledge to retrieve, clean, process, analyse and extract knowledge from structured and unstructured real-time data. Moreover, the interdisciplinary nature of the programme will prepare students to handle the data-related tasks from diverse sources to help an individual or an organisation to make informed decisions. The programme also aims to produce students who are not only skilled in the practicalities of dealing with complex data in the real-world but also proficient to understand the professional conducts concerning data ethics. Graduates of the programme will be skillful and knowledgeable in making sense of massive data generated from both offline and online sources.

To connect classroom learning with hands-on experience, students will be attached to some industries for a duration of two months immediately after Year 4, Semester I. After returning from industrial attachment students will continue to work on the capstone project. Through the capstone project, the programme aims to produce graduates with greater confidence in applying knowledge and skills acquired in the first three years in carrying out projects on real-world problems. Also, the programme will equip students with portable skills to present the results of a (research) project to relevant stakeholders. Furthermore, the four-year degree will prepare students to pursue higher studies in related fields.

Unlike Statistics degrees which focus on using numerical data to address business issues, Data Science degrees teach students how to find business insights rooted in statistical theory and technical skills. Data Science careers typically require strong technical skills so subject matter is often related to complex coding and systematic languages. Skills required in programming, machine learning and algorithms is relatively more for Data Science graduates than graduates in statistics.

In order to provide wholesome education and personal development to the students some breadth modules are included in the programme. Through these modules, the students will gain oral and written communication skills, awareness about local and global cultural values, leadership and interpersonal skills.

1.2.2 Learning Outcomes of the Programme

Upon completion of the programme, graduates will be able to:

1. Perform data processing such as data retrieval, cleaning, processing, analysing and storage of data.
2. Develop broad understanding of knowledge extraction from data.
3. Apply data science technologies and tools to process and analyse structured and unstructured data.
4. Identify appropriate methods to determine the solutions to real-world problems.
5. Use a suitable programming language to store, manipulate and visualise data.
6. Analyse quantitative and qualitative aspects of data analysis for decision-making and situational analysis.
7. Design and implement data analysis using statistical programming tools.
8. Draw statistical inferences based on scalability and performance.
9. Communicate findings of a project through documentation, discussion and/or presentation.
10. Apply principles of data governance and ethics in handling real-time data.
11. Apply the knowledge and skills for carrying out projects on real-world problems.
12. Take-up entrepreneurship as a career option in the field of data science
13. Exhibit logical thinking, structured reasoning and synthesis skills.
14. Demonstrate ability to work as a team, listen and respond to others, and use negotiation and conflict-resolution skills.

1.3 Career Related Opportunities

Graduates of the programme can find employment opportunities as data analysts in various sectors of the government, financial institutions, stock trading, private business concerns, consultancy services etc. both within and outside the country. The skills and knowledge gained would not only be helpful in getting employed in government and corporate agencies, but also be very useful for self-employment as successful entrepreneurs.

1.4 Programme Structure

Table 1: Programme Structure

Year	Sem	Modules				
Year 1	**I**	PLT102 Programming in Python	DST101 Descriptive Statistics	CAL111 Calculus	CAE102 Computer Organisation and Programming	DZG101 རྩོད་ཁ་བད་དོན་སྒོད་ལེན།
	II	AMT114 Discrete Structures	DST106 Probability and Distributions	ADS104 Data Structures	ALG108 Matrix Analysis and Vector Spaces	ACS101 Academic Skills
Year 2	**I**	PLT203 R Programming for Data Analysis	CAE201 Database Systems	IST303 Statistical Inference	IST302 Regression Analysis	ECO101 Fundamentals of Economics
	II	AMT210 Numerical Methods	ADS205 Algorithm Analysis	IST306 Multivariate Analysis	DSC201 Machine Learning	GSE101 Analytical Skills
Year 3	**I**	RSM301 Research Methods	APC317 Data Warehousing and Data Mining	IST201 Sampling Theory	DSC302 Unsupervised Learning	PMT301 Project Design and Management
	II	AMT311 Introduction to Operations Research	DSC303 Ethics and Security	AMT312 Stochastic Process	APC318 Deep Learning	EDP101 Entrepreneurship
Year 4	**I**	APC419 Computer Vision	DSC404 Big Data Technologies and Applications	APC420 Natural Language Processing	DSC405 Geospatial Data Analysis	DSC406 Communication for Data Science
	II	RSM413 Data Science Internship	RSM414 Capstone Project			

The programme is of four years duration comprising eight semesters. Like all other programmes, the Year 1, Semester I and Semester II is a foundation year. The students will be presented with topics that will be built on in the subsequent semesters. Except for Year 4, semester II, the programme will have five modules in each semester.

In the initial semesters, the students will learn university-wide modules (e.g. Academic Skills, Dzongkha Communications) and foundational core modules in Mathematics, Statistics and Computer Science. As they progress to the second and third year, the students will get to learn some advanced concepts in the above three areas and their applications in data science. The Capstone Project in the fourth year (internship worth 12 credits and a project work worth 48 credits) will equip students with skills to work independently and collaboratively. The internship is intended to take place during the winter vacation after completion of Semester I of Year 4 for two and half months (between second week of December and end of February the next year). The project will be based on real-time data they have selected as the focus of the study.

At the end of four years, the students must complete 40 modules translating to 480 credits including general education modules and open electives. The programme structure is designed in a progressive manner with ten modules in each year.

1.4.1 Modules Categorized by the Subdivision of the Programme

The module codes have been devised as follows, depending on the nature of the module:
DSC: Data Science
DZG: Dzongkha
GSE: Global Skills Enhancement
CSD: Communication Skills Development
ECO: Economics
LAN: Language
ACS: Academic Skills
CAL: Calculus
APC: Applied Computer Science
ADS: Algorithms and Data Structures
CAE: Computer Architecture and Engineering
PLT: Programming Language Techniques
DST: Descriptive Statistics
IST: Inferential Statistics
ALG: Algebra
AMT: Applied Mathematics
RSM: Research Methods
PMT: Project Management
EDP: Entrepreneurship Development

1.4.2 Modules Borrowed from Other Programmes

To ensure the development of other skills and knowledge significant to university students, ten modules have been borrowed from other programmes.

Table 2: Modules borrowed from other programmes

Sl. No.	Module	Borrowed from
1.	ACS101 Academic Skills	University-wide module

2.	DZG101 Dzongkha Communication	University-wide module
3.	GSE101 Analytical Skills	University-wide module
4.	PLT102 Programming in Python	BSc in Statistics
5.	CAL111 Calculus	BSc in Statistics
6.	ALG108 Matrix Analysis and Vector Spaces	BSc in Statistics
7.	PLT203 R Programming for Data Analysis	BSc in Statistics
8.	CAE201 Database Systems	BSc in Statistics
9.	IST303 Statistical Inference	BSc in Statistics
10.	IST302 Regression Analysis	BSc in Statistics
11.	AMT210 Numerical Methods	BSc in Statistics
12.	IST306 Multivariate Analysis	BSc in Statistics
13.	RSM301 Research Methods	BSc in Environmental Science
14.	AMT311 Introduction to Operations Research	BSc in Statistics
15.	IST201 Sampling Theory	BSc in Statistics
16.	AMT312 Stochastic Process	BSc in Statistics
17.	DST101 Descriptive Statistics	BSc in Statistics
18.	ECO101 Fundamentals of Economics	BA in Population and Development Studies
19.	EDP101 Entrepreneurship	University-Wide Module
20.	PMT301 Project Design and Management	BA in Population and Development Studies

1.5 Learning and Teaching Approach

The programme aims to promote student-centered learning process. This will be achieved through lectures, laboratory sessions, presentations, and tutorial sessions. Learning and teaching will be predominantly carried out using lecture method and hands-on practice will be conducted during laboratory sessions. To test their theoretical understanding they will be made to solve pre-posted questions during weekly tutorial sessions. As the academic level of students' progresses, the learning and teaching approach is also changed with greater emphasis on independent learning and less contact hours for traditional teaching. Effective use of ICT will be encouraged and the tutors will be expected to make full use of the VLE platform for sharing the available free e-resources. To synergize supervised and independent learning, the students will be continuously monitored and will be expected to show progress on the works given to them to try on their own.

For most of the modules there will be either a tutorial class or a laboratory session scheduled every week. During the tutorial class, the students will be divided into smaller groups of not more than six and they will be asked to work on pre-assigned theorems and problem sets. The tutor will supervise and assist the students by giving them hands on problem-solving techniques. They will be encouraged to come up with problems they face in that module which will be subsequently addressed by the concerned tutor. If necessary, one member of the group will be made to present the results in the class. During laboratory sessions, students will work on programming tasks, hands-on practice with open-source software packages, pre-assigned project works etc. In these supervised lab sessions, students will be able to apply the concepts taught in the class with the help of statistical software and tools.

In addition, students are expected to explore and work independently towards achieving the learning outcomes of the programme. This will include peer collaboration, reading, assignments, and virtual learning through use of educational technology.

1.5.1 Placements/Work-based learning

The College Management will establish linkages with relevant industries/organisations and decide on the number of students to be attached to each industry/organisation and other logistics. Each student must undergo two and half months of internship training in relevant government agencies or private organisations where data is central to their business operation.

A faculty member of the programme will be nominated to coordinate the internship and liaise with organisations or agencies for monitoring. After returning from the internship, the students must continue to work on the capstone project. More details are available in the module descriptor of RSM413 Data Science Internship.

1.6 Assessment Approach

The assessment will be consistent with the assessment rules and regulation as per the Wheel of Academic Law (WAL) of RUB ensuring the principles of Validity, Reliability, Transparency and Fairness in the proposed programme. The assessment is divided into two components - Continuous Assessment (CA) and Semester-end Examination (SE). It will assess students' learning, skill acquisition, and academic achievement of the module objectives prescribed in the definitive programme document (DPD) through continuous and summative assessments. The forms of assessment will include but not limited to:

a. Continuous Assessment- assessments such as written assignment, laboratory assignments, oral presentation, tutorial session, VLE quiz, case studies, term project, and capstone project as specified in the module descriptor will be carried out to assess the students' learning.
b. Summative Assessment- assessments such as class tests, mid-semester examinations and Semester-end examinations as specified in the module descriptor will be carried out to assess the students' learning.

The weightings for the components of assessment may be different for different modules based on the nature of the subject matter and the learning outcomes of the modules. As they progress to the higher semesters, continuous assessments will be geared towards

assessing the students' independent work rather than emphasizing on conventional summative assessment.

1.7 Regulations

1.7.1 Entry Requirements

The admission for the programme will be routed through the University's common online admission system. The candidates will be selected through merit ranking based on the eligibility criteria given below (Table 3). The weighting assigned to a subject is based on the relevance of the subject to the programme. The selection criteria are subject to revision as per the admission policy guideline of the RUB and Sherubtse College.

Applicants who have not studied Dzongkha in class XII should have passed it in class X. A proficiency test in Dzongkha will determine the eligibility of Bhutanese applicants who have not studied Dzongkha in either class X or XII. The test shall be administered by the Registry, RUB. Marks secured in the proficiency test/class X will be scaled down to 40% and applied for merit ranking.

Table 3: Entrance requirements for the Programme

Eligibility Criteria	Ability Rating
Class XII pass Science students with a minimum of 55% in Mathematics.	Mathematics - 5 4 other subjects – 1

1.7.2 Assessment and Progression Requirements

To pass a module a student must obtain a minimum of 50% overall including both the Continuous Assessment (CA) and Semester-end examination (SE). However, a student must obtain a minimum of 40% each in CA and SE.

Re-assessment

A student is eligible for re-assessment if the number of failed modules is less than 30% of the total number of modules prescribed for the semester. Upon passing the failed module(s) in re-assessment, a student will not be awarded more than 50%, this being the minimum pass mark.

Repeat Module(s)

A student may repeat a failed module any number of times within the normal registration period for completing an award where he or she:

- has failed in the re-assessment of a module(s). In such an event, the student shall meet all assessment requirements of those modules, both CA and SE. For students under this category, attendance in lectures is not mandatory.
- has failed more than 30% of the total number of modules prescribed for that semester. In such an event the student shall meet all teaching, learning and assessment requirements of the failed modules. For students under this category, attendance in lectures is mandatory.

- For any particular semester, a student cannot register for more than 2 repeat modules in addition to the modules prescribed for the semester

Role of Programme Board of Examiners

The Board of Examiners shall, in the light of the University's assessment regulations spelt out in the Wheel of Academic Law and the programme specific regulations, determine, for each module, the mark to be assigned to each student's performance. The Programme Board of Examiners shall determine whether each student shall:

- be eligible for an award
- be allowed to continue on the programme, possibly with provision for re-assessment in certain modules and/or for the repeat of certain modules, or
- be required to withdraw from the programme

A detailed information on progression criteria is available in "D1 Assessment Regulations" of "The Wheel of Academic Law" (www.rub.edu.bt).

1.8 Planned Student Numbers

The student intake for the next five years for the programme is reflected in the following table.

Year	2021	2022	2023	2024	2025
Year 1	30	30	30	30	30
Year 2		30	30	30	30
Year 3			30	30	30
Year 4				30	30
Total	30	60	90	120	120

1.9 Programme Management, Quality Assurance and Enhancement

To ensure quality and proper monitoring both in terms of delivery and assessments, a closely coordinated system of programme management mechanisms are in place that is in line with the guidelines spelt in "The Wheel of Academic Law" (www.rub.edu.bt) and the "Academic Affairs Guidelines" (www.sherubtse.edu.bt). This is enabled through specific responsibilities delegated to dedicated staff and institution of several committees, and through timely student feedback, as outlined below.

The following sections provide an overview of the mandatory university-wide quality assurance and enhancement procedures:

Dean of Academic Affairs (DAA)

DAA maintains and implements academic regulations/guidelines, supports and implements innovative approaches to teaching-learning and quality across all programmes in the College. DAA is responsible for maintaining the health of all programmes at the College.

Head of the Department (HoD)

HoD provides organisational and academic leadership for programmes offered by the department, and is accountable to the DAA in the day-to-day operations. HoD is responsible for maintaining the health of programmes within the department.

Programme Leader (PL)
PL provides organisational and academic leadership for the programme and is directly accountable to the HoD. PL is responsible for maintaining the health of the programme and reporting issues related to the programme to appropriate committees. PL is also responsible for drafting the Annual Programme Monitoring Report and its reporting to the Programme Committee and College Academic Committee and final submission to the Programmes and Quality Committee.

Module Tutor (MT)
MT is directly accountable to PL and is responsible for teaching and assessment of a particular module as per the semester plan agreed upon with the PL. Importantly, MT evaluates the relevancy and currency of the module descriptor and recommends need for updates to the PL. When there are multiple tutors teaching the same module, a Module Coordinator (MC) is appointed, who is responsible for ensuring the health of the module as per the definitive programme document. At the conclusion of the semester, the MT (or the MC) prepares module reports for the modules taught in the semester. Module reports provide insights on issues related to the delivery of modules and plan of actions for the subsequent year, which are reported in the Annual Programme Monitoring Report.

College Academic Committee (CAC)
CAC is the highest decision-making body for all matters related to academic affairs in the College. In particular, CAC serves as the guarantor of academic standards and quality in respect of the design, delivery, development and promotion of best practice in curricula, programmes, general educational matters and research within the College. It is responsible for implementation of the University academic quality assurance policies and procedures covering the development and the monitoring of taught programmes, learning and teaching and the academic support of students within the College. CAC is chaired by the DAA, with members constituting of President, Dean of Student Affairs, Dean of Research & Industrial Linkages, elected staff representatives (HoDs & selected PLs), elected student representatives (two student leaders), representative of other groups of staff (Librarian & ICT Officer), an external member and one senor academic as a secretary.

Programme Committee (PC)
At the programme level, PC is responsible for the effective conduct, organisation and development of the programme, including appointment of module tutors, allocation of teaching-learning resources required for the semester. PC is chaired by the PL, with all MTs/Module Coordinators of the programme and at least three students of the programme representing different cohorts, constituting the membership.

Student Consultative Meetings (SCM)
The purpose of the SCM is to involve students in the operation of programmes and in improving the effectiveness of their own education. The SCM is convened at the Departmental level once in the mid-semester. The meeting provides a forum for the students to provide feedback on all elements of the programme such as the delivery of

the modules; the subject matter of the modules; the effectiveness of the teaching, learning and assessment approaches; the adequacy of teaching learning resources; progression and achievement; guidance and support as well as examples of good practice. The SCM is chaired by DAA, with HoD, PLs and at least two student representatives from each year of the programme constituting the membership.

Student Module Evaluation (SME)

SME is another quality assurance and enhancement mechanism in which students are engaged in the assurance and enhancement of their educational experience. Through this mechanism, the College seeks feedback from all the students enrolled for the module in terms of module delivery, resources available, quality of learning and teaching, relevance of assessment methods, and the professionalism of module tutors. SME is done at the end of semester through use of RUB-wide standard module evaluation form, integrated to the VLE. The line managers (Programme Leaders, Head of Departments and Dean of Academic Affairs) at the College review the feedback for every module and actions are taken when the new semester starts.

Annual Programme Monitoring Report (APMR) and Module Report (MR)

APM is a key component of the University's quality assurance and enhancement processes which provides assurance of the continued quality, standards and relevance of programmes in operation. APMR ensures that programmes leading to an award of the University meet their aims and learning outcomes effectively, while at the same time, it strives to enhance the quality of learning and teaching at the University. It is a continuous process of appraising the performance of programmes throughout the year culminating in a consolidated Annual Programme Monitoring Report at the end of the academic year. Programme Leader is responsible for compiling the APMR.

The MR provides a critical appraisal of the delivery of a module by reviewing its current strengths and weaknesses, and provides evidence upon which to plan the improvement of the module. It feeds the APMR by providing informed, evidence-based action points for the programme of which the module forms a part. Module tutor (or the module coordinator) is responsible for producing the module report.

Programme Board of Examiners (PBoE)

As outlined in “The Wheel of Academic Law”, the PBoE ensures that module assessments are in compliance with the validated/reviewed module descriptors and the progression of students to the next level is assessed in compliance with RUB regulations. PBoE is chaired by a senior member of the staff cognisant of the programme but not closely involved in it, with HoD, PLs, staff with assigned responsibility for assessments, and an external examiner appointed by the Academic Board constituting the membership.

Moderation of Assessments

Moderation is a quality assurance process to ensure assessment is accurate, consistent and fair. It also assures that the results are an accurate reflection of performance and can be relied upon by students and staff within the university, as well as by external stakeholders. As required by the RUB regulation: “D8 Moderation of Assessments” of “The

Wheel of Academic Law" (www.rub.edu.bt), at Sherubtse College all 'assessment tasks' and 'assessed student works' constituting 20% or more of the total assessment weighting of a module are moderated through internal and external moderation process. Internal moderation is done by a moderation committee, consisting of tutors from same discipline, who may or may not teach the module. External moderation is the review of examination questions and a representative sample of answer scripts, and assessed components of Continuous Assessment tasks for a module by the external examiner(s) for a programme.

Role of External Examiners

External examiners are independent advisers for a programme in operation. They contribute to the quality assurance and enhancement of a programme by providing an external view on assessments, student achievements, academic standards and a range of academic matters related to the delivery of a programme. The roles and responsibilities of external examiners are outlined in regulation D3 "External Examiners" in The Wheel of Academic Law.

In addition to the university-wide mandatory procedures, Sherubtse has instituted the following internal quality assurance and enhancement initiatives:

Department Academic Committee (DAC)

Given multiple programmes offered by each department at Sherubtse College, the DAC of Mathematics and Computer Science is mandated with the responsibility of promoting academic quality and standards of the department, besides implementing resolutions of the CAC at the departmental level. The committee is chaired by the HoD, with PLs and all MTs constituting the membership.

Semester Guide (SG)

At Sherubtse College one staff is appointed as SG from among the module tutors for each cohort of students for every programme. SG is responsible for giving guidance on both academic and non-academic matters, which includes counseling/parenting, advising on class attendance records, and reporting specific academic needs of students to MT and PL.

1.10 Academic Staff

The existing faculty members in the Department of Mathematics and Computer Science will be sufficient to teach the first two years of the programme.

The following table shows the staff profile of the Department of Mathematics and Computer Science faculties.

Table 4 Overall faculties in DoMCS and their qualification

Department of Mathematics and Computer	Name	Designation	Qualification	No. of years in a teaching position	Status

Science (DoMCS)					
Mathematics	Mr. R. Balamurugan	Assistant Professor	MSc, MPhil Mathematics	30	Fixed-term
	Mr. Thinley Namgyel	Lecturer	MSc Mathematics	12	Regular
	Mr. Pema Tshering	Associate Lecturer	MSc Mathematics	10	Regular
	Mr. Pema Wangdi	Associate Lecturer	MSc Mathematics	10	Regular
	Mr. Dechen Lhendup	Assistant Lecturer	MSc Mathematics	4	Regular
	Mr. Ugyen Samdrup Tshering	Assistant Lecturer	BSc Mathematics	1	Pursuing Master's degree
Computer Science	Mr. P. Paulraj	Assistant Professor	MCom, MPhil, ANC (NIIT)	19	Fixed term
	Mr. Sangay Thinley	Lecturer	MCS	15	Regular
	Mr. J. Gurubalan	Lecturer	MSc (CS), BEd	8	Fixed-term
	Mr. Karma Dorji	Associate Lecturer	MCS	10	Regular
	Mr. Phub Namgay	Associate Lecturer	MCA	6	Pursuing PhD
Statistics	Mr. Dawa Wangchuk Gyelpo	Assistant Lecturer	BSc Mathematics	1	Pursuing Master's degree
	Ms. Samten Choden	Associate Lecturer	MSc Mathematics	10	Regular
	Ms.Chimi Lhazom	Assistant Lecturer	BSc in Statistics	1	Regular

Before the CPL faculty leaves, the College is planning to recruit additional faculty member to teach the programme. As the cohort progresses to the higher level, the College has to recruit additional 2 faculty members specialized in data science (1 by Autumn 2023 and 1 by Autumn 2024) so that the programme is delivered as expected.

The details of modules offered in various semesters and recruitment plan is given in Table 5.

Table 5: Recruitment Plan

Cohort	2021-22	2022-23	2023-24	2024-25
Student Number	**30**	**60**	**90**	**120**
S-F ratio	2	4	6	8
Regular	2	2	2	5
Fixed	2	2	2	3
Excess	2	0	0	0
Short	0	0	2	0
Recruitment Plan		2(1-R 1-FT) June 2023		
Qualification		MSc/BSc		
Specialization		Computer Vision/AI and Machine Learning		

Based on the gaps in knowledge and skills identified by the programme team, the following short-term professional development programmes are planned as shown in Table 6 below.

Table 6: Short-term Professional Development Programmes

Sl. No.	Broad Area	No. of faculty members to be trained	Timeline	Estimated Cost	Source of Funding
1.	Machine Learning	3	Before Autumn 2022	300,000	College HRD Budget
2.	Deep Learning	3	Before Autumn 2023	300,000	College HRD Budget
3.	Big Data Technologies	3	Before Spring 2024	350,000	College HRD Budget
4.	Computer Vision	3	Before Spring 2024	350,000	College HRD Budget
5.	Natural Language Processing	3	Before Spring 2024	350,000	College HRD Budget

1.11 Resource needs

1.11.1 Accommodation

With the phasing out of programmes, there would not be a problem in terms of availability of classrooms and other academic facilities. Due to the same reason enough space will be available in the hostels to accommodate the students of this programme.

1.11.2 Equipment

For effective delivery of modules, the College has committed budget to procure a high-end server with a minimum storage capacity of 50TB by the end of June 2022. Another server with the same/more capacity will be put in place once the programme is fully operational. Each server is estimated to cost Nu. 1.2 million.

1.11.3 Overall Staff Support

The high-end server(s) which will be procured for the programme will be configured and administered by the faculty members. There are sufficient support staff to provide network and ICT-related services. No additional support staff is required to run the programme.

1.11.4 Library Support

The College has an annual budget of Nu. 1.61 million for library book procurement (Details in **Appendix B**). The college will invest 0.5 million for procurement of books for the programme every year for the first three years and the budget will reduce to 0.10 million in the subsequent years.

Table 7: Semester-Wise Funding Required for Library Books/References

Year & Semester	Cost (Nu.)	Acquire Before/ Timeline	Source of Funding
Year 1 Semester I	153554	Spring 2021	**College Library Fund**
Year 1 Semester II	69637	Autumn 2021	
Year 2 Semester I	466749.6	Spring 2022	
Year 2 Semester II	373014.28	Autumn 2022	
Year 3 Semester I	129088.26	Spring 2023	
Year 3 Semester II	150636	Autumn 2023	
Year 4 Semester I	259012.05	Spring 2024	
Year 4 Semester II	8002	Autumn 2024	
Total	**1609693.19**		

The College has committed to buy the required books/references. The College has also made a commitment to subscribe to at least two important journals of Data Science.

1.11.5 Computing Support

The College has four computer labs. There are 30 computers in each lab with Internet connectivity. The College has an internet bandwidth of 84 Mbps. The labs are also equipped with Wi-Fi facilities, so that students can use their personal laptops. With the Computer Science programme being phased out, the existing labs are more than sufficient to accommodate the Data Science students.

1.11.6 Other Support Facilities

The College has the following support facilities for the smooth functioning of the programme:

- Reading Space: There are spacious reading places in the library, student service centre and around the academic block of the College for the students' use after class hours.
- Wi-Fi connections are available around all the academic blocks, student service centre, library and the administrative building. Further, there is a dedicated lab with Wi-Fi facilities available from 8.30 am to 10 pm for the students' use.
- Student service centre also has a counselling centre with three trained counsellors who are available as and when students require. Further, a toll-free number 6006 can be contacted any time for availing counseling related services.
- Student service centre also has Happiness and Wellbeing Centre that frequently conducts Mindfulness and Yoga sessions which can be attended by interested students. It also has a mini library with books related to mindfulness and personality development which available for students.

- Student Service Centre also has a reprographic centre where students can avail printing and photo copying services with nominal charges.

2 Module Descriptors

Year 1 Semester I

PLT102: Programming in Python

Module Code & Title: PLT102 Programming in Python
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Sangay Thinley

General Objective

This module introduces the fundamental principles of computer programming in Python. Students will learn how to use Python programming language for writing simple programs and for data analysis. Further, the module will prepare the students to use other special-purpose programming languages like R for use in solving statistical problems and data analysis.

Learning Outcomes

On completion of the module, students will be able to:

1. Draw flow charts based on algorithms for computer programs.
2. Design and program simple applications using Python.
3. Apply loop structures and Booleans to navigate through Python programming structures.
4. Outline the purpose and application of loops and decision statements.
5. Employ functions in Python for reuse.
6. Import data from online sources.
7. Describe the core objects of Python, such as strings, lists, tuples and dictionaries.
8. Compute statistical measures using Python.
9. Write simple object-oriented program in Python.
10. Import useful modules from the Internet and create one's own modules.
11. Store and manipulate data using NumPy library
12. Manipulate data using Pandas library
13. Determine the correct use of Numpy or Pandas based on a condition

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**90**
	Laboratory session	3	
Independent Study	Self-study	2	**30**
Total			**120**

Assessment Approach

A. **VLE quiz: 5%**
VLE quiz (with 20 multiple-choice questions worth 20 marks) will be conducted after the mid-semester examination with at least 80% of the subject matters being covered and before laboratory examination for a duration of 30 minutes. The marks obtained will then be converted to 5%.

B. **Lab Assessment: 35%**
Lab Assessment will be assessed based on the following components:

- Laboratory Assignments: 15%.
- Laboratory Examination: 10%
- Viva Voce: 10% (Laboratory assignments: 7%, Laboratory examination: 3%).

Laboratory Assignments (a total of 10 with one assignment for first three units and one each for remaining units) will be based on the concepts taught in the theory class. Students will be provided a set of programming questions every week covering the concepts taught on each unit. The students will have to submit the solutions at the end of the lab session through the VLE.

The following rubrics will be used to assess the set of programming questions. Each set will be assessed out of 100 marks and converted to 1.5%.

Program (100)	**(Excellent)** **100-81%**	**(Good)** **80-61%**	**(Satisfactory)** **(60-41% of the points)**	**(Unsatisfactory)** **(<=40% of the points)**
Requirements and Delivery **(50 points)**	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution **(20 points)**	Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic **(10 points)**	• Program is logically well designed.	• Program has slight logic errors that do no	• Program has significant logic errors.	• Program is incorrect.

		significantly affect the results		
Coding Standards **(10 points)**	• Includes name, date, and assignment title. • Excellent use of white space. • Creatively organized work. • Excellent use of variables (no global variables, unambiguous naming).	• Includes name, date, and assignment title. • Good use of white space. • Organized work. • Good use of variables (no global variables, unambiguous naming)	• Includes name, date, and assignment title. White space makes program fairly easy to read. • Organized work. • Good use of variables (few global variables, unambiguous naming).	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines). • Disorganized and messy. • Poor use of variables (many global variables, ambiguous naming).
Documentation **(10 points)**	• Clearly and effectively documented including descriptions of all class variables. • Specific purpose noted for each function, control structure, input requirements, and output results.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function and control structure.	• Basic documentation has been completed including descriptions of all class variables. • Purpose is noted for each function.	• Very limited or no documentation included. • Documentation does not help the reader understand the code.

Laboratory examination will be conducted after the completion of the module and before semester-end examination and will cover all the units. Each student will be given two questions and will be graded out of 100 based on the rubric mentioned above. The marks obtained will then be converted out of 10%.

Viva voce for the laboratory assignments will be conducted at least two times to assess the students' understanding of the laboratory assignments. Viva voce will also be conducted for the laboratory examination. The following rubrics will be used for viva voce. Each viva voce (maximum 5 minutes per student) will be assessed out of 100 marks and converted to 7% for laboratory assignments and 3% for lab examination respectively.

Program (100)	(Excellent) 85-100%	(Good) 61-84%	(Satisfactory) (<=60% of the points)
Knowledge in programming **(50 points)**	Demonstrates deep knowledge, answer the questions with explanation and elaboration.	Adequate knowledge of most topics, answer the questions but fails to elaborate.	Superficial knowledge of topic, only able to answer basic questions.
Problem-solving ability **(30 points)**	Efficient mapping of theory concepts with practical problem solving approaches.	Moderate mapping of theory concepts with practical problem solving approaches.	Improper mapping of theory concepts with practical problem solving approaches.
Critical Thinking **(20 points)**	Can predict and defend problem outcomes.	Approximately predicts and defends problem outcomes.	Is unable to predict problem outcomes for the given input data set.

C. **Mid-Semester Examination: 10%**

Mid-semester Examination will be conducted (out of 40) in the mid of semester with at least 60% of the subject matters being covered for the duration of 2 hours. The marks obtained will then be converted out of 10%.

D. **Semester-end Examination: 50%**

Semester end Examination for the duration of two and half hours will be conducted. The exam will be conducted out of 50.

Overview of Assessment Approaches and Weighting

Areas of Assignments	Quantity	Weighting

A. VLE Quiz		1	5
B. Laboratory Assessment	Laboratory assignments	10	15%
	Laboratory examination	1	10%
	Viva-voce	3	10%
C. Mid-Semester Examination		1	10%
D. Semester-end Examination		1	50%
Total			100%

Prerequisite: None

Subject Matter

Unit I: Fundamentals of Programming

1.1 Algorithm Basics
1.2 Flow Chart
1.3 Introduction to Programming
1.4 Programming Languages

Unit II: Introduction to Python

2.1 Software Development Process
2.2 Write Simple Python Program
2.3 Output Statements
2.4 Elements of Programs
2.5 Definite Loops

Unit III: Variables, expressions and statements

3.1 Values and Types
3.2 Variables
3.3 Variable names and Keywords
3.4 Statements
3.5 Evaluating Expressions
3.6 Operators and operand
3.7 Order of operations
3.8 Operations on Strings
3.9 Composition
3.10 Comments

Unit IV: Defining Functions

4.1 Function of Functions
4.2 Recursive functions
4.3 Arguments
4.4 Return values
4.5 Local Variables
4.6 Functions that Modify Parameters
4.7 Functions and Program Structure

Unit V: Iteration and Conditional Execution

5.1 If-Else Statement
5.2 For Loops
5.3 While Loops
5.4 Indefinite Loops
5.5 Common Loop Patterns
5.6 Computing with Booleans
5.7 Other Common Structures

Unit VI: Strings

6.1 Concatenation and repetition
6.2 Indexing
6.3 Slices
6.4 Looping
6.5 String Methods
6.6 Escape Characters

Unit VII: Data Collections

7.1 Lists and Arrays
7.2 List Operations
7.3 Statistics with Lists
7.4 List of Records
7.5 Designing with Lists and Classes

Unit VIII: Dictionaries

8.1 Dictionary examples
8.2 Working with dictionaries
8.3 Counting words

Unit IX: Objecting-Oriented Programming with Python

9.1 Creating your own classes
9.2 Inheritance
9.3 Polymorphism
9.4 Method Overloading and Overriding
9.5 Examples (Tic-tac-toe)

Unit X: Modules

10.1 Importing Modules
10.2 Dates and Times
10.3 Working with files and directories
10.4 Zip Files

10.5 TXT Files
10.6 Getting files from the Internet
10.7 Creating your own modules
10.8 HTTP Protocol (request and response)
10.9 JSON

Unit XI: Introduction to NumPy

11.1 Understanding Data Types in Python
11.2 The Basics of NumPy Arrays
11.3 Computation on NumPy Arrays: Universal Functions
11.4 Aggregations; Comparisons
11.5 Masks, and Boolean Logic
11.6 Fancy Indexing; Sorting Arrays
11.7 Structured Data

Unit XII: Data Manipulation with Pandas

12.1 Introducing Pandas Objects
12.2 Data Indexing and Selection
12.3 Operating on Data in Pandas
12.4 Handling Missing Data
12.5 Hierarchical Indexing
12.6 Juggling between Numpy and Pandas, Knowing when to use Numpy and Pandas

List of Laboratory Sessions

1. Program to use operators and operand
2. Program to swap two variables
3. Program to find factorial of a number (including recursion)
4. Program to make a calculator
5. Program to work with dictionaries (e.g. merge two dictionaries)
6. Program to apply string methods
7. Program to demonstrate difference between lists and arrays
8. Program to implement inheritance and polymorphism
9. Programs to implement Pandas data frame
10. Programs to implement NumPy arrays
11. Program to import and use modules (including downloading files from the Internet)
12. Program to create Python dictionary from JSON response

Reading List:

Essential Reading

Deitel, P., Deitel, J. (2019*). Introduction to Python for Computer Science and Data Science Learning to Program with AI, Big Data and the Cloud*. Pearson.

Heinold, B. (2012). *A Practical Introduction to Python Programming.* Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.

VanderPlas, J. (2016). *Python Data Science Handbook*. OReilly

Zelle, J.M. (2017). *Python programming: An Introduction to Computer Science* (3rd.). Tom Sumner.

Additional Reading

Lutz, M. (2013). *Learning Python* (5th ed.). OReilly.

Rao, R. N. (2017). *Core python programming*. Dreamtech press.

Sedgewick, R., Wayne, K., & Dondero, R. (2015). *Introduction to programming in Python: An interdisciplinary approach*. Addison-Wesley.

Date: March, 2022

DST101 Descriptive Statistics

Module Code & Title: DST101 Descriptive Statistics
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

The objective of this module is to introduce the basic information about variables in a data set and emphasise the potential relationship between variables and descriptive statistical techniques that would help the students in understanding the importance and need of statistics. It would help them in understanding the concepts involved in data presentation, analysis, and interpretation using suitable statistical data and measures. Students are expected to use Spreadsheet Package during Laboratory sessions.

Learning Outcomes

On completion of the module, students will be able to:

1. Define the basic concepts and scope of statistics.
2. Explain the descriptive statistical techniques and its scope in describing the real world problems.
3. Present data in a meaningful manner after the proper arrangement using Spreadsheet Package.
4. Interpret data using appropriate statistical measures using statistical functions.
5. Compute different measures from the data collected using statistical functions.
6. Evaluate the values of the different measures of dispersion by making use of an appropriate formula.
7. Calculate the moments from the datasets using statistical functions.
8. Measure the skewness of the datasets and interpret the “lack of symmetry” from the value computed.
9. Find the inter-class correlation of the datasets using Data Analysis Tools.
10. Apply the concept of attributes and consistency in real-world problems.
11. Apply the coefficient of association and colligation in practical problems.

Teaching and learning approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Laboratory session	1	
Independent study	Written assignment	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. **Problem-solving Assignment (10%)**
Individual assignment will be given during the first half of the semester. The assignment will comprise of problem-solving based on the subject matters of Unit I and Unit II to assess the students' ability in data representation skills, data visualizing skills and interpretation of descriptive statistics. The assignments will be assessed using following criteria:

- Correctness of the solutions: 40%
- Logical flow in the process: 20%
- Uniqueness of the solution: 20%
- Use of appropriate formulae and symbols: 20%
-

The final marks will be converted out of 10%.

B. **Laboratory Assessment (20%)**
Laboratory assessment will be carried out as follows:

- **Laboratory report- 5%**

Students will submit a weekly record of the laboratory session of 250-500 words to keep track of his/her progress. Each weekly report will be assessed for 5 marks. At the end of all the laboratory sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. The students will be required to submit minimum of 10-laboratory. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD. The laboratory report will be assessed using the following criteria:

- Objective: 1
- Methods: 2
- Results: 1
- Conclusions: 1

- **Laboratory Exam- 15%**
  A laboratory exam for the duration of 1 ½ hour will be conducted towards the end of the 13th week of regular teaching to assess their ability in describing and interpreting real world problems using Spreadsheet Package. Laboratory exam will be assessed using the following criteria:

- Correctness of the solutions: 40%
- Use of appropriate Spreadsheet functions/formulae/graphs 30%
- Inference on findings 30%
-

The final marks scored by the students will be converted out of 15%.

C. **Mid-Semester Examination (20%)**

Mid-semester examination will be conducted in the mid of the semester with at least 60% of the subject matters being covered for the duration of 2 ½ hours to evaluate their ability in interpreting descriptive statistics. The exam will be conducted out of 50 marks and the final marks will be converted out of 20%.

D. **Semester-end Examination (50%)**
Semester-end Examination will cover all of the subject matters for the duration of 2 ½ hours to evaluate their ability in interpreting descriptive statistics. The exam will be conducted out of 50 marks and the final marks will be directly taken into the account.

Overview of Assessment Approaches

Areas of assignments		Quantity	Weighting
A. Problem-solving Assignment		1	10%
B. Laboratory Assignment	Laboratory report	10	5%
	Laboratory exam	1	15%
C. Mid-Semester Examination		1	20%
D. Semester-end Examination		1	50%

Prerequisite: None

Subject Matter

Unit I: Introduction and Data Presentation

1.1 Introduction
- 1.1.1 Definition and scope of Statistics
- 1.1.2 Descriptive statistics vs Inferential statistics
- 1.1.3 Concepts of statistical population and sample
- 1.1.4 Types of data: nominal, ordinal, interval and ratio

1.2 Data presentation
1.2.1 Frequency Distribution: type and construction of frequency distribution tables
1.2.2 Graphical representation of frequency distribution: Histogram, Frequency Polygon, Cumulative frequency curve or the Ogives and Lorenz curves.
1.2.3 Classification and tabulation

Unit II: Analysis of univariate data

2.1 Measures of central tendency
2.1.1 Mean, Median, and Mode and the Relationship between Mean, Median and Mode
2.1.2 Arithmetic mean, Geometric mean, Harmonic mean and their properties
2.1.3 Partition values: Quartiles, Deciles and Percentiles
2.2 Measure of dispersion
2.2.1 Measures of location
2.2.2 The Range, Quartile Deviation and Mean Deviation
2.2.3 Variance and Properties of Variance
2.2.4 Standard Deviation and Application of Standard Deviation
2.2.5 Concept of dispersion, Absolute and Relative measures of dispersion, Requirements of a good measure of dispersion.
2.2.6 Relationship between the measures of dispersion
2.2.7 Coefficient of variation.
2.2.8 Moment generating function, characteristic function, cumulant generating, function
2.2.8.1 their properties
2.2.8.2 moments
2.2.8.3 measures of locations
2.2.8.4 dispersion
2.2.8.5 Skewness and Kurtosis for discrete and continuous variates.

Unit III: Analysis of bivariate data

3.1 Bivariate data and scatter diagram
3.2 Covariance and properties
3.3 Simple correlation and properties
3.4 Correlation coefficients: Pearson, Kendall, Spearman
3.5 Correlation ratio
3.6 Intra-class correlation
3.7 Simple linear regression
3.8 Properties of regression coefficients
3.9 Partial and multiple correlation and regression.

Unit IV: Theory of attributes

4.1 Fundamental set of frequencies
4.2 Consistency of data
4.3 Conditions for consistency
4.4 Contingency table
4.5 Association of attributes
4.6 Measures of associations
4.7 Yule's coefficient of association

4.8 Coefficient of colligation

List of Laboratory Sessions

1. Construction of univariate and bivariate frequency distributions with samples of size not exceeding 200 using COUNTIFS () / FREQUENCY () functions and Pivot Table tool.
2. Diagrammatic and Graphical representation of data and Frequency distributions - Pie diagram, Bar diagram, Multiple Bar diagrams, Sub divided Bar diagrams, percentage bar diagrams, Line diagram, Frequency polygon, Frequency curve, Histogram, Ogives, Lorenz curves.
3. Numerical computations of measures of location and dispersion (absolute and relative) moments, measures of Skewness and Kurtosis for both grouped and ungrouped data - Sheppard's correction for grouped data using the following functions:
   AVERAGE (), MEDIAN (), MOD (), STDEV (), SLOPE (), INTERCEPT (), SKEW (), and KURT ().
4. Fitting of first degree, second degree and exponential curves by method of least squares using Data Analysis Tools.
5. Computation of correlation coefficient, regression lines for raw and grouped data. Rank correlation coefficient, linear predictions using CORREL (), PEARSON (), and FORECAST () functions.
6. Computation of partial regression, partial correlation and multiple correlation coefficients given total correlation coefficients (involving 3 variables) using Data Analysis Tools.
7. Construction of contingency tables using Pivot tables and testing the consistency of data.
8. Computations of various measures of association of attributes using Table tool.

Reading list

Essential Reading

Miller, I., Miller, M., & Freund, J. E. (2014). *John E. Freund's mathematical statistics with applications*. Boston: Pearson.

Goon, A. M., Dasgupta, B., & Gupta, M. (1991). *Fundamentals of Statistics Vol. 1*. The world press.

Gupta, S. P. (2002): *Statistical Methods.* Sultan Chand and Sons, New Delhi.

Linneman, T. J. (2021). *Social Statistics Managing Data, Conducting Analyses, Presenting Results*. Taylor and Francis Group.

Additional Reading

Gupta, S. C., and Kapoor, V. K. (2007). *Applied Statistics*, Sultan Chand and Sons, New Delhi.

Gupta, S. C., & Kapoor, V. K. (2020). *Fundamentals of mathematical statistics*. Sultan Chand & Sons.

Hogg, R. V., and Criag, A. T. (2013). *Introduction to Mathematical Statistics*, Macmillan Publishing, New York.

Mood, A. M., Graybill, F. A., and Bose D. C. (2001). *Introduction to the theory of Statistics*, Tata McGraw Hill, New Delhi.

Sheldon, R. (2007). *Introduction to Probability Models*, Academic Press, New Delhi.

Walpole, R. E., Myers, R. H., Myers S. L., and Ye, K. (2001). *Probability and Statistics for Engineers and Computer Scientists*, Prentice Hall, New Delhi.

Date: December, 2021

CAL111 Calculus

Module Code & Title: CAL111 Calculus
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Pema Tshering

General Objective

The module introduces fundamental concepts of mathematical typesetting using LaTex. The module will also help students to explore concepts, properties and aspects of differential and integral calculus of single and multivariable functions.

Learning Outcomes
On completion of the module, students will be able to

1. Use LaTeX for typesetting mathematical documents and journal articles formats.
2. Label equations while typing mathematical articles.
3. Use beamer class as an alternative for presentations with numerous mathematical symbols and equations.
4. Differentiate functions using different rules of differentiation.
5. Find limits of multivariable functions.
6. Explain the concepts of partial differentiation.
7. Explain the rules and methods of integration
8. Explain the concepts of double and triple integration.
9. Use double and triple integrations to find expected values of a probability distribution function.

Learning and teaching approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	4	**75**
	Tutorial	1	
Independent study	Written assignment	1	**45**
	Self-study	2	
Total			**120**

Assessment Approach

A**: Assignments: 10%)**

An individual assignment will be given during the first half of the semester. The assignment will comprise of problem solving and questions will be given from the units II, III and IV. The assignment will be evaluated broadly on the following criteria:

- Evaluation of limit, checking continuity and finding derivative 2 marks

- Finding extreme values 2 marks
- Checking continuity and finding partial derivatives 3 marks
- Finding extreme values of a multivariate function 3 marks

B**: Class Test: (10%)**

Class test will be conducted to give students the preview of semester-end examinations and to maintain check and balance in the assessment process. A class test of one hour duration will be conducted before the semester-end examinations. The questions will be included from the subject matters covered after mid-semester exam.

C**: Mid-Semester Examination: (20%)**

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the module being covered for the duration of one and half hours. The total weighting of the examination paper should be 40 and the marks obtained in the examination will be converted into 20%

Semester-end Examination: (60%)

Semester end Examination will be conducted for the duration of three hours and questions will be from all the units of the subject matter of the module. The total weighting of the examination will be 60. Test blue print of the question paper and an answer key for grading will also be prepared along with the question paper

Overview of Assessment Approaches

Areas of assignments	Quantity	Weighting
A. Assignment	1	10%
B. Class Test	1	10%
C. Mid-Semester Examination	1	20%
D. Semester-end Examination	1	60%

Pre-requisite: None
Subject Matter
Unit I: Introduction to LaTex

1.1 Introduction to mathematical typesetting with LaTeX
1.2 LaTeX processing modes
1.3 Typesetting packages
1.4 Mathematics fonts: alphabetic and symbolic, hyphens, dashes, minus signs, dots, delimiters, math accents
1.5 Commands with arguments and some mathematical environments
1.6 Typesetting mathematical formulae
1.7 Grouping in Math mode
1.8 Building Blocks of mathematical Formula

1.9 Math spacing, vertically aligned material, Phantoms, Math font size, Theorems, Laws, Bold Symbols and list of mathematical symbols.
1.10 Bibliography, Indexing
1.11 The Verbatim Package
1.12 creating Presentations with the beamer class.

Unit II: Limit and continuity of functions of one variable

2.1 Introduction to the concept of limit
2.2 Rules for finding limits
2.3 Continuity of a function
2.4 Types of discontinuity

Unit III: Derivatives and Applications

3.1 Differentiability of a function
3.2 Relation between continuity and differentiability of a function
3.3 Implicit differentiation
3.4 Maximum and minimum values of a function
3.5 Greatest and least values of a function
3.6 Increasing or decreasing functions on an interval
3.7 First and second derivatives
3.8 Concavity
3.9 Optimisation problems
3.10 Linearization of functions
3.11 Shapes and graphs of elementary functions

UNIT IV: Multivariable functions and partial derivatives

4.1 Functions of several variables
4.2 Limit and continuity of multivariable functions
4.3 Differentiable functions
4.4 Partial derivatives
4.5 Euler's theorem
4.6 Linearization of functions of two variables
4.7 Chain rules
4.8 Extreme values
4.9 Saddle points
4.10 Lagrange multipliers
4.11 Error formula for linear approximations

Unit V: Integration and Application

5.1 Integration by partial fraction
5.2 Integration by trigonometric substitution
5.3 Integration of rational functions
5.4 Integration of irrational functions

5.5 Definite integrals and its properties
5.6 Fundamental theorem of integral calculus
5.7 Area under a curve
5.8 Surface area and volume of revolution of a curve about a straight line
5.9 Relation between area under a curve and probability

Unit VI: Multiple Integrals

6.1 Double Integrals over rectangles
6.2 Iterated integrals
6.3 Double integral over general regions
6.4 Area in the plane using double integral
6.5 Double integral as volume
6.6 Double integral in polar form
6.7 Average value of a function of two variables
6.8 Change of variable in double integral
6.9 Definition of triple integral
6.10 Iterated integrals
6.11 Volume of a region in space using triple integral
6.12 Average value of a function in space
6.13 Jacobian
6.14 Change of variable in triple integral

Reading List
Essential Reading

Oetiket, T., et al. (2005). The Not So Short Introduction to LaTeX. Free Software Foundation, Inc., 675 Mass Ave Cambridge, MA 02139, USA

Thomas, G.B and Finney, R.L. (2012). *Calculus and Analytic Geometry* (12th ed.) New Delhi: Dorling Kindersly (India) Pvt. Ltd.

Kreyszig, E. (2006). *Advanced Engineering Mathematics*. (8thedition). John Wiley and Sons.

Salas, S.L., Hille, E. & Etgen, G.J. (2003). *Calculus: One and Several Variables*. (9th edition). John Wiley and Sons.

Additional Reading

Apostal, T.M. (2005). *Calculus: Volume I*. John Wiley and Sons.

Spiegel, M.R. (1981). *Advanced Calculus*. (Schaum's Outline Series, Asian Student Edition). McGraw Hill Book Company.

Frank, A. Jr., & Mendelson, E. (1992). *Differential and Integral Calculus*. (Schaum's Outline series, 3rd edition). McGraw Hill Book Company.

Steward, J. (1999). *Calculus* (4th ed.) Singapore: Brooks/Cole Publishing Company.

Date: December, 2021

CAE102 Computer Organisation and Programming

Module Code and Title: CAE102 Computer Organisation and Programming
Programme: BSc in Data Science
Credit Value: 12
Module Tutors: J. Gurubalan
Module Coordinator:

General Objective:

This module introduces the organization of computer systems and usage of assembly language for optimization. It also covers internal representation of data at machine level, and show how data is stored and accessed in memory. It also covers different number systems, input/output (I/O) systems which brings the processor and memory together with a wide range of devices. Finally, students will be introduced to write programming in an assembly language.

Learning Outcomes:

On completion of the module, students will be able to:

1. Describe various number systems and convert numbers into different number systems.
2. Describe the fundamental organization of a computer system.
3. Articulate the relationship amongst the functional units of the processor.
4. Describe the functions of CPU (Central Processing Unit) and its components.
5. Create, assemble, and execute assembly language programs along with a basic understanding of the assembly, linker, and loader processes.
6. Write programs using the capabilities of the stack, the program counter, the status register and show how these are used to execute a machine code program.
7. Explain organization of memory hierarchies including the basics of cache design.
8. Compare the basic input/output functions with program controlled I/O and interrupt I/O.

Learning and Teaching Approach:

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	90
	Laboratory sessions	3	
Independent reading	Assignment/Homework	1	30
	Self-study	1	
Total			120

Assessment approach:

A. VLE Quiz: 10%

One Quiz will be conducted on VLE upon coverage of at least 40% of the subject matter to assess the students' overall understanding of the subject matter from first three units. This quiz will contain 30 Multiple Choice Questions, each question will carry 1 mark and later it will be scale-down to 10%. The test will be conducted for 30 minutes duration.

B. Class test: 10%

One class test will be conducted upon coverage of at least 50% of the subject matter. This is a written conceptual test conducted within the class room for a duration of 60 minutes to assess their overall understanding of the subject matter from first four units. The students have to write answers on a sheet of paper in the class room within the predefined time. This class test will be conducted for 25 marks and later it will be scaled down to 10%.

C. Laboratory Assignments: 20%

The first laboratory assignment will be assigned on the 7th week to students to evaluate their practical skills on number system, register organization and addressing mode concepts in Computer Organization. This assignment will be conducted for 100 marks and later it will be scaled down to 10%.

The second laboratory assignment will be assigned on the 14th week to students to enhance and evaluate their practical skills on assembly language concepts in Computer Organization. This assignment will be conducted for 100 marks and later it will be scaled down to 10%. Assessment of the assignments will be done based on the following criteria:

1. Viva voce (2%)
2. Correctness of result (5%)
3. Logic (3%)

D. Laboratory Examinations: 20%

Two practical examinations will be conducted in a semester where the first exam will be conducted on the 7th week from first four units and second will be conducted on the 14th week from the entire subject matter. Each exam will be given a weighting of 10%. Both practical examinations will be conducted for a minimum of two hours and 20 marks and later it will be scaled down to 10%.

E. Semester-End Examination: - 40%

The exam at the end of the semester will be a closed book for a minimum of 2 hours and 40 marks. The exam will evaluate the students' overall understanding and critical thinking ability with regards to the module. All the units of the subject matter will be included for the examination.

Areas of Assignments	Quantity	Weighting
VLE Quiz	1	10%
Class Test	1	10%
Laboratory Assignments	2	20%
Laboratory Examinations	2	20%
Semester-End Examination	1	40%
Total		100%

Pre-requisites: None

Subject Matter:

Unit I: Computer Evolution

1.1. Structure of a Computer System
1.2. Von Neumann Architecture
1.3. Number Systems
1.4. Computer Structure
    1.4.1. Arithmetic Logic Unit
    1.4.2. Control Unit
    1.4.3. Bus Structure

Unit II: Central Processing Unit and Instructions

2.1. Instruction Characteristics
2.2. CPU with Single BUS
2.3. Types of Operands
    2.3.1. Addresses
    2.3.2. Numbers
    2.3.3. Characters
    2.3.4. Logical data
2.4. Types of Operations
    2.4.1. Arithmetic
    2.4.2. Data transfer
    2.4.3. Logical
    2.4.4. Conditional branch
    2.4.5. Bitwise
2.5. Addressing Modes
    2.5.1. Immediate
    2.5.2. Direct
    2.5.3. Register
    2.5.4. Indirect
    2.5.5. Implied
2.6. Instruction Formats
2.7. Graphics Processing Unit
    2.7.1. Computational Functions

Unit III: Processor Organization

3.1. Parallelism and Computer Arithmetic
3.2. Computer arithmetic associativity
3.3. Floating Point in the 8086
3.4. Programmers Model of 8086
3.5. Max/Min Mode, Minimum mode
    3.5.1. Maximum mode, Register Organization
    3.5.2. 8086 general purpose Registers
    3.5.3. 8086 segment registers
    3.5.4. 8086 special purpose registers
3.6. Instruction Cycles
    3.6.1. Read Write cycles, Read cycle, Write cycle

Unit IV: Programming the Basic Computer

4.1. Introduction
4.2. Machine Language
4.3. Assembly Language
  4.3.1. Assembler
  4.3.2. Program loops
  4.3.3. Programming Arithmetic and logic operations
  4.3.4. Subroutines
  4.3.5. I-O Programming

Unit V: Memory Organization

5.1. Characteristics of Memory Systems,
5.2. Main Memory
  5.2.1. Types of Random-Access Memory and ROM
  5.2.2. Organization
5.3. Static and dynamic memories
5.4. Memory system considerations
5.5. Memory interleaving

Unit VI: High Speed Memories

6.1. Cache Memory
  6.1.1. Principles of cache memory
  6.1.2. Structure of cache and main memory
  6.1.3. Performance using cache memory
  6.1.4. Elements of Cache Design
6.2. Mapping functions
  6.2.1. Replacement algorithms

Unit VII: Secondary Memory

7.1. Magnetic Disk and Tape
7.2. Digital Audio Tape (DAT)
7.3. Redundant Array of Independent Disks
  6.3.1. RAID 0
  6.3.2. RAID 1
  6.3.3. RAID 2
  6.3.4. RAID 3
  6.3.5. RAID 4
  6.3.6. RAID 5
7.4. Optical memory
7.5. External Memory
7.6. Virtual memory

Unit VIII: I/O Organization

8.1. Need of I/O Module
8.2. External Devices
  8.2.1. Input / Output Module
  8.2.2. I/O Module Function
  8.2.3. I/O Module Decisions
8.3. Input Output Techniques
  8.3.1. Programmed I/O

8.3.1.1. I/O commands
8.3.1.2. I/O instructions

8.4. Interrupt Driven I/O
8.4.1. Basic concepts of an Interrupt
8.4.2. Response of CPU to an Interrupt
8.4.3. Design Issues,
8.4.4. Priorities
8.4.5. Interrupt handling
8.4.6. Types of Interrupts

8.5. DMA (Direct Memory Access)

List of Laboratory sessions (Using TASM/MASM)

1. Compute arithmetic Operations of numbers.
2. Transfer / exchange block of data using string Operations.
3. Determine bit / byte value using Logical Operations.
4. Store flag register values using branch operations.
5. Determine bit value using shift and rotate instruction Operations.
6. Convert a Hexadecimal, Binary number into Decimal number.

Reading List:

Essential Reading:

Yadin, A. (2016). *Computer systems architecture*. CRC Press.

Mano, M. M. (2005). *Computer system architecture*. Prentice-Hall of India.

Chaudhuri, P. P. (2008). *Computer organization and design*. PHI Learning Pvt. Ltd..

Additional Reading:

Kumar, N. S., Saravanan, M., & Jeevananthan, S. (2011). *Microprocessors and microcontrollers*. Oxford University Press, Inc.

Stallings, W. (2003). *Computer organization and architecture: designing for performance*. Pearson Education India.

Hamacher, C., Vranesic, Z., & Zaky, S. (2002). *Computer organization*. McGraw-Hill.

Date: March, 2022

DZG101 རྩོང་ཁ་བརྟེ་དོན་སྦྱོང་ལེན།

སྦྱོང་ཚན་གྱི་མིང་དང་སྦྱོང་ཚན་ཨང་འགས་རིམ་	DZG101 རྩོང་ཁ་བརྟེ་དོན་སྦྱོང་ལེན།
	གཙུག་ལག་སློབ་སྟེ་ཡོངས་ཁྱབ།
སྦྱོང་འབྲས་	༡༢
སྦྱོང་ཚན་སློབ་སྟོན་པ།	རྩོང་ཁའི་ལེགས་བཤད་པ།
སྦྱོང་ཚན་འགོ་འདྲེན་པ།	Kelzang Namgyal

སྦྱིར་བཤད་ལས་དོན།

སྦྱོང་ཚན་འདི་གི་དམིགས་ཡུལ་གཙོ་བོ་ར་ སློབ་སྦྱོང་པ་ཚུ་གིས་ རང་གི་མི་ཚེ་ནང་ ལྟ་གཡོག་དང་འཛེལ་བའི་གནད་དོན་ གི་ཅིའི་ཐད་ལས་འབད་རུང་ རྩོང་ཁའི་ཐོག་ལུ་ བརྟེ་དོན་སྦྱོང་ལེན་ ཚུལ་དང་ མཐུན་ཉོག་ཉོ་སྟེ། འབད་ཚུགས་ནའི་དོན་ལུ་ཨིན། དེ་མ་ཚང་ སྦྱོང་ཚན་འདི་གིས་ རྩོང་ཁའི་ཐོག་ལུ་ ཤེས་ཡོན་འབྲི་རྒྱ་ལས་ལུགས་དང་འཁྱིལ་ཏེ་ ཐི་ཚུགས་ནི་ལུ་ཡང་དམིགས་གཏད་བཟོདཔ་ཨིན།

སློབ་སྦྱོང་གྲུབ་འབྲས།

སྦྱོང་ཚན་འདི་མཇུག་བསྡུས་པ་ སློབ་ཕྲུག་ཚུ་གིས་

༡	རྩོང་ཁའི་སྐད་ཡིག་གི་འཛུང་རབས་དང་ རྩོང་ཁ་ལྟར་དགོ་པའི་ཁྱངས་དང་དགོས་པ་ སྐབས་ཚུགས།
༢	སྒོག་རིག་ནང་ཐུབ་སྟེ་ རྩོང་ཁའི་མཇུག་སྦྱོན་ ཆུག་ཚུགས།
༣	འབྲེལ་སྐུ་དང་ཐྲིབ་སྐུ་ ལྟག་བཅས་ ཚིག་གྲོགས་ འབྲེལ་ཚིག་ ཚབ་ཤད་ཚུ་ དབྱེ་དཔྱད་འབདཝ་དེ་ ཨ་འཛོལ་བར་ ལག་ལེན་འཐབ་ཚུགས།
༤	ཡུལ་ དུས་ གནས་སྐབས་དང་བསྟུན་ཏེ་ རྩོང་ཁའི་ཐོག་ལུ་ ཡན་སྐབས་འབད་ཚུགས།
༥	རྩོང་ཁའི་ཐོག་ལུ་དེས་ཡོད་ཚེ་ཚུ་ ངག་གཤིས་དང་མཐུན་ཐེ་ལྟག་ཚུགས།
༦	གནས་སྐབས་དང་འཁྱིལ་བའི་དབྱེ་གཏམ་ ལག་ལེན་འཐབ་ཚུགས།
༧	རྩོང་ཚིག་དང་དོན་མཚམས་ ཚུལ་མཐུན་ཐེ་དེ་ཚུགས།
༨	གཞུང་སྐེར་གྱི་ཡིག་རིགས་དེ་ཚུགས།
༩	རྩོང་ཁའི་ནང་ཡོད་པའི་ཐོག་ཤོག་ཁྲ་ ལག་ལེན་ཐབས་ཚུགས།
༡༠	ཡིག་རིགས་ཚུ་ ལུགས་མཐུན་དང་ལུགས་ཡངས་གཞིས་ཆ་རའི་ཐོག་ལུ་དེ་ཚུགས།
༡༡	ལུང་འདེན་དང་ཆུབ་རྐེན་གྱི་ཐོ་ བོས་འབད་ཐུན་ཏོག་ཏོ་སྟེ། ལག་ལེན་འཐབ་ཚུགས།

སློབ་སྟོན་ཐབས་ལམ།

དབྱེ་ཞིབ་ལམ།	དབྱེ་བ།	ཐབས་ལམ།	བདུན་རྒྱག་གཅིག་ནང་ཚུ་ཚོད།	སྦྱོང་འབྲས་ཚུ་ཚོད།
	དངོས་འབྲེལ།	གསལ་བཤད།	༢	༤༠
		སྦྱོང་ལྟ།	༡	
		སྤྱན་ལུ།	༡	
སྦྱོང་ཚན་དོན་ཡུ་དབྱེ་ཞིབ་སྦྱོང་ལེན།	རང་སྦྱོང་།	ལས་འགུལ་བྱེ་ནི།	༢	༤༠
		དབེ་མཛོད། ལྷག་དེབ་ལྷག་ནི།	༢	
	སྦྱོང་ཚན་འདི་གི་དོན་ཡུ་ ཡོངས་བསྡོམས་ ཆུ་ཚོད་			༡༢༠

ཐབས་ལམ་

དབྱེ་ཞིབ་ གཞིས་ཆ་རའི་ཐོག་ལས་ དབྱེ་ཞིབ་འབད་དགོཔ་ཨིན།

ཀ **ལས་འགུལ་དང་པ།** **ཉན་ནིའི་རིག་རྩལ་དབྱེ་ཞིབ། (སྐྱབས་ ༢༠%)**

སློབ་སྦྱོན་པ་གིས་འོས་འབབ་ལྷན་པའི་དོན་ཚན་ (ཐུམ་བཟང་ཡོད་མེ་ཅིག་) གཙམ་པ་ཕྱིས་སྟེ་ རང་ཕྱང་གི་ཐོག་ལས་ ཁུ་སེམ་སེམ་སྟེ་ཉན་བསྡུག ངོ་རྟང་ ཚར་ཅིག་ཉན་ཅིག་དང་གཅིག་ཁར་ཟེན་ཏེས་བཀབ་བསྡུག་ཞིན་མས་ ཉན་མེ་འདི་ ཚོག་ཐབ་འབད་དེ་ ཏི་བསྡུག དེ་གི་དམིགས་ཡུལ་གཙོ་བོ་ ཉན་ཀེ་ཀ་གོ་ནི་དང་ ཉན་ཀེ་ཏི་ཚུགས་མེ་ཚུགས་དབྱེ་ཞིབ་ འབད་ནིའི་དོན་ལུ་ཨིན།

སྐྱབས་ཀྱི་ཚད་གཞི།

- གོ་དོན་ལེན་ཏེ་ རང་ཚིག་ནང་བཤད་པ་རྒྱབ་ཐངས། ༡༠%
- ཐེབ་སྒྲོར། ༤%
    - ཉན་ཏེ་བྲིས་མི་དེ་གི་དབྱེ་དཔྱད་སྙན་ཞུ། ༣%
    - རང་གི་ཉན་ནིའི་རིག་རྩལ་ཡར་རྒྱས་ཀྱི་གྲོས་འཆར། ༣%

དྲན་གསོ། དབྱེ་དཔྱད་སྙན་ཞུ་དང་ གྲོས་འཆར་ བསྡོམས་ཏེ་ ཤོག་ལེབ་གཅིག་བྲི་དགོ།

ཁ **ལས་འགུལ་གཉིས་པ། སླབ་ནིའི་རིག་རྩལ་དབྱེ་ཞིབ། (སྐུགས་ ༢༠%)**

སློབ་སྟོན་པ་གིས་ སྤྱིར་བཏང་མི་སྡེའི་ནང་ རྫོང་ཁ་བརྡ་དོན་སྤྲོད་ལེན་ནང་ལུ་ དཀའ་ངལ་ག་ཅི་ར་འབྱུང་དོ་ཡོདཔ་ཨིན་ན་གནདམ་ཁ་རྒྱབ་ཞིནམ་ལས་ འཐུས་མི་ ༥ རེ་འབད་མིའི་སྡེ་ཚན་ནང་སྦེ་ སློབ་ཁང་ནང་ལུ་ སྐར་མ་ ༡༥ འི་རིང་ལུ་ སྨྲུན་ཞུ་འབད་བཅུག་དགོ། འཐུས་མི་རེ་གིས་ སྐར་མ་ ༣ རེ་གསལ་བཤད་འབད་དགོ། དེ་གི་དམིགས་ཡུལ་གཙོ་བོ་སླབ་ནིའི་རིག་རྩལ་གྱི་དབྱེ་ཞིབ་འབད་ནིའི་དོན་ལུ་ཨིན།

སྐུགས་ཀྱི་ཚད་གཞི།

- དོན་ཚན་དང་འབྲེལ་བ། ༤%
- སྐད་ཡིག་སྦྱོར་བ། ༣%
- གསལ་བཤད་འབད་ཐངས
    - ཐོགས་ཆགས་མེད་པ། ༣%
    - རྗོད་སྒྲ་དག་པ། ༣%
    - སྐད་ཀྱི་སེང་ཕབ། ༢%
    - གཟུགས་ཀྱི་རྣམ་འགྱུར། ༡%
    - སྤོ་སྤོབས། ༢%
- ཉན་མི་ཚུ་འབྲེལ་གཏོགས་འབད་ཐངས། ༢%

ག **ལས་འགུལ་གསུམ་པ། ལྷག་ནི་དང་བྲི་ནིའི་རིག་རྩལ་དབྱེ་ཞིབ། (སྐུགས་ ༣༠%)**

ཚིག་འབྲུ་ ༡༥༠༠-༢༠༠༠ གི་བར་ན་ཡོད་པའི་རྩོམ་བྲིས་ཅིག་ སློབ་དཔོན་དང་སློབ་ཕྲུག་མཉམ་རུབ་ཀྱི་ཐོག་ལས་ གདམ་ཁ་རྒྱབ་ཞིནམ་ལས་ རང་རྐྱང་གི་ཐོག་ལས་ དབྱེ་དཔྱད་སྙན་ཞུ་བྲི་ནི། དེ་གི་དམིགས་གཏད་འདི་ སློབ་ཕྲུག་ཚུ་གིས་ ལྷག་སྟེ་གོ་དོན་ལེན་ཏེ་ བརྩད་དོན་ཏ་གོ་ནི་དང་ དབྱེ་དཔྱད་འབད་ཐངས་ཀྱི་རིག་རྩལ་ཐོབ་སྟེ་ ཡིག་ཐོག་ལུ་ བཀོད་ཚུགས་ནིའི་དོན་ལུ་དང་ ལུང་འདྲེན་དང་རྒྱབ་རྟེན་ལག་ལེན་འཐབ་ཐངས་ཚུ་ དབྱེ་ཞིབ་འབད་ནིའི་ཨིན། སློབ་ཕྲུག་རེ་གིས་ རྩོམ་བྲིས་ཀྱི་བརྩད་དོན་རེ་དང་ དབྱེ་དཔྱད་སྙན་ཞུ་རེ་བྲི་དགོཔ་ཨིན། དབྱེ་དཔྱད་སྙན་ཞུ་འདི་ ཚིག་འབྲུ་ ༢༠༠༠ ཀྱི་ནང་འཁོད་བྲི་དགོཔ་ཨིན།

སྐུགས་ཀྱི་ཚད་གཞི།

- རྩོམ་བྲིས་ཀྱི་བརྩད་དོན་གཙོ་བོ་ཚུ་ མ་འཛོལ་བར་ བཀོད་ཐངས། ༥%
- དཔྱད་པ་གི་ཐོག་ལས་ དབྱེ་དཔྱད་འབད་ཐངས། ༦%
- གནད་དོན་ཚུ་གོ་རིམ་སྒྲིག་ཐངས། ༣%
- འོས་འབབ་དང་འབྲེལ་བ་ཡོད་པའི་མིང་ཚིག་ལག་ལེན་འཐབ་ཐངས། ༣%
- ལུགས་མཐུན་དང་འཁྲིལ་ཏེ་འབྲི་ཐངས། ༤%
- ཐེབ་སྒྲོར་དག་པ། ༢%
- ཚིག་སྦྱོར་ཚུལ་མཐུན་ལག་ལེན་འཐབ་ཐངས། ༢%
- ལུང་འདྲེན་དང་རྒྱབ་རྟེན་ལག་ལེན་འཐབ་ཐངས། ༣%
- འབྲི་བཀོད། ༢%

ང་ **སློབ་དུས་ཚོས་རྒྱུགས། (སྐུགས་ ༣༠%)**

ལས་རིམ་འདི་ནང་བཅའ་མར་གཏོགས་མི་ཚུ་གིས་ སློབ་དུས་མཇུག་བསྡུ་ཝ་ད་ དུས་ཡུན་ཆུ་ཚོད་ ༢ འི་རིང་ལུ་ སྐུགས་ ༣༠ འི་འབྲི་རྒྱུགས་ཕུལ་དགོཔ་ཨིན།

དབུ་གསོ། འདི་རྒྱུགས་ཀྱི་སྐྱགས་དང་དུས་ཡུན་རྒྱ་ཚོང་གྲུ་ རང་སོའི་མཐོ་རིམ་སློབ་གྲྭའི་ལས་ལུགས་དང་འཁྱལ་ཏེ་ལེན་ཞིན་ལས་ སྐྱགས་བརྒྱ་ཚ་ ༡༠ ནང་ལུ་ ཕད་དགོ་པ་ཨིན།

དབྱེ་ཞིབ་ཐབས་ལམ་དང་ཚྲིད་ཚད་ཀྱི་བཀོད་རིས།

ཐབས་ལམ།	དབྱེ་ཞིབ་ཀྱི་དབྱེ་བ།	གྲངས་ཀ།	སྐྱགས་ཀྱི་བརྒྱ་ཚ།
དུས་རྒྱུན་དབྱེ་ཞིབ།	ཀ། ཉན་ནི་རིའི་རིགས་རྣལ་དབྱེ་ཞིབ།	༡	༢༠%
	ཁ སྐབས་རིའི་རིགས་རྣལ་དབྱེ་ཞིབ།	༡	༢༠%
	ག ལྷག་ནི་བྱེ་རིའི་རིགས་རྣལ་དབྱེ་ཞིབ།	༡	༣༠%
སྐོབ་དུས་ཚོས་རྒྱུགས།	ང་ འབྱེ་རྒྱུགས།	༡	༣༠%
	སྐྱགས་བསྡོམས་ ༡༠༠		

སྡོན་ཚིག། ཨེད།

ནང་དོན།

ལས་ཚན་བདུན་པ། ཡིག་འགྲུལ།

๗.๑ ཡིག་ཆུང་འཕྲི་ཐངས།
๗.๒ མཚོན་ཞུ་འཕྲི་ཐངས།
๗.๓ གཏུང་ཡིག་འཕྲི་ཐངས།
๗.๔ ཞུ་ཡིག་དང་ཞུ་ཚོགས་/བཤེར་ཡིག་ འབྲི་ཐངས།
๗.๕ གཅན་ཡིག་འབྲི་ཐངས།
๗.๖ ཁྱབ་བསྒྲགས་ཀྱི་རིགས་འབྲི་ཐངས།

ལས་ཚན་བརྒྱད་པ། ཤེས་ཡོན་འབྲི་རྩལ།

๘.๑ ཡུགས་མཐུན་དང་ལུགས་ཡངས་ཀྱི་འབྲི་ཐངས།
๘.๒ ལུང་འདྲེན་ ལག་ལེན་འཛུབ་ཐངས།
๘.๒.๑ ཐད་ཀར་ལུང་འདྲེན།
๘.๒.๒ ཚོག་སྐྱར་ལུང་འདྲེན།

๘.๓ རྒྱབ་རྟེན་ཐོ་བཀོད་འབད་ཐངས།
๘.๓.๑ དབེ་དེབ་ཀྱི་རིགས་ཐོ་བཀོད་འབད་ཐངས།
๘.๓.๒ གཅན་དེབ་ཀྱི་རིགས་ཐོ་བཀོད་འབད་ཐངས།

སྒྲུག་དགོ་པའི་དབེ་ཐོ།

བེས་པར་དུ་སྒྲུག་དགོ་པའི་དབེ་ཐོ།

ཀུན་བཟང་རྡོ་རྗེ། (*๑๙๗๒*) *རྩོང་ཁའི་རྩོད་སྐུ་ངག་རྒྱུན་དུག་པའི་ཁྲབས་རྒྱ།* ཐིམ་ཁུ། རྩོང་ཁ་གོང་ཟླན་ཚོགས།
ཀུན་བཟང་རྡོ་རྗེ། (*๑๙๗๫*) *དབེ་ཁ་ཏམ་དོན་ཀྱི་རྒྱན་རྟ།* ཐིམ་ཁུ། རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས།
རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས། (*๑๙๗๓*) *རྩོང་ཁའི་བརྟ་ཁ་ཤྲེང་སྐང་བའི་སྐྱོན་ཚེ།* ཐིམ་ཁུ། རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས།
རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས། (*๑๙๗๔*) *འབྲུག་ཡི་ཡིག་བསྐྱར་རྩམ་ཁ་ཤ་ཁ་ཤེས།* ཐིམ་ཁུ། རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས།
གཙུག་ལག་སློབ་སྡེའི་སློབ་སྦྱོང་ཉེ་བ། (*๑๙๗๙*) ལུང་འདྲེན་དང་ རྒྱབ་རྟེན་འབད་ཐངས། ཐིམ་ཁུ། འབྲུག་རྒྱལ་འཛིན་གཙུག་ལག་སློབ་སྡེ།

ཁ་སྐོང་སྒྲུག་དགོ་པའི་དབེ་ཐོ།

སྐལ་བཟང་དཔང་ཕྱུག (*๑๙๙๒*) *རྩོང་ཁ་བརྟ་དོན་རྒྱུན་འཛེམ།* བཞམ་ཉེ།
རྩོང་གོང་འཛེམ་ཟླན་ཚོགས། (*๑๙๗๗*) *ཤལ་སྐང་དང་ཞེ་ཁའི་རྩམ་གཤེག་སྐར་མའི་འོད་ཟེར།* ཐིམ་ཁུ། རྩོང་ཁ་གོང་འཛེམ་ཟླན་ཚོགས།

སྤྱི་ཚེས། *๑๑/๗๑/๑๙๗๙*།

Year 1 Semester II

AMT114 Discrete Structures

Module Title & Title: AMT114 Discrete Structures
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Pema Wangdi

General Objective

This module introduces the fundamental concepts in discrete mathematics such as logic, proofs, lattices, graphs and recurrence relations. The module is designed to provide students with skills and techniques to deal with logical arguments.

Learning Outcomes

On completion of the module, students will be able to:

1. Translate English statements into logical statements and vice versa.
2. Test the validity of the arguments.
3. Apply various methods to prove the theorems and problems.
4. Identify various function types with examples.
5. Use the exclusion-inclusion principle to solve problems.
6. Verify if a given set is a partially ordered set (POSET) or not.
7. Determine if a POSET is a lattice along with the Hasse diagram.
8. Examine the isomorphism between POSETs and as lattices.
9. Employ graph techniques to solve problems.

Learning and Teaching approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	60
	Tutorial	1	
Independent Study	Written Assignment	1	60
	Self-study	3	
Total			**120**

Assessment Approach

A. Assignments: (30%)

Two written assignments, each worth 15%, will be given in the semester, one before mid-semester examination and one after the mid-semester examination. The assignments will consist a set of problems (at least 10 but not exceeding 15 questions) from the subject matter. The rationale is to assess the understanding of the subject matter taught in the class and their ability to arrive at an acceptable solution. (S)he may be called to attend a viva-voce to justify his work. The final report shall be assessed based on the following rubric:

- Completeness: 5%
- Originality: 5%

- Exposition (language, references, layout, graphics, etc.): 5%

B. Mid-Semester Examination: (20%)

Mid-Semester examination will be conducted upon completing at least 50% of the subject matter and test items set as per the Bloom's taxonomy. The two-hour long exam will be set out of 50 marks with equal weighting for all the Units covered and then be converted to 20%.

C. Semester-end Examination: (50%)

Semester-end Examination will be a comprehensive examination covering all the subject matters prescribed for the module and test items set as per the Bloom's Taxonomy. It will be set out of 50 marks and conducted for a duration of two and half hours.

Overview of assessment approaches and weighting:

Areas of Assessment	Quantity	Weighting
A. Assignments	2	30%
B. Mid-Semester Examination	1	20%
C. Semester-end Examination	1	50%

Prerequisites: None

Subject Matter

Unit I: Predicate Calculus

1.1 Sets
- 1.1.1 Definition and examples
- 1.1.2 Empty set and power sets
- 1.1.3 Cardinality
- 1.1.4 Cartesian product

1.2 Propositions
1.3 Conditional Propositions and logical equivalence
1.4 Syntax and Semantics
1.5 Truth table
1.6 Validity of arguments
1.7 Quantifiers and Nested Quantifiers

Unit II: Induction Principle

2.1 Basic proof techniques
- 2.1.1 Inductive,
- 2.1.2 Deductive
- 2.1.3 Proof by contradiction

2.2 Fundamental theorem of arithmetic
- 2.2.1 Review of properties of the integers
- 2.2.2 The greatest common divisor (GCD) and the least common multiple (LCM)
- 2.2.3 Definition of prime and prime factorization
- 2.2.4 The Division algorithm

2.2.5 The prime divisibility property with proof
2.2.6 Fundamental theorem of arithmetic with proof

2.3 Euclidean Algorithm
2.3.1 Basic algorithm with proof
2.3.2 Extended algorithm without proof

2.4 Continued fractions
2.4.1 Basic formula and examples
2.4.2 Calculating continued fraction representations
2.4.3 Finite continued fractions
2.4.4 Reciprocals

2.5 Modular arithmetic
2.5.1 Congruence
2.5.2 Theorem (without proof) and examples

2.6 Linear congruence and Diophantine equations
2.6.1 Linear Diophantine equations in two variables
2.6.2 Linear Diophantine equations in several variables

2.7 Chinese Remainder Theorem

Unit III: Pigeonhole Principle

3.1 Relations and Functions
3.1.1 Definition of relations and examples
3.1.2 Properties: reflexive, symmetric, antisymmetric, transitive
3.1.3 Equivalence relation and examples
3.1.4 Definitions of functions with examples
3.1.5 Injections, surjections and bijections

3.2 Basic principles of counting
3.3 Number of functions, injections, bijections
3.4 Permutations and Combinations
3.5 Number of subsets, binomial coefficients and their properties
3.6 Binomial and multinomial theorems
3.7 Inclusion-exclusion principle with applications (surjections, derangements, Euler's totient function)
3.8 The Pigeonhole Principle (with proof)

Unit IV: Partially Ordered Sets and Lattices

4.1 Definition and examples of Partially Ordered Sets and Hasse Diagrams
4.2 Maximal and Minimal elements, Greatest and Least Elements
4.3 LUB and GLB
4.4 Lattices & Isomorphic lattices
4.5 Properties of lattices and their proofs
4.6 Bounded and Complemented Lattices
4.7 Distributive lattices & Sub-lattices

Unit V: Recurrence Relations

5.1 Introduction of Recurrence relations
5.2 Linear recurrence relation with constant coefficients
5.3 Homogeneous solutions:

5.3.1 distinct roots,
5.3.2 repeated roots, and
5.3.3 mixture of both distinct and repeated roots

5.4 Particular solutions
5.5 Total solutions
5.6 Generating functions
5.7 Solution by the method of generation functions

Unit VI: Graphs

6.1 Vertex degree
6.2 Bipartite graphs
6.3 Matchings
6.4 Connected Components
6.5 Trees & Spanning trees
6.6 Isomorphisms of graphs
6.7 Eulerian and Hamiltonian paths
6.8 Directed Graphs and DeBruijn graphs

Reading List

Essential Reading

Johnsonbaugh, R. (2018). *Discrete Mathematics* (8th ed.). New York**:** Pearson Education Inc.

Liu, C.L. (2011). *Elements of Discrete Mathematics* (4th ed.). New Delhi: McGraw-Hill.

Tremblay, J.P., & Manohar, R. (2007). *Discrete Mathematical Structures with Applications to Computer Science*. New Delhi: Tata McGraw Hill.

W. Stein. (2008). Elementary Number Theory: Primes, Congruences, and Secrets, Springer-Verlag.

White, R, T. & Ray, A. T. (2021). Practical Discrete Mathematics: Discover math principles that fuel algorithms for computer science and machine learning with Python. Packt Publishing.

Additional Reading

Cormen, T. H., Leiserson, C. E., & Rivest., R. L. (2001). *Introduction to Algorithms* (2nd ed). New Delhi: Prentice-Hall India.

Fletcher, N.R. (1985). *Discrete Structures: an introduction to mathematics for computer science*. New Jersey: Prentice-Hall Inc.

Kolman, B., Busby, R, C., & Ross, S. (1996). *Discrete Mathematical Structures*. (3rd ed.). New Jersey: Prentice-Hall Inc.

Lipschutz, S., & Lipson., M (2012). *2000 Solved Problems in Discrete Mathematics. (*2nd ed). New Delhi: TMH.

Rosen, K.H. (2011). *Discrete Mathematics and its Applications* (7th ed.). New Delhi: Tata McGraw Hill.

Date: Jan, 2022

DST106 Probability and Distributions

Module code & title: DST106 Probability and Distributions
Programme : BSc in Data Science
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

The module explores the basic concepts of modern probability theory and its applications for decision making in various fields. This module will demonstrate how distribution theory offer useful techniques for quantifying the uncertainties in economics business and so on. The module will also enable students to use calculators and tables to perform simple statistical analyses for small samples and use Python to perform simple and sophisticated analyses for large samples.

Learning Outcomes

On completion of the module, students will be able to:

1. Calculate the probabilities by applying probability laws and theoretical results.
2. Identify an appropriate probability distribution for a given discrete or continuous random variable.
3. Compute probabilities using the properties of an identified probability distribution.
4. Use transform techniques to generate data from various distributions.
5. Derive the probability density function of transformed random variables.
6. Compute the expected value and variance for a distribution.
7. Interpret the area under the density curve for a continuous distribution.
8. Apply results from large-sample theory and the Central Limit Theorem to approximate a sampling distribution.
9. Develop test statistics and test criteria.
10. Implement basic simulations methods using Python to investigate probabilistic scenarios.

Learning and Teaching Approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	3	**60**
	Laboratory Sessions	1	
Independent Study	Written Assignment	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. Weekly homework: (10%)

Weekly problem set will be given to an individual student. These assignments will be due in class the following week. Solutions of the problem sets will be provided by collecting the best answers submitted by the students. Average score of only ten problem sets will be considered after dropping the lowest score. The solution will be assessed using following criteria:

- Use of right methods 3%
- Critical Analysis of the problem 4%
- Correctness of the final answer 3%

B. Multiple Choice Quizzes: (10%)

Two multiple choice VLE quizzes will be conducted for a duration of one and half hour. First quiz will be conducted during first half of the semester after covering 50% of the subject matter and second quiz will be conducted towards the end of the semester covering remaining 50% of the subject matter.

C. Laboratory: (20%)

- **Laboratory report – 5%**

  Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Laboratory Exam – 15%**

A lab exam for the duration of 3 hour will be conducted towards the end of a semester after the last day of regular teaching.

D. Mid-Semester Examination: The portion of Final Marks: (20%)

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 20% of the scores will be taken as final score.

E. Semester-end Examination: Portion of Final Marks: (40%)

Semester end Examination for the duration of two and half hours will be for 50 marks. All the units of the subject matter will be included for the examination

Overview of assessment approaches

Areas of assignment		Quantity	Weighting
A. Homework		10	10%
B. Quiz		2	10%
C. Laboratory	Laboratory Report	10	5%
	Laboratory Exam	1	15%
D. Mid-Semester Examination		1	20%
E. Semester-end Examination		1	40%

Pre-requisite: PLT102 Programming in Python

Subject Matter

Unit I: Probability Theory

1.1 Probability concept
- 1.1.1. Random experiment, sample space, event
- 1.1.2. Classical definition, axiomatic definition and relative frequency definition of probability
- 1.1.3. Concept of probability measure
- 1.1.4. Addition and multiplication theorem
- 1.1.5. Conditional probability and Bayes Theorem-numerical problems.

1.2 Random Variables
- 1.2.1. Definition of probability distribution of a random variable
- 1.2.2. Expectation of random variables
- 1.2.3. Definition of discrete and continuous random variables
- 1.2.4. Probability mass function
- 1.2.5. Probability density function and (cumulative) distribution function and their properties
- 1.2.6. Expectation of random variables and its properties

Unit II: Standard discrete distributions

- 2.1. Point distribution and their properties, moment generating functions, characteristic functions
- 2.2. Power series and their properties, moment generating functions, characteristic functions
- 2.3. Discrete Uniform and their properties, moment generating functions, characteristic functions

2.4. Binomial and their properties, moment generating functions, characteristic functions
2.5. Poisson and their properties, moment generating functions, characteristic functions
2.6. Geometric and their properties, moment generating functions, characteristic functions
2.7. Hypergeometric and their properties, moment generating functions, characteristic functions
2.8. Multinomial and their properties, moment generating functions, characteristic functions
2.9. Negative Binomial distributions and their properties, moment generating functions, characteristic functions for the above distributions

Unit III: Standard continuous distributions

3.1 Definition, mean, variance, properties, moment generating function, and characteristic function of:
- 3.1.1. Uniform
- 3.1.2. Exponential
- 3.1.3. Gamma
- 3.1.4. Bet a
- 3.1.5. Normal
- 3.1.6. Lognormal distributions

Unit IV: Limit Laws

4.1 Chebyshev's inequality: Statement and proof
4.2 Cauchy-Schwarz inequality: Statement and proof
4.3 Markov inequality: Statement and proof
4.4 Jensen inequality: Statement and proof
4.5 convergence in probability: Statement and proof
4.6 Weak Law of Large Numbers (W.L.L.N.): Statement and proof
4.7 Definition of almost sure convergence
4.8 Strong Law of Large Numbers (S.L.L.N.): Statement, proof and applications

Unit V: Convergence of distributions

5.1 Binomial to Poisson
- 5.1.1 Deriving Poisson distribution from Binomial distribution

5.2 De-moivre's - Lajilace theorems
- 5.2.1 State and prove
- 5.2.2 Properties

5.3 Central Limit Theorem (CLT) due to Lindeberg
- 5.3.1 Necessary and sufficient conditions with proof by Lindeberg
- 5.3.2 Dependent Lindeberg CLT
- 5.3.3 Properties

5.4 Levy for iid Statements of continuity theorem
- 5.4.1 Statement and Proof
- 5.4.2 Connection of convergence in distribution

5.5 Uniqueness theorem and inversion theorem.
- 5.5.1 Statement and Proof
- 5.5.2 Examples

Unit VI: Hypothesis Testing

6.1 Hypothesis testing and p-values
6.2 One-tailed and two-tailed test
6.3 Chi-Square test
6.4 Z-statistics and t-statistics
6.5 Large sample confidence interval for a population mean
6.6 Large sample confidence interval for a population proportion
6.7 Determining the sample size
6.8 Large sample hypothesis test for a population mean
6.9 Observed significance levels
6.10 Large sample hypothesis test for a population proportion
6.11 Large sample inferences concerning the difference in two population means based on independent samples
6.12 Large sample inferences concerning the difference in two population proportions
6.13 Small sample inferences about a population mean
6.14 Small sample inferences concerning the difference in two population means based on independent samples

List of Laboratory Sessions

1. Calculate the simple probability of all types of experiments
2. Calculate the following probabilities:
    2.1. Probability that a normal random variable with mean x and variance y lies between z_1 and z_2
    2.2. Probability that in x tosses of a fair coin the head comes up
    2.3. Estimate the probability of getting X heads over four independent flips.
    2.4. A random variable X has Poisson distribution with mean y. Find the probability that x is less than z
3. Constructing a probability distribution for random variables
4. Calculate the value of the cumulative distribution function at (or the probability to the left of) a given number.
5. Plotting pmf and pdf of a distribution
6. Fitting of distributions and testing goodness of fit for binomial, Poisson Negative Binomial and Normal
7. Drawing random samples of size not exceeding 25 from (1) Binomial (2) Poisson (3) Uniform distribution, (4) Cauchy distribution (5) Normal distribution and (6) Exponential with known mean and variance, using random number tables.
8. Numerical problems involving derivation of marginal and conditional probability density function and related measures of moments.
9. Convergence and asymptotic results
10. Testing of Hypothesis for both large and small samples (using data from UCI Machine Learning Repository)
    10.1. Hypothesis testing in Economics
    10.2. Hypothesis testing in Finance
    10.3. Hypothesis testing in Business
    10.4. Hypothesis testing in medical research
11. Construction of confidence intervals for the parameter or statistics

Reading List

Essential Reading

Mood, A. M., Graybill, F. A., & Boes, D. C. (2017). *Introduction to the theory of statistics*. Chennai: McGraw Hill Education (India) Pvt. Ltd.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2020). *Introduction to mathematical statistics*. Harlow, Essex: Pearson.

Goon, A. M., Gupta, M. K., & Dasgupta, B. (1998). *An outline of statistical theory*. Calcutta: World Press Private.

Rohatgi, V. K., & Ehsanes, S. A. (2015). *An introduction to probability theory and mathematical statistics*. Hoboken, NJ: John Wiley & Sons.

Additional Reading

Hoel, P. G. (2005). *Introduction to Mathematical Statistics*. New York: John Wiley & Sons.

Spiegel, M. R. (1997). *Schaum's outline of theory and problems of probability and statistics*. New York etc.: McGraw-Hill.

Lipschutz, S., & Lipson, M. (2000). *Schaum's outline of theory and problems of probability*. New York: McGraw-Hill.

Fisz, M. (2018). Probability Theory and Mathematical Statistics. Dehli: Scientific International.

Chung, K. L. (2003). *Elementary probability theory with stochastic processes*. New York: Springer.Feller, W. (2010). *An introduction to probability theory and its applications*. New York: Wiley.

Date: March, 2022

ADS104 Data Structures

Module Code and Title: ADS104 Data Structures
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Sangay Thinley

General Objective
This module will familiarize the students with basic data structures and their use in fundamental algorithms. This module aims to equip the students with the knowledge of the organisation of data in computer memory for the purpose of programming. Python programming language will be used as a tool for the delivery of this module.

Learning Outcomes
On completion of the module, students will be able to:

1. Explain the need of data structures as a programmer
2. Use static and dynamic lists
3. Select appropriate data structure to implement an algorithm
4. Solve real world problems using stack and queue
5. Identify and choose appropriate data structure for the given application
6. Implement sorting and searching algorithm
7. Use string manipulation functions
8. Solve network-based problems using graph

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	90
	Laboratory Session	3	
Independent study	Self-study	2	30
Total			120

Assessment Approach

A. Assignment: (10%)
Individual students are required to do the assignment during the first half of the semester after completion of 30 percent of the syllabus content. Set of questions based on the topics covered will be provided to the students. Students will have to exhibit their understanding of the concepts through their solutions to the questions. Assessment of the assignment will be done based on the following criteria:

4. Presentation (2%)
5. Correctness of result (3%)
6. Logic (5%)

B. Lab Assessment: (40%)

Lab Assessment will be assessed based on the following components:

- Lab Assignments: 15%
- Lab Examination: 20%
- Viva Voce: 5% (Lab Assignment: 2%, Lab Examination: 3%)

Lab Assignments (a total of 9) will be based on the concepts taught in the theory class. Students will be provided a set of programming questions every week covering the concepts taught on each unit. The students will have to submit the solutions at the end of the lab session through the VLE.

The following rubrics will be used to assess the set of programming questions. Each set will be assessed out of 100 marks and converted to 1.5%.

Program (100)	**(Excellent)** **100-81%**	**(Good)** **80-61%**	**(Satisfactory)** **(60-41% of the points)**	**(Unsatisfactory)** **(<=40% of the points)**
Requirements and Delivery **(50 points)**	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution **(20 points)**	• Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic **(10 points)**	• Program is logically well designed.	• Program has slight logic errors that do no significantly affect the results	• Program has significant logic errors.	• Program is incorrect.
Coding Standards **(10 points)**	• Includes name, date, and assignment title. • Excellent use of white space.	• Includes name, date, and assignment title. • Good use of white space. • Organized work.	• Includes name, date, and assignment title. White space makes program fairly easy to read.	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines).

	• Creatively organized work. • Excellent use of variables (no global variables, unambiguous naming).	• Good use of variables (no global variables, unambiguous naming)	• Organized work. • Good use of variables (few global variables, unambiguous naming).	• Disorganized and messy. • Poor use of variables (many global variables, ambiguous naming).
Documentation **(10 points)**	• Clearly and effectively documented including descriptions of all class variables. • Specific purpose noted for each function, control structure, input requirements, and output results.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function and control structure.	• Basic documentation has been completed including descriptions of all class variables. • Purpose is noted for each function.	• Very limited or no documentation included. • Documentation does not help the reader understand the code.

Lab examination will be conducted after the completion of the module and before semester-end examination and will cover all the units. Each student will be given two questions and will be graded out of 100 marks based on the rubric mentioned above. The marks obtained will then be converted out of 20%.

Viva Voce for the lab assignments will be conducted once to assess the students' understanding of the lab assignments. Viva Voce will also be conducted for the lab examination. The following rubrics will be used for viva voce. Each viva voce will be assessed out of 100 marks and converted to 2% for lab assignments and 3% for lab examination respectively.

Program (100)	**(Excellent)** **85-100%**	**(Good)** **61-84%**	**(Satisfactory)** **(<=60% of the points)**
Knowledge in programming **(50 points)**	Demonstrates deep knowledge, answer the questions with explanation and elaboration.	Adequate knowledge of most topics, answer the questions but fails to elaborate.	Superficial knowledge of topic, only able to answer basic questions.
Problem-solving ability **(30 points)**	Efficient mapping of theory concepts with practical	Moderate mapping of theory concepts with practical	Improper mapping of theory concepts with practical

	problem solving approaches.	problem solving approaches.	problem solving approaches.
Critical Thinking **(20 points)**	Can predict and defend problem outcomes.	Approximately predicts and defends problem outcomes.	Is unable to predict problem outcomes for the given input data set.

C. Semester-end examination: Portion of Final Marks: **(50%)**

Semester-end Examination for the duration of two and half hours will be conducted covering all units of the module. The exam will be conducted out of 50.

Overview of assessment approaches and weighting:

Areas of Assessment		**Quantity**	**Weighting**
A. Assignment		1	10%
B. Lab Assessment	Lab Assignment	9	15%
	Lab Examination	1	20%
	Viva Voce	2	5%
C. Semester-end Examination		1	50%
Total			100%

Pre-requisites: PLT102 Programming in Python

Subject Matter

Unit I: Arrays and Introduction To Data Structures

1.1. Abstract Data type
- 1.1.1. Definitions, examples and scope

1.2. Array/List in Python
- 1.2.1. Storage Structure
- 1.2.2. Accessing elements
- 1.2.3. Implementation examples
- 1.2.4. Multi-Dimensional Array
- 1.2.5. Limitations

1.3. Dynamically allocated array
- 1.3.1. Allocation Scheme
- 1.3.2. Implementation examples

1.4. Searching and sorting algorithm
- 1.4.1. Linear and Binary Search
- 1.4.2. Bubble Sort, Insertion Sort, Selection Sort, Heap sort

1.5. Polynomial representation and its evaluation
1.6. String manipulation
    1.6.1. Implementation of String Operations (Functions)

Unit II: Stack and Queue

2.1. Stack operation's implementation using dynamic array
    2.1.1. Access Mechanism
    2.1.2. Push, Pop, Top and is Empty operations
    2.1.3. Swapping elements between stacks
    2.1.4. Implementations of real world examples
2.2. Queue operation's implementation using dynamic array
    2.2.1. Access mechanism
    2.2.2. Enqueue, Dequeue and is Empty Operations
    2.2.3. Implementation of real world examples
2.3. Circular Queue implementation
    2.3.1. Access Mechanism
    2.3.2. Real world examples
2.4. Evaluating expression
    2.4.1. Infix, prefix and postfix expression and their inter-conversions

Unit III: Linked-Lists

3.1. Concept of pointers
    3.1.1. Concepts, Usability and Implementations and scope
    3.1.2. Access mechanism
    3.1.3. Double pointers
    3.1.4. Pros & Cons
3.2. Implementation of stack and queue using dynamic linked list
3.3. Doubly Linked List
    3.3.1. Concept, Usability and Implementations and scope
    3.3.2. Access mechanism
    3.3.3. Pros & Cons

Unit IV: Trees and Graphs

4.1. Introduction and Terminology
    4.1.1. Definition and examples
4.2. Representation of tree and Binary Tree
    4.2.1. Definitions and examples
    4.2.2. Physical and logical representation
    4.2.3. Complete and Full binary tree
    4.2.4. Relationship between number of Nodes and Height of binary tree
    4.2.5. m-ray tree (Definitions)
    4.2.6. Applications examples

4.3. Properties of binary tree and binary search tree
4.4. Insertion and deletion into binary search tree

4.5. AVL Tree
    4.5.1. Properties and scope
    4.5.2. Different Rotations
    4.5.3. Insertion and deletion
4.6. Graph traversals Techniques.
    4.6.1. Preorder, Post-order and In-order traversal
    4.6.2. Depth First Search (DFS) and Breadth First Search (BFS) traversal

List of Laboratory Sessions (Using Python)

1. **Strings:**
   a) Write a menu driven program to compare, concatenate, copy strings and find the length of a string.
   b) Write a menu driven program to find the index of a pattern in a given string and to extract a substring.

2. **Lists**
   a) Write a program to insert and delete an element(s) in one dimensional list.
   b) Write a program to insert and delete an element(s) in two dimensional list.

3. **Searching Techniques:**
   a) Write a program to implement Linear Search with sentinels.
   b) Write a program to implement Binary Search using recursion.

4. **Sorting techniques:**
   a) Write a menu driven program to implement insertion sort
   b) Write a menu driven program to implement selection sort.
   c) Write a menu driven program to implement quick sort using recursion
   d) Write a menu driven program to implement merge sort using recursion.

5. **Singly linked list:**
   Write a menu driven program to implement singly linked lists creation, insertion and deletion

6. **Stack:**
   Write a menu driven program to implement different operations on a stack using an array and linked list.

7. **Queue:**
   Write a menu driven program to implement different operations on a queue using an array and linked list.

8. **Binary Search Trees:**
   Write a menu driven program to create a binary search tree and to perform Insertion and different types of traversal.

9. **Graphs:**
   - Write a menu driven program to implement breadth first search (bfs)

- Write a menu driven program to implement depth first search (dfs)

Reading List
Essential Reading

Bradley N. Miller, David L. Ranum. (2011*) Problem Solving with Algorithms and Data Structures Using Python [2nd ed]*. Franklin, Beedle & Associate.

Goodrich, T. M., Tamassia, R., & Goldwasher, H, M. (2016). *Data Structures and Algorithms in Python.* Wiley.

Additional Reading

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2017). *Introduction to algorithms*. MIT Press.

Tremblay J.P and Sorenson P.G. (2002). *An Introduction to Data Structures with Applications* 2nd edition. TMH.

Date: March, 2022

ALG108 Matrix Analysis and Vector Spaces

Module Code & Title: ALG108 Matrix Analysis and Vector Spaces
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Thinley Namgyel

General objective

Important concepts in mathematics like matrices and vector spaces that can be used for storing and analyzing data will be taught in this module. The concepts learned through this module will help the students for data processing, data transformation and data evaluation. Therefore, the module will introduce matrices and their operations followed by use of matrices in various areas of Linear Algebra like solving system of linear equations, inner product spaces and eigenvalues and eigenvectors. To understand the concepts better, explanation will be done with the use of Numpy package in Python. GeoGebra software package will also be used in demonstrating the concepts taught.

Learning outcomes

On completion of the module, students will be able to:

1. Use matrices to store various information and draw meaning out of it.
2. Identify different types of matrices.
3. Apply matrix operations and use their properties in various situations.
4. Solve system of linear equations using Gaussian elimination method and Gauss-Jordan elimination method.
5. Explain the meaning of consistent system of linear equations with three variables.
6. Define vector spaces with various examples.
7. Find basis and dimension of vector spaces.
8. Find rank and nullity of a given matrix.
9. Compute different powers of the given matrix.
10. Determine orthogonal and orthonormal basis of a given matrix using Gram-Schmidt process.
11. Compute powers of matrices using orthogonal diagonalization method.
12. Compute different geometrical structures in an inner product spaces like norm of a vector and angle between vectors.
13. Perform matrix operations (addition, subtraction, multiplication, transpose, determinant, trace, power, inverse) using Numpy package in Python.
14. Solve system of linear equations using Numpy package
15. Use GeoGebra software to plot the given linear equations and check for consistency or inconsistency of the solution of system of linear equations.

Learning and Teaching Approach:

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Tutorial	1	
Independent study	Written assignment	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach:

A: Assignment: 15%
The assignment will be completed during the first half of the semester. The main objective of this assignment is to verify whether students have understood the concepts taught in the class. Set of questions based on the topics covered will be provided to the students. Students will have to exhibit their understanding of the concepts through their solutions to the questions. It is an individual work and each student will have to make their submission. In order to avoid plagiarism, different sets of questions with same difficulty level will be given to the students. The assignment will be evaluated based on the following criteria and the marks achieved out of 100 will be converted to 10%.

- Correctness of the solutions: 40%
- Logical flow in the process: 20%
- Uniqueness of the solution: 20%
- Use of appropriate terms and symbols: 20%

B: Lab exam on the use of software packages: 10%
The tutor will demonstrate the use Numpy package in Python and GeoGebra software in delivering Unit I and Unit II of the subject matter. To ensure the use of software packages, a laboratory exam with weighting of 10% will be carried out towards the end of unit II. Students will be assessed on whether they can use the software packages to write simple commands and programmes to compute solutions to the given problems. Set of questions covering subject matter from Unit I and Unit II will be provided to the students where they will be required to use Python and GeoGebra software to compute the solutions. This part of the assessment will verify the achievement of learning outcome 13, 14 and 15. The lab exam will be evaluated based on the following criteria and the mark will be converted to 10%.

- Use of appropriate commands 25%
- Correctness of solutions 25%
- Analysis of solutions 50%

C: Mid-Semester Examination: 15%
Mid-semester Examination of 2.5 hours duration will be conducted in the mid of semester with at least 60% of the subject matter being covered.

D: Online Quiz: 10%
To encourage the students to use Virtual Learning Environment (VLE) and to emphasize the importance of online assessment, students will be made to take part in online quiz through VLE. The duration of the quiz will be for 1 hour and it will be conducted either during one of the classes or after class hour, depending on the availability of computer labs. The quiz will be conducted towards the end of the semester and will cover questions from all the topics included in the subject matter. Most of the questions will be MCQs or short answer questions and the marks will be reflected instantly by VLE.

E: Semester-end Examination: Portion of final marks: (50%)
Semester end Examination of 50% will be conducted for the duration of 2.5 hours. The format and pattern of the question paper will be as per the academic norm of the College. The blue print of the question paper and answer key will also be prepared along with the question paper.

Overview of Assessment Approaches and weighting

Areas of assessments	Quantity	Weighting
A. Assignment	1	15%
B. Lab exam on use of software packages	1	10%
C. Mid-Semester Examination	1	15%
D. Online Quiz	1	10%
E. Semester-end Examination	1	50%

Prerequisite: None

Subject Matter

Unit I: Introduction and Type of Matrices

(The treatment of unit I will be at an introductory level only.)

1.1 Introduction to Matrices: Use and scope of Matrices with real time examples
1.2 Type of Matrices:
- 1.2.1 Row matrix
- 1.2.2 Column matrix
- 1.2.3 Zero Matrix
- 1.2.4 Rectangular Matrix
- 1.2.5 Square Matrices

1.2.6 Diagonal Matrices
1.2.7 Scalar Matrix
1.2.8 Identity Matrix
1.2.9 Triangular matrix
1.2.10 Symmetric Matrices
1.2.11 Singular Matrix

1.3 Operations on Matrices (Only up to 3 x 3 matrices will be used for manual computations however, concepts will be extended to n x n matrices using software packages)
1.4 Equality of Matrices
1.5 Matrix Addition and associated properties
1.6 Matrix Subtraction
1.7 Scalar Multiplication
1.8 Matrix Multiplication and associated properties
1.9 Transpose of a Matrix and associated properties
1.10 Powers of a Matrix
1.11 Determinant of a square matrix
1.11.1 Properties of determinants (reflection property, all zero property, proportionality property, switching property, scalar multiple property, sum property, property of invariance, Triangle property)
1.11.2 Singular and non-singular matrices
1.12 Finding inverse of a matrix using cofactor method

Unit II: System of Linear Equations

2.1 Matrix representation of system of linear equations (homogeneous and non-homogeneous)
2.2 Substitution method for solving system of linear equation
2.3 Gaussian elimination method for solving system of linear equations
2.4 Gauss-Jordan elimination method for solving system of linear equations
2.5 Consistency (exactly one solution, infinitely many solutions) and inconsistency of system of linear equations (both algebraic and geometrical meaning)
2.6 Elementary Matrices and method for finding the inverse of a matrix.

Unit III: Vector Spaces

3.1 Definition and examples of vector spaces
3.2 Vector subspaces and examples
3.3 Linear combination, Linear Span, Linear independence
3.4 Basis and Dimension of Vector spaces
3.5 Row Space, Column Space and Null Space
3.6 Rank and Nullity of a Matrix

Unit IV: Inner Product Spaces

4.1 Definition and examples of Inner product spaces
4.2 Norm of a vector
4.3 Angle and orthogonality
4.4 Cauchy-Schwarz Inequality

4.5 Orthonormal basis: Gram-Schmidt process
4.6 QR-decomposition

Unit V: Eigen Values and Eigen Vectors

5.1 Definition and examples
5.2 Diagonalization
5.3 Computing powers of a Matrix
5.4 Orthogonal Diagonalization

Reading List

Essential Reading

Anton. H. & Rorres. C. (2005). *Elementary Linear Algebra* (9 ed.). USA: John Wiley & Sons.
Biswas, S. (2012). Textbook of Matrix Algebra (3 ed.). New Delhi: PHI Learning Pvt. Ltd.
Lang, S. (1988). *Introduction to Linear Algebra* (2 ed.). New York: Springer Verlag Inc.

Additional Reading

Cohen, A.M, Cuypers, H., & Sterk, H. (1999). *Algebra Interactive*. New Yok: Springer Verlag Inc.
Hoffman, K., & Kunze, R. (2006). *Linear Algebra* (2 ed.) New Delhi: Dorly Kindersly (India) Pvt. Ltd.
Kolman, B., & Hill, D.R. (2006). *Introduction to Linear Algebra with applications* (7 ed.) New Delhi: Dorly Kindersly (India) Pvt. Ltd.
Kreyszig, E. (2005). *Advanced Engineering Mathematics* (8 ed.) Singapore: John Whiley & Sons (Asia) Pvt. Ltd.

Date: March, 2022

ACS101 Academic Skills

Module Code and Title: ACS101 Academic Skills
Programme: University-wide Module
Credit Value: 12
Module Tutor: Faculty member from English
Module Coordinator: Wangmo

General Objective

This module aims to develop the knowledge and understanding of a range of academic skills required for study at university level. The module will focus on the development of academic writing, oral presentation, as well as listening skills to enable students to communicate effectively in both spoken and written forms. The module will enhance their learning throughout their studies at university and beyond, through close reading, discussions and critiquing of academic texts. Further, it will also enhance students' capacity to critically reflect on their own learning.

Learning Outcomes

On completion of the module, students will be able to:

1. use effective note taking skills to extract relevant information from a range of academic texts.
2. lead and participate productively in group situations.
3. apply features of academic writing in academic discourses.
4. apply learned strategies to avoid the consequences of academic dishonesty.
5. employ a range of strategies and techniques to read academic texts.
6. demonstrate information retrieval and analysis skills by identifying, assessing and using appropriate sources i.e. author, publisher or website.
7. identify the content, viewpoint and relevance of articles and reports on a wide range of topics.
8. write academic papers using a process approach: planning, drafting, eliciting feedback and revising, following consistent academic standards.
9. construct a coherent and substantiated argument that integrates appropriate source material, and uses appropriate research and APA referencing conventions in clear and correct language in the form of an essay.
10. produce academic essays using process approach: planning, drafting, eliciting feedback and revising using appropriate terminology and a consistent academic style.
11. plan, organise and deliver a clear, well-structured academic oral presentation.

Teaching and Learning Approach

Tutors will employ an interactive, student-centred approach, integrating language and critical thinking skills using the following strategies: demonstrations/modelling, lab exercises and activities, group work (discussions, problem-solving activities, collaborative and individual tasks, peer feedback and debates), academic essay writing (process learning with diagnosis, feedback and remediation), oral presentation, portfolio, independent study and VLE discussions over the 120 credit hours.

Approach	Hours per week	Total credit hours
Demonstrations/Modelling	1	15
Lab exercises and group works	2	30
Academic essay writing	1	15
Oral presentation	0.5	7.5
Portfolio	1.5	22.5

Independent study and VLE discussions	2	30
Total		**120**

Assessment Approach

Since the module is entirely assessed through Continuous Assessment, a student must complete all five components of the assessment outlined below and get an aggregate mark of 50% in order to pass. Assessment will be carried out on a continuous basis through the following tasks:

A. Academic Essay: Portion of the Final Mark (30%)

Students have to write one 800 to 1000-word academic essay following the rules of academic standards, essay writing, APA referencing and mechanics of language in order to practice and develop academic writing skills at the university level. The academic essay will be written in three drafts; the first draft to be peer reviewed, the second and final essay to be assessed based on the following criteria:

Second Draft (10%)	**Final Draft (20%)**
Content (4%)	Content (10%) (*Introduction-3%, Body-5%, Conclusion-2%)*
Language (2%)	Language (4%)
References (2%)	References (4%)
Format (2%)	Format (2%)

B. Presentation: Portion of the Final Mark (15%)

Each student has to make one 5-7 minute presentation. This will help them acquire the skills necessary for carrying out effective oral presentations during the course of their university study. The students can choose one presentation topic related to their Academic Skills module, programme or an evidence-based subject that interests them for this task. The presentations will be assessed based on the following criteria:

Greetings (3%)

- *Introduction*
- *Topic*
- *Overview*

Content (4%)

- *Clarity*
- *Discussion*
- *Evidence*
- *Coherence*

Delivery (5%)

- *Pronunciation*
- *Grammar*
- *Tone and pitch*

- *Body language*

Visual Aids (2%)

- *Effectiveness*
- *Relevance*

Time Management (1%)

- *Coverage*
- *Conclusion*

C. Portfolio: Portion of the Final Mark (25%)

Each student has to maintain a portfolio containing series of exercises from both within and outside the class. This is to ensure the development of independent study, skills and ability to work with other students. The portfolio will be assessed based on the following:

- Organisation (5%)
- Class Work (8%)
- Class Notes (5%)
- Homework (7%).

D. Class Test: Portion of the Final Mark (20%)

Students have to write one class test towards the end of week seven. The test will mainly focus on referencing skills.

E. VLE Discussion: Portion of the Final Mark (10%)

Students will contribute to VLE discussions on selected topics assigned by tutors.

- Frequency (5%)
- Relevance (5%)

An overview of the assessment approaches and weighting:

Areas of assessment	Quantity	Weighting
A. Academic essay	1	30%
B. Oral presentation	1	15%
C. Portfolio	1	25%
D. Class test	1	20%
E. VLE discussion	2-5	10%

Pre-requisite: None

Subject Matter

Unit I: Academic Standards

1.1. Definition
1.2. Purpose of Academic Activities
1.3. Ethics and Integrity

Unit II: Note-taking

2.1. Basics of note-taking
  2.1.1. Storing information during lecture sessions
2.2. Types of notes and strategies

2.2.1. Pattern Notes or Mind Maps
2.2.2. The Cornell Method
2.2.3. The Outlining Method
2.2.4. Symbol and Abbreviation Method

2.3. Listening and note-taking
2.3.1. Practicing Listening with the partners
2.3.2. Listening to BBC service podcasts
2.3.3. Listening to IELTS test samples

Unit III: Academic Writing

3.1 Academic Writing
3.1.1 Definition
3.1.2 Importance of academic writing
3.1.3 Identifying various academic texts
3.1.4 Applying academic features in writing for academic purposes

3.2 Features of academic writing
3.2.1 Formality
3.2.2 Structure
3.2.3 Logic
3.2.4 Evidence and sources
3.2.5 Objectivity
3.2.6 Precision

3.3 Types of academic writing
3.3.1 Essays
3.3.2 Reports
3.3.3 Exam responses
3.3.4 Academic assignments
3.3.5 Proposals (Research and project)

3.4 Academic argument
3.4.1 Definition
3.4.2 Distinction between academic argument and everyday argument
3.4.3 Facts, opinions and beliefs

Unit IV: Referencing Techniques and APA format

4.1 Types of referencing styles
4.1.1 Documentary note styles
4.1.2 Parenthetical styles or author-date styles
4.1.3 Numbered styles
4.1.4 Why and when to cite

4.2 Introduction to using source materials
4.2.1 Defining sources
4.2.2 Critical evaluation of resources

4.3 Using source materials for in-text citation
4.3.1 Direct and Indirect/Reported voice

4.4 Making end-text/reference lists
4.4.1 Writing references for books, newspapers, websites and scholarly journals

4.5 Referencing and academic integrity
4.5.1 Understanding plagiarism and its consequences
4.5.2 Maintenance of academic standards
4.5.3 Honesty and rigor in academic writing and publishing
4.5.4 Following academic ethics

Unit V: Academic Essay Writing

5.1 Writing Process
- 5.1.1 Pre-writing, Drafting, Revising, Editing and Publishing

5.2 Understanding Written Assignments
- 5.2.1 Instruction words
- 5.2.2 Content words
- 5.2.3 BUG method

5.3 Academic Essay
- 5.3.1 Purpose and features of academic essays

5.4 Essay Format/Structure
- 5.4.1 Introduction- Opening statement, background information and thesis statement
- 5.4.2 Body paragraphs
- 5.4.3 Conclusion

Unit VI: Academic Reading

6.1. Text features and organisation
- 6.1.1. Textual Features
- 6.1.2. Graphic Aids
- 6.1.3. Informational Aids
- 6.1.4. Organisational Aids

6.2. Reading Techniques
- 6.2.1. Skimming
- 6.2.2. Scanning
- 6.2.3. SQ3R

6.3. Introduction to Using Source Materials
- 6.3.1. Locating, evaluating and selecting information
- 6.3.2. Internet Source- Web endings

6.4. Summarizing and Paraphrasing academic texts

6.5. Critical reading (author viewpoints/biases, reading for detail)

Unit VII: Oral Presentations

7.1 Basics of oral presentation
- 7.1.1 Definition and Examples
- 7.1.2 Tips to Overcome Anxiety in Oral Presentation (Controlling Nervousness, Controlling Physical Nervousness, Capitalizing on the Law of Attraction)
- 7.1.3 Organising the Content (Introduction, Body, Conclusion)

7.2 Strategies for delivering an effective presentation
- 7.2.1 Signposting (Introducing topic of presentation, outlining the structure of presentation, indicating the start of new section, concluding)
- 7.2.2 Using Visual Aids
- 7.2.3 Sense of Humour
- 7.2.4 Body Language
- 7.2.5 Tone and Pitch

Reading List

Essential Reading

American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: Author.

Department of Academic Affairs. (2019). Students' materials for academic skills. Thimphu: Royal University of Bhutan.

Department of Academic Affairs. (2019). Tutors' materials for academic skills. Thimphu: Royal University of Bhutan.

Additional Reading

Bailey, S. (2011). Academic writing: A handbook for international students (3rd ed.). Abingdon, Oxford: Routledge.

Butler, L. (2007). Fundamentals of academic writing. New York, NY: Pearson Longman.

Gillet, A. (2013, January 15). UEFAP (Using English for academic purposes): A guide for students in higher education. Retrieved from http://www.uefap.com

Gillet, A., Hammond, A., & Martala, M. (2009). Inside track successful academic writing. England: Pearson Education.

Hogue, A. (2007). First steps in academic writing. New York: Pearson Education ESL.

Oshima, A., & Hogue, A. (2005). Writing academic English (4th ed.). White Plains, NY: Pearson Education.

Oshima, A., & Hogue, A. (2006). Introduction to academic writing (3rd ed.). New York: Pearson Longman.

Ramsey-Fowler, H., & Aaron, J. E. (2010). The little brown handbook (11th ed.). New York, NY: Pearson Longman.

Date: 29 June 2018

Year 2 Semester I

PLT203 R Programming for Data Analysis

Module Code & Title: PLT203 R Programming for Data Analysis
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

The module will enable students to use R for data analysis. The students will be introduced to R programming language and learn the basic syntax, semantics, grammars, and vocabulary of R programming language to aid in data analysis. Further, they will learn how to expand R by importing and configuring software packages necessary for statistical analysis and data visualization.

Learning Outcomes

On completion of the module, students will be able to:

1. Use R syntax and semantics in writing programs in R.
2. Apply the concepts of objects and assignment.
3. Demonstrate the concepts of vector and data types.
4. Load a Workspace containing an R data frame, edit the dataset, and save the Workspace.
5. Access online resources for R and import new function packages into the R Workspace.
6. Import, review, manipulate and summarize data-sets in R.
7. Explore data-sets to create testable hypotheses and identify appropriate statistical tests.
8. Perform appropriate statistical tests using R.
9. Create and edit visualizations with R.

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**75**
	Lab session	2	
Independent Study	Self-study	3	**45**
Total			**120**

Assessment Approach

A. **Lab Assessment: (35%)**
Lab Assessment will be assessed based on the following components:

- Lab Assignments: 15%.

- Lab Examination: 10%
- Viva Voce: 10% (Lab assignments: 7%, Lab examination: 3%).

Lab Assignments (a total of 10 with one assignment for first three units and one each for remaining units) will be based on the concepts taught in the theory class. Students will be provided a set of programming questions every week covering the concepts taught on each unit. The students will have to submit the solutions at the end of the lab session through the VLE.

The following rubrics will be used to assess the set of programming questions. Each set will be assessed out of 100 marks and converted to 1.5%.

Program (100)	**(Excellent)** **100-81%**	**(Good)** **80-61%**	**(Satisfactory)** **(60-41% of the points)**	**(Unsatisfactory)** **(<=40% of the points)**
Requirements and Delivery **(50 points)**	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution **(20 points)**	• Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic **(10 points)**	• Program is logically well designed.	• Program has slight logic errors that do no significantly affect the results	• Program has significant logic errors.	• Program is incorrect.
Coding Standards **(10 points)**	• Includes name, date, and assignment title. • Excellent use of white space. • Creatively organized work. • Excellent use of variables (no global	• Includes name, date, and assignment title. • Good use of white space. • Organized work. • Good use of variables (no global variables,	• Includes name, date, and assignment title. White space makes program fairly easy to read. • Organized work. • Good use of variables (few	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines). • Disorganized and messy. • Poor use of variables (many

	variables, unambiguous naming).	unambiguous naming)	global variables, unambiguous naming).	global variables, ambiguous naming).
Documentation **(10 points)**	• Clearly and effectively documented including descriptions of all class variables. • Specific purpose noted for each function, control structure, input requirements, and output results.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function and control structure.	• Basic documentation has been completed including descriptions of all class variables. • Purpose is noted for each function.	• Very limited or no documentation included. • Documentation does not help the reader understand the code.

Lab examination will be conducted after the completion of the module and before semester-end examination and will cover all the units. Each student will be given two questions and will be graded out of 100 based on the rubric mentioned above. The marks obtained will then be converted out of 10%.

Viva voce for the lab assignments will be conducted at least two times to assess the students' understanding of the lab assignments. Viva voce will also be conducted for the lab examination. The following rubrics will be used for viva voce. Each viva voce will be assessed out of 100 marks and converted to 7% for lab assignments and 3% for lab examination respectively.

Program (100)	**(Excellent)** **85-100%**	**(Good)** **61-84%**	**(Satisfactory)** **(<=60% of the points)**
Knowledge in programming **(50 points)**	Demonstrates deep knowledge, answer the questions with explanation and elaboration.	Adequate knowledge of most topics, answer the questions but fails to elaborate.	Superficial knowledge of topic, only able to answer basic questions.
Problem-solving ability	Efficient mapping of theory concepts	Moderate mapping of theory concepts	Improper mapping of theory concepts

(30 points)	with practical problem solving approaches.	with practical problem solving approaches.	with practical problem solving approaches.
Critical Thinking **(20 points)**	Can predict and defend problem outcomes.	Approximately predicts and defends problem outcomes.	Is unable to predict problem outcomes for the given input data set.

B. **Mid-Semester Examination: (15%)**
Mid-semester examination will be conducted in the mid of the semester with at least 60% of the subject matters being covered for the duration of 2 hours with maximum marks of 40. The final marks will be converted out of 15.

C. **Semester-end Examination: (50%)**
Semester end Examination for the duration of two and a half hours will be conducted with maximum marks of 50 and final marks will be taken without conversion.

Overview of Assessment Approaches

Areas of Assignments		**Quantity**	**Weighting**
A. Laboratory Assessment	Weekly lab assignments	10	15%
	Lab examination	1	10%
	Viva-voce	1	10%
B. Mid-Semester Examination		1	15%
C. Semester-end Examination		1	50%
Total			100%

Prerequisite: PLT102 Programming in Python

Subject Matter

Unit I: Preliminaries

1.1 The R environment
1.2 R and statistics
1.3 Using R interactively
1.4 Getting help with functions and features
1.5 R commands, case sensitivity, etc.
1.6 Recall and correction of previous commands
1.7 Executing commands from or diverting output to a file

1.8 Data permanency and removing objects

Unit II: Simple manipulations: numbers and vectors

2.1 Vectors and assignment
2.2 Vector arithmetic
2.3 Generating regular sequences
2.4 Logical vectors
2.5 Missing values
2.6 Character vectors
2.7 Index vectors; selecting and modifying subsets of a data set
2.8 Other types of objects:
2.8.1 Ordinal factors
2.8.2 Nominal factors

Unit III: Objects, their modes and attributes

3.1 Intrinsic attributes: mode and length
3.2 Changing the length of an object
3.3 Getting and setting attributes
3.4 The class of an object
3.5 Ordered and unordered factors

Unit IV: Arrays and matrices

4.1 Arrays
4.2 Array indexing. Subsections of an array
4.3 Index matrices
4.4 The array () function
4.5 Mixed vector and array arithmetic.
4.6 Recycling rule
4.7 The outer product of two arrays
4.8 Generalized transpose of an array
4.9 Matrix facilities:
4.9.1 Matrix multiplication
4.9.2 Linear equations and inversion
4.9.3 Eigenvalues and eigenvectors
4.9.4 Singular value decomposition and determinants
4.9.5 Least squares fitting and the QR decomposition)
4.10 Forming partitioned matrices, cbind() and rbind()
4.11 The concatenation function, c(), with arrays
4.12 Frequency tables from factors and grouped data

Unit V: Lists and data frames

5.1 Lists:
5.1.1 Constructing and modifying lists
5.1.2 Concatenating lists
5.2 Data frames:
5.2.1 Making data frames
5.2.2 attach() and detach()
5.2.3 Working with data frames

5.2.4 Attaching arbitrary lists
5.2.5 Managing the search path
5.3 Data cleaning
5.3.1 Renaming column names
5.3.2 Formatting the output
5.3.3 Reshape
5.3.4 Remove empty columns and rows
5.3.5 Remove duplicate records
5.3.6 Imputation method for missing values

Unit VI: Reading data from files

6.1 The read.table() function
6.2 The scan() function
6.3 Accessing built in datasets
6.4 Loading data from other R packages
6.5 Editing data

Unit VII: Grouping, loops and conditional execution

7.1 Grouped expressions
7.2 Control statements
7.3 Conditional execution: if statements
7.4 Repetitive execution: for loops, repeat and while

Unit VIII: Writing your own functions

8.1 Simple examples
8.2 Defining new binary operators
8.3 Named arguments and defaults
8.4 Assignments within functions
8.5 More advanced examples:
8.5.1 Efficiency factors in block designs
8.5.2 Dropping all names in a printed array
8.5.3 Recursive numerical integration
8.6 Scope
8.7 Customizing the environment
8.8 Classes, generic functions and object orientation
8.9 Reusable functions:
8.9.1 Calling function

Unit IX: Graphical Procedures

9.1 High-level plotting commands:
9.1.1 The plot() function
9.1.2 Displaying multivariate data
9.1.3 Display graphics

List of Laboratory Sessions

1. Use the R Commander to input data, then using read.table() directly to achieve the same result

1.1 Determine number rows, columns, missing values of data
1.2 Check the type of data
1.3 Print the name of variables

2 Use the functions mean() and range() to find the mean and range of:

2.1 the numbers 1, 2, . . . , 21
2.2 the sample of 50 random normal values, that can be generated from a normaL distribution with mean 0 and variance 1 using the assignment y <- rnorm(50).
2.3 the columns height and weight in the data frame women. [The datasets package that has this data frame is by default attached when R is started.]
2.4 Repeat (2.2) several times, on each occasion generating a new set of 50 random numbers.

3 Extract the following subsets from the data frame ais: (a) Extract the data for the rowers. (b) Extract the data for the rowers, the netballers and the tennis players. (c) Extract the data for the female basketabllers and rowers.

4 Create a function that will return the sum of 2 integers.

5 Create a function what will return TRUE if a given integer is inside a vector.

6 Create a function that given a data frame will print by screen the name of the column and the class of data it contains (e.g. Variable1 is Numeric).

7 Create the function unique, which given a vector will return a new vector with the elements of the first vector with duplicated elements removed.

8 Create a function that given a vector will print by screen the mean and the standard deviation, it will optionally also print the median.

9 Check the class of each of the columns of the data frame cabbages (MASS). Do side by side plots of HeadWt against Date, for each of the levels of Cult.

Reading List

Essential Reading

Crawley, M.J. (2012). *The R book*. John Whiley & Sons.

Matloff, N. (2011). *The Art of R programming: A tour of statistical software design*. No Starch Press.

Additional Reading

Braun, W. J., Murdoch, D. J. (2017). *A First Course in Statistical Programming with R*. Cambridge University Press. New York.

Gardener, M. (2012). *Beginning R: The Statistical Programming Language.* Wiley Publications.

Date: December, 2021

CAE201 Database Systems

Module Code and Title: CAE201 Database Systems
Programme: BSc in Statistics
Credit Value: 12
Module Tutors: J.Gurubalan

General Objectives:

This module presents a comprehensive introduction to relational database management systems with an emphasis on how to create tables, store, retrieve and update data efficiently and effectively. In addition, this module also covers more general issues such as database design principles, data independence, integrity, security, back-up and recovery, transaction, concurrency control techniques, and locking protocol. The students are expected to implement the database systems using Structured Query Language (SQL) in MySQL and spreadsheet package. This module will also introduce students to eXtensible Markup Language (XML) to structure data for storage, transport, and display.

Learning Outcomes:

On completion of the module, students will be able to:

1. Discuss the need of databases, its architecture, and schemas.
2. Design Entity Relationship (ER) models to represent database schemas.
3. Convert ER model to relational tables.
4. Populate relational database.
5. Formulate SQL queries on data.
6. Demonstrate functional dependency, primary key and relationship among tables.
7. Refine database design by normalization.
8. Construct database queries using SQL to create tables, insert and manipulate data.
9. Apply transaction in database for isolation, and authorization.
10. Design and implement database as a backend for database-centric applications.
11. Discuss distributed database systems architecture and design.
12. Discuss how to create XML documents and to store, transport and display data.

Learning and Teaching Approach:

<table>
<tr><th>Type</th><th>Approach</th><th>Hours per week</th><th>Total credit hours</th></tr>
<tr><td rowspan="2">Contact</td><td>Lecture</td><td>3</td><td rowspan="2">90</td></tr>
<tr><td>Laboratory sessions</td><td>3</td></tr>
<tr><td rowspan="2">Independent Study</td><td>Assignment/Project</td><td>1</td><td rowspan="2">30</td></tr>
<tr><td>Self-Study</td><td>1</td></tr>
<tr><td colspan="3">Total</td><td>120</td></tr>
</table>

Assessment approach:

F. VLE Quiz: 10%

One quiz will be conducted on VLE upon coverage of at least 40% of the subject matter to assess the students' overall understanding of the subject matter. This quiz will contain 30 Multiple Choice Questions, each question will carry 1 mark and it will be scale-down to 10%. The test will be conducted for 30 minutes duration.

G. Class Test: 10%

One class test will be conducted upon coverage of at least 60% of the subject matter. The test will include the ER model and Normalization. The test will be conducted for 90 minutes duration.

H. Laboratory Examinations: 20%

The first laboratory examination will be conducted on the 7th week for students for a minimum of two hours and 20 marks which will be scaled down to 10% is to evaluate SQL such as Create, Alter, drop, select, Where, AND, OR, NOT, Between, Is Null, Like, In, Exists, Aggregate Functions, group by, Order by and Having clauses in the DBMS.

The second laboratory examination will be conducted on the 14th week for students for a minimum of two hours and 20 marks which will be scaled down to 10% is to evaluate SQL such as Join, subquery, correlated queries, Union, Union All, Intersect, and Minus, Triggers, Stored Procedures PL/SQL Stored Functions and Embedded SQL in the DBMS.

I. Project: 20%

One project will be given to the students in groups (3 or 4) to test students' understanding of the design and implementation of the relational database systems such as requirement analysis, conceptual diagrams, logical design, and relational design and finally implement using SQL.

The project will be evaluated in two parts:

- Presentation: 5%
- Written report: 15%

The following marking criteria will be used to evaluate the two components:

Presentation:

i. Communication skills and self-confidence : 10%
ii. Contents : 10%
iii. Viva voce : 10%

Written report:

iv. Referencing and formatting : 5 %
v. Requirement analysis : 20%
vi. ER –diagram : 10%
vii. Normalized Tables : 15%
viii. Relational schema and SQL queries : 20%

Each project will be awarded 100 marks and will be scaled accordingly.

J. Semester End Examination: 40%

The exam at the end of the semester will be a closed book for a minimum of 2 hours and 40 marks. The exam will evaluate the students' overall understanding and critical thinking ability with regards to the module.

Area of Assignments	Quantity	Weighting
A. VLE Quiz	1	10%
B. Class Test	1	10%
C. Laboratory examinations	2	20%
D. Project	1	20%
E. Semester End Examination	1	40%
Total		100%

Pre-requisites: Nil

Subject Matter

Unit I: Introduction & DBMS Architecture

1.5. Introduction- Data, Database, Database management system
- 1.5.1. Characteristics of the database approach
- 1.5.2. Role of Database administrators
- 1.5.3. Role of Database Designers, End Users
- 1.5.4. Advantages of Using a DBMS and When not to use a DBMS.

1.6. DBMS Architecture

1.7. Data Models
- 1.7.1. Categories of Data models, Schemas, Instance, and Database states

1.8. DBMS Architecture and Data Independence
- 1.8.1. The Three schema architecture
- 1.8.2. Data Independence

1.9. DBMS language and interface

1.10. Classifications of Database Management Systems.

Unit II: Data Modelling Using Entity-Relationship Model

2.8. Using high level conceptual Data models for Database Design

2.9. An Example Database Application

2.10. Entity types

2.11. Entity Sets

2.12. Attributes and Keys

2.13. Relationships

2.14. Relationship types

2.15. Roles and Structural constraints

2.16. Weak Entity Types

2.17. Drawing E- R Diagrams.

Unit III: Database Design

3.7. Introduction to Logical Database Design

3.8. Relational Data Model

3.9. Super Key, Candidate Keys, Primary Key

3.10. Codd's Rules

3.11. Relational Algebra

3.12. Integrity Constraints

3.13. Transforming ER diagrams into relations

3.14. Functional Dependencies

3.15. Normalization – 1NF, 2NF, 3NF, BCNF and 4NF

Unit: IV System Implementation & Transaction Processing

4.4. Introduction to SQL
  4.4.1. Inserting, Updating, and Deleting data
  4.4.2. Processing Single Tables
  4.4.3. Manipulation Data in SQL
4.5. Processing Multiple Tables
  4.5.1. Joining Multiple Tables (Equi Joins)
  4.5.2. Joining a Table to itself (self Joins)
  4.5.3. Sub queries Union
  4.5.4. Intersect & Minus Clause
4.6. Creating view
  4.6.1. Renaming the Column of a view
  4.6.2. Updating, Selection, Destroying view
  4.6.3. Granting Permissions
4.7. PL/SQL Constructs Views
  4.7.1. Creating Indexes
  4.7.2. Triggers
4.8. Page rank

Unit: V Transaction Processing Concepts and Concurrency Control

5.6. Transaction and System concepts
  5.6.1. Desirable properties of Transactions
  5.6.2. Schedules and Serializability Issues
  5.6.3. Recoverability
5.7. Lock-Based Protocols
  5.7.1. Locks, Granting of Locks
  5.7.2. Two phase locking protocol.
5.8. Introduction to MapReduce programming model

Unit: VI XML

6.1. Introduction to XML
6.2. Uses of XML
6.3. XML key components
6.4. Document Type Definition (DTD) and Schemas

Unit: VII Distributed Databases

6.3. Distributed database concepts
6.4. Data fragmentation
6.5. Replication
6.6. Allocation Techniques for Distributed database design
6.7. Types of Distributed database systems

List of Laboratory sessions:

1. Create a sample database (e.g., employee, bank, university).
2. Use Data Definition commands to create tables and constraints in tables.

3. Use Data Manipulation commands for insert, delete, update, and list rows from table.
4. Use where, order, having, join clauses with logical operators and special operators to retrieve data from tables.
5. Use relational operators and sub query operators to join tables.
6. Write queries using R/Excel to retrieve, view, insert and analyze data from the backend database.
7. Use Procedural SQL to automate query, transform and update data in the database.
8. Create XML document to represent data.

Reading List:

Essential Reading:

Chaudhri, A. B., Rashid, A., & Zicari, R. (2003). XML data management: Native XML and XML-enabled database systems. Addison-Wesley Professional.

Elmasri, R., & Navathe, S. B. (2016). Fundamentals of Database Systems (7 ed.). Pearson.

Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019). Database System Concepts (7 ed.). McGraw-Hill.

Additional Reading:

Deitel, P., & Deitel, H. (2007). Internet & world wide web: how to program. Prentice Hall Press.

Ramakrishnan, R., & Gehrke, J. (2000). Database management systems. McGraw Hill.

Williams, K., Brundage, M., Dengler, P., Gabriel, J., Hoskinson, A., Kay, M. R. & Vanmane, M. (2000). Professional XML databases. Birmingham, UK: Wrox press.

Date: March, 2022

IST303 Statistical Inference

Module Code & Title: IST303 Statistical Inference
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dr. Bishal Gurung

General Objective

The module will enable students to learn different techniques of estimation and its properties. The module also enable students to understand the problems related to point and confidence interval estimation and testing of hypothesis using R along with inequalities and theorems related to inference.

Learning Outcomes

On completion of the module, the students will be able to:

1. Implement the basic concepts of various properties of estimators.
2. Explain the properties of the estimators.
3. Use the various inequalities to find the lower bound on the variance of estimators.
4. Transform an estimator using the Rao–Blackwell theorem to make a better estimator
5. Test the hypothesis using various tests and lemma.
6. Compute the confidence interval of estimated parameters.
7. Develop robust model with proper handling of outliers in the data
8. Apply the knowledge of Bayesian techniques for getting posterior estimates.
9. Execute randomized and non-randomized tests where one can view how to make a decision in hypothesis testing.

Learning and teaching approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Tutorial	1	
Independent Study	Written assignment/Quizzes	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. **Assignment: (10%)**

A major written assignment will be given in a semester. In written assignments, a set of problems/Theorems (not over 15) covering first four units will be given to the students, which are to be solved by the students independently. The submission of assignment would be two weeks after the assigned date. The assignment will be assessed based on following criteria:

- Correctness of the solutions: 4%
- Logical flow in the process: 2%
- Uniqueness of the solution: 2%

- Use of appropriate terms and symbols: 2%

B. **Laboratory: The portion of final mark: (15%)**
The lab assessment will be carried out as follows:

- **Lab report -5%**
  Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Lab Exam -10%**

A lab exam for the duration of 2 hour will be conducted towards the end of a semester after the last day of regular teaching covering all the subject matter.

C. **Mid-Semester Examination: The portion of final mark: (15%)**
Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 20% of the scores will be taken as final score.

D. **Semester-end Examination: Portion of Final Mark: (60%)**
Semester end Examination for the duration of three hours will be for 60 marks. All the units of the subject matter will be included for the examination

Overview of Assessment Approaches

Areas of assignments	Approach	Quantity	Weighting
A. Assignment		1	10%
B. Lab	Lab Report	10	5%
	Lab Exam	1	10%
C. Mid-Semester Examination		1	15%
D. Semester-end Examination		1	60%

Pre-requisite: PLT203 R Programming for Data Analysis

Subject Matter

Unit I: Estimators and its Properties

1.1 Notion of a parameter
1.2 Parameter space
1.3 General problem of estimation
1.4 Point and interval estimation
1.5 Frechet-Cramer-Rao inequality
1.6 Rao-Blackwell theorem
1.7 Basu's theorem.
1.8 Properties of estimators:
1.8.1 Unbiasedness
1.8.2 Consistency
1.8.3 efficiency and
1.8.4 sufficiency

Unit II: Methods of estimation

a. Maximum likelihood
b. Ordinary least squares
c. minimum χ^2
d. minimum distance
e. moments
f. Robust estimation and robust tests.
g. Asymptotic techniques.
h. Kernel method for univariate data: Rosenblatt's naïve estimator.
i. Consistency of general Kernel estimators, MSE and IMSE.
j. Concepts of consistency and asymptotic normality (CAN).
k. Cramer-Huzurbazar theorem, method of scoring.
l. Robust estimation and tests
m. Asymptotic techniques
n. Estimation of density function
o. Conditional inference

Unit III: Hypothesis Testing

a. Randomized and non-randomized tests
b. Neyman-Pearson lemma
c. Consistency and relative efficiency of tests
d. MP test and region
e. Power function
f. Uniformly most powerful tests and their constructions
g. Unbiased tests
h. Likelihood ratio tests
i. Confidence-interval estimation
j. Sequential analysis
k. Sequential probability ratio test

l. Elements of decision theory and Bayesian inference

Unit IV: Statistical Models for advanced inference

4.1 Detection and handling of outliers in statistical data.
4.2 Loglinear models.
4.3 Saturated models.
4.4 Hierarchical models.
4.5 Analysis of multi - dimensional contingency tables.

Unit V: Statistical Decision Problem

5.1 Decision problem and 2-person game
5.2 Non-randomized decision rules
5.3 Mixed and randomized decision rules
5.4 Decision problem for finite parameter space
5.5 Test of simple hypothesis against a simple alternative from decision theoretic view point.

Unit VI: Bayesian Estimations

6.1 Bayes theorem and computation of posterior distribution
6.2 Bayesian point estimation as a prediction problem from posterior distribution
6.3 Bayes estimators for absolute loss function, squared loss function, 0-1 loss function
6.4 Evaluation of estimates in terms of the posterior risk.
6.5 Bayesian interval estimation
6.6 Bayesian testing of hypothesis
6.7 Bayes factor for various types of testing hypothesis problem depending upon whether the null hypothesis and the alternative hypothesis are simple or composite
6.8 Bayesian prediction problem.

List of Laboratory Sessions

1. Verify the properties of the estimator such unbiasness, consistency, efficiency, and sufficiency
2. Compute the point and interval estimates
3. Compute and Compare the estimates using different method of estimation
4. Test the hypothesis using data and construct the confidence intervals
5. Perform the sequential analysis in the data and apply the sequential probability test
6. Identification of outliers in the statistical data
7. Building the model from advance inference methods

Reading Lists

Essential Reading

Casela, G., and Berger, R. L. (2017). *Statistical Inference*, Duxbury Thompson Learning, New Delhi.

Christensen, R. (2013). *Log Linear Models*, Springer, New Delhi.
Conover, W. J. (2007). *Practical Non-parametric Statistics*, John Wiley, New Delhi.
Hogg, R. V., and Craig, T. T. (2019). *Introduction to Mathematical Statistics*, Prentice-Hall, New Delhi.
Mood, A. M. Graybill, F. A., and Boes, D. C. (2017). *Introduction to Theory of Statistics*, Cambridge University Press, New Delhi.

Additional Reading

Box, G. E. P., and Tiao, G. C. (2011). *Bayesian Inference in Statistical Analysis*, Addison Wesely, New Delhi.
Dudewicz, E. J., and Mishra, S. N. (1988). *Modern Mathematical Statistics*. John Wiley, New Delhi.
Rao, C. R. (2001). *Linear Statistical Inference and its Applications*, John Wiley, New Delhi.
Rohatgi, V. K., and Saleh, A. K. Md. E. (2005). *An Introduction to Probability and Statistics*, John Wiley, New Delhi.
Wald, A. (2013). *Sequential Analysis*, Dover Publications, New Delhi.

Date: December, 2021

IST302 Regression Analysis

Module Code & Title: IST302 Regression Analysis
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

In this module students will learn how to derive multiple linear regression models, how to use R software to implement them, and what assumptions underlie the models. Students will also learn how to test whether their data meets those assumptions, what can be done when those assumptions are not met, and strategies to build and understand useful models.

Learning Outcomes

On completion of the module, students will be able to:

1. Calculate both simple and multiple regression models.
2. Visualize univariable linear regression model fits.
3. Implement and infer Ordinary Least Square (OLS) regression using R.
4. Compare linear and nonlinear regression models.
5. Assess parameter estimates of linear models using Gauss-Markoff's theorem.
6. Explain the concept of a best linear unbiased estimator (BLUE).
7. Implement Polynomial regression for data fitting problems.
8. Test homoscedasticity and normality of residuals.
9. Execute the regression diagnostics to non-normal errors, non-constant error variances, correlated observations and nonlinearity of the model.
10. Carry out variable selection and assess model accuracy.
11. Decipher output from R software package.

Learning and Teaching Approaches

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	60
	Laboratory sessions	1	
Independent Study	Written assignment	1	60
	Self-study	3	
Total			120

Assessment Approach

A. Assignment: (10%)

The assignment will be given during the first half of the semester. The assignment will comprise of problem-solving based on some questions provided to them. Assignments will be assessed using following criteria:

- Correctness of the solutions: 4%
- Logical flow in the process: 2%
- Uniqueness of the solution: 2%
- Use of appropriate terms and symbols: 2%

B. Laboratory sessions: (15%)

The lab assessment will be carried out as follows:

Laboratory report - **5%**

Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

Laboratory Exam - **10%**

A lab exam for the duration of 2 hour will be conducted towards the end of a semester after the last day of regular teaching.

C. Data Analysis Project: 10%)

Students will be divided into groups of 2-5 members depending upon the cohort. Each group will either be provided a data set along with the description of the data and variables, or choose its own dataset after review and approval from the tutor. The group members will collaborate and work together to develop research and analytical plans, conduct analyses and computation, explain and present results in the class. They will be assessed using the following marking criteria:

Objective	- 2 marks
Materials and methods	- 3marks
Results and Discussion	- 3 marks
Conclusion	- 2 marks

D. Mid-Semester Examination: (15%)

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 20% of the scores will be taken as final score.

E. Semester-end Examination: (50%)

Semester end Examination for the duration of three hours will be for 60 marks. All the units of the subject matter will be included for the examination

Overview of Assessment Approaches

Areas of assessments	Approach	Quantity	Weighting
A. Assignment		2	10%
B. Laboratory	Lab Report	10	5%
	Lab Exam	1	10%
C. Data Analysis Project		1	10%
D. Mid-Semester Examination		1	15%
E. Semester-end Examination		1	50%

Pre-requisites: DTS103-Probability Distribution I

Subject Matter

Unit I: Linear regression with one regressor

1.1 Simple linear regression
  1.1.1 Definition
  1.1.2 Assumptions of simple linear regression
1.2 Estimating the coefficient of the linear regression model
  1.2.1 The Ordinary least square method
  1.2.2 Restricted least squares: special cases of one- and two-way classifications.
1.3 Measures of fit
  1.3.1 The coefficient of determination
  1.3.2 The standard error of the regression
  1.3.3 Application to the test score data
1.4 The least square assumptions
1.5 The sampling distribution of the OLS estimator
1.6 Hypothesis tests and confidence Interval
  1.6.1 Testing two-sided hypotheses concerning the slope coefficient
  1.6.2 Confidence interval for regression coefficients
1.7 Using the t-statistic in regression when the sample size is small
1.8 R Regression when X is a binary variable
  1.8.1 The Gauss-Markov theorem (BLUE)
  1.8.2 Aitken's transformation

Unit II: Regression Model with multiple regressor

2.1 Multiple regression model
2.2 Omitted variable bias
2.3 OLS assumptions in multiple regression
2.4 The measure of fit in multiple regression
2.5 The distribution of the OLS estimator
2.6 Hypothesis testing and confidence interval
  2.6.1 Hypothesis tests and confidence intervals for a single coefficient
  2.6.2 Joint hypothesis testing using the F-statistic
  2.6.3 Confidence sets for multiple coefficients
2.7 Model specification
2.8 Assessing studies based on multiple regression
  2.8.1 Internal and external validity
  2.8.2 Threats to internal validity
  2.8.3 Internal and external validity when the regression is used for forecasting
  2.8.4 Diagnostics
2.9 Multicollinearity, influential observations
  2.9.1 Problems of multicollinearity
  2.9.2 Methods of detecting multicollinearity
  2.9.3 Ridge regression as a remedial for multicollinearity
2.10 Selection of independent variables
  2.10.1 Cross-Validation
  2.10.2 Stepwise variable regression

2.10.3 selection of variables with ridge regression

Unit III: Nonlinear Regression functions

3.1 A general strategy for modelling nonlinear regression functions
3.2 Nonlinear functions of a single independent variable
  3.2.1 Polynomials
  3.2.2 Logarithms
3.3 Interaction between independent variables
  3.3.1 Interactions between two binary variables.
  3.3.2 Interactions between a binary and a continuous variable.
  3.3.3 Interactions between two continuous variables.

Unit IV: Special Types of Regression

4.1 Regression with panel data
  4.1.1 Panel Data
  4.1.2 Panel data with two time period: before and after comparison
  4.1.3 Fixed effects regression
  4.1.4 Regression with time fixed effects
  4.1.5 The fixed effect regression assumptions and standard error for fixed effect
  4.1.6 Random effects model
  4.1.7 Mixed effects model
4.2 Regression with a Binary Dependent variable
  4.2.1 Binary variable and linear probability model
  4.2.2 Probit and Logit regression
  4.2.3 Estimation and Inference in the Logit and Probit model
4.3 Instrumental Variable regression
  4.3.1 The Instrumental variables with a single regression and a single instrument
  4.3.2 The general instrumental variable regression
  4.3.3 Checking instrument validity

List of Laboratory Sessions

1. Testing hypothesis in linear models
2. Analysis of variance and Analysis of covariance
3. Fitting Polynomial regression
4. Lab session on Indicator variable technique, Regression with ordinal data, Non-linear regression models
5. Testing the various regression diagnostics
6. Multicollinearity, Ridge regression, principal component regression and robust regression

Reading List

Essential Reading

Kleinbaum, D. G., and Kupper, L. L. (2014). *Applied Regression analysis and other Multivariate Methods*, Duxbury Press, Massachusetts, USA.

Draper, N. R., and Smith, H. (1998). *Applied Regression Analysis*, John Wiley, New Delhi.

Bapat, R. B. (2012). *Linear Algebra and Linear Models*, Springer-Verlag, New Delhi.

Barnett, V., and Lewis, T. (1998). *Outliers in Statistical Data*, John Wiley, New Delhi.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2004). *Regression Diagnostics-Identifying Influential Data and Sources of Collinearity*, John Wiley, New Delhi.

Additional Reading

Graybill, F. A. (2000). *Theory and Application of the Linear Model*, Duxbury, North Scituate, USA.

Joshi, D. D. (1987). *Linear Estimation and Design of Experiments*, Wiley Eastern, New Delhi.

Montgomery, D. C., Peck, E., and Vining, G. (2012). *Introduction to Linear Regression Analysis*, John Wiley, New Delhi.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models*, Chapman and Hall, New Delhi.

Rao, C. R. (2001). *Linear Statistical Inference and Its Applications*, Wiley Eastern, New Delhi.

Searle, S. R., and Gruber, H.J. (2016). *Linear Models*, John Wiley, New Delhi.

Searle, S. R. Casella, G. McCulloch, C.E. (1998). *Variance Components*, John Wiley, New Delhi.

Scheffé, H. (1999). *The Analysis of Variance*, John Wiley, New Delhi.

Date: December, 2021

ECO101 Fundamentals of Economics

Module Code & Title: ECO101 Fundamentals of Economics
Programme: BA in Population and Development Studies
Credit Value: 12
Module Tutor: Rinzin Dema

General Objective

This module aims to provide students with an understanding of the fundamental principles of economics. Students will learn how to apply economic theory to questions of government, business and household management, with how markets work and how economic relations affect the wider organisation of society.

Learning Outcomes

On completion of the module, students will be able to:

1. Explain the basic economic concepts
2. Describe the basic components of economic theories
3. Relate the demand and supply theories in current market context
4. Explain the role of financial, fiscal and monetary policy
5. Illustrate the linkages between poverty, unemployment and inflation
6. Explain the economic and development policies of Bhutan
7. Discuss the concept of welfare state, fiscal spending and revenue

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Tutorial	1	
Independent study	Written Assignment	2	**30**
	Self-study	2	
Total			**120**

Assessment Approach

Assessment will be carried on a continuous basis through the following assignments:

A. **Assignment: 20%**

Students have to write two assignments. Each assignment carries 10 percent. The assignment topics will be determined by the concerned module tutor.

- Content (relevancy of the content and originality):: 5%
- Organisation and language (clarity and flow, use of appropriate words and grammar, presentation): 3%
- Proper Referencing and Citation: 2%

B. **Class Test: 10%**
Class test will be conducted after completion of 50% of the module for duration of one hour.

C. **Quiz: 10%**
Quiz will conducted after completion of the module. The question should cover all the units prescribed in the module descriptor. This will help students to prepare for Semester End Examination.

D. **Semester-end Examination: The portion of Final Marks: (60%)**
Students have to write end-semester with students of other programme.

Overview of assessment approaches

Area of Assignments	Quantity	Weighting
A. Assignment	2	20%
B. Class Test	1	10%
C. Quiz	1	10%
D. Semester-end Examination	1	60%

Prerequisite: None

Subject Matter

Unit I: Introduction

1.1 Introduction to economic concepts and terminologyDendogram analysis
1.2 Demand, supply and market
- 1.2.1 Demand and Supply Curve Cluster validation
- 1.2.2 An individual demand and supply for produc
- 1.2.3 The market demand and supply curve
- 1.2.4 Movement along and shifts in demand and supply curve

1.3 Elasticity of demand and supply
- 1.3.1 Price elasticity of Demand and Supply
- 1.3.2 Coefficient of Price elasticity
- 1.3.3 Elasticity and total revenue
- 1.3.4 Factors affecting price elasticity
- 1.3.5 Price, cross-elasticity of demand
- 1.3.6 Income elasticity of demand
- 1.3.7 Calculation of marginal revenue
- 1.3.8 The geometry of marginal revenue determinants
- 1.3.9 The MRC lies below the demand curve
- 1.3.10 Elasticity and Marginal Revenue

Unit II: Microeconomics

2.1 Consumer Choice and demand decision
2.2 Supply decisions
2.3. Costs and supply
2.4. Perfect competition and pure monopoly
2.5. Market structure and imperfect competition
2.6. The labor market
2.7. Different types of labor
2.8. Factor markets and income distribution
2.9. Risk and information

Unit III: Macroeconomics

3.1. Introduction to macroeconomics
3.2. Output and aggregate demand
3.3. Fiscal policy and foreign trade
3.4. Money and Banking
3.5. Interest rates and monetary transmission
3.6. Monetary and fiscal policy Page 23 of 158
3.7. Aggregate supply, prices and the adjustment to shocks
3.8. Inflation, expectations, and credibility
3.9. Unemployment
3.10. Exchange rates and balance of payments
3.11. Open economy macroeconomics
3.12. Economic growth
3.13. Business cycles

Unit IV: Welfare State

4.1. Introduction to welfare economics
4.2. Government spending and revenue
4.3. Industrial policy and competition
4.4. Natural monopoly: Public or private

Reading List

Essential Reading

Begg, D., Fischer, S., & Dornbusch, R. (2008). Economics (9th edition). McGraw-Hill Contemporary.

Bodman, D. (2002). Macroeconomics. McGraw-Hill Europe.

Charlé, E. (1983). Macroeconomics of developing countries. Tata McGraw-Hill.

Lipsey, R. G., & Chrystal, K. A. (1999). Principles of Economics (9 edition). Oxford University Press.

Mankiw, N. G. (2015). Macroeconomics. New York: Worth Publishers.

Mankiw, N. G. (2014). Principles of Macroeconomics. Australia: South-Western College Pub.

NSB (2017). Consumer price Index bulletin January 2017. National Accounts and Price Division, NSB

Additional Reading

Katz M. L., & Rosen, H. S. (1991). Microeconomics.-Richard D. McGraw Hill. Koutsoyiannis, A. (2001). Modern Micro Economics. ELBS.

Pindyck, R. S., & Rubinfeld, D. (1999). Micro Economics. McGraw Hill.

Varian, H. (1998). Intermediate Micro Economics. WW Norton

Date: December, 2021

Year 2 Semester II

AMT210 Numerical Methods

Module Code & Title: AMT210 Numerical Methods
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Pema Tshering

General Objective

The module provides the numerical methods of solving the non-linear equations, interpolations of polynomials, differentiation, and integration. The module will help the students' to improve skills in numerical methods by using Python programming.

Learning Outcomes

On completion of the module, students will be able to:

1. Apply the concepts of error and its sources in solving problems.
2. Apply various numerical methods to solve equations.
3. Identify different methods for solving non-linear equations.
4. Solve a function using interpolation techniques.
5. Calculate derivative of a function numerically.
6. Identify appropriate methods to integrate a function.
7. Examine direct and iterative methods to solve system of linear equations.
8. Transform symmetric matrices into tridiagonal matrices using different methods.
9. Evaluate Eigen values of the tridiagonal matrices.
10. Evaluate Eigenvalues and eigenvectors using symmetric orthogonal matrices.

Learning and Teaching Approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	4	**75**
	Lab sessions	1	
Independent Study	Written assignment	1	**45**
	Self-study	2	
Total			**120**

Assessment Approach

A. Assignment: (15%)

An individual assignment will be given to the students before mid-semester examination. The assignment will comprise of problem solving and it will be evaluated broadly on the following criteria:

- Solving non-linear equations mathematically and using python 4 marks
- Solving functions using interpolation techniques 4 marks
- Fitting linear and non-linear curves 3 marks
- Integration and differentiation of functions using various methods 4 marks

B. Class Test: (10%)

A class test of one hour duration will be conducted before the semester end examination and the questions will be from the subject matters covered after the mid semester examination.

C. Lab Exam: (10%)

A lab exam for the duration of 2 hours will be conducted towards the end of a semester after the last day of regular teaching. The objective of the lab exam is to test the understanding of the concepts taught. Students will have to use Python to evaluate numerical solutions to some of the concepts listed in the list of laboratory sessions. The evaluation will be based on the solution to lab exam questions.

D. Mid-Semester Examination: (15%)

Mid-Semester examination will be of one and half hours duration with at least 60% of subject matter included. The total weighting of the examination paper will be 40 and the marks obtained will be converted into 15.

E. Semester-end Examination: (50%)

Semester end Examination will be conducted for the duration of two and half hours and questions will be from all the units of the subject matter of the module. The total weighting of the examination paper will be 50. Test blue print of the question paper and an answer key for grading will also be prepared along with the question paper.

Overview of Assessment Approaches

Areas of assignments	Quantity	Weighting
A. Assignment	1	15%
B. Class Test	1	10%
C. Lab exam	1	10%
D. Mid-Semester Examination	1	15%
E. Semester-end Examination	1	50%

Prerequisite: PLT102: Programming in Python

Subject matter

Unit I: Errors

1.1 Definition and source of errors
1.2 Significant error
1.3 Error in numerical computations

Unit II: Solutions of non-linear equations (Algorithm and Problem solving)

2.1 The Bisection method
2.2 Method of false-position
2.3 The Secant method
2.4 Muller's method
2.5 Fixed point iterative method
2.6 Newton-Raphson's method
2.7 Graffe's method

Unit III: Interpolation and curve fitting

3.1 Lagrange's interpolation (Algorithm and Problem solving)
3.2 Newton's divided difference interpolation (Algorithm and Problem solving)
3.3 Newton's forward and backward difference interpolation (Algorithm and Problem solving)
3.4 Central difference interpolation:(Discuss algorithm and Problem solving)
    3.4.1 Bessel's interpolation
    3.4.2 Stirling's interpolation
    3.4.3 Gauss' interpolation
    3.4.4 Everett's interpolation
3.5 Hermite interpolation (Algorithm and Problem solving)
3.6 Cubic Spline interpolation (Algorithm and Problem solving)
3.7 Chebyshev Interpolation polynomial (Algorithm and Problem solving)
3.8 Least-squares curve fitting (Algorithm and Problem solving)
3.9 Fitting linear and non-linear curves (Algorithm and Problem solving)
3.10 Polynomial Approximations (Algorithm and Problem solving)
3.11 Orthogonal Polynomials (Algorithm and Problem solving)

Unit IV: Numerical Differentiation and Integration

4.1 Numerical Differentiation(Discuss algorithm and Problem solving)
    4.1.1 Newton's forward differentiation
    4.1.2 Newton's backward differentiation
    4.1.3 Central formula differentiation
4.2 Numerical Integration: (Algorithm and Problem solving)
    4.2.1 Rectangular rule
    4.2.2 Trapezoidal rule
    4.2.3 Simpson rule
    4.2.4 Boole's rule

4.2.5 Weddle's rule
4.2.6 Gaussian quadrature formula
4.2.7 Errors in Gaussian quadrature formula

Unit V: System of Linear Equations

5.1 Direct Methods:(Algorithm and Problem solving)
5.1.1 Gaussian Elimination method
5.1.2 Gauss-Jorden Method
5.1.3 LU Decomposition method
5.1.4 Cholesky method
5.1.5 Matrix inversion method
5.2 Iterative Methods:(Algorithm and Problem solving)
5.2.1 Jacobi's Method
5.2.2 Gauss-Seidel Method
5.2.3 Relaxation method

Unit VI: Eigen Values and Eigen vectors (Algorithm and Problem solving)

6.1 Jacobi's method
6.2 Power method
6.3 Given's method
6.4 Householder's method
6.5 The QR method
6.6 Lanczo's method

Unit VII: Ordinary Differential Equations

7.1 Picard's method
7.2 Taylor's series method
7.3 Euler's method
7.4 Modified Euler's method
7.5 Ranga Kutta Methods
7.6 Adams-Bashford Methods
7.7 Adams-Moulton methods
7.8 Milne-Simpson's Method

List of Laboratory Sessions

The lab sessions will help the students to understand the concepts well. They will be taught how to compute and analyze solutions to some of the numerical problems using Python. Students are expected to use Python to evaluate the following problems:

1. Zero of nonlinear functions.
2. Roots of polynomials using Newton Raphson model.
3. Interpolation with polynomials.
4. Interpolation with splines.

5. Interpolation and approximation with trigonometric polynomial.
6. Integration with simple and composite newton's cotes formula.
7. Integration with Gaussian quadrature formula.
8. Gaussian Elimination.
9. Triangular factorization.
10. Gauss-Seidel method.
11. Relaxation method.
12. Jacobi method.
13. Transformation of symmetric matrix to tridiagonal matrix.

Reading List

Essential reading

Gerald, C.F. & Wheatley, P. O. (2013): *Applied Numerical Analysis*. Dorling Kindersley (India) Pvt. Ltd.

Greenbaum A. &Chartier T. P. (2012). *Numerical Methods: Design, Analysis and Computer Implementation of algorithms*. Princeton University Press.

Burden, R.L. & Faires, J.D: *Numerical Analysis.* (9 ed.). Brooks/Cole Publishing Company

Additional reading

Atkinson, K.E. An Introduction to Numerical Analysis. (2 ed.). John Wiley & Sons.

Sastry, S. S. (2000): *Introductory Methods of Numerical Analysis*. Asoke K. Ghosh, Prentice-Hall of India Private Limited

Prasad, D. (2012): *An Introduction to Numerical Analysis*. Narosa Publishing House Pvt. Ltd.

Date: December, 2021

ADS205 Algorithm Analysis

Module Code and Title: ADS205 Algorithm Analysis
Programme: BSc in Data Science
Credit Value: 12
Module Tutors: Sangay Thinley

General Objective

This module aims at providing the students with an understanding of methods for analyzing and evaluating the performance of an algorithm in terms of computational efficiency. It also introduces the different techniques to design efficient algorithms. Python programming language will be used as a tool for the delivery of this module.

Learning Outcomes

On completion of the module, students will be able to:

1. Use appropriate data structures to solve a problem
2. Design algorithmic solution to search and sort problems
3. Determine time-space complexity of an algorithm in terms of asymptotic notations
4. Derive and solve recurrence relations
5. Apply the divide-and-conquer technique to solve real world problems
6. Integrate dynamic and recursive approach to improve the efficiency of an algorithm
7. Implement parallelism of algorithm on computer machine
8. Analyze several other algorithms of importance such as string matching, NP completeness, Approximation algorithm etc.

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	90
	Laboratory Sessions	3	
Independent study	Self-study	2	30
Total			**120**

Assessment Approach

A. Assignment: (10%)

Individual students are required to do the assignment during the first half of the semester after completion of 30 percent of the syllabus content. Set of questions based on the topics covered will be provided to the students. Students will have to exhibit their understanding of the concepts through their solutions to the questions. Assessment of the assignment will be done based on the following criteria:

1. Presentation (2%)
2. Correctness of result (3%)
3. Logic (5%)

B. Lab Assessment: (30%)

Lab Assessment will be assessed based on the following components:

- Lab Assignments: 15%
- Lab Examination: 10%
- Viva Voce: 5% (Lab Assignment: 2%, Lab Examination: 3%)

Lab Assignments (a total of 7) will be based on the concepts taught in the theory class. Students will be provided a set of programming questions every week covering the concepts taught on each unit. The students will have to submit the solutions at the end of the lab session through the VLE.

The following rubrics will be used to assess the set of programming questions. Each set will be assessed out of 100 marks and converted to 1.5%.

Program (100)	**(Excellent)** **100-81%**	**(Good)** **80-61%**	**(Satisfactory)** **(60-41% of the points)**	**(Unsatisfactory)** **(<=40% of the points)**
Requirements and Delivery **(50 points)**	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution **(20 points)**	• Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic **(10 points)**	• Program is logically well designed.	• Program has slight logic errors that do no significantly affect the results	• Program has significant logic errors.	• Program is incorrect.
Coding Standards **(10 points)**	• Includes name, date, and assignment title. • Excellent use of white space. • Creatively organized work.	• Includes name, date, and assignment title. • Good use of white space. • Organized work. • Good use of variables (no	• Includes name, date, and assignment title. White space makes program fairly easy to read. • Organized work.	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines). • Disorganized and messy.

	• Excellent use of variables (no global variables, unambiguous naming).	global variables, unambiguous naming)	• Good use of variables (few global variables, unambiguous naming).	• Poor use of variables (many global variables, ambiguous naming).
Documentation **(10 points)**	• Clearly and effectively documented including descriptions of all class variables. • Specific purpose noted for each function, control structure, input requirements, and output results.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function and control structure.	• Basic documentation has been completed including descriptions of all class variables. • Purpose is noted for each function.	• Very limited or no documentation included. • Documentation does not help the reader understand the code.

Lab examination will be conducted after the completion of the module and before semester-end examination and will cover all the units. Each student will be given two questions and will be graded out of 100 marks based on the rubric mentioned above. The marks obtained will then be converted out of 10%.

Viva Voce for the lab assignments will be conducted once to assess the students' understanding of the lab assignments. Viva Voce will also be conducted for the lab examination. The following rubrics will be used for viva voce. Each viva voce will be assessed out of 100 marks and converted to 2% for lab assignments and 3% for lab examination respectively.

Program (100)	**(Excellent)** **85-100%**	**(Good)** **61-84%**	**(Satisfactory)** **(<=60% of the points)**
Knowledge in programming **(50 points)**	Demonstrates deep knowledge, answer the questions with explanation and elaboration.	Adequate knowledge of most topics, answer the questions but fails to elaborate.	Superficial knowledge of topic, only able to answer basic questions.
Problem-solving ability **(30 points)**	Efficient mapping of theory concepts with practical problem solving approaches.	Moderate mapping of theory concepts with practical problem solving approaches.	Improper mapping of theory concepts with practical

			problem solving approaches.
Critical Thinking **(20 points)**	Can predict and defend problem outcomes.	Approximately predicts and defends problem outcomes.	Is unable to predict problem outcomes for the given input data set.

Mid-Semester Examination: (10%)

Mid-semester Examination will be conducted (out of 40) in the mid of semester with at least 60% of the subject matters being covered for the duration of 2 hours. The marks obtained will then be converted out of 10%.

C. Semester-end Examination: (50%)

Semester end Examination for the duration of two and half hours will be conducted covering all the units of the module. The exam will be conducted out of 50.

Overview of assessment approaches and weighting:

Areas of Assessment		**Quantity**	**Weighting**
A. Assignment		1	10%
B. Lab Assessment	Weekly Lab Assignment	7	10%
	Lab Examination	1	15%
	Viva Voce	2	5%
C. Mid-Semester Examination		1	10%
D. Semester-end Examination		1	50%
Total			100%

Pre-requisites: PLT102 Programming in Python

Subject Matter

Unit I: Introduction

1.1. Algorithm Definition and Specification, Asymptotic notations
1.2. Average, worst case, best case and amortized analysis
1.3. Analysis of different searching and sorting algorithms in terms of time and space complexity
  1.3.1. Linear search
  1.3.2. Insertion-sort
  1.3.3. Bubble-sort
  1.3.4. Selection-sort
  1.3.5. Heap-sort
  1.3.6. Shell-sort
  1.3.7. Bucket-sort

1.3.8. Count-sort
1.3.9. Radix-sort

Unit II: Divide & Conquer Strategy
2.1. Introduction
2.2. Recurrence and different methods to solve recurrence
2.3. Multiplying large integer problem
2.4. Binary Search algorithm and its analysis
2.5. Max-Min Problem
2.6. Sorting algorithm (Merge-sort and Quick-Sort) analysis
2.7. Matrix Multiplication, Exponential

Unit III: Greedy Strategy
3.1. General Characteristics of greedy approach
3.2. Activity Selection Problem
3.2.1. Problem Definitions and its solution
3.3. Minimum Spanning Tree problem (Kruskal's and Prim's algorithm)
3.3.1. Problem Definition and its solution
3.4. Knapsack problem
3.4.1. Problem definition and its solution,
3.5. Job Scheduling Problem
3.6. Huffman code
3.6.1. File Compression

Unit IV: Dynamic Programming
4.1. Introduction
4.2. Principle of Optimality
4.3. Calculating binomial coefficient
4.4. All pair Shortest Path
4.4.1. Dijkstra's algorithm
4.5. Assembly line scheduling
4.6. Matrix Chain Multiplication
4.7. Longest Common Subsequence

Unit V: Branch & Bound Technique and Backtracking
5.1. Introduction
5.2. 0/1 Knapsack problem
5.2.1. Problem definition and its solution
5.3. Travelling Salesman Problem
5.3.1. Problem definition and its solution
5.4. N-Queen Problem
5.4.1. Problem definition and its solution

Unit VI: Pattern Match
6.1. Brute force algorithm
6.2. Knuth Morris Pratt algorithm
6.3. Boyer Moore Algorithm

Unit VII: Advanced Techniques

7.1. Approximation Algorithm
7.2. PRAM Algorithms
7.3. Mesh Algorithm
7.4. Hypercube Algorithm

List of Laboratory Sessions

Each of the following sessions will be supervised by the module tutor and each programming exercise will be based on real world problems assigned by module tutor. Each of the implementations would be done under parallel environment on computer machine. Running time should be one of the outputs of all the programs.

1. Implement the following sorting algorithms to compare their efficiencies in terms running time with same set of input data.
    a. Bubble sort
    b. Insertion sort
    c. Selection Sort
    d. Heap-Sort
    e. Shell –Sort
    f. Bucket Sort
    g. Count Sort

2. Implement linear and binary search algorithm and compare their efficiencies in terms of running time with same set of input data.
3. Write and execute a program to accomplish the following using iterative and recursive approach:
    a. Find Factorial of an Integer
    b. Generate Fibonacci sequence
4. Write a program to implement 0/1 knapsack problem using
    a. Dynamic programming
    b. Branch & Bound technique.
5. Write a program to find the maximum profit subset of jobs such that no two jobs in the subset overlap (Job Scheduling Algorithm).
6. Write a program to determine longest common subsequence (LCS) between two given strings.
7. Implement N-Queen problem using backtracking.
8. Write a program to multiply a chain of N number of matrices using dynamic programming. Obtain optimal sequence and total no of operations as output of the program.
9. Write a program to find minimum spanning tree in a graph using
    a. Prim’s Algorithm.
    b. Kruskal’s Algorithm.
10. Write a program to find a match of a substring in a given string using
    a. Brute Force Method.
    b. Boyer Moore algorithm.

Reading List

Essential Reading

Coremen T H, Leiserson C E, Rivest R L and Stein, Clifford. (2022). *Introduction to Algorithms* (4 ed.). MIT Press .

Enrst L. Leiss (2007). *A Programmer's Companion to Algorithm Analysis* (1 ed.). Chapman & Hall/CRC.

Horowitz E, Sahni S, Rajasekaran S. (2010). *Fundamentals of Computer Algorithms* (2 ed.). Universities Press Pvt Ltd.

Additional Reading

Donald E. Knuth. (1997). *The Art of Computer Programming Volume 3, Sorting and Searching*, (2 ed.). Pearson Education, Addison-Wesley.

Pai G.A.V. (2008). *Data structures and Algorithms*, Tata McGraw Hill.

Date: March, 2022.

IST306 Multivariate Analysis

Module Code & Title: IST306 Multivariate Analysis
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dr. Bishal Gurung

General Objective

The module will introduce students to multivariate data structure, multinomial and multivariate normal distribution, estimation and testing of parameters, various data reduction methods. The module also helps the students in having a better understanding of research data, its presentation and analysis. The module will be delivered through the use of R.

Learning Outcomes

On completion of the module, students will be able to:

1. Determine marginal and conditional distributions of multivariate normal distribution.
2. Apply the Wishart distribution for appropriate datasets.
3. Apply Hotelling's T^2 for testing hypothesis for multivariate data.
4. Determine the distance between clusters using Mahalanobis' D^2 statistics.
5. Apply the concepts of Wilk's Lambda for testing appropriate hypothesis.
6. Allocate different source of variation using MANOVA.
7. Classify items into groups by using discriminant analysis.
8. Make groups and allocate items into groups using cluster analysis.
9. Find the correlation between two sets of variables using canonical correlations (CC).
10. Apply principal component analysis (PCA) as a bridge between two analyses
11. Conduct factor analysis and describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables.
12. Apply PCA and CC as a data reduction technique.

Learning and teaching approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	3	**60**
	Laboratory sessions	1	
Independent Study	Written assignment/Quizzes	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. Assignment: (10%)
A major written assignment will be given to an individual student during second half of the semester to test the understanding of basic techniques employed while analyzing multivariate data for the purpose of grouping, classifying or data reduction.

The assignment will be assessed based on following criteria:

- Correctness of the use of the appropriate multivariate technique: 4%
- Use of appropriate examples: 2%
- Presentation of the assignment: 4%

B. Laboratory Assignment: (25%)

- **Laboratory report – 5%**

Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Lab exam – 20%**

A lab exam for the duration of 2 hour will be conducted towards the end of a semester after the last day of regular teaching.

C. Mid-Semester Examination: (15%)

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 50 marks and 15% of the scores will be taken as final score.

D. Semester-end Examination: (50%)

Semester end Examination for the duration of two and half hours will be for 50 marks. All the units of the subject matter will be included for the examination.

Overview of Assessment Approaches

Areas of assignments		**Quantity**	**Weighting**
A. Assignment		1	10%
B. Lab	Lab Report	10	5%

	Lab Exam	1	20%
C. Mid-Semester Examination		1	15%
D. Semester-end Examination		1	50%

Prerequisites: ALG108 Matrix Analysis and Vector Spaces, DST101 Descriptive Statistics and PLT203 R Programming for Data Analysis

Subject Matter

Unit I: Multivariate data

1.1 Concept of random vector
1.2 Concept of probability mass/density functions of random vector
1.3 Concept of distribution function
1.4 Concept of mean vector & dispersion matrix
1.5 Concept of marginal distribution
1.6 Concept of conditional distribution

Unit II: Introduction to Multivariate Distribution

2.1 Concept of multivariate normal distribution
2.2 Probability density function of multivariate normal distribution: Proof
2.3 Properties of multivariate normal distribution
2.4 Derivation of marginal probability density function of multivariate normal distribution
2.5 Derivation of conditional probability density function of multivariate normal distribution
2.6 Expectation and variance-covariance matrix
2.7 Joint distributions of multivariate normal distribution
2.8 Conditional distributions and independence of random vectors
2.9 Multinomial distribution
2.10 Sample mean vector and its distribution
2.11 Maximum likelihood estimates of mean vector and dispersion matrix

Unit III: Distributions and Testing of Hypothesis

3.1 Wishart distribution.
3.2 Hotelling's T^2 and null distribution of Hotelling's T^2
3.3 Tests of hypothesis about mean vector
3.4 Mahalanobis' D^2 statistics
3.5 Rao's U statistics and its distribution
3.6 Wilks' Lambda criterion and statement of its properties
3.7 Multivariate analysis of variance (MANOVA) and covariance (MANCOVA).

Unit IV: Techniques of Multivariate Analysis

4.1 Concepts of discriminant analysis

4.2 Derivation of linear discriminant function (LDF)
4.3 Classification between k multivariate normal populations based on LDF and Mahalanobis D^2
4.4 Cluster analysis: k-means and Hierarchical clustering

Unit V: Data reduction techniques

5.1 Concept and derivation of formula of canonical correlations
5.2 Concept of principal component analysis and derivation of principal components
5.3 Concept of factor analysis and derivation of factor loadings using different methods
5.4 Multi-dimensional scaling
5.5 Correspondence analysis.

List of Laboratory Sessions

1. Maximum likelihood estimates of mean-vector and dispersion matrix in Excel and R language.
2. Testing of hypothesis on mean vectors of multivariate normal populations in Excel and R language.
3. Cluster analysis, Discriminant function, Canonical correlation, and Principal component analysis in Excel and R language.
4. Factor analysis. Multivariate analysis of variance and covariance, multidimensional scaling in Excel and R language.

Reading List

Essential Reading

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, John Wiley, New Delhi.

Rao, C. R. (2001). *Linear Statistical Inference and its Applications*, John Wiley, New Delhi

Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*, John Wiley, New Delhi.

Johnson, R. A., and Wichern, D. W. (2008). *Applied Multivariate Statistical Analysis*, Prentice Hall, New Delhi.

Rencher, A. C. (2002). *Methods of Multivariate Analysis*, John Wiley, New Delhi.

Additional Reading

Srivastava, M. S., & Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New Delhi.

Chatfield, C., & Collins, A. J. (1982). *Introduction to Multivariate Analysis*, Prentice Hall, New Delhi.

Date: December, 2021

DSC201 Machine Learning

Module Code & Title: DSC201 Machine Learning
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

The objective of this module is to introduce formal definition of Machine Learning and continues on with explanations for the various machine learning and training techniques. Module also review both Supervised and Unsupervised learning, showcasing the main differences between each type of learning method.

Learning Outcomes

On completion of the module, students will be able to:

1. Explain the different Machine Learning training techniques.
2. Explain the difference between Supervised and Unsupervised training.
3. Differentiate between Classification and Regression.
4. Apply Neural Network in supervised learning.
5. Interpret results from scikit-learn estimators
6. Explain the tradeoffs inherent in different machine learning methods: speed, accuracy, complexity of hypothesis space, etc.
7. Explain issues of algorithmic bias, transparency, fairness in supervised machine learning applications.
8. Apply regression and classification objects (estimators) in scikit-learn.
9. Apply a wide variety of evaluation metrics to supervised learning scenarios.

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**75**
	Laboratory Session	2	
Independent study	Written Assignment	1	**45**
	Self-study	2	
Total			**120**

Assessment Approach

A. **Problem-solving Assignment: (10%)**

The assignment will be given during the first half of the semester. The assignment will comprise of problem-solving based on some questions provided to them. Assignments will be assessed using following criteria:

- Correctness of the solutions: 4%
- Logical flow in the process: 2%
- Uniqueness of the solution: 2%

- Use of appropriate terms and symbols: 2%

B. **Laboratory: (35%)**

Lab assessment will be carried out as follows:

- **Laboratory report- 10%**

  Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

  At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Laboratory Exam- 25%**

  A lab exam for the duration of 3 hour will be conducted towards the end of a semester after the last day of regular teaching.

C. **Mid-Semester Examination: The portion of Final Marks: (15%)**

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 15% of the scores will be taken as final score.

D. **Semester-end Examination: The portion of Final Marks: (40%)**

Semester end Examination for the duration of two and half hours will be for 50 marks. All the units of the subject matter will be included for the examination.

Overview of assessment approaches:

Areas of Assessment		**Quantity**	**Weighting**
A. Problem-solving Assignment		1	10%
B. Lab Assignment	Lab report	10	10%
	Lab exam	1	25%
C. Mid-Semester Examination		1	15%
D. Semester-end Examination		1	40%

Prerequisite: PLT102 Programming in Python, DST101 Descriptive Statistics, IST302 Regression Analysis

Subject Matter

Unit I: Introduction and Basic Concepts

1.3 Definition of Machine Learning
1.4 History/evolution of Machine Learning
1.5 Taxonomy of Machine Learning: A simplistic view based on tasks
1.6 Definition of supervised learning with examples
1.7 Supervised Machine learning setup with examples
1.8 Different types of supervised machine learning (Definition and examples): Regression, Classification, Naïve Bayesian model, Random forest model, Neural Networks, and Support vector machines.
1.9 Regression vs Classification
1.10 Definition of unsupervised learning with examples
1.11 Unsupervised machine learning setup with examples
1.12 Different types of unsupervised machine learning (Definition and examples): Clustering, association
1.13 Definition of reinforcement learning with examples

Unit II: Regression

2.1 Linear regression with one variable:
- 2.1.1 Model Representation
- 2.1.2 Cost function
- 2.1.3 Cost function-intuition
- 2.1.4 Gradient descent
- 2.1.5 Gradient descent -intuition
- 2.1.6 Gradient descent for linear regression

2.2 Linear regression with multiple variables
- 2.2.1 Multiple features
- 2.2.2 Gradient descent for multiple variables
- 2.2.3 Gradient descent in practice: Feature scaling, learning rate
- 2.2.4 Features and Polynomial regression
- 2.2.5 Normal equation
- 2.2.6 Normal equation: Non invertibility

2.3 Regularization
- 2.3.1 Problem of overfitting
- 2.3.2 Cost function
- 2.3.3 Regularized linear regression
- 2.3.4 Hypothesis representation
- 2.3.5 Decision boundary
- 2.3.6 Cost function
- 2.3.7 Simplified cost function and Gradient descent
- 2.3.8 Optimisation
- 2.3.9 Multiclass classification: one -vs-all

Unit III: Classification

3.1 Classification and Representation
3.2 Hypothesis representation
3.3 Decision boundary
3.4 Cost function
3.5 Simplified cost function and Gradient descent
3.6 Optimisation

3.7 Multiclass classification: one-vs-all
3.8 Regularization
3.8.1 Problem of overfitting
3.8.2 Cost function
3.8.3 Regularized logistic regression
3.9 Naïve Bayes
3.9.1 Classifier based on Bayes Rule
3.9.2 Naïve Bayes Algorithm
3.10 K-Nearest Neighbor
3.11 Decision Tree
3.12 Random Forest
3.12.1 Bagging
3.12.2 From Bagging to random forest
3.12.3 Relationship to nearest neighbors
3.13 Support Vector Machine
3.13.1 Optimisation objective
3.13.2 Large margin intuition
3.13.3 Mathematics behind Large margin classification
3.13.4 Kernels

Unit IV: Neural Network

4.1 Neuron models and basic learning rules
4.2 McCulloch-Pitts neuron model
4.3 Terminologies: Weights, effective input, threshold, activation function
4.4 Generalization of the neuron model: discrete neurons and continuous neurons
4.5 Activation function of continuous neuron: sigmoid function
4.6 A neuron model with augmented input
4.7 Single layer neural network and multi-layer neural network
4.8 Basic steps for using a neural network
4.8.1 Learning
4.8.2 Recall
4.8.3 Basic diagram of learning and Recall
4.9 General learning rule for one neuron
4.10 Perception learning rule
4.11 Delta Learning rule
4.12 Multilayer neural networks and back- propagation

List of Laboratory Sessions

1. Explore the relationship between model complexity and generalization performance, by adjusting key parameters of various supervised learning models. Part 1 of this lab will look at regression and Part 2 will look at classification.
2. Write a function that fits a polynomial Linear Regression model on the *training data* `X_train` for any degree.
3. Train two models: a non-regularized Linear Regression model (default parameters) and a regularized Lasso Regression model (with parameters `alpha=0.01`, `max_iter=10000`) on polynomial features of any degree. Return the R squared score for both the Linear Regression and Lasso model's test sets.

4. Using any two training data, train a DecisionTreeClassifier with default parameters and random_state=0. Then identify important features found by the decision tree.
5. Write Program for Perception learning and Delta rule.
6. Write program to Classify data using one neuron.
7. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
8. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set

Reading List

Essential Reading

Albon, C. (2018). Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.".

Coelho, L. P., & Richert, W. (2015). Building machine learning systems with Python. Packt Publishing Ltd.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). Data mining: practical machine learning tools and techniques (Morgan-Kaufman Series of Data Management Systems).

Additional Reading

Bowles, M. (2015). Machine learning in Python: essential techniques for predictive analysis. John Wiley & Sons.

Garreta, R., & Moncecchi, G. (2013). Learning scikit-learn: machine learning in python. Packt Publishing Ltd.

Date: March, 2022.

GSE101 Analytical Skills

Module Code and Title **:** GSE101 Analytical Skills
Programme **:** University-wide module
Credit **:** 12
Module Tutors **:** Sangay Thinley

General Objective

The module aims at developing critical and analytical thinking skills of students to enhance their creativity and ability to think laterally that will aid problem solving and decision making abilities. With these essential analytical thinking and problem solving skills students gain an edge in a competitive world.

Learning outcomes

On completion of the module, students will be able to:

- articulate thinking paradigms;
- explain creativity and barriers to creative thinking;
- apply creative thinking skills to spot unnoticed opportunities;
- describe problem solving process;
- apply appropriate problem solving tools to a given issue;
- evaluate issues to make informed decisions;
- generate creative solutions by using appropriate methods.

Teaching and learning approaches

Approach	Hours per week	Total Credit Hours
Lectures	1	15
Group and Panel Discussions, Presentations, Case Study	1	15
Role Plays/Demonstrations, Mock sessions, Audio visuals	2	30
Independent Study, Reflection, Written Assignments, Project Work , Individual Reading	4	60
Total		**120**

Assessment approach

A. Written Assignment: Portion of final Marks-20%

Students will be required to complete one written assignment on the contemporary issue of a subject. The required data and contextual information will be provided to students. Students will be required to read, analyse and interpret the data and contextual information, and communicate the result to the intended audience. Wherever there is a need, students should substantiate the existing data with their own data collection. The length of the assignment should be anywhere between 1000 and 1500 words.

Criteria:
4%-Originality and creativity
2% -Clarity of the points and opinions
4% -Reliability of data and accuracy of data interpretation
8%- Analysis of the issue
2% -Overall effectiveness of writing style

B. Class Participation: Portion of the final Marks-10%

Students will participate in class discussions, contributing their ideas and opinions about the methods and tools being taught in the module.

Criteria:
2%-frequency of participation in class
3%-quality of comments –involving critical thinking and analysis of information and reasoning
5%-contribution in a group discussion in class –understanding of group dynamics and processes

C. Case Analysis and Presentation: Portion of Final Marks-30%

Students will solve one case study in a group which will be assessed in two components. The case can be related to any field of knowledge such as engineering, climate change, biotechnology, sustainable development, procurement, production, marketing, strategic management, human resource and current economic and social development.

1. **Written**
   Criteria:
   5% identifying the problem
   10% choosing the right approach for the analysis and solving the problem
   5% drawing the correct conclusion with a recommendation
2. **Presentation**
   Criteria:
   2% Creativity in delivery of the presentations;
   2% Visual appeal
   2% Confidence
   4% Content analysis

D. Panel Discussion: Portion of Final Mark-20%

A group of students will be required to discuss a topical issue such as climate change, green procurement, disruptive innovation, and big data moderated by a peer.

Criteria:
5% Preparedness on the topic
5% Relevance of the argument
5% Respect for other panelist views
5% Coherent and logical flow of ideas

E. Debate: Portion of the Final Mark-20%

Students in groups of four or five will debate on a given topic against another group.
Criteria:
5%-Language Proficiency
5%-Intelligence, ability and competence
5%-Logical thinking and reasoning
5%-Ability to use appropriate information

Overview of the assessment approaches and weighting

Areas of Assignment	**Quantity**	**Weighting**
A. Written Assignment	1	20%
B. Class Participation	NA (non-definite/should participate in the class discussion at least 5 times)	10%
C. Case Analysis & Presentation	1 + 1	30%
D. Panel Discussion	1	20%
E. Debate	1	20%
TOTAL		**100%**

Pre-requisite: None

Subject matter

UNIT I: Thinking process & Reflection

1.1. Introduction to the Thinking Process & Reflection
1.2. Concept of mind mapping
1.3. Metacognition and thinking about thinking
1.4. Thinking Paradigms: Lateral and Vertical thinking
- 1.4.1. Whole brain (system 1 and system 2)
- 1.4.2. Analytical
- 1.4.3. Critical
- 1.4.4. Creative
- 1.4.5. Logical
- 1.4.6. Scientific
- 1.4.7. Statistical
- 1.4.8. Systems
- 1.4.9. Visual
- 1.4.10. Ethical

UNIT II: Overview of analytical thinking skills

2.1. Concept of analytical skills
2.2. Competencies of analytical thinking
2.3. Benefits of analytical thinking
2.4. Analytical thinking process

2.5. Tools and techniques for analytical skills
2.6. Application of analytical thinking
2.7. Validity and strength in arguments

UNIT III: Creative Thinking

3.1. Definition of creativity
3.2. Creative thinking – Self-Assessment
3.3. Characteristics of a creative person,
3.4. Barriers to creativity and overcoming the barriers
3.5. Ways to enhance creative thinking (e.g. brain storming)
3.6. Methods of creativity

UNIT IV: Problem solving process

4.1 Understanding problem analysis
4.2 Conventional problem solving process
- 4.2.1 Present the problems
- 4.2.2 Ask solutions
- 4.2.3 Shoot down ideas
- 4.2.4 Make consensus

4.3 Creative problem solving process
- 4.3.1 Problem definition
- 4.3.2 Problem analysis
- 4.3.3 Generating possible solutions
  - 4.3.3.1 Brain storming process and rules
  - 4.3.3.2 Fishbone Analysis
  - 4.3.3.3 Mind mapping
- 4.3.4 Analysing the solutions
- 4.3.5 Selecting the best solution
- 4.3.6 Implementing the best solution
- 4.3.7 Planning the next course of action

4.4 Questioning techniques

UNIT V: Decision making process

5.1. Introduction to Decision making process
5.2. Six Thinking Hats
5.3. SWOT Analysis
5.4. Decision Tree analysis/what-if analysis
5.5. Pareto chart
5.6. Logical Framework Analysis

Reading List

Essential Reading

Bano, E. d. (2000). *Six Thinking Hats* (2nd ed.). New Delhi, India: Penguin India.

Michalko, M. (2006). *Thinkertoys: A handbook of creative-thinking techniques* (2nd ed.). Ten Speed Press.

Puccio, G.J., Mance, M. & Switalski, L.B. (2017). *Creativity Rising Creative Thinking and Creative Problem Solving in the 21st Century.* ICSC Press, International Center for Creativity, US

Treffinger, D. J. (2006). *Creative Problem Solving: An introduction* (4th ed.). Prufrock.

Additional Reading

Bano, E. d. (2008). *Creativity workout: 62 exercises to unlock your most creative ideas.* Ulysses Press.

Bano, E. d. (2009). *Lateral Thinking.* e-Penguin.

Bono, E. d. (2005). *Thinking course (Revised Edition).* Bernes and Nobel

Chopra, R. (n.d.). *Logical Critical Analytical Reasoning.* Galgoba Publications Pvt Ltd.

Eiffert, S. D. (1999). *Cross-train your brain: a mental fitness program for maximizing creativity and achieving success.* Amacom.

Kahneman, D. (2015). *Thinking fast and slow.* New York: Farrar, Straus and Giroux.

Scott, J. W. (2016). *Critical Thinking: Proven strategies for improving your decision making skills, retaining information longer and analyzing situations with simple logic ---- Logical thinking and critical thinking skills.* New Familiar Publishing.

Updated: January 2018

Year 3 Semester I

RSM301 Research Methods

Module Code and Title: RSM301 Research Methods
Programme: BSc in Environmental Science
Credit Value: 12
Module Tutor: Dr. Bishal Gurung

General Objective
This module will introduce students to basic research methods. It will provide an overview of quantitative and qualitative research methods. The module will teach design issues, finding and assessing research papers. It will provide skills and understanding of research process to write the research proposal.

Learning Outcomes
On completion of the module, students will be able to:

1. Discuss principles of quantitative and qualitative research.
2. Identify criteria involved in choosing an appropriate design.
3. Distinguish between research purpose, research question and objective.
4. Compare different data sources for its strengths and weakness.
5. Distinguish between various sampling methods.
6. Find and assess research articles.
7. Identify different elements in a research article.
8. Identify types of literature review.
9. Use reference manager tool in research process.
10. Apply ethical principles in a research proposal and literature review.
11. Understand how to organize a literature review.
12. Develop a preliminary research proposal.

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	4	60
Independent study	Written assignment	3	45
	Self-study	1	15
Total			**120**

Assessment Approach
Assessments will be carried out on a continuous basis through the following assignments:

A. Research Article Review: Portion of Final Marks: (15%)

Each student will select two research articles from journals identified by the tutor for review. The review should summarise and evaluate the elements of a research paper such as research problem, purpose, objectives, key variables, design, instrument, data source, scope and the main

ideas of the paper. Each review should be presented within 500 words, which will contribute to 7.5% of the assessment weighting, evaluated based on the following criteria.

Criteria	85 - 100%	71 - 84%	60 - 70%	<60%	Marks Assigned (%)
Content of Review	In depth analysis of the content (all major points discussed). Excellent summary. Communicates the key ideas/ themes/ findings with a high degree of clarity and insight. Engaging introduction and conclusion, both indicate the overall focus of the paper	Comprehensive analysis of the content (most of the major points discussed). Adequate summary. Communicates the key ideas/findings with considerable clarity, but lacks insight. Engaging introduction and conclusion, paper focus inconsistently supported.	Minimal analysis of the content (some of the major points discussed). Basic summary. Communicates the key ideas/themes/findings with little clarity or insight. Introduction and conclusion do not indicate the overall focus of the paper.	No content (missed all major points of the content). Incomplete summary. Key ideas/findings are not communicated clearly and/or missed in analysis. There is no clear introduction or conclusion.	/40
Reflection	Demonstrates thorough understanding of the article by listing all key findings and reflecting upon their implications.	Demonstrates considerable understanding of the article by listing all of the key findings.	Demonstrates some understanding of the article by listing some of the key findings but documentation is lacking in completeness.	Demonstrates little understanding of the article with few or no key findings reported.	/30
Organisation & Development of Ideas	Logical development of ideas through well- developed paragraphs, good use of transitions.	Logical organisation, paragraph development not perfected.	Logical organisation, paragraphs not fully developed.	No evidence of structure or organisation.	/15
Mechanics	Meets length requirement. 1-2 grammatical errors. 1-2 punctuation errors. APA Guidelines are meticulously followed. Adheres to Font/Spacing Guidelines.	Length requirement is met with adequate content 3-4 grammatical errors. 3-4 punctuation errors. Adheres consistently to APA guidelines, however, one error present. Adheres to	Length requirement is not met; minimal content 5-6 grammatical errors 5-6 punctuation errors Demonstrates little ability to adhere to APA guidelines, more than two errors present. Adheres to neither Font nor Spacing Guidelines.	Length requirement is not met; poor content Numerous grammatical errors (distracting) Numerous punctuation errors Neither style is used Does not adhere to	/15

		either Font or Spacing Guidelines, but not both.		Font/Spacing Guidelines	

B. Literature Review: Portion of the Final Marks: (20%)

Each student will conduct a brief literature review on the chosen topic of interest. The review could add on to the “Literature Review” section of the research proposal assignment (Assessment C). It must provide an overview of the topic, identify research problems and question to address. The literature review should be of 1500 words and it will be assessed based on the following criteria:

Criteria	80 - 100%	60 - 79%	<60%	Marks Assigned (%)
Title and Structured Abstract	The abstract has appropriate sections and concisely reflects the content of the review.	The abstract reflects the content of the review, but is unclear or could be more concise.	Structured abstract not included.	/5
Introduction	Provides excellent background information, including relevance of the review. Research question and purpose is clearly articulated. Context and rationale for the review is clearly described.	Research question is articulated, but relevance and context are not adequately described.	Relevance of the review is not explained. Research question is poorly articulated or absent.	/15
Methods	Search methods are appropriate for the topic and clearly described. Search strategy includes appropriate search terms and relevant synonyms. Inclusion/exclusion criteria are articulated with sound rationale.	Search methods are mostly appropriate for the topic. Description of search methods missing some details. Search includes most relevant synonyms. Article inclusion/exclusion criteria adequately articulated.	Description of search methods inadequate or absent. Search is inadequate (missing key databases or search terms). Article inclusion/exclusion criteria not listed or unclear.	/10
Results/ Discussion	Synthesizes and analyzes the literature accurately. Focuses on central ideas, and critically evaluates sources & perspectives. Identifies gaps & controversies, as well as study limitations if relevant.	Summarizes the literature, but misses some key points. Does not adequately evaluate sources and perspectives. Does not relate findings to the research question.	Lists findings, instead of synthesizing literature. Provides summary instead of critique. Too few sources discussed. Sources are inappropriate for the research question.	/35

Conclusion	Clearly summarizes literature review results. Identifies implications for practice and/or research. Provides a compelling argument/ conclusion that is appropriately qualified (given study limitations).	Summarizes literature review results. Conclusion is weak or unsupported by evidence.	Lacks a final summary / conclusion. Implications for research or practice not articulated.	/15
Referencing	All sources are correctly cited in an existing bibliographic style (APA).	A few minor errors in citation formatting.	Cittation format inconsistent. Citations incomplete or missing.	/10
Quality of Writing	Prose is clear. Thoughts are well organized. Style and vocabulary are appropriate for a research paper. No typos or grammatical errors.	Writing is acceptable, but could be more polished. Style is too informal for a research paper. Inadequate proofreading.	Poorly written. Frequent grammatical or spelling errors.	/10

C. Research Proposal Assignment: Portion of the Final Marks: (25%)

Each student will write a research proposal on the chosen topic. The research proposal should include introduction (purpose, problem statement and objectives), literature review, and methods. The research proposal should not be less than 2000 words. It will be assessed based on the following criteria:

Criteria	**<60%**	**60 - 79%**	**80 - 100%**	**Marks Assigned (%)**
Introductory Matters: Title and Abstract	Title or abstract lacks relevance or fails to offer appropriate details about the educational issue, variables, context, or methods of the project.	Title and abstract are relevant, offering details about the research project.	Title and abstract are informative, succinct, and offer specific details about the educational issue, variables, context, and proposed methods of the study.	/5
Introduction: Problem, Significance, & Purpose of the Study	Research issue is identified, but statement is too broad or fails to establish the importance of the problem. The research purpose, questions, hypotheses, definitions or variables and controls are poorly formed, ambiguous, or not logically connected	Identifies a relevant research issue. Research questions are succinctly stated, connected to the research issue, and supported by the literature. Variables and controls have been identified	Presents a significant research problem related to the chemical sciences. Articulates clear, reasonable research questions given the purpose, design, and methods of the project. All variables and controls have been appropriately defined. Proposals are clearly supported from the	/15

	to the description of the problem. Unclear connections to the literature.	and described. Connections are established with the literature.	research and theoretical literature. All elements are mutually supportive.	
Literature Review	A key component was not connected to the research literature. Selected literature was from unreliable sources. Literary supports were vague or ambiguous.	Key research components were connected to relevant, reliable theoretical and research literature.	Narrative integrates critical and logical details from the peer-reviewed theoretical and research literature. Each key research component is grounded to the literature. Attention is given to different perspectives, threats to validity, and opinion vs. evidence.	/15
Methods: Research Design	The purpose, questions, and design are mutually supportive and coherent. Attention has been given to eliminating alternative explanations and controlling extraneous variables. Appropriate and important limitations and assumptions have been clearly stated.	The research design has been identified and described in sufficiently detailed terms. Some limitations and assumptions have been identified.	The research design is confusing or incomplete given the research questions and strategy. Important limitations and assumptions have not been identified.	/20
Methods: Instruments	Descriptions of instruments and techniques included. Evidence of the validity and reliability was presented.	Instruments and observation protocols were identified by name and described.	Description of the instruments and techniques were confusing, incomplete, or lacked relevance to the research questions and variables.	/15
Methods: Procedures	Procedures were thorough, coherent, and powerful for generating valid and reliable data. Procedures were replicable.	Procedures for project were identified and described in an appropriate fashion.	Procedures were confusing, incomplete, or lacked relevance to purpose, research questions, or sampling strategy.	/20

Methods: Data Analysis	Analytical methods were sufficiently specific, clear, and appropriate for the research questions.	Both descriptive and statistical methods were identified. Level of significance was stated.	Descriptive or statistical methods were confusing, incomplete or lacked relevance to the research questions, data, or research design.	/10

D. Semester-end Examination: Portion of the Final Marks: (40%)

Semester-end examination will cover the subject matter taught in the class. The exam will assess students' knowledge of the content using Blooms Taxonomy. A common end-semester examination will be conducted for all programmes offering the module.

Overview of the assessment approaches and weighting:

Areas of assignments	Quantity	Weighting (%)
A. Research Article Review	2	15
B. Literature Review	1	20
C. Research Proposal Assignment	1	25
D. Semester-end Examination	1	40

Pre-requisites: None

Subject Matter:

Unit I: Introduction

1.1. Meaning of Research
1.2. Objectives of Research
1.3. Motivation of Research
1.4. Types of Research: Descriptive, Analytical, Applied, Fundamental, Qualitative and Quantitative, Conceptual and Empirical Research
1.5. Research Approaches
1.6. Significance of Research
1.7. Research method vs. Methodology
1.8. Research and Scientific Method
1.9. Importance of Knowing How Research is Done
1.10. Research Process
1.11. Criteria of Good Research

Unit II: Research Problem

2.1. What is Research Problem
2.2. Selecting the Research Problem
2.3. Necessity of Selecting the Problem
2.4. Technique Involved in Defining a Problem

Unit III: Review of Literature

3.1. Concept and need for review of literature
3.2. Locating resources for literature review
3.3. Organising and cite literature using reference manager (Zotero)
3.4. Formulation research questions and hypotheses
3.5. Types of research hypothesis

Unit IV: Research Design

4.1. Meaning of Research Design
4.2. Need for Research Design
4.3. Features of a Good Design
4.4. Important Concepts Relating to Research Design
4.5. Differences quantitative and qualitative research designs
4.6. Quantitative research designs
  4.6.1. Basic principles of experimental designs (replication, randomisation and control)
  4.6.2. Important experimental designs (before-and-after without control, after-only with control, before- and after with control)
  4.6.3. Quasi-experimental designs
  4.6.4. Cross-sectional and longitudinal study designs
  4.6.5. Correlational designs
4.7. Qualitative study design - case study, ethnography, phenomenology, grounded theory

Unit V: Measurement and Sampling

5.1. Measurement Process: Conceptualization and Operationalization
5.2. Reliability and validity
5.3. Levels of measurement
5.4. Designing instruments
5.5. Types of Sampling Design: Probabilistic and non-probabilistic sampling

Unit VI: Ethics and Data Collection

6.1. Ethical principles in research
6.2. Secondary and primary data
6.3. Qualitative data collection: Individual interviews, focus groups, observations
6.4. Quantitative data collection: Survey, lab and data logger

Reading List

Essential Reading

Bhattacherjee, A. (2012) *Social science research: Principles, methods, and practices*. Published under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Open access book available at: https://open.umn.edu/opentextbooks/textbooks?term=research&commit=Go

Bryman, A. (2016). *Social research methods*. Oxford university press.

Jhangiani,R. S., I-Chant A. Chiang, I. A., Cuttler, C & Leighton, D. C (2019). *Research methods in psychology*. Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Open access book available at: https://open.umn.edu/opentextbooks/textbooks?term=research&commit=Go

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.
Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners*. Sage Publications Limited.

Additional Reading

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
Ghosh, B. N. (2007). Scientific Method and Social Research (Revised edition). New Delhi: Sterling Publishers Pvt. Ltd.
Grix, J. (2002). The foundations of Research. Palgrave Foundations. New York. (accession no.001.4 GRI).
Patten, M. L., & Newhart, M. (2017). *Understanding research methods: An overview of the essentials*. Taylor & Francis.
Punch, K. F. (2005). Introduction to Social research: Qualitative and Quantitative Approaches. London: Sage Publication.
Wolf, H. K. and Pant, P. R. (2002). Social Science Research and Thesis Writing. Kathmandu: Buddha Academic.

Date: November, 2020

APC317 Data Warehousing and Data Mining

Module Code and Title: APC317 Data Warehousing and Data Mining
Programme: BSc in Data Science
Credit Value: 12
Module Tutors: J.Gurubalan

General Objectives:

This module introduces the multidisciplinary field of data mining and introduces techniques for preprocessing the data before mining. It also introduces, data warehousing and on-line analytical processing (OLAP) to manage data with dimensional models. It gives an overview of the principles and applications of market basket analysis, various techniques for frequent item set mining, also introduces techniques for mining multilevel association rules and multidimensional association rules. It also describes methods for data classification and prediction, including decision tree induction, Bayesian classification, and rule-based classification.

Course Outcomes:

On completion of the module, students will be able to:

1. Identify the main characteristics of different data mining techniques through observation of their operations.
2. Identify various data mining functionalities that can be applied to different real-world problems.
3. Identify the main characteristics of different data warehousing techniques through observation of their operations.
4. Apply OLAP (Online Analytical Processing) and OLTP (Online Transaction Processing) models to analyse database information.
5. Analyse data characteristics and outcome of data mining process.
6. Implement the main algorithms in data warehousing and data mining in a computationally efficient way.
7. Analyze the data and carry out supervised, un-supervised learning processes.
8. Formulate new solutions for data warehousing and data mining problems by improving and combining current techniques.
9. Extract data from data warehouse and transform into refined/actionable information using data mining tool.
10. Produce clean data after processing.

Learning and Teaching Approach:

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**90**
	Laboratory Session	3	
Independent reading	Assignment	1	**30**
	Self-study	1	
Total			**120**

Assessment approach:

A. Assignment: 10%

This assignment is focused on identifying and characterizing a dataset and practice some data preprocessing techniques on it. Individual assignment will be given to students to develop their critical thinking and analytical skills. It will be assessed for 50 marks and later it will be scale down to 10%.Each student will have to submit a report of maximum of 500 words.

Marking criteria:

- Describe datasets & discuss data mining technique : 4%
- Data preprocessing : 4%
- Writing report : 2%

The report shall be assessed based on the following criteria:

- Assignment Objective, results and conclusion : 1%
- Exposition (language, references, layout, graphics) : 1%

B. Case Study: 40%

The first case study will be given to students to evaluate their practical and data warehouse implementation skills. Groups of 3 students will be formed to carry out this case study.

Marking criteria: Case study-I (Group) -20%

- Problem analysis & Select interesting data for analysis : 6%
- Extract data from OLTP : 6%
- Build data warehouse schema : 4%
- Select & Aggregate measurement and build cube : 4%

The second case study is an extension of the previous one. Same groups should continue working to discover patterns that improve decision making process and analytical thinking. Each group will have to submit a report of maximum of 1000 words.

Marking criteria: Case Study-II (Group)-20%

- Find relation and interdependence of data : 6%
- Implementation : 7%
- Writing report : 3%
- Presentation /viva : 4 %

C. Class test: 10%

One class test will be conducted upon coverage of at least 60% of the subject matter. This is a written conceptual test conducted within the classroom for a duration of 60 minutes. Students have to write answers on a sheet of paper in the classroom within the predefined time. This class test will be conducted for 25 marks and later it will be scaled down to 10%.

D. Semester-End Examination: 40%

The exam at the end of the semester will be a closed book for a minimum of two hours and evaluated for 40 marks. The exam will evaluate the students' overall understanding and critical thinking ability with regards to the module. All the units of the subject matter will be included for the examination.

Areas of Assignments	Quantity	Weighting
A. Assignment	1	10%
B. Case Study	2	40%
C. Class Test	1	10%
D. Semester-end Examination	1	40%
		100%

Pre-requisites: CAE201 Database Systems

Subject Matter

Unit 1: Introduction to Data mining

1.1. Introduction- Data Mining
  1.1.1. Motivation for Data Mining
  1.1.2. Data Mining-Definition and Functionalities
1.2. Data Features
  1.2.1. Attribute types
  1.2.2. Basic statistical description
  1.2.3. Measuring data similarity and dissimilarity.
1.3. DM task primitives
1.4. Integration of a Data Mining system with a Database or a Data Warehouse
1.5. Issues in DM – KDD Process

Unit II: Data Pre-processing

2.1. Purpose of pre-process data
2.2. Tidy data
  2.2.1. Data Structure
  2.2.2. Data Semantics
  2.2.3. Handle messy data
2.3. Data cleaning
  2.3.1. Missing Values
  2.3.2. Noisy Data
  2.3.3. Outlier Detection.
2.4. Data Integration and transformation
2.5. Data Reduction
  2.5.1. Data cube aggregation
  2.5.2. Dimensionality reduction
  2.5.3. Data Compression
  2.5.4. Numerosity Reduction
2.6. Data Mining Primitives
2.7. Languages and System Architectures
  2.7.1. Task relevant data
  2.7.2. Kind of Knowledge to be mined
  2.7.3. Discretization and Concept Hierarchy

2.8. Data pipeline

Unit III: Overview and concepts Data Warehousing

3.1. Why reporting and Analyzing data
  3.1.1. Raw data to valuable information
  3.1.2. Lifecycle of Data
3.2. The building Blocks: Defining Features
  3.2.1. Data warehouses and data marts
  3.2.2. Overview of the components
  3.2.3. Metadata in the data warehouse
3.3. Need for data warehousing
3.4. Basic elements of data warehousing
3.5. Trends in data warehousing

Unit: IV Architecture of Data Warehouse

4.1. Introduction OLAP (Online analytical processing) definitions
4.2. Difference between OLAP and OLTP
4.3. Dimensional analysis
  4.3.1. Purpose of Cubes
  4.3.2. Drill-down and roll-up
  4.3.3. Slice and dice or rotation
4.4. OLAP models
  4.4.1. ROLAP versus MOLAP
4.5. Defining schemas
  4.5.1. Stars, snowflakes and fact constellations
4.6. HOLAP

Unit: V Concept Description and Association Rule Mining

5.1. Concept description
  5.1.1. Data Generalization and summarization
  5.1.2. Based characterization
5.2. Attribute relevance
5.3. Class comparisons Association Rule Mining
  5.3.1. Market basket analysis
  5.3.2. Basic concepts
  5.3.3. Finding frequent item sets: Apriori algorithm
  5.3.4. Jaccard similarities
  5.3.5. Generating rules
  5.3.6. Improved Apriori algorithm
5.4. Incremental ARM
5.5. Associative Classification
5.6. Rule Mining

Unit: VI Classification and Prediction

6.1. Classification and prediction
6.2. Issues regarding Classification and prediction
6.3. Classification methods
  6.3.1. Decision tree
  6.3.2. Bayesian Classification
  6.3.3. Rule based
  6.3.4. CART
6.4. Introduction of tools such as DB Miner /WEKA/DTREG DM Tools

List of Laboratory sessions:

1. Demonstration of preprocessing on dataset using student.arff, student.arff datasets.
2. Load the “weather.arff ” dataset in Weka and run the ID3 classification algorithm. What problem do you have and what is the solution?
3. Demonstration of Association rule process on dataset contactlenses.arff ,test.arff using apriori algorithm
4. To implement A-priori algorithm to find the frequent patterns in the given dataset. Students can use programming language of their choice to code.
5. Demonstration of classification rule process on dataset student.arff using j48 Algorithm.
6. Demonstration of classification rule process on dataset employee.arff using naïve bayes algorithm
7. Use of WEKA tool to use various association mining algorithms on datasets and evaluate them based on pattern evaluation measures.
8. Demonstration of clustering rule process on dataset iris.arff using simple k-means
9. Use of WEKA tool to use various classification algorithm
10. Demonstration of clustering rule process on dataset student.arff using simple k- means on datasets and evaluate them on the basis of accuracy and other parameters.
11. Use of WEKA tool to use various clustering algorithms on datasets and evaluate them based on cluster quality and other parameters.

Reading List:

Essential Reading:

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Bhatia, P. (2019). *Data mining and data warehousing: principles and practical techniques*. Cambridge University Press.

Inmon, W. H. (2005). *Building the data warehouse*. John wiley & sons.

Additional Reading:

Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India.

Mourya, S. K., & Gupta, S. (2012). *Data mining and data warehousing*. Alpha Science International, Ltd.

Date: March, 2022

IST201 Sampling Theory

Module Code & Title:	IST201 Sampling Theory
Programme:	BSc in Statistics
Credit Value:	12
Module Tutor:	Dr. Bishal Gurung

General Objective

This module is meant to expose the students to the techniques of drawing representative samples from various populations and then preparing them on the mathematical formulations of estimating the population parameters based on the sample data using Spreadsheets and R. The students would also be exposed to the real-life applications of sampling techniques and estimation of parameters.

Learning Outcomes

On completion of the module, students will be able to:

1. Compare basic concepts of various sampling techniques.
2. Explain the importance of sampling over complete enumeration in many situations of practical importance.
3. Apply simple random sampling for real datasets.
4. Differentiate ratio, difference and regression estimators for estimation of population mean or population total.
5. Distinguish between cluster sampling, stratified and systematic sampling procedures.
6. Estimate the proportions and confidence intervals.
7. Determine the sample size of a sample.
8. Implement the concepts of inverse sampling for rare phenomena.
9. Differentiate between sampling and non-sampling error.

Learning and teaching approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	3	**60**
	Laboratory sessions	1	
Independent Study	Written assignment/Quizzes	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. **Assignment: (10%)**

The assignment will be completed during the first half of the semester. Set of questions (not over 10) based on the first 3 units covered will be provided to the students. Students will have to exhibit their understanding of the concepts through their solutions to the questions. It is an individual work and the date of submission will be two weeks from the given date. In order to avoid plagiarism different sets of questions with same difficulty level will be given to the students. The assignment will be evaluated based on the following criteria:

- Correctness of the solutions: 4 marks
- Logical flow in the process: 2 marks
- Uniqueness of the solution: 2 marks
- Use of appropriate terms and symbols: 2 marks

If it is confirmed that a student has plagiarized the assignment, then as per the academic norm, the student will be awarded zero mark or made to redo the assignment based on the convenience

of the tutor. In the event where plagiarism is suspected but could not be confirmed, the student in question will be called for viva-voice or asked for oral presentation of the solution. Based on the viva-voice or oral presentation, the tutor will either maintain the initial mark or reduce it.

B. **Laboratory Assessment: (15%)**
The lab assessment will be carried out as follows:

- **Laboratory report -5%**

Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Laboratory Exam -10%**

A laboratory examination for the duration of 2 hours will be conducted towards the end of the semester after the last day of regular teaching.

C. **Mid-Semester Examination: (15%)**
Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of 3 hours. Examination will be conducted out of 60 marks and one-fourth of the scores will be taken as final score.

D. **Semester-end Examination: (60%)**
Semester end Examination for the duration of three hours and will be for 60 marks. All the units of the subject matter will be included for the examination.

Overview of Assessment Approaches

Areas of assignments	Approach	Quantity	Weighting
A. Assignment		1	10%
B. Lab	Laboratory Report	10	5%

	Laboratory Exam	1	10%
C. Mid-Semester Examination		1	15%
D. Semester-end Examination		1	60%

Pre-requisites: PLT203 R Programming for Data Analysis

Subject Matter

Unit I: Introduction to Sampling

1.1 Notion of sample and population
1.2 Complete enumeration versus sampling
1.3 Types of sampling: non-probability and probability sampling
1.4 Basic principle of sample survey

Unit II: Simple Random Sampling

2.1 Definition and procedure of selecting a sample
- **2.1.1** Population and unit, finite and infinite population
- **2.1.2** Sampling frame
- **2.1.3** Parameter and parameter space
- **2.1.4** Complete enumeration and sample survey
- **2.1.5** Sampling and non-sampling errors

2.2 Simple random sampling (SRS) with replacement
- **2.2.1** Sampling scheme
- **2.2.2** Estimation of population mean and variance
- **2.2.3** Estimation of population proportion

2.3 Simple random sampling without replacement
- **2.3.1** Sampling scheme
- **2.3.2** Estimation of population mean and variance
- **2.3.3** Estimation of population covariance
- **2.3.4** Estimation of population proportion
- **2.3.5** Estimation of domain mean and total

2.4 Sample size determination
- **2.4.1** Consideration of cost of a survey
- **2.4.2** Consideration of the efficiency of estimator with given variance or coefficient of variation or margin of permissible error

Unit III: Sampling Techniques

3.1 Stratified random sampling technique
- **3.1.1** Definition and advantages of stratified random sampling
- **3.1.2** Estimation of population mean
- **3.1.3** Estimation of population proportion

3.2 Different allocations and their comparison with SRS
3.3 Systematic Sampling technique
- **3.3.1** Linear systematic sampling
- **3.3.2** Efficiency of systematic sampling

3.3.3 Circular systematic sampling

3.4 Estimates of population mean and total
3.5 Variances of estimates of population mean and total
3.6 Comparison of systematic sampling with SRS and stratified sampling

Unit IV: Other Sampling Techniques

4.1 Inverse sampling technique
4.2 Cluster sampling technique
4.3 Ratio, difference and regression estimators
4.4 Multi-stage sampling with equal probability
4.5 Double sampling
4.6 Successive sampling

Unit V: Sampling and Non-sampling errors

5.1 Sampling errors: sources and classification
5.2 Non-sampling errors: sources and classification
5.3 Non-response survey techniques
5.4 Imputation techniques
5.5 Measurement errors
5.6 Repeated measurement techniques

List of Laboratory Sessions

1. Estimation of proportions and confidence interval of estimators
2. Determination of Sample size in various sampling
3. Lab session on Inverse sampling, stratified sampling, Cluster sampling
4. Non-response survey techniques, Imputation techniques

Reading List

Essential Reading

Cassel, C. M., Sarndal, C. E., and Wretman, J. H. (1992). *Foundations of Inference in Survey Sampling*, John Wiley, New Delhi.

Chaudhari, A., and Stenger, H. (2005). *Survey Sampling Theory and Methods*, Chapman and Hall, New Delhi.

Chaudhari, A., and Voss, J. W. E. (1988). *Unified Theory and Strategies of Survey Sampling*, North Holland, New Delhi.

Cochran, W. G. (2005). *Sampling Techniques*, John Wiley, New Delhi.

Murthy, M. N. (1977). *Sampling Theory and Methods*, Statistical Publishing Society, Kolkatta.

Additional Reading

Hansen, M. H., Hurwitz, W. H., and Madow, W. G. (1993). *Sample Survey Methods and Theory*, Vol. I and Vol. II., John Wiley, New Delhi.

Hedayat, A. S., and Sinha, B. K. (1991). *Design and Inference in Finite Population Sampling*, John Wiley, New Delhi.

Singh, D., and Chaudhary, F. S. (2020). *Theory and Analysis of Sample Survey Designs*, New Age International Pvt. Ltd., New Delhi.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984). *Sampling Theory of Surveys with Applications*, Indian Society of Agricultural Statistics, New Delhi.

Thompson, S. K. (2012). *Sampling*, John Wiley, New Delhi.

Date: December, 2021

DSC302 Unsupervised Learning

Module Code & Title: DSC302 Unsupervised Learning
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

This module introduces student to one of the main types of Machine Learning: Unsupervised Learning. Students will learn how to find insights from data sets that do not have a target or labeled variable. Students will learn several clustering and dimension reduction algorithms for unsupervised learning as well as how to select the algorithm that best suits your data. The hands-on section of this module focuses on using best practices for unsupervised learning with python.

Learning Outcomes

On completion of the module, students will be able to:

8. Identify the kind of problems suitable for Unsupervised Learning approaches.
9. Explain the curse of dimensionality, and how it makes clustering difficult with many features.
10. Describe and use common clustering and dimensionality-reduction algorithms.
11. Apply clustering points where appropriate.
12. Improve the model using feature selection methods.
13. Interpret results from clustering methods in scikit-learn, including k-means, agglomerative clustering, hierarchical clustering, and DBSCAN.
14. Demonstrate how unsupervised learning can be used to improve supervised prediction.
15. Explain the tradeoffs and assumptions inherent in different clustering techniques.
16. Explain the taxonomy of association rules.

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**75**
	Laboratory Session	2	
Independent study	Written Assignment	1	**45**
	Self-study	2	
Total			**120**

Assessment Approach

E. **Problem-solving Assignment: (10%)**
The assignment will be given during the first half of the semester. The assignment will comprise of problem-solving based on some questions provided to them. Assignments will be assessed using following criteria:

- Correctness of the solutions: 4%

- Logical flow in the process: 2%
- Uniqueness of the solution: 2%
- Use of appropriate terms and symbols: 2%

F. **Laboratory: (35%)**

Lab assessment will be carried out as follows:

- **Laboratory report- 10%**

Students will submit a weekly record of the lab session to keep track of his/her progress. Each weekly report will be assessed for 10 marks. At the end of all the lab sessions, the average of scores in each lab report will be taken as the final score for lab report. This will be finalized before their lab exam.

- **Laboratory Exam- 25%**

  A lab exam for the duration of 3 hour will be conducted towards the end of a semester after the last day of regular teaching.

G. **Mid-Semester Examination: The portion of Final Marks: (15%)**

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 15% of the scores will be taken as final score.

H. **Semester-end Examination: The portion of Final Marks: (40%)**

Semester end Examination for the duration of two and half hours will be for 50 marks. All the units of the subject matter will be included for the examination.

Overview of assessment approaches

Area of Assignments		**Quantity**	**Weighting**
E. Problem-solving Assignment		1	10%
F. Lab Assignment	Lab report	10	10%
	Lab exam	1	25%
G. Mid-Semester Examination		1	15%
H. Semester-end Examination		1	40%

Prerequisite: DSC201 Machine Learning

Subject Matter

Unit I: Fundamentals of clustering techniques

1.4 Agglomerate Clustering
1.5 Dendogram analysis
1.6 K-Means clustering
  1.6.1 Defining number of clusters
  1.6.2 Defining cluster K-Mean algorithm: Lloyd's algorithm
  1.6.3 Cluster validation
1.7 Hierarchical clustering
  1.7.1 agglomerative hierarchical clustering
  1.7.2 Clustering distance metric
  1.7.3 Linkage methods: Single linkage, Complete linkage, Average linkage
  1.7.4 Shortcoming of each linkage type
  1.7.5 Hierarchical clustering with factors
  1.7.6 Choosing a hierarchical clustering algorithm
1.8 K-means vs Hierarchical clustering
1.9 Expectation
1.10 Anonomy detection
1.11 Gaussian Mixture Models (Expectation maximization clustering)

Unit II: Association Rules

2.1 Fundamental of association rules
  2.1.1 Representation
2.2 A taxonomy of Association Rules
2.3 Measures of Interest
  2.3.1.1 , itemsets, support, confidence, lift etc.
  2.3.1.2 Minlen, Maxlen, target
  2.3.1.3 Frequent itemset generation
  2.3.1.4 General rule
  2.3.1.5 Association mining rule
2.4 Types of Association Rules
  2.4.1 Binary association rules
  2.4.2 Quantitative association rules
  2.4.3 Fuzzy association rules
2.5 Apriori
  2.5.1 Appriori Algorithm
  2.5.2 Rule generation in apriori
  2.5.3 Market basket analysis and real-world problems
  2.5.4 Challenges in association rules mining

Unit III: Dimensionality Reduction

3.1 compression and visualization as motivation for the dimension reduction
3.2 The curse of dimensionality Data
3.3 Feature extraction
  3.3.1 Principal Component Analysis (PCA)
  3.3.2 Independent Component Analysis (ICA)
  3.3.3 Singular Value Decomposition (SVD)

Unit IV: Feature Selection Methods

4.1 Feature selection techniques
- 4.1.1 Expert-knowledge-based techniques
- 4.1.2 Feature ranking
- 4.1.3 Subset selection techniques
  - 4.1.3.1 Embedded methods
  - 4.1.3.2 Wrapper methods
  - 4.1.3.3 Filter methods

Unit V: Advanced Methods

5.1 Reinforcement learning (RL)
- 5.1.1 Definition with examples
- 5.1.2 Formulating RL
- 5.1.3 RL agent: policy, value function and model
- 5.1.4 Example: Maze, Tic-Tac-Toe
- 5.1.5 RL and Tic-Tac-Toe
- 5.1.6 Exploration vs Exploitation
- 5.1.7 Markov Decision Problems: formulation

5.2 Neural network for unsupervised learning
- 5.2.1 CNN for image processing

5.3 Ensembles method
- 5.3.1 Construction of Ensembles: Bayesian voting, Manipulating the training examples, Manipulating input features and output target, injecting randomness.

5.4 Introduction to TensorFlow: One-liner TensorFlow

List of laboratory sessions

1. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering.
2. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions.
3. Write a program to implement Hierarchical clustering algorithms
4. Write a program to use association rules in data mining
5. Write your own program to use PCA and ICA for reducing the dimension of the dataset
6. Write program to implement Apriori algorithm.

Reading List

Essential Reading

Albon, C. (2018). Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.".

Hinton, G. E., & Sejnowski, T. J. (Eds.). (1999). Unsupervised learning: foundations of neural computation. MIT press.

Additional Reading

Bowles, M. (2015). Machine learning in Python: essential techniques for predictive analysis. John Wiley & Sons.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

Garreta, R., & Moncecchi, G. (2013). Learning scikit-learn: machine learning in python. Packt Publishing Ltd.

Date: March, 2022.

PMT301 Project Design and Management

Module Code and Title: PMT301 Project Design and Management

Programme: B.A Population and Development Studies

Credit Value: 12

Module Tutor: Tshering Dorji, Tashi Dorji

General Objective

This module will enable students to acquire knowledge on the principle aspects of project management and equip them with project management skills. The module willintroduce project management concepts and terminologies. It will also enable students to solve practical problems and apply some of the project managementtools and techniques.

Learning Outcomes

On the completion of this module, students will be able to:

1. Explain the key concepts and terminologies of project management
2. Discuss the steps and process of project management cycle
3. Discuss reasons for initiating projects or programme
4. Identify major milestones in each stage of the project cycle
5. Identify major management functions including planning and supervision,leadership, teamwork and effective field based monitoring of progress
6. Apply Logical Framework Approach (LFA) for result-based planning
7. Develop project proposal
8. State the differences between monitoring and evaluation
9. Carry out a case study to apply the skills and knowledge of projectmanagement

Learning and Teaching Approach

Approach	Hours per Week	Total Credit Hours
Lecture	3	45
Tutorials	1	15
Case Study	2	30
Independent Study	2	30
Total		**120**

Assessment approach

Assessments will be carried out on a continuous basis through the following assignments:

A. Case Study/Project Analysis: Portion of Final Mark: 35%

Students will be given two choices to conduct case study/project analysis. Students attempting first choice, have to visit a problem area to understand and identify thetype of project that

would benefit the community. The project proposal will be graded based on the following criteria:

- Business case development (clarity of problem statement and feasibility of solutions): **3.5**
- Planning (Resource, quality, communication and procurement planning): **3.5**
- Execution or implementation of project (Monitoring and Control using various management mechanism): **3.5**
- Project closure (presentation of deliverables): **1.75**

Students attempting 2nd choice have to visit project initiated areas to assess project framework, management and monitoring process. Students have to critical review all these aspects and marks will be provided based on following criteria:

- Understanding project design, management and monitoring process : **3.5**
- Critical evaluation of project design, management and monitoring process(Identifying key challenges/issues): **5.25**
- Recommendation of best practices: **3.5**

B. Presentation: Portion of Final Mark: 15%

Based on the case study, student will make one presentation for duration of 20 minutes. The presentation will be assessed based on the following criteria:

- Content: **7**
- Discussion: **5**
- Presentation skills: **3**

C: Semester End Examination: Portion of Final Mark: 50%

Semester end examination carries 50%. It will be conducted for the duration 3 hours.

Overview of the assessment approaches and weighting

Area of Assessment	**Quantity**	**Weighting**
A. Case Study	1	35%
B. Presentation	1	15%
C. Semester End Examination	1	50%
Total		**100%**

Pre-requisite: None

Subject Matter

Unit I: Introduction to Project Management

1.1. Definition and concepts of Project, Programme and Portfolio
1.2. Programme, project and portfolio management
1.3. Project life cycle

1.4. Project and programme manager and related concepts
1.5. Responsibilities and Competencies
1.6. Importance of initiating project

Unit II: Project Initiation

2.1. Develop Business case
2.2. Undertake Feasibility study
2.3. Establish Terms of Reference
2.4. Appoint Project Team
2.5. Set up Project Office

Unit III: Project Planning

3.1. Project plan
3.2. Resource plan
3.3. Financial plan
3.4. Quality plan
3.5. Risk plan
3.6. Acceptance plan
3.7. Communication plan
3.8. Procurement plan
3.9. Identifying Supplier

Unit IV: Project Execution

4.1. Building deliverables
4.2. Monitor and Control
4.3. Time Management
4.4. Cost and Quality Management
4.5. Change and Risk Management
4.6. Issue and Procurement Management
4.7. Acceptance and Communication Management

Unit V: Project Closure

5.1. Perform Project Closure
5.2. Review Project Completion
5.3. Evaluation of Project; meanings, concepts and its importance

Reading List

Essential Reading

Charvat, J. (2003). *Project management methodologies: selecting, implementing, and supporting methodologies and processes for projects*. John Wiley & Sons.

Duncan, W. R. (1996). A guide to the project management body of knowledge.Horine, G. (2009). *Absolute beginner's guide to project management*. Pearson

Education.

Kerzner, H. R. (2013). *Project management: a systems approach to planning,scheduling, and controlling*. John Wiley & Sons.

Khatua, S. (2011*). Project management and appraisal*. Oxford University Press.

Meredith, J. R., Mantel, S. J., & Shafer, S. M. (2016). *Project management: amanagerial approach*. Singapore: Wiley.

Potts, D. (2002). *Project planning and analysis for development*. Boulder: LynneRienner

Project Management Institute. (2014). *A guide to the project management body ofknowledge - pmbok guide.* Project Management Inst.

Additional Reading

A guide to the project management body of knowledge: PMBOK® guide. (2013).

Newton Square: Project Management Institute.

Hansen, J. R. (1986). *Guide to practical project appraisal: social benefit-cost analysisin developing countries*. Vienna: United Nations Industrial Development Organization.

Kerzner, H. R. (2002).*Strategic planning for project management using a project management maturity model*. John Wiley & Sons.

Pinto, J. K., & Pearson. (2016). *Project management: achieving competitiveadvantage*. Boston: Pearson.

Year 3 Semester II

AMT311 Introduction to Operation Research

Module Code & Title: AMT311 Introduction to Operations Research
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

This module introduces mathematical methods for formulating, solving and providing quantitative data for managerial decision making under the given constraints of resources. It also introduces decision making under deterministic and probabilistic situations. Linear programming models, transportation and assignment models, network models and decision trees will be discussed in detail. The students will also be trained to use TORA, Excel Solver and R to solve optimisation problems.

Learning Outcomes

On completion of the module, the students will be able to:

1. Formulate a given real-life optimisation problem in to a Linear Programming Problem (LPP) with a clear objective function and constraints.
2. Solve a given LPP using appropriate method and interpret the result.
3. Obtain the optimal solution to a given transportation model/ assignment model and estimate the least cost for such a model.
4. Draw a network diagram for a given project schedule and find the optimum duration using Critical Path Method (CPM).
5. Solve a network model using Programme Evaluation and Review Technique when the duration of each activity has the most optimistic and most pessimistic completion time.
6. Draw a decision tree for a decision making problem and use expected value criterion to make optimize cost/ profit.
7. Solve an integer programming problem using Branch and Bound algorithm or Cutting-Plane algorithm.
8. Formulate a goal programming and solve it using weights method or preemptive method.
9. Use TORA, Excel Solver and R to solve decision making problems.

Learning and Teaching Approaches

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Tutorial	1	
Independent Study	Written Assignment	1	**60**
	Self-study	3	
Total credit hours for the semester			**120**

Assessment Approach

A. Group Assignment: The portion of Final Marks: (15%)
Six problems from the first three units will be given to the students. Groups of three randomly selected students will be formed and each group is expected to solve the problems and submit the solutions. Grading will be done based on the choice of solution method, procedure, correctness of solution and interpretation of results.

B. Individual Assignment: The portion of Final Marks: (15%)
Five problems from the last three units will be given to the students. They are expected to solve the problems individually and submit the solutions. Grading will be based on the solution procedure and interpretation of results.

C. Computer Solution to Optimisation Problems The portion of Final Marks: (10%)
Six problems – one from each unit will be given. Students are expected to solve each problems using TORA or Excel Solver or R individually and upload the results in VLE as a PDF file. Grading will be based on the choice of appropriate package for each problem and the interpretation of results.

D. Mid-Semester Examination: The portion of Final Marks: (15%)
Mid-semester Examination will be conducted for a duration of two hours and the first four units of the subject matter will be included.

E. Semester-end Examination: Portion of Final Marks: (45%)
Semester-end Examination will be conducted for the duration of 2.5 hours and questions will be from all the units of the subject matter. Test blue print of the question paper and an answer key for grading will also be prepared along with the question paper.

Overview of assessment approaches and weighting:

Areas of Assessments	Quantity	Weighting
A. Group Assignment	1	15%
B. Individual Assignment	1	15%
C. Computer Solution	1	10%
D. Mid-semester Examination	1	13%
E. Semester-end Examination	1	45%

Pre-requisite: PLT203 R Programming for Data Analysis

Subject Matter

Unit I: Linear Programming Problems (LPP)

1.1 Introduction to LPP
1.2 Formulation of LPP
1.3 Types of solutions of LPP
1.4 Solution of LPP:
  1.4.1 Graphical Method
  1.4.2 Simplex Method
  1.4.3 Use of Artificial Variables in simplex method:
    1.4.3.1 Charnes' Big M method
    1.4.3.2 Two Phase Method
1.5 Degeneracy
1.6 Infeasible solutions
1.7 Duality in LPP
1.8 Dual Simplex Method
1.9 Sensitivity analysis
1.10 Computer solution to LPP with TORA, Excel Solver and R

Unit II: Transportation Problems (TP) and Assignment Problems (AP)

2.1 Introduction to Transportation Problem
2.2 TP as a case of LPP
2.3 Methods to obtain initial basic feasible solution to a TP:
  2.3.1 North West Corner Rule
  2.3.2 Matrix Minima Method
  2.3.3 Vogel's Approximation Method
  2.3.4 Solution of the TP by MODI method
  2.3.5 Degeneracy in TP
  2.3.6 Unbalanced transportation problems and their solutions.
2.4 Introduction to AP
2.5 AP as a complete degenerate form of TP
2.6 Hungarian Method for solving AP
2.7 Unbalanced Assignment Problems and their solutions
2.8 AP with restrictions
2.9 Computer solution to TP and AP using TORA, Excel Solver and R

Unit-III: Network Analysis

3.1 Network representation of a Project Rules for construction of a Network
3.2 Use of Dummy activity
3.3 The critical Path method (CPM) for constructing the time schedule for the project
3.4 Float (or shack) of an activity and event
3.5 Programme Evolution and Review Technique (PERT)
3.6 Probability considerations in PERT
3.7 Probability of meeting the scheduled time
3.8 PERT Calculation
3.9 Distinctions between CPM and PERT
3.10 Computer solution to network problems

Unit IV: Decision Analysis

4.1 Decision making under certainty
4.2 Decision making under risk
  4.2.1 Decision trees
  4.2.2 Expected value criterion
  4.2.3 Variations of the expected value criterion
4.3 Decision making under uncertainty

Unit V: Integer Linear Programming (ILP)

5.1 Examples of Integer Linear Programming
  5.1.1 Capital budgeting
  5.1.2 Self-covering problem
  5.1.3 Fixed-charge problem
  5.1.4 Either-or and If-then constraints
5.2 Integer programming algorithms
  5.2.1 Branch-and-Bound algorithm
  5.2.2 Cutting-Plane algorithm
  5.2.3 Computational considerations in ILP

Unit VI: Goal Programming

6.1 Goal Programming Formulation
6.2 Goal Programming Algorithms
  6.2.1 The Weights Method
  6.2.2 The Preemptive Method

Reading List
Essential reading

Taha, H. A. (2013). *Operations research: An introduction*. (9th ed.). Pearson Education India, New Delhi.

Hillier, F. S. (2021). *Introduction to operations research*. (11th ed.). Tata McGraw-Hill Education, New Delhi.

Vanderbei, R. J. (2020). Linear programming: foundations and extensions (Vol. 285). Springer Nature, New Delhi.

Additional reading

Bernard W. Taylor III. (1993). *Introduction to Management Sciences*. (4th ed.). Allyn and Bacon, New York.

Kanti Swarup, Manmohan, Gupta. (2018). *Operations Research*. (10th ed.). Sultan Chand and Sons Publishing Co., New Delhi:

Manmohan & Gupta P.K. (1987). *Operations Research and Statistical Analysis*. (3rd ed.). Sultan Chand and Sons Publishers, New Delhi.

Date: December, 2021

DSC303 Ethics and Security

Module Code and Title: DSC303 Ethics and Security
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Sangay Thinley

General Objective

This module introduces ethics and security with respect to data and data science. The module will teach students the importance of applying ethical principles and respecting privacy while handling data. Further, the module will equip students with knowledge on key data management principles and data governance.

Learning Outcomes
On completion of the module, students will be able to:

1. Utilize best practices for data ethics while dealing with data.
2. Apply ethical principles to respect privacy of data.
3. Explain the importance of data security.
4. Present an analysis of case studies in breach of security and data ethics.
5. Describe data analysis problem in structures framework.
6. Apply principle of data protection to their work.
7. Describe key data management principles.
8. Explain the importance of data governance.

Learning and Teaching Approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	4	75
	Presentation	1	
Independent study	Written Assignment	2	45
	Self-study	1	
Total			**120**

Assessment Approach

A. Assignments: 15%
Two assignments worth 7.5% each (with word count between 1500 and 2000) will be given to the students in weeks 6 (Based on topics from Units II – IV) and 12 (Based on Units V – VIII). For each assignment, two random topics will be chosen by the tutor. The students

will be expected to find notes from the Internet and submit a summary of the notes. Students will be given at least a week's time to submit the assignments. Each assignment will be graded out of 100 marks and then be scaled down to 7.5%.

The following rubrics will be used to assess the two assignments.

	Excellent (85-100%)	**Good (61-84%)**	**Satisfactory (<=60%)**
Completeness (25 points)	Completely followed all directions.	Mostly followed all directions.	Did not follow directions for complete responses.
Correctness (40 points)	All the information is correct and accurate	Some of the information is correct & accurate	Most information is not correct or accurate.
Neatness/Legibility (20 points)	Very neat. Writing illustrates a lot of thought and preparation.	Mostly neat and legible. Writing illustrates some thought and preparation	Slightly legible. Ideas expressed are difficult to understand.
Time Management (15 points)	The assignment was turned in on time.	The assignment was turned in one day late.	The assignment was turned in more than two days late.

B. Mid-Semester Examination: 15%

Mid-semester Examination will be conducted (out of 40) in the mid of semester with at least 60% of the subject matters being covered for the duration of 2 hours. The marks obtained will then be converted out of 15%.

C. Presentation: 20%

Students will work in groups (maximum of 4 students per group) to analyse/critically review case studies related to ethical and security issues of data and present their findings and opinion to the class. Based on the group work, each student will make a 10-15 minutes (40 minutes to one hour per group) presentation about the findings of the case study.

At least 10% (2% overall) of this component will be graded by their peers. The following criteria will be used for the presentation:

Criteria	**Weightage**
Quality of Content	8%
Clarity	2%
Audience Engagement	2%
Time Management	2%
Speaking skills, Confidence, Eye Contact	2%
Peer Grading	4%

To ensure that the group work is assessed fairly, the following peer evaluation form (Adapted from a peer evaluation form developed at Johns Hopkins University (October, 2006)) will be used. The total marks obtained from the peer evaluation will be converted to 2%.

Peer Evaluation Form for Group Work

Your name _____

Write the name of each of your group members in a separate column. For each person, indicate the extent to which you agree with the statement on the left, using a scale of 1-4 (1=strongly disagree; 2=disagree; 3=agree; 4=strongly agree). Total the numbers in each column.

Evaluation Criteria	**Group member:**	**Group member:**	**Group member:**
Attends group meetings regularly and arrives on time.			
Contributes meaningfully to group discussions.			
Completes group assignments on time.			
Prepares work in a quality manner.			
Demonstrates a cooperative and supportive attitude.			
Contributes significantly to the success of the project.			
TOTALS			

Feedback on team dynamics:

1. How effectively did your group work?
2. Were the behaviors of any of your team members particularly valuable or detrimental to the team? Explain.
3. What did you learn about working in a group from this project that you will carry into your next group experience?

D. Semester-end Examination: 50%

Semester end Examination for the duration of two and half hours will be conducted covering all the units of the module. The exam will be conducted out of 50.

Overview of assessment approaches and weighting:

Areas of assessment	**Quantity**	**Weighting**
A. Assignments	2	15%

B. Mid-semester Examination	1	15%
C. Presentation	1	20%
D. Semester-end Examination	1	50%

Pre-requisites: None

Subject Matter:

Unit I: Introduction
1.1 Introduction to Privacy, Security and Ethics of Data
1.2 Definitions and Concepts: Rights and interests, Personal data, A responsible organisation

Unit II: Ethics of Big Data
2.1. Values and Actions
2.2. Turning Values Into Actions
2.3. Elements of Data Ethics
2.4. Benefits of Ethical Inquiry
2.5. Ethical Decision Points
2.6. Privacy, Security, Ownership, Evidence-based Decision-Making
2.7. Intellectual Property

Unit III: Best Practices for Data Ethics
3.1 Keep Data Ethics in the Spotlight
3.2 Consider the Human Lives and Interests Behind the Data
3.3 Focus on Downstream Risks and Use of Data
3.4 Treat Data as a Conditional Good
3.5 Establish Chains of Ethical Responsibility and Accountability
3.6 Practice Data Disaster Planning and Crisis Response
3.7 Promote Values of Transparency, Autonomy, and Trustworthiness
3.8 Consider Disparate Interests, Resources, and Impacts
3.9 Invite Diverse Stakeholder Input
3.10 Make Ethical Reflection & Practice Standard, Pervasive, Iterative, and Rewarding

Unit IV: Security of Data Science & Data Science for Security
4.1 Types of Security Breaches
4.2 Infrastructure Security
    4.2.1 Network Security
4.3 Data Protection
4.4 Software Security
4.5 Data Anonymisation
4.6 Anomaly Detection
4.7 Malware Detection
4.8 Threat Detection

Unit V: Cryptographic Technologies for Big Data
5.1 Basics of Cryptography (encryption, decryption, private key, public key, hash function)
5.2 Construct System to Search, Filter for Encrypted Data
5.3 Apply Relational Encryption to Enable Comparison of Encrypted Data

5.4 Reconcile Authentication and Anonymity
5.5 Implement Identity-based Encryption
5.6 Use Oblivious RAM for Privacy Preservation
5.7 Consider Convergent Encryption for Deduplication

Unit VI: Data Provenance
6.1. Develop Infrastructure Authentication Protocol
6.2. Ensure Accurate, Periodic Status Updates
6.3. Verify Data Integrity
6.4. Implement Effective Encryption Methods
6.5. Use Access Control
6.6. Establish Flexible Revocation Mechanisms

Unit VII: Data Protection
7.1 Introduction to Data protection
7.2 Importance of Data Protection
7.3 GDPR: General Data Protection Regulation
7.4 CCPA: The California Consumer Privacy Act
7.5 Information, Communication and Media Act of Bhutan

UNIT VIII: DATA GOVERNANCE

8.1 Introduction to data governance
- 8.1.1 Anatomy of data governance
- 8.1.2 Data governance framework
- 8.1.3 Benefits of data governance

8.2 Data governance tools
- 8.2.1 Metadata – Dublin Core and Darwin Core
- 8.2.2 Data Management plans (DMPs) – H2020 Programme DMPs
- 8.2.3 Enterprise Information Management (EIM) maturity model
- 8.2.4 Data sharing platforms– figshare and re3data

8.3 Data governance challenges
- 8.3.1 Data protection and privacy
- 8.3.2 Citizen generated data

8.4 Licences for data governance
- 8.4.1 Creative Commons (CC)
- 8.4.2 Open Data Commons Attribution Licence (ODB-by)
- 8.4.3 Open Database Licence (ODbL)

UNIT IX: DATA MANAGEMENT AND STEWARDSHIP

9.1 FAIR data principles
- 9.1.1 Introduction to FAIR
- 9.1.2 Data FAIRification process
- 9.1.3 Interpretation and implementation of FAIR

9.2 TRUST principles
- 9.2.1 Introductions to TRUST principles
- 9.2.2 CoreTrustSeal certification

9.3 Five-Star Linked Open Data
- 9.3.1 Introduction to Linked Data

9.3.2 Five-Star deployment scheme
9.3.3 Five-Star vs FAIR data principles
9.4 Case Study: Health Information Exchange

Reading List:

Essential Reading

Davis, K. (2012). Ethics of Big Data. OReilly

Ladley, J. (2019). Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program. Academic Press.

Lee, W.W., Zankl, W., Chang, H. (2016). An Ethical Approach to Data Privacy Protection. ISACA Journal.

Mashima, D., Rajan,S.P. (2016). Big Data: Security and Privacy Handbook. Cloud Security Alliance.

Perry, Y. (2019, September 17). Meeting Data Compliance with a Wave of New Privacy Regulations. Retrieved from https://cloud.netapp.com/blog/data-compliance-regulations-hipaa-gdpr-and-pci-dss

Tellenbach, B., Rennhard, M., & Schweizer, R. (2019). Applied Data Science. Springer

Vallor, S. (2020, May 10). An Introduction to Data Ethics. Santa Clara University. Retrieved from https://www.scu.edu/media/ethics-center/technology-ethics/IntroToDataEthics.pdf

White, G., Ariyachandra, T. (2016). Big Data Analytics: Examining the Grey Areas of Big Data Analytics. Issues of Information Systems.

Additional Reading

Bunnik, A.,Cawley, A.,Mulqueen, M.,Zwitter, A. (2016). Big Data Challenges. Palgrave Macmillan

Freeman, L.A, Peace, A.G. (2005). Information Ethics: Privacy and Intellectual Property. Information Science Publishing.

Date: March, 2022.

AMT312 Stochastic Process

Module Code & Title: AMT312 Stochastic Processes
Programme: BSc in Statistics
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

This module will consider Markov processes in discrete and continuous time. The module also prepares students for undertaking research in the field of operation research, biology and economy.

Learning Outcomes

On completion of the module, students will be able to:

1. Apply the basic concepts of stochastic processes for various real-life problems.
2. Classify stochastic process according to various state space and time domain approach.
3. Apply the different variants of Markov Chain for finding sequence of possible events.
4. Employ Chapman-Kolmogorov equations in identity relating the joint probability distribution of different sets of coordinates on a stochastic process.
5. Construct Ergodic state and Ergodic chain concept which allows to go from every state to every other state with positive probability.
6. Evaluate Random walk and gamblers ruin problem.
7. Apply Birth and death processes to calculate the current size of a population.
8. Apply Queuing processes to model waiting lines so that queue length and waiting times can be predicted.
9. Evaluate the concepts of Epidemic processes as a time-dependent process the state transitions are caused by exposure to some influence.
10. Apply forward and backward Kolmogorov diffusion equations.

Learning and teaching approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**60**
	Laboratory sessions	1	
Independent Study	Written assignment	1	**60**
	Self-study	3	
Total			**120**

Assessment Approach

A. Assignment: (10%)

One written assignment will be given in a semester. The written assignment will be given before the mid-semester examination. A set of problems/Theorems (Minimum 7 and maximum 10) spread over at least 50% of the subject matter will be given to the students, which are to be solved by the students independently.

The assignment will be evaluated based on the following criteria:

- Correctness of the solutions: 4%
- Logical flow in the process: 2%
- Uniqueness of the solution: 2%
- Use of appropriate terms and symbols: 2%

If it is confirmed that a student has plagiarized the assignment, then as per the academic norm, the student will be awarded zero mark or made to redo the assignment based on the convenience of the tutor. In the event where plagiarism is suspected but could not be confirmed, the student in question will be called for viva-voice or asked for oral presentation of the solution. Based on the viva-voice or oral presentation, the tutor will either maintain the initial mark or reduce it.

B. Lab: (30%)

The lab assessment will be carried out as follows:

- **Lab Assignment (10%)**

One lab assignment worth 10% will be assigned for students to do independently. The module tutor will provide the data on which they will apply the statistical data analysis techniques learned in the lab classes and generate a report. The report will be assessed based on the following criteria:

- Abstract 2%
- Diagrammatic/graphical representation 2%
- Identification and calculation of necessary descriptive 3%
- Interpretation of the descriptive and inferential statistics 2%
- Structure of the report 1%

- **Lab report (5%)**

Students will submit a weekly record of the lab session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the lab sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam. Template for laboratory report is given in Appendix A of the DPD.

- **Lab Exam (15%)**

A lab exam for the duration of 3 hour will be conducted towards the end of a semester after the last day of regular teaching.

C. Mid-Semester Examination: (10%)

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 50 marks and 10% of the scores will be taken as final score.

D. Semester-end Examination: (50%)

Semester end Examination for the duration of two and half hours will be for 50 marks. All the units of the subject matter will be included for the examination.

Overview of Assessment Approaches

Areas of assignments		Quantity	Weighting
A. Assignment		1	10%
B. Lab	Lab Assignment	1	10%
	Lab Report	10	5%
	Lab Exam	1	15%
C. Mid-Semester Examination		1	10%
D. Semester-end Examination		1	50%

Prerequisite: PLT203 R Programming for Data Analysis

Subject Matter

Unit I: Introduction to Stochastic process

1.1 Definition and examples of Stochastic Processes (SPs)
1.2 Classification of random processes according to
  1.2.1 State space
  1.2.2 Parameter space
1.3 Types of SPs
  1.3.1 Weakly stationary and strongly stationary processes
  1.3.2 Independent and Identically distributed (iid) process
  1.3.3 White Noise process
  1.3.4 Random Walk process
  1.3.5 Relation between strong and weak Stationarity

Unit II: Discrete-time Markov Chains (DTMCs)

2.1 Definition and examples of Markov Chains
2.2 Transition probability matrix

2.3 Chapman-Kolmogorov equations
2.4 Calculation of n-step transition probabilities
2.5 Limiting probabilities
2.6 Classification of states
2.7 Ergodicity
2.8 Stationary distribution
2.9 Transient Markov Chain
2.10 Random walk and gambler's ruin problem, applications.

Unit III: Continuous-time Markov Chains (CTMCs)

3.1 Kolmogorov- Feller differential equations
3.2 Infinitesimal generator
3.3 Poisson process
3.4 Birth-death process
3.5 Stochastic Petri net

Unit IV: Brownian Motion and Martingale

4.1 Wiener process as a limit of random walk
4.2 Process derived from Brownian motion
4.3 Stochastic differential equation
4.4 Stochastic integral equation
4.5 Ito formula
4.6 Some important Stochastic differential equations (SDEs) and their solutions
    4.6.1 Geometric Brownian Motion
4.7 Applications to finance.
    4.7.1 Direct application
    4.7.2 Ornstein-Uhlenbeck process
    4.7.3 Vasicek interest rate model
4.8 Martingale
    4.8.1 Conditional expectations
    4.8.2 Definition and examples of martingales.

Unit V Renewal and Branching Process

4.1 Renewal function and its properties
4.2 Renewal theorems, cost/rewards associated with renewals
4.3 Markov renewal and regenerative processes
4.4 Non-Markovian queues
4.5 Applications of Markov regenerative processes.
4.6 Definition and examples branching processes
4.7 Probability generating function
4.8 Mean and variance
4.9 Galton-Watson branching process
4.10 Probability of extinction.

Unit VI: Simulation

5.1 Random Number Generation
5.2 Simulation of random variables: Discrete and continuous
6.2.1 The Method of Inverse Functions
6.2.2 The Box-Muller method
5.3 Monte Carlo integration

List of Laboratory Sessions

1. Calculation of probability of moving from one state to other state in Markov chain process
2. Exact simulation of birth and death process
3. A simple simulation of Brownian motion
4. Brownian Motion with Exponential Arrive Displacement
5. Estimation of renewal function
6. Calculate the probability of extinction
7. Simulation using method of inverse functions and Box-Muller method
8. Estimation of estimate using Monte Carlo simulation method

Reading List

Essential Reading

Bhat, B. R. (2000). *Stochastic Models*: *Analysis and Applications*, New Age International, New Delhi.
Medhi, J. (2001). *Stochastic Processes*, Wiley Eastern Ltd., New Delhi.
Ross, S. (2008). *Stochastic Processes* (2 ed.). Wiley.

Additional Reading

Adke, S. R., and Manjunath, S. M. (1984). *An Introduction to Finite Markov Processes*, John Wiley, New Delhi.
Bailey, N. T. J. (1990). *Elements of Stochastic Processes with Applications to the Natural Sciences*, Wiley Eastern Ltd., New Delhi.

Date: December, 2021

APC318 Deep Learning

Module Code and Title: APC318 Deep Learning
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: J.Gurubalan
Module coordinator:

General Objectives:
This module introduces the fundamentals of deep learning and build simple learning systems to understand their mathematical foundations. It also covers to build and train deep neural networks, identify key architecture parameters, and implement efficient neural networks It also introduces the Tensorflow for Deep Learning, including how to perform basic computation. It also covers how to use Python libraries such as PyTorch for Deep Learning applications and build Deep Neural Networks using PyTorch.

Learning Outcomes:
On completion of the module, students will be able to:

1. Identify key concepts related to deep learning.
2. Compute the gradient of the cost function using back propagation.
3. Design Feedforward Neural networks and apply neural networks to real-world problems
4. Implement multi-layer perceptron to solve a classification problem.
5. Implement standard neural machine translation methods.
6. Discuss deep learning methodologies to process image based datasets.
7. Solve problems using deep learning techniques independently.
8. Use a convolutional neural network to classify images.
9. Develop and train models using TensorFlow.

Learning and Teaching Approach:

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	105
	Laboratory	3	
	Assignment/Project	1	
Independent Study	Self-study	1	15
Total			**120**

Assessment approach:
A. Assignment: 20%
These assignments will help to understand the deep learning fundamentals covered in class to give hands-on experience on developing and/or using various networks and prepare students to work on the term projects. Two assignments will be given in a semester where the first assignment will be given on the 4th week and second assignment will be given on the 9th week. Students will be provided a set of programming questions and datasets will be uploaded on machine learning platform. Each assignment will be given a weighting of 10%. It will be assessed for 100 marks and later it will be scale down to 10%. Students will submit their answers and preditions on machine learning platform. The following rubrics will be used to assess the set of programming questions.

Program (100)	(Excellent) 100-81%	(Good) 80-61%	(Satisfactory) (60-41% of the points)	(Unsatisfactory) (<=40% of the points)
Requirements and Delivery (50 points)	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution (20 points)	• Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic (10 points)	• Program is logically well designed.	• Program has slight logic errors that do no significantly affect the results	• Program has significant logic errors.	• Program is incorrect.
Coding Standards (10 points)	• Includes name, date, and assignment title. • Excellent use of white space. • Creatively organized work. • Excellent use of variables (no global variables, unambiguous naming).	• Includes name, date, and assignment title. • Good use of white space. • Organized work. • Good use of variables (no global variables, unambiguous naming)	• Includes name, date, and assignment title. White space makes program fairly easy to read. • Organized work. • Good use of variables (few global variables, unambiguous naming).	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines). • Disorganized and messy. • Poor use of variables (many global variables, ambiguous naming).
Documentation (10 points)	• Clearly and effectively documented including descriptions of all class variables.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function	• Basic documentation has been completed including descriptions of all class variables.	• Very limited or no documentation included. • Documentation does not help the reader

	• Specific purpose noted for each function, control structure, input requirements, and output results.	and control structure.	• Purpose is noted for each function.	understand the code.

B. Class Test: 10%

One class test will be conducted in a semester where the exam will be conducted on 8th week from selected units .The questions will be cumulative coverage of the syllabus. This is a written conceptual test conducted within the class for the duration of 60 minutes. This class test will be conducted for 25 marks and later it will be scaled down to 10%.

C. Project and Presentation: Portion of Final Marks - 40%

The students will form a group of 3 and undertake a project. Each project group will prepare a project proposal, progress report, final report and a presentation. Each project will be expected to explore novel applications of contemporary deep learning techniques or develop novel deep learning techniques. The students' work will be evaluated in two parts: Presentation and Project; each project will be evaluated for 40 marks. Each group will have to submit a report of maximum of 1000 words. Each group will make a 30-45 minutes presentation about the findings of the project.

Marking criteria:

a. Writing Report/Presentation – 10 marks
   i. Language, Communication skills and self-confidence: 2%
   ii. Report- Referencing : 5%
   iii. Q & A : 3%

b. Project – 30 marks
   i. Define Problem and collect data : 7%
   ii. Prepare data : 8%
   iii. Build model : 10%
   iv. Visualize the solution : 5%

D. Semester End Examinations: Portion of Final Marks - 30%

The exam at the end of the semester will be a closed book for a minimum of 2 hours and 30 marks. The exam will evaluate the students' overall understanding and critical thinking ability with regards to the module. The questions will be from all the units of the module.

Overview of the assessment approaches and weighting:

Areas of assessment	Quantity	Weighting
A. Assignment	2	20%
B. Class Test	1	10%
C. Project	1	40%
D. Semester-end Examination	1	30%

Pre-requisites: PLT102 Programming in Python, CAL111 Calculus, ALG108 Matrix Analysis and Vector Spaces, DSC201 Machine Learning, DSC302 Unsupervised Learning
Subject Matter

Unit: I Introduction to Deep learning

1.1. Deep Learning primitives
  1.1.1. Fully connected Layer
  1.1.2. Convolutional Layer
  1.1.3. Recurrent Neural Network Layer
  1.1.4. Long short-term memory cells
1.2. Deep Learning Frameworks
1.3. Feed Forward Neural network
  1.3.1. Data and Task
  1.3.2. Model
  1.3.3. Loss Function
  1.3.4. Evaluation

Unit: II TensorFlow Primitives

2.1. TensorFlow graphs
2.2. TensorFlow session
2.3. TensorFlow constants
2.4. TensorFlow variables
2.5. TensorFlow feeding data to the graph
2.6. Tensorboard
2.7. TensorFlow Save/restore models

Unit: III Shallow Neural Networks

3.1. Neural Networks Overview
  1.1.1. Neural Networks Representation
3.2. Computing a Neural Network's Output
3.3. Vectorized Implementation
3.4. Activation functions
3.5. Derivations of Activation functions
3.6. Gradient Descent for Neural Network
3.7. Back propagation Intuition
3.8. Random Initialization

Unit: IV Convolutional Neural Networks

4.1. The Convolution operation
  4.1.1. 1D Convolution operation
  4.1.2. 2D Convolution operation
4.2. Convolutional Neural Networks
  4.2.1. Convolution operation related to Neural Networks
  4.2.2. Understanding the input/output dimensions
  4.2.3. Sparse Connectivity and Weight Sharing
  4.2.4. Max Pooling and Non-Linearities
4.3. CNN Architectures
  4.3.1. Number of computations in a convolution layer
  4.3.2. 1x1 Convolutions

4.3.3. The GoogleNet Architecture
4.3.4. LeNet
4.3.5. AlexNet
4.3.6. ResNet

4.4. Visualizing CNN
4.4.1. Receptive field of a neuron
4.4.2. Identifying images which cause certain neurons to fire
4.4.3. Visualizing filters
4.4.4. Occlusion experiments

Unit: V Encoder and Decoder Models

5.1. Introducing Encoder Decoder Model
5.1.1. Connecting to data,task,model, and loss function
5.1.2. Connecting to Learning algorithm and Evaluation

5.2. Artificial Neural Networks
5.2.1. McCulloch-Pitts model
5.2.2. MCP error correction

5.3. Attention Mechanism
5.3.1. Object Detection
5.3.2. RCNN - Region Proposal
5.3.3. RCNN - Feature Extraction
5.3.4. RCNN - Classification
5.3.5. RCNN - Regression
5.3.6. RCNN- Training

5.4. Long Short-Term memory Networks

List of Laboratory Sessions:

1. Solve classification problem using PyTorch.
2. Train and predict hidden layer network using gradient descent.
3. Test FeedForward Neural Network with PyTorch.
4. Develop two dimensional image representation using CNN with PyTorch.
5. Predict sequence problems (text data) using RNN with PyTorch.
6. Predict sequence problems (time series or speech data) using LSTM

Reading List:

Essential Reading

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Nelli, F. (2018). Python data analytics: with pandas, numpy, and matplotlib. Apress.

Subramanian, V. (2018). Deep Learning with PyTorch: A practical approach to building vneural network models using PyTorch. Packt Publishing Ltd.

Additional Reading:

Chollet, F. (2018). Deep learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into deep learning. arXiv preprint arXiv:2106.11342.

Nielsen, M. A. (2015). Neural networks and deep learning (Vol. 25). San Francisco, CA: Determination press.

Date: March, 2022

EDP101 Entrepreneurship

Module Code and Title : EDP101 Entrepreneurship
Programme : University-wide module
Credit : 12
Module Tutors **:** Sonam Dendup & Karma Yoezer

General objective
The module will enable students to kindle the spirit of enterprise, evaluate and develop their skills, and motivate them to consider entrepreneurship as a career option. The module intends to enable students to assess the Bhutanese labour market, economy, and equip them with ability to identify business ideas, spot business opportunities, develop business model and business plan/proposal.

Learning outcomes
On completion of the module, students will be able to:

1. explain the Bhutanese labor market and the role of entrepreneurship
2. map out the Bhutanese entrepreneurship ecosystem
3. explain policies related to entrepreneurship in Bhutan
4. evaluate entrepreneurial competencies
5. develop entrepreneurial competencies
6. identify solution driven business ideas
7. evaluate business opportunities
8. use value chain analysis to generate business ideas
9. apply business model to develop business plan
10. develop a feasible business plan
11. pitch and present business plans

Teaching and learning approaches

Approach	Hours per week	Total credit hours
Lectures	2	30
Class discussions, exercises, presentations, role plays, seminar, entrepreneur talk, debate and case studies	2	30
Independent study and library research, assignments, project work	4	60
Total		**120**

Assessment Approach A. Case writing/analysis: Portion of Final Marks-10%
Each student will be required to write a case about an entrepreneur and the enterprise in the locality (maximum 1200 words). The case must document the entrepreneurial journey from idea to starting business to challenges, opportunities and way forward. The students must

produce documentary evidence such as interview recordings, minutes, and images to support their work.

Criteria:

3% organization of ideas

5% content

2% evidence and validity

OR

Each student will be required to analyze a case and write a case analysis report of maximum 1200 words.

Criteria:

2% organization of ideas

6% analysis (identification of issues, analysis of decision alternatives, recommendations/suggestions)

2% referencing

B. Group Work: Business Opportunity identification and selection: Portion of Final Marks-10%

Students in groups of 3-5 members or individually will engage in ideation exercise and generate ideas as well as select a business opportunity. The students will be required to submit idea evaluation report of maximum 1200 words. Criteria:

2% Environmental scanning
3% generating ideas
3% evaluation of ideas
1% selection of ideas
1% evidence

C. Group Work: Business Model Development Portion of Final Marks-10%

Students in groups of 3-5 members will develop business model for the opportunity selected. Criteria:

7% Business Model and explanation of building blocks

3% Business Model Canvas

D. Project: Business Plan Development (2 parts and final version): Portion of Final Marks - 50%

Students will work in groups of five to develop a business plan specific for the Bhutanese market. There will also be a group presentation (approximately 10-15 minutes) that includes slides.

Written report (30%):

10% Part A (In-depth overview, market analysis and marketing plan):

10% Part B (Production and operations, Human Resource and financial plan):
10% Part C Final draft:

1% Executive summary
1% Project Description
1% Market analysis
1% marketing plan
1% operation plan
1% human resources plan
1% financial plan
1% referencing
1% appendices
1% general structure

10% Individually assessed process score (contribution to the group output part A 3%, Part B 3% and Part C 4%) Presentation (10%):

Group presentation mark (6%)

Cohesiveness: 2%
Organization: 1%
Level of professional delivery: 2%
Questioning & Answering (2%)
4% Individual presentation mark
Clarity: 1%
Conciseness: 1%
Content: 1%
Tone: 1%

E. Semester-End Examination: Portion of Final Marks-20%

The module will have a semester-end examination for 1 hour covering the entire syllabus. Cases will also be used to test the levels of knowledge.

Areas of assignments	Quantity	Weighting
A. Case writing/analysis	1	10%
B. Group Work	2	20%
C. Project: Business Plan Development (2 parts and final report)	3	40%
Business Plan Presentation	1	10%
Total Continuous Assessment (CA)		**80%**
D. **Semester-End Examination (SE)**	1	**20%**
TOTAL		**100%**

Pre-requisites: None

UNIT I: Introduction to Entrepreneurship

1.1 Definition and concept of Entrepreneurship

1.2 Types of entrepreneur

1.3 Characteristics of an entrepreneur

1.4 Entrepreneurial functions

1.5 The entrepreneurial method

1.6 Bhutanese entrepreneurship ecosystem (course pack - Entrepreneurship related strategies in Bhutan)

1.7 Role and scope of entrepreneurship in Bhutan

UNIT II: The Entrepreneurial Mind-Set and competencies

2.1 The Entrepreneurial Thinking

2.2 Grassroot Innovations (course pack – include emerging trends/opportunities in Bhutan)

2.3 Problem solving techniques

2.4 Entrepreneurial motivation

2.5 Evaluation of entrepreneurial competencies using GETT/SRQ (course pack should include networking, negotiation, motivation etc.)

2.6 Emotional resilience and entrepreneurial discipline

UNIT III: Business Opportunity Identification (BOI) and selection

3.1 Meaning of business opportunity

3.2 BOI process

3.2.1 Environmental scanning

3.2.2 Generating ideas (course pack – include process & techniques of idea generation)

3.2.3 Identifying and evaluating ideas/opportunities

3.2.4 Selecting ideas/opportunities

3.3 Value chain analysis and business development

UNIT IV: Business Model

4.1 Business model canvas

4.2 Sustainable business models

UNIT V: Building up a Business Plan

5.1 Overview

5.1.1 Need and importance of business plan

5.1.2 Audience of business plan

5.1.3 Components of a business plan

5.2 In-depth business overview

5.2.1 Business profile

5.2.2 Business background

5.2.3 Vision

5.2.4 Mission

5.2.5 Objectives

5.2.6 Keys to success and USPs

UNIT VI: Market Analysis

6.1 Business environment analysis

6.1.1 SWOT

6.1.2 STEEPLES

6.2 Market research

6.3 Target market

6.4 Target customers

6.5 Demand analysis

6.6 Competitor analysis

6.7 Price determination

6.8 Income and revenue forecasting

UNIT VII : Marketing Plan

7.1 Marketing Mix (including costing)

UNIT VIII: Production and Operations

8.1 Product/Service description

8.2 Production process

8.3 Plant and equipment (including costing)

8.4 Inventory (including costing)

8.5 Administration (including costing)

UNIT IX: Human Resource Management

9.1 A brief introduction to human resources

9.2 Organizational chart

9.3 Owner(s) and CEO

9.4 Required staff

9.5 Recruitment and selection options

9.6 Training programs

9.7 Human resource costing

UNIT X: Financials

10.1 Estimated project cost

10.2 Financing the project

10.3 Capital cost

10.4 Depreciation schedule

10.5 Financing cost: Cost of equity, cost of debt, and weighted average cost of capital, loan amortization schedule.

10.6 Projected cash flow statement

10.7 Projected income statement

10.8 Projected financial position

10.9 Economic Analysis

10.9.1 Break-even analysis

10.9.2 Pay-back period

10.9.3 NPV

Reading List Essential Reading

Hisrish, R., Peters, M., & Shepherd, D. (2016). *Entrepreneurship* (10th ed.). McGraw Hill Education.

Hsieh,T. (2014). *Delivering Happiness: A path to profits, passion and purpose*. Grand Central Publishing

Isaacson, W. (2011). *Steve Jobs.* Simon & Schuster.

Kuratko, D. F. (2016). *Entrepreneurship: Theory, process and practice* (10th ed.). South Western College Publication.

Mycoskie, B. (2012). *Start Something That Matters.* Random House Inc.

Additional Reading

Barringer, B.R., & Ireland, R.D. (2015). *Entrepreneurship: Successfully launching new ventures* (5th ed.). Pearson.

Drucker, P.F. (2006). *Innovation and entrepreneurship.* Harper Business.

Hisrish, R., Peters, M., & Shepherd, D. (2016). *Entrepreneurship* (10th ed.). McGraw Hill Education.

Johnson, K.D. (2013). *The entrepreneur mind: 100 essential beliefs, characteristics, and habits of elite entrepreneurs.* Johnson Media Inc.

Kumar, A., (2012). *Entrepreneurship: Creating and leading an entrepreneurial organization*. Pearson (India).

Kuratko, D. F. (2016). *Entrepreneurship: Theory, process and practice* (10th ed.). South Western College Publication.

Osterwalder, A & Pigneur, Y. (2017). *Business Model Generation.* Self-published.

Royal Government of Bhutan. (2007). *Labour and Employment Act of Bhutan.* National Assembly of Bhutan.

Royal Government of Bhutan. (2012). *Cottage Small and Medium Industry policy.* Royal Government of Bhutan.

Royal Government of Bhutan. (2016). *Companies Act of Bhutan.* Natioional Assembly of Bhutan.

Retrieved October 24, 2017, from http://www.nab.gov.bt/en/business/acts

Royal Government of Bhutan. (2017). *12th Five-Year Plan.* Royal Government of Bhutan.

Royal Government of Bhutan. (2017). *Economic Development Policy of Bhutan.* Ministry of Economic Affairs.

Royal Government of Bhutan. (n.d.). *BAS for SMEs.* Riyal Government of Bhutan.

Royal Government of Bhutan. (n.d.). *Cottage Small and Medium Industry Development Strategies (2012-2020)n..* Royal Government of Bhutan

Scarborough, N.M. (2014). *Entrepreneurship and effective small business management* (11th ed.). Pearson.

Shankar, R. (n.d.). *Entrepreneurship: Theory and practice* (1st ed.). McGraw Hill Education.

Year 4 Semester I

APC419 Computer Vision

Module Code and Title: APC419 Computer Vision
Programme: BSc in Data Science
Credit Value: 12
Module Tutors: Manish Kumar Jha

General Objective
This module aims at providing a basic foundation towards digital image processing and video analytics. This module is intended to provide knowledge of fundamental concepts necessary to analyze and interpret the visible objects in the application area of computer vision. Python programming language will be used as a tool for the delivery of this module.

Learning Outcomes
On completion of the module, students will be able to:

1. Describe image and video in terms of pixels
2. Use the tools for analysis of visual data
3. Conceptualize the fundamental principles of image and video analysis
4. Apply image and video compression
5. Enhance and restore image and video quality
6. Explore application area of image processing in the field of automation
7. Implement feature and object detection

Learning and Teaching Approach

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	3	90
	Laboratory Session	3	
Independent study	Self-study	2	30
Total			**120**

Assessment Approach

A. Laboratory Assessment: (40%)

Laboratory Assessment will be assessed based on the following components:

- Laboratory Assignments: 15%
- Laboratory Examination: 20%
- Viva Voce: 5% (Laboratory Assignment: 2%, Laboratory Examination: 3%)

Laboratory Assignments (a total of 10) will be based on the concepts taught in the theory class. Students will be provided a set of programming questions every week covering the concepts taught on each unit. The students will have to submit the solutions at the end of the laboratory session through the VLE.

The following rubrics will be used to assess the set of programming questions. Each set will be assessed out of 100 marks and converted to 1.5%.

Program (100)	**(Excellent)** **100-81%**	**(Good)** **80-61%**	**(Satisfactory)** **(60-41% of the points)**	**(Unsatisfactory)** **(<=40% of the points)**
Requirements and Delivery **(50 points)**	• Completed between 90-100% of the requirements. • Delivered on time, and in correct format.	• Completed between 80-90% of the requirements. • Delivered on time, and in correct format.	• Completed between 70-80% of the requirements. • Delivered on time, and in correct format.	• Completed less than 70% of the requirements. • Delivered on time but not in correct format.
Program execution **(20 points)**	• Executes without errors excellent user prompts, good use of symbols, spacing in output. • Thorough and organized testing or input validation has been completed.	• Executes without errors. • User prompts are understandable, minimum use of symbols or spacing in output. • Most testing or input validation completed.	• Executes without errors. • User prompts contain little information, poor design. • Some testing or input validation has been completed.	• Does not execute due to errors. • User prompts are misleading or non-existent. • No testing has been completed, or no input validation.
Design of logic **(10 points)**	• Program is logically well designed.	• Program has slight logic errors that do no significantly affect the results	• Program has significant logic errors.	• Program is incorrect.
Coding Standards **(10 points)**	• Includes name, date, and assignment title. • Excellent use of white space. • Creatively organized work. • Excellent use of variables (no global variables,	• Includes name, date, and assignment title. • Good use of white space. • Organized work. • Good use of variables (no global variables, unambiguous naming)	• Includes name, date, and assignment title. White space makes program fairly easy to read. • Organized work. • Good use of variables (few global variables,	• No name, date, or assignment title included. • Poor use of white space (indentation, blank lines). • Disorganized and messy. • Poor use of variables (many global variables,

	unambiguous naming).		unambiguous naming).	ambiguous naming).
Documentation **(10 points)**	• Clearly and effectively documented including descriptions of all class variables. • Specific purpose noted for each function, control structure, input requirements, and output results.	• Clearly documented including descriptions of all class variables. • Specific purpose is noted for each function and control structure.	• Basic documentation has been completed including descriptions of all class variables. • Purpose is noted for each function.	• Very limited or no documentation included. • Documentation does not help the reader understand the code.

Laboratory examination will be conducted after the completion of the module and before semester-end examination and will cover all the units. Each student will be given two questions and will be graded out of 100 marks based on the rubric mentioned above. The marks obtained will then be converted out of 20%.

Viva Voce for the laboratory assignments will be conducted once to assess the students' understanding of the laboratory assignments. Viva Voce will also be conducted for the laboratory examination.

B. Project Work: (10%)

Students will be required to furnish a minor project (on video analysis, video and image compression) work assigned by tutor during mid of the semester. It can be done either individually or in group. Nature of the work would be practical as well as theoretical in line with learning outcome. Assessment criteria would be as follows:

Presentation: 3%

Accuracy of Output: 5%

Viva Voce: 2%

C. Semester-end Examination: 50%

Semester end Examination for the duration of two and half hours will be conducted covering all units of the module. The exam will be conducted out of 50.

Overview of assessment approaches and weighting:

Areas of Assessment	**Quantity**	**Weighting**

A. Laboratory Assessment	Laboratory Assignment	1	20%
	Laboratory Examination	10	15%
	Viva Voce	2	5%
B. Project Work		1	10%
C. Semester-end Examination		1	50%
Total			100%

Pre-requisites: ADS104 Data Structures, ADS205 Algorithm Analysis

Subject Matter

Unit I: Introduction to Digital Image and Video Processing

1.1 Digital image representation
1.2 Sampling and Quantization
1.3 Types of Images
    1.3.1 Vector
    1.3.2 Raster
    1.3.3 Binary
    1.3.4 Gray-scale
    1.3.5 Color
    1.3.6 Multispectral
1.4 Basic Relations between Pixels
    1.4.1 Neighbors
    1.4.2 Connectivity
    1.4.3 Distance Measures between pixels
    1.4.4 Linear and Non Linear Operations
1.5 Introduction to Digital Video
    1.5.1 Sampled Video
    1.5.2 Video Transmission
1.6 Gray-Level Processing
    1.6.1 Image Histogram
    1.6.2 Linear and Non-linear point operations on Images
    1.6.3 Arithmetic Operations between Images
    1.6.4 Geometric Image Operations
1.7 Binary Image Processing
    1.7.1 Image Thresholding
    1.7.2 Region labeling
    1.7.3 Binary Image Morphology

Unit II: Image and Video Enhancement and Restoration

2.1 Spatial domain
    2.1.1 Linear and Non-linear Filtering
    2.1.2 Morphological filtering
2.2 Frequency domain
    2.2.1 Homomorphic Filtering

2.3 Blotch Detection and Removal
    2.3.1 Blotch Detection
    2.3.2 Motion Vector Repair and Interpolating Corrupted Intensities
2.4 Intensity Flicker Correction
    2.4.1 Flicker Parameter Estimation
2.5 Brief introduction towards Wavelets
    2.5.1 Wavelet based image denoising
    2.5.2 Basic methods for image restoration using deconvolution filters.

Unit III: Image Analysis
3.1 Image Compression
    3.1.1 Huffman coding
    3.1.2 Run length coding
    3.1.3 LZW coding
    3.1.4 Lossless Coding
    3.1.5 Wavelets based image compression.

Unit IV: Video Analysis
4.1 Video Compression
    4.1.1 Basic Concepts and Techniques of Video Coding and the H.264 Standard
    4.1.2 MPEG-1 and MPEG-2 Video Standards

Unit V: Feature Detection and Description
5.1 Introduction to feature detectors, descriptors, matching and tracking
5.2 Basic edge detectors
    5.2.1 canny, sobel, prewitt etc.
5.3 Image Segmentation
    5.3.1 Region Based Segmentation
    5.3.2 Region Growing and Region Splitting and Merging
5.4 Thresholding
    5.4.1 Basic global thresholding
    5.4.2 Optimum global thresholding using Otsu's Method.

Unit VI: Object Detection and Recognition
6.1 Object detection and recognition in image and video
6.2 Basic texture descriptors
    6.2.1 GLCM, LBP and its applications in image and video analysis
6.3 Object tracking in videos.

List of Laboratory Sessions
Students are required to implement the following programs on computer machine:

1. basic gray-scale and binary processing - image histogram, image labeling, image thresholding
2. Extraction of frames from videos and analyzing frames
3. spatial domain - linear and non-linear filtering
4. domain – homomorphic filtering on gray scale and color images

5. image restoration methods on images
6. flicker correction on video datasets
7. multi-resolution image decomposition and reconstruction using wavelet
8. image compression using wavelets
9. image segmentation using thresholding
10. Local Binary Pattern texture descriptor

Reading List

Essential Reading

Bovik, A. (2005). *Handbook of Image and Video Processing* (2 ed.). Academic Press.

Gonzalez, R.C., Woods, R.E. (2018). *Digital Image Processing* (4 ed.). Pearson Education.

Szelisk, R. (2011). *Computer Vision – Algorithms and Applications*, Springer.

Additional Reading

Jain, A.K. (2011). *Fundamentals of Digital Image Processing*, PHI.

Marques, O. (2011). *Practical Image and Video Processing Using MatLab*, Wiley.

Woods, J.W. (2006). *Multidimensional Signal, Image, Video Processing and Coding*, Academic Press.

Date: March, 2022

DSC404 Big Data Technologies and Applications

Module Code and Title: DSC404 Big Data Technologies and Applications
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Karma Dorji

General Objective

This module is an introduction to big data architectures and technologies. Hadoop is the de facto open-source framework used in the industry for large scale, parallel and distributed data processing. This module covers in-depth knowledge on Big Data and Hadoop Ecosystem tools such as HDFS, YARN, MapReduce, Hive, and Pig. Some advanced job patterns and their applications are also discussed and Python will be used as the main language. The module will enhance the skills of the learners by offering comprehensive knowledge of the Hadoop framework, and the required hands-on experience for solving real-time industry-based Big Data projects.

Learning Outcomes

On completion of the module, students will be able to:

1. Install and configure Hadoop eco system.
2. Describe the Hadoop eco system.
3. Apply the concepts of HDFS, YARN
4. Work with Hadoop storage & resource management.
5. Develop YARN applications.
6. Manage job execution in Hadoop Environment using Oozie.
7. Work on a real-time Hadoop cluster
8. Differentiate between structured and unstructured data.
9. Apply data ingestion techniques using Sqoop and Flume
10. Implementing Partitioning, Bucketing, and Indexing in Hive

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	90
	Laboratory Session	3	
Independent Study	Written Assignment	1	30
	Self-study	1	
Total			**120**

Assessment Approach

A: Project: (30%)

For this project the students can work in group of 3 or 4. The objective of the project is to let the student use the tools and techniques taught and apply these tools to perform data analytics on real data set. The datasets are available freely online on various websites. The size of the data set will be on the discretion of the module tutor.

a. Project proposal: (**10)**
    a. Introduction: What will be done? Team members list. 1
    b. Goal of the project: Why the project is being carried out? 2
    c. Relation to the topics studied in the module 2
    d. Literature review 3
    e. Methods to be used 2

b. Project milestones: (**6)**
    a. Approach, methods and tools used 3
    b. Provide the general description of the code 3

c. Project report (**14)**
    a. Final 2000 – 2500 words report to be submitted

        A detailed report consisting of 2000-2500 word to be submitted by the group. The report should cover a detailed writings on the following:

        a. Abstract 1
        b. Introduction 1
        c. Related work 2
        d. Approach and method used 2
        e. Experimental details 4
        f. Result Analysis 4

B: Laboratory Assessments: (20%)

Students are required to complete every laboratory task. Laboratory sessions will be of 3 hours once a week. Students are required to submit the lab report of the completed task and write a short reflection of 150 – 200 once a week. The number of reports should equal the number of laboratory sessions conducted in the semester.

a. Completion of tasks during the laboratory session **1**
b. Lab report submission **1**
c. Weekly submission of laboratory session reflection **3**
    a. The reflection should include

C: Critique a research article: (10%)

Individual students are to be provided peer reviewed research article on the field of Big Data by the faculty. Students are required to submit a report consisting of 1000 words. The following guidelines should be followed:

a. Draw 3 to 6 keywords that encapsulates the main topic of the research **1**
b. Introduction **3**
    a. Literature review: appraising the work done by others rather than describing their work
    b. Aims of the study
    c. Rationale for conducting the study
c. Methods used **3**
    a. Sampling technique used

b. Strength and weakness of the methods used
d. Results and discussions **3**

D: Mid-Semester Examination: (10%)
Students will write mid semester examination covering the theoretical aspect of the module. The exam duration will be of 2½ hours. The exam will be closed book exam. The exam will include important and relevant questions from the content covered. The exam shall contain 40% of the module covered. The exam will be awarded 50 marks and shall be converted to 10% of the total weightage.

E: Semester-end Examination: (30%)
Students will write a Semester-end examination for a duration of 2½ hours. It will be a closed book examination of what has been exposed in the class.5% of the question shall be from the topics covered before the midterm exams and the rest 90% of the question shall be administered from the topics covered after midterm exams.

Overview of assessment approaches and weighting:

Areas of Assessment	Quantity	Weighting
A. Project	1	30%
B. Laboratory Assessments	1 per week	20%
C. Critique a research article	1	10%
D. Mid-Semester Examination	1	10%
E. Semester-end Examination	1	30%

Pre-requisites: CAE201 Database Systems, PLT102 Programming in Python

Subject Matter

Unit I: Introduction to Big Data
1.1 Introduction to Big Data
1.2 Hadoop vs Traditional systems
1.3 Limitations of traditional data analytic architectures
1.4 Types of data

Unit II: Hadoop
2.1 Introduction to Hadoop
2.2 Hadoop clusters and eco systems
2.3 Hadoop core components
2.4 Hadoop installations and configurations
2.5 HDFS file systems and features
    2.5.1 HDFS name node
    2.5.2 HDFS data node
    2.5.3 HDFS data blocks
2.6 HDFS commands

Unit III: Hadoop Map Reduce Framework
3.1 Configuring Map Reduce into Hadoop
3.2 Overview of Map Reduce framework

3.3 MapReduce architecture
3.4 Job scheduling, shuffle and sort
3.5 MapReduce types and formats
3.6 Task executions
3.7 MapReduce using Python

Unit IV: Hive and Pig
4.1 Traditional Extract, Transform vs Load (ETL) and Hadoop ELT
4.2 Advantages and disadvantages of ETL processes
4.3 Introduction to Hive and Pig
4.4 How to use Hive for ELT
4.3 introduction to Hive Data definition language and Hive Data manipulation language
4.4 Difference between Hive and Relational Database Management Systems.
4.5 Hive partitions
4.8 Data movement with Flume and Sqoop
4.9 Running and working with operators in Pig
4.10 Working with functions in Pig

Unit V: NoSQL Database
5.1 Unstructured data vs structured data
5.2 Introduction to HBase
5.3 HBase Components
5.4 HBase storage model
5.5 Introduction to MongoDB
5.6 Indexing
5.7 Aggregation operators
5.8 Advantages of MongoDB over RDBMS

Unit VI: YARN
6.1 YARN architecture
    6.1.1 Resource manager, application master, node manager and YARN clients
6.2 Developing YARN applications
6.3 Monitoring YARN
    6.3.1 Job scheduling in YARN
6.4 YARN commands
    6.4.1 User commands
    6.4.2 Administration commands

List of Laboratory Sessions (Creating a single node cluster)

1. **Preinstallation setup.**
    a. Create groups and users(hdadmin) for Hadoop to isolate Hadoop file system from Unix file system
    b. Install Java (version 1.8 above) and set Java Home path
    c. Disable IPV6 (verify if this has to be done with the latest version)
    d. Disable SELinux and Firewall
    e. Change dynamic to static IP address
    f. Define (hdadmin) as the root user
    g. Enable hdadmin to make changes without any password
    h. Add static IP address and map it to name nodes and data nodes
    i. Generate SSH keys for password less SSH connection

2. **Hadoop Installation**
    a. Download and install Hadoop
    b. Configuring Hadoop
        1. Set path to HADOOP_PREFIX
        2. Add JAVA_HOME to HADOOP_ENV.sh
        3. Editing the Core-site.xml to define the name node
        4. Define the replication of blocks in hdfs-site.xml
        5. Create the HDFS file system
        6. Starting HDFS service

3. **Working with HDFS file system and HDFS commands**

    a. Creating directory in HDFS
    b. Removing a file
    c. Copying files from local machine to HDFS file system
    d. Copying files from HDFS file system to local machine
    e. Deleting files from data directory and HDFS data directory

4. **Configuring MapReduce and YARN**

    a. Copy the marped-sitexml.template and rename it to marped-site.xml
    b. Make appropriate changes in the xml file to configure YARN as the resource manager
    c. Start the dfs service

5. **Laboratory sessions using PIG**
    a. Join operation,
        i. self-join, inner join and outer join operation
        ii. Join operation using multiple keys
    b. Split operations
    c. Filter operations
    d. Distinct operations
    e. Arithmetic operations
6. **Laboratory sessions using Hive**
    a. Creating a database
    b. Loading data from local file system and HDFS
    c. Querying database
        i. Select query
        ii. Aggregation
        iii. Grouping
        iv. Inserting
    d. Relational operation
    e. Arithmetic and logical operations

7. **Laboratory session using SPARK**
    a. Spark SQL
    b. Creating data frame from JSON file

c. Operations supported by data frames
    i. Filter
    ii. Join
    iii. Groupby

Reading List

Essential Reading

DT Editorial Services. (2016). Big data black book: Covers Hadoop 2, Mapreduce, Hive, Yarn, Pig, R and data visualization. Dreamtech Press.

Karanth, S. (2014). Mastering Hadoop. Packt Publishing Ltd.

Kunigk, J., Buss, I., Wilkinson, P., & George, L. (2018). Architecting modern data platforms: A guide to enterprise Hadoop at scale. O'Reilly Media, Inc.

Additional Reading

Bengfort, B., & Kim, J. (2016). Data analytics with Hadoop: An introduction for data scientists. O'Reilly Media, Inc.

Holmes, A. (2014). Hadoop in practice: Includes 104 techniques. Manning.

White, T. (2015). Hadoop: The definitive guide. O'Reilly.

Date: February, 2022

APC420 Natural Language Processing

Module Code and Title: APC420 Natural Language Processing
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Karma Dorji

General Objective:

This module provides a critical introduction to key topics in natural language processing with hands-on experience of using existing software tools and developing applications to process texts and access linguistic resources. This module present programming concepts by example, in the context of a linguistic processing task. The students will learn to write programs to access text from local files and from the web, in order to get hold of an unlimited range of language materials. This module also cover some fundamental techniques in Natural Language Processing, including sequence labeling, n-gram models and tagging. The students will be exposed to some important machine learning techniques, including decision trees, naive Bayes' classifiers, and maximum entropy classifiers.

Learning Outcomes:

On completion of the module, students will be able to:

1. Explain the use of software tools such as corpus readers, stemmers, taggers and parsers.
2. Manipulate texts using regular expressions and classify texts into classes using predictive models.
3. Use NLP methods to analyze sentiment of a text document.
4. Critically appraise existing Natural Language Processing (NLP) applications such as chatbots and translation systems.
5. Write programs to access text from local files and web, analyze and produce formatted output.
6. Identify lexical categories and how they are used in language processing.
7. Use grammars to describe the structure of an unlimited set of sentences, represent the sentence and build a syntax tree using parsers.
8. Compare and contrast the use of parsing techniques for context-free grammar problems.
9. Build a system that extracts structured data, such as tables, from unstructured text by applying techniques such as chunking and named-entity recognition.
10. Investigate problem, limitations and challenges in natural language.

Learning and Teaching Approach:

Type	**Approach**	**Hours per week**	**Total credit hours**
Contact	Lecture	3	90
	Laboratory session	3	

Independent study	Written Assignment	1	15
	Self-study	1	15
Total			**120**

Assessment approach:

A: Project: (30%)

The assignment will be a group project. The group will consist of four students each. The project will offer students a chance to apply their newly acquired skill. The project will be awarded 100 marks and it shall be converted to 30% of the total weightage offered for the module. The project should consist of 2500–3000 words report.

Marking criteria for the project:

Project proposal: **(5)**

- Describe the main goal of the project 1
- What NLP task will be addressed 1
- What data set are to be used 1
- What methods will be used 1
- How the results will be evaluated 1

Project milestones: **(5)**

- Abstract 2
- Approach and experiments 2
- Future work 1

Project poster/video **8**

- Briefly explain the problem being addressed and its importance 2
- Explain the data and task 2
- Approach: methods and techniques used 2
- Result and analysis 2

Project report **12**

Each group shall submit a project report worth 60% of the overall grade. The following key information should be in the report:

- Title of the project
- Team members name with email-id and student number and their contribution to the project

- **Abstract** **1**
    - In between 250-300 words, concisely describe your aims and objective, motivation for the project and brief information on the findings.
- **Introduction** **1**
    - Elaborated version of the abstract. Explain the importance of the project, key ideas of the work and the approach, the failure or success of the project etc..
- **Related Work** **1**
    - Inspiration of your work. Do not focus on a single source of paper. Discuss the paper that you have used as the basis of your work, discuss how your work is different than that of the papers that you have read.
- **Approach/Method Used/Experiments** **4**
    - Specific description of your approaches – equations or figures to be included
    - Specify if there are any or part in the project which is of your own original approaches
    - Provide references or links to materials or codes that is not original or copied
    - Describe the data sets used in the experiments, be precise on the form of input and output
    - Describe the evaluation metrics and include any other details to understand your evaluation
    - Report of how the experiments were run
- **Results/Conclusions** **5**
    - Report of the results you have obtained after completion of the project. Was the result as expected or it was otherwise – comments on your results. Clear analysis of your findings and outcomes, whether it was a success or a failure.

B: Critique a research article: (10%)
Students are to be provided peer reviewed research articles on the field of Natural Language Processing by the faculty. Students are required to submit a report consisting of 1000 words. The following guidelines should be followed and the marks for each are provided at the right:

a. Draw 3 to 6 keywords that encapsulates the main topic of the research **1**
b. Introduction **3**
    a. Literature review: appraising the work done by others rather than describing their work
    b. Aims of the study
    c. Rationale for conducting the study
c. Methods used **3**
    a. Sampling technique used
    b. Strength and weakness of the methods used

d. Results and discussions **3**

C: Mid-Semester Exam (20%)

Students will write mid semester examination covering the theoretical aspect of the module. The exam duration will be of 2 ½ hours. The exam will be closed book exam. The exam will include important and relevant questions from the content covered. The exam shall contain 40% of the module covered. The exam will be awarded 50 marks and shall be converted to 10% of the total weightage.

D: Semester-end Examination: (40%)

Students will write a Semester-end examination for a duration of 2½ hours. It will be a closed book examination of what has been exposed in the class.10% of the question shall be from the topics covered before the Mid-semester exams and the rest 80% of the question shall be administered from the topics covered after midterm exams.

An overview of assessment approaches and weighting:

Areas of assessment	Quantity	Weighting
A. Project	1	30%
B. Critique a Research Article	1	10%
C. Mid-semester Examination	1	20%
D. Semester-end Examination	1	40%

Pre-requisites: DSC201 Machine Learning, DSC302 Unsupervised Learning, PLT102 Programming in Python

Subject Matter:

Unit I: Introduction to Natural Language Processing

1.1 Natural language and philosophy
1.2 Language syntax and structure
1.3 Grammar
1.4 Word order topology
1.5 Language semantics
1.6 Text corpora and text analytics
1.7 Computing with language
1.9 Computing with language: Simple statistics
1.10 Automatic and natural language understanding
1.11 Automatic natural language understanding

Unit II: Accessing Text Corpora and Lexical Resources

2.1 Accessing text corpora
2.2 Loading your own text corpora
2.3 Conditional frequency distribution
2.4 Lexical resources
    2.4.1 Wordlist corpora
    2.4.2 Comparative wordlists

2.4.3 Shoebox and toolbox lexicons
2.5 Wordnet
2.5.1 Senses and synonyms
2.5.2 The wordnet hierarchy

Unit III: Processing Raw Text and Writing Structured Programs
3.1 Accessing text from the web and local disk
3.2 Processing strings
3.3 Encoding and decoding texts
3.4 Using regular expressions to detect word patterns and tokenizing texts
3.5 Writing structured programs
3.6 Algorithm design
3.6.1 Recursion
3.6.2 Space time trade off
3.7 Python libraries
3.7.1 Natural Language Toolkit (NLTK)
3.7.2 TextBlob
3.7.3 spaCy

Unit IV: Categorizing and tagging words
4.1 Using a tagger
4.2 Tagged corpora
4.2.1 Representing tagged tokens
4.2.2 Reading tagged corpora
4.2.3 Part of speech tag set
4.3 Mapping words to properties using python dictionaries
4.4 Automatic tagging
4.5 N-Gram tagging
4.5.1 Unigram tagging
4.5.2 General N-Gram tagging
4.6 Combining taggers
4.6.1 Tagging the token with the bigram tagger
4.6.2 Using the default tagger and unigram tagger to combine tagger
4.7 Storing tagger
4.8 Bag of Words model
4.8.1 N-grams
4.8.2 Apply TF-IDF vectorizer

Unit V: Classifying and Extracting Information from text
5.1 Supervised classifications
5.1.1 Decision trees
5.2 Evaluation
5.2.1 Naïve Bayes classifiers
5.3 Information extraction architectures
5.4 Chunking
5.4.1 Chunking with regular expressions
5.4.2 Non phrase chunking

5.5 Developing and evaluating chunkers
5.6 Regression in linguistic structure
5.7 Named entity recognition
    5.7.1 Identifying named entities
5.8 Ensemble Classifiers
    5.8.1 Stacking
    5.8.2 Blending
    5.8.3 Bagging
    5.8.4 Boosting

Unit VI: Analyzing Sentence Structures and Grammars

6.1 Context free grammars
    6.1.1 Writing your own grammar
    6.1.2 Recursion in syntactic structure
6.2 Parsing with context free grammars
    6.2.1 Recursive descent and shift reduce parsers
6.3 Grammar development
6.4 Grammatical features
6.5 Processing feature structures
6.6 Extending feature-based grammars

Unit VII: Analyzing the Meaning of Sentences and Managing Linguistic Data

7.1 understanding natural languages
7.2 Propositional logic
7.3 First order logic
7.4 Semantics of English sentences
7.5. Semantics of Dzongkha sentences
7.6 Discourse semantics
7.7 Life cycle of a corpus
7.8 Using XML for linguistic structures
7.9 Obtaining data from word processor files

Unit VIII NLP for language processing

8.1 Reading and exploring dataset
8.2 Pre-processing data
    8.2.1 Remove punctuations
    8.2.2 Tokenization
    8.2.3 Remove stopwords
    8.2.4 Using Stemming
    8.2.5 Using a lemmatizer
8.3 Vectorizing data
    8.3.1 Bag-of-words or CountVectorizer
    8.3.1 N-grams
    8.3.2 Apply TF-IDF vectorizer
8.4 Feature creations
    8.4.1 Create features of texts using the domain knowledge
8.5 Building ML classifiers
    8.5.1 Support vector machines

List of Laboratory Sessions:

1. Development of Corpus
2. Preprocessing of Text
3. Implementation of text representation using statistic and non-statistical model
4. Implementation of text representation application
5. Implementation of Features extraction
6. Implementation of Tokenization and Segmentation
7. Implementation of POS taggers
8. Implementation of Syntactic Analysis
9. Implementation of Semantic Analysis
10. Implementation Text classification
11. Development of Named Entity Recognition System

Reading List:

Essential Reading:

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with python: Analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Sarkar, D. (2019). Text analytics with python: A practitioner's guide to natural language processing. Apress.

Additional Reading:

Eisenstein, J. (2019). Introduction to natural language processing. MIT Press.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of machine learning research, 12(ARTICLE), 2493-2537.

Norbu, T., & Namgyel, T. (2019). Languages and technology in Bhutan. Proceedings of the Language Technologies for All (LT4ALL), 235-238.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Date: February, 2022

DSC405 Geospatial Data Analysis

Module Code & Title: DSC405 Geospatial Data Analysis
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Dechen Wangdi

General Objective

This module introduces students to automated geospatial data collection, analysis, and visualization. Students will be also exposed to the concept underlying spatial data management, processing, and visualization using python.

Learning Outcomes

On completion of the module, students will be able to:

1. Identify and manage appropriate data models to represent spatial data.
2. Apply Python as a GIS computer language
3. Read and write spatial data from or to different file format.
4. Perform different geometric operations and geocoding.
5. Visualize spatial data and create interactive maps.
6. Handle data with different projections.
7. Classify data based on different criteria.
8. Perform operations using different geometric objects.
9. Create maps and images to communicate spatial data in a meaningful way to others

Learning and Teaching Approach

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	**90**
	Laboratory Session	3	
Independent study	Written Assignment	1	**60**
	Self-study	1	
Total			**120**

Assessment Approach

A. Assignment: (10%)
Individual assignment will be given during the first half of the semester. The assignment will comprise of problem-solving based on the subject matters of Unit I and Unit II to assess the students' ability in data representation skills, data visualizing skills and interpretation of descriptive statistics. Assignments will be assessed using following criteria:

- Correctness of the solutions: 4%
- Logical flow in the process: 2%
- Uniqueness of the solution: 2%

- Use of appropriate terms and symbols: 2%

B. Lab: The portion of Final Marks: (40%)

Lab assessment will be carried out as follows:

- **Laboratory report- 15%**

Students will submit a weekly record of the laboratory session to keep track of his/her progress which will be assessed for 5 marks. Each student will have to submit 10 reports with 250 to 500 words based on the nature of the practical. They will be assessed using the following marking criteria:

Objective	- 1 mark
Method description	- 2 marks
Results	- 1 mark
Interpretation/ conclusion	- 1 mark

At the end of all the laboratory sessions, the average of scores in each laboratory report will be taken as the final score for laboratory report. This will be finalized before their laboratory exam.

- **Laboratory Exam- 25%**

A laboratory exam for the duration of 3 hour will be conducted towards the end of a semester after the last day of regular teaching covering all the subject matter. It will be out of 60 marks and 20% of the score will be taken as final score.

C. Mid-Semester Examination: 10%

Mid-semester Examination will be conducted in the mid of semester with at least 60% of the subject matter being covered for the duration of two and half hours. Examination will be conducted out of 40 marks and 10% of the scores will be taken as final score.

D. Semester-end Examination: 40%

Semester end Examination for the duration of two hours will be for 40 marks. All the units of the subject matter will be included for the examination

Overview of assessment approaches and weighting:

Areas of Assessment		**Quantity**	**Weighting**
A. Assignment		1	10%
B. Laboratory Assignment	Laboratory report	10	15%
	Laboratory exam	1	25%
C. Mid-Semester Examination		1	10%
D. Semester-end Examination		1	40%

Prerequisite: PLT102 Programming in Python

Subject Matter

Unit I: Geometric Objects

1.1 General Attributes and Methods
1.2 Points
1.3 Linestrings
1.4 LinearRings
1.5 Polygons
1.6 Collections
1.7 Collections of Points
1.8 Collections of lines
1.9 Collections of polygon
1.10 Empty features
1.11 Coordinate sequences
1.12 Linear Referencing methods

Unit II: Geospatial Data

2.1 Data Structures
- 2.1.1 Common traits
- 2.1.2 Geo-location
- 2.1.3 Subject information
- 2.1.4 Metadata
- 2.1.5 File structure

2.2 Vector data
- 2.2.1 Shapefiles
- 2.2.2 CAD files
- 2.2.3 Tag and markup-based formats
- 2.2.4 GeoJSON

2.3 Raster Data
- 2.3.1 TIFF files, JPEG, GIF, BMP and PNG
- 2.3.2 Compressed format
- 2.3.3 ASCII GRIDS

2.4 Point cloud data

Unit III: Predicates and Relationships

3.1 Unary Predicates
3.2 Binary Predicates
3.3 DE-9IM Relationships
3.4 Spatial Analysis Methods:
- 3.4.1 Set-theoretic methods
- 3.4.2 Constructive Methods

3.5 Affine Transformations

Unit IV: Operations

4.1 Merging Linear Features
4.2 Efficient Unions
4.3 Delaunay triangulation

4.4 Nearest Points
4.5 Snapping
4.6 Shared paths
4.7 Splitting
4.8 Substring
4.9 Prepared Geometry
4.10 Operations
4.11 Diagnostics
4.12 Polylabel
4.13 STR-packed R-tree

Unit V: Interoperation

5.1 Well-known formats
5.2 Numpy and Python Arrays
5.3 Python Geo Interface
5.4 Performance

Unit VI: Maps

6.1 Loading and preprocessing data
6.2 Making map using basemap
6.3 Adding a background map to the plot
6.4 Show a Kernel Density estimation (KDE) plot of the spatial distribution
6.5 Color the markers with the KDE information
6.6 Use more specific symbols as map markers
6.7 Raster overlay on interactive map

List of Laboratory Sessions

1. GDAL: Fundamental package for processing vector and raster data formats (many modules below depend on this) and used for raster processing.
2. Geopandas: Working with geospatial data in Python made easier, combines the capabilities of pandas and shapely.
3. Shapely: Python package for manipulation and analysis of planar geometric objects (based on widely deployed GEOS).
4. Fiona: Reading and writing spatial data (alternative for geopandas).
5. Pyproj: Performs cartographic transformations and geodetic computations (based on PROJ.4).
6. Pysal: Library of spatial analysis functions written in Python.
7. Geopy: Geocoding library: coordinates to address and address to coordinates.
8. GeoViews: interactive Maps for the web.
9. Networkx: Network analysis and routing in Python (e.g., Dijkstra and A* -algorithms)
10. Cartopy: Make drawing maps for data analysis and visualization as easy as possible.
11. Scipy.spatial: Spatial algorithms and data structures.
12. Rtree: Spatial indexing for Python for quick spatial lookups.
13. Rasterio: Clean and fast and geospatial raster I/O for Python.
14. RSGISLib: Remote Sensing and GIS Software Library for Python.

Reading list

Essential Reading

Zandbergen, P. A. (2015). *Python scripting for ArcGIS*. Esri press.
Lawhead, J. (2013). *Learning geospatial analysis with Python*. Packt Publishing Ltd.
Westra, E. (2013). Python geospatial development. Packt Publishing Ltd.

Additional Reading

Pimpler, E. (2013). *Programming ArcGIS 10.1 with Python Cookbook*. Packt Publishing Ltd.
Tateosian, L. (2015). Python For ArcGIS. Springer.
Kropla, B. (2006). Beginning MapServer: open source GIS development. Apress.

Date: March, 2022

DSC406 Communications for Data Science

Module Code and Title: DSC406 Communications for Data Science
Programme: BSc in Data Science & BSc in Statistics
Credit Value: 12
Module Tutors: P. Paulraj & Ms. Anju Chhetri & Ms. Tenzin Dema
Advisor: Prof. Yuju Kuo

General Objective:

The module is designed to integrate both the technical knowledge, and the ability to communicate effectively, for the purpose of problem-solving in the corporate environment. After establishing competence with basic spreadsheet applications, the module introduces the elements of effective human communication, which impact the collection of requirements for the application, as well as the application's layout and presentation. The module then covers the technical aspects of the choices available for presenting the data/results, based on the intent of the application, and the needs of the audience. Human factors involved in the live presentation of results are then introduced. The group project gives students the opportunity to put their theoretical learning into practice. Particular emphasis is given to spreadsheet applications involving model-building and simulations.

Software Requirement: To teach this module one of the following Opensource tools will be used:

Apache OpenOffice Calc/Google Sheet.

These tools do not require much/any coding and manages to deliver better results than the paid versions like Oracle Data Mining, Rapid Insight Veera, Sisense, Periscope Data, and Microsoft SQL Server Integration Services and so on.

Learning Outcomes:

On completion of the module, students will be able to:

1. explain corporate data terminologies and the Spreadsheet Application.
2. apply conversation skills to foster interactive dialogue
3. identify and apply appropriate data simulation process to generate necessary data
4. create table to organize and analyze related data.
5. arrange and summarize complex data using Pivot table.
6. assemble information from several sources to create a consolidated data set.
7. create data visualization models using appropriate tools.
8. design interactive dashboards to display real-time key metrics and performance indicators of a corporate.

9. tackle (examine, or analyze) business problems using Spreadsheet data modeling.
10. evaluate a decision with Monte Carlo Simulation using Solver tool of Spreadsheet.
11. apply facilitation strategies for getting results and building trust

Learning and Teaching Approach:

Type	Approach	Hours per week	Total credit hours
Contact	Lecture	3	90
	Laboratory Session	3	
Independent study	Group Project & Presentation	1	30
	Self-Study	1	
Total			120

Assessment approach:

1. Continuous Assessment: Portion of the final mark: 60%

A: Practical Test: Portion of the final mark: 20%

The tutor will administer two practical tests on Data Organization, Data Visualization, Get & Transform Data from different data sources, Spreadsheet models, and Monte Carlo Simulation.

i. Lab Assignment 1: Portion of the final mark: 10%

The tutor will conduct first practical test on Data Organization, Data Presentation, and Patterns and trends on Big Data upon coverage of at least 40% of the subject matter. The test will be conducted for a duration of 60 minutes to assess the students' data visualizing skills trend setting ability. The division of marks among the units covered are given below:

Usage of Spreadsheet functions – 5 Marks
Usage of Data organization and Data visualization techniques – 3 Marks
Knowledge on Conditional Formatting – 2 Marks

ii. Lab Assignment 2: Portion of the final mark: 10%

The tutor will conduct second practical test on Spreadsheet Models, and Monte Carlo Simulation upon coverage of at least 90% of the subject matter. The test will be conducted for a duration of 60 minutes to appraise the students' data modeling and decision-making skills. The division of marks among the units covered are given below:

Usage of advanced Spreadsheet functions – 3 Marks
Usage of Power Query/Pivot Table in Spreadsheet – 4 Marks
Knowledge on Analytical Solver Platform – 3 Marks

B: Group project Work & Presentation: Portion of the final mark: 40%

Upon completion of 80% of the subject matter students will be asked to work on a two weeks-long group project (4-5 students) to analyse big data. Each project will be expected to uncover hidden patterns, correlations and meaningful insights about trends. The group work will be evaluated in two parts: Presentation & Project; each project will be graded for 30 marks which includes project evaluation for 25 marks and group work assessment for 5 marks. Each group will have to submit a project report of minimum of 1000 words and the maximum of not exceeding 1500 words and present their findings for a duration of 30 minutes during the 14th week of the semester. This communication skill will be evaluated for 10 marks.

Marking Criteria: Project & Presentation

1. Project Assessment Rubric		
Task	**Expectations**	**Marks**
Data Collection	Choose an appropriate data set from the listed data sources for analysis	**2**
Data Cleaning	Use an appropriate tool/s in Spreadsheet to carry out the following tasks: 1. Removal of unwanted observations 2. Fixing structural errors 3. Managing unwanted outliers 4. Handling missing data	**3**
Data Analysis	Use Analysis Tool Pack to analyse the data	**4**
Data Visualization	Use Pivot table to design an interactive dashboard to unveil the meaningful insights of a concern.	**6**
Project Report	Fulfil the following elements: 1. Background of the company - 1 Mark 2. Data Dictionary – 1 Mark 3. Tables & Charts – 2 Marks 4. Data interpretation - 4 Marks 5. Spelling & Grammar - 2 Marks	**10**

Total Marks	25
2. Group Work Assessment Rubric	

Skills	Expectations	Marks	Final Mark
Contributions & Attitude	1. Always willing to help and do more. Routinely offered useful ideas. Always displays positive attitude.	**4**	
	2. Cooperative. Usually offered useful ideas. Generally, displays positive attitude.	**3**	
	3. Sometimes cooperative. Sometimes offered useful ideas. Rarely displays positive attitude.	**2**	
	4. Seldom cooperative. Rarely offers useful ideas. Is disruptive.	**1**	
Cooperation with Others	1. Did more than others–highly productive. Works extremely well with others. Never argues.	**4**	
	2. Did their part of the work-cooperative. Works well with others. Rarely argues.	**3**	
	3. Could have done more of the work–has difficulty. Requires structure, directions, and leadership. Argues sometimes.	**2**	
	4. Did not do any work–does not contribute. Does not work well with others. Usually argues with teammates.	**1**	
Focus, Commitments	1. Tries to keep people working together. Almost always focused on the task and what needs to be done. Is very self-directed.	**4**	
	2. Does not cause problems in the group. Focuses on the task and what needs to be done most of the time. Can count on this person.	**3**	
	3. Sometimes not a good team member. Sometimes focuses on the task and what needs to be done. Must be prodded and reminded to keep on task.	**2**	
	4. Often is not a good team member. Does not focus on the task and what needs to be done. Let's others do the work.	**1**	
	1. Participated in all group meetings. Assumed leadership role as necessary. Did the work that was assigned by the group.	**4**	
	2. Participated in most group meetings. Provided leadership when asked. Did most of the work assigned by the group.	**3**	

Team Role Fulfilment	3. Participated in some group meetings. Provided some leadership. Did some of the work assigned by the group.	**2**	
	4. Participated in few or no group meetings. Provided no leadership. Did little or no work assigned by the group.	**1**	
Ability to Communicate	1. Always listens to, shares with, and supports the efforts of others. Provided effective feedback to other members. Relays a great deal of information–all relates to the topic.	**4**	
	2. Usually listens to, shares with, and supports the efforts of others. Sometimes talks too much. Provided some effective feedback to others. Relays some basic information–most relates to the topic.	**3**	
	3. Often listens to, shares with, and supports the efforts of others. Usually does most of the talking–rarely listens to others. Provided little feedback to others. Relays very little information–some relates to the topic.	**2**	
	4. Rarely listens to, shares with, or supports the efforts of others. Is always talking and never listens to others. Provided no feedback to others. Does not relay any information to teammates.	**1**	
Total Marks			
Reduce the Total Marks (out of 5)			**5**
Grand Total (1 + Reduced Total Marks)			**30**

Project Presentation Rubric (Group)					
	2 (40%)	**1.5 (30%)**	**1 (20%)**	**.5 (10%)**	**Marks**
Content	The presentation was a concise summary of the topic with all questions answered. Comprehensive and complete coverage of information.	The presentation did a good job of summarizing the subject. Most crucial information is covered; minimal unimportant information.	The presentation was informative but several elements went unanswered. Much of the information irrelevant; coverage of some of major points.	The presentation was a brief look at the topic but many questions were left unanswered. Majority of information irrelevant and significant points left out.	
Visual Appeal	There are no errors in spelling, grammar and punctuation. Information is clear and concise on each slide. Visually appealing/engaging.	There are a few spelling, grammatical, and punctuation mistakes. On two or more slides, there is too much information. considerable visual appeal.	There are many errors in spelling, grammar and punctuation. Too much information was contained on many slides. Minimal effort made to make slides appealing or too much going on.	There are many errors in spelling, grammar and punctuation. The slides were difficult to read and too much information had been copied onto them. No visual appeal.	
Comprehension	Extensive knowledge of topic. Members showed complete understanding of assignment. Accurately answered all questions posed.	Most demonstrated a solid grasp of the subject. Any queries from the audience can be answered by all members.	Few members showed good understanding of some parts of topic. Only some members accurately answered questions.	Presenters didn't understand topic. Majority of questions answered by only one member or majority of information incorrect.	

Presentation Skills	Regular/constant eye contact, the audience was engaged, and presenters held the audience's attention. Appropriate speaking volume & body language.	Most members spoke to majority of audience; steady eye contact. The audience was engaged by the presentation. Majority of presenters spoke at a suitable volume. Some fidgeting by member(s).	Members focused on only part of audience. Sporadic eye contact by more than one presenter. The audience was distracted. Speakers could be heard by only half of the audience. Body language was distracting.	Minimal eye contact by more than one member focusing on small part of audience. The audience was not engaged. Majority of presenters spoke too quickly or quietly making it difficult to understand. Inappropriate/disinterested body language.	
Preparedness/ Participation/ Group Dynamics/Use of visual aids	All presenters knew the information, participated equally, and helped each other as needed. Extremely prepared and rehearsed. Slide contents are short & simple with appropriate usage of visual effects.	Slight hegemony of one speaker. Members aided one another. Good preparation. Simple and brief slide content is there, but extraneous visual effects are not.	Significant controlling by some members with one minimally contributing. Primarily prepared but with some dependence on just reading off slides. Slide contents are large and no visual effects.	Unbalanced presentation or tension resulting from over-helping. Multiple group members not participating. Evident lack of preparation/rehearsal. Dependence on slides. Lack of technical skills in preparation of the slides.	
Total Marks					**10**

2. Semester-End Examination (Practical): Portion of Final Mark: 40%

Students will appear for semester-end practical examination for a duration of 2 hours. It will involve a comprehensive examination of what has been exposed to them in the class (lectures/practical and discussions) to assess their ability in visualizing data, apply data modeling tools and employ Pivot table and Power Query to analyse the Big Data and suggest pertinent solutions. The division of marks are given below:

From Unit I to Unit II: 2 questions on data organization, data visualizing and data summarization - 20 marks.

From Unit III to Unit IV: 2 questions on Spreadsheet models and Monte Carlo Simulation – 20 marks.

Areas of assignments	Quantity	Weighting
A. Practical Test	2	20%
B. Group Project & Report	1	30%
C. Group Presentation	1	10%
D. Semester-End Examination (Practical)	1	40%
Total		100%

Pre-requisites: None

Subject Matter:

Unit I: Introduction to Spreadsheet Application

1.1 Basic concepts of Spreadsheet
1.2 Highlight information using Conditional Formatting
1.3 Formulae, Functions and Name Manager
- 1.3.1 – Financial Functions
- 1.3.2 – Logical Functions
- 1.3.3 – Text Functions
- 1.3.4 – Date & Time Functions
- 1.3.5 Lookup & References Functions
- 1.3.6 Math or Trigonometry Functions
- 1.3.7 Statistical Functions

1.4 Protect a range, Worksheet and Workbook.

Unit 2: Communication Skills & Barriers to Communication

2.1 Stress

2.2 Lack of focus

2.3 Body Language
2.4 Breaking barriers to effective communication
2.5 Soft Skills
2.5.1 Emotional Intelligence

2.5.1.1 Self-Awareness

2.5.1.2 Self-Regulation

2.5.1.3 Social Skills

2.5.1.4 Empathy

2.5.1.5 Motivation

2.5.2 Active Listening Skills:

2.5.2.1 Paying Attention

2.5.2.2 Withholding Judgement

2.5.2.3 Reflecting

2.5.2.4 Clarifying

2.5.2.5 Summarizing

2.5.2.6 Sharing

2.5.3 Communicating under stress

2.5.3.1 Understanding stress

2.5.3.2 Assess the situation

Unit III: Data Visualization Spreadsheet Model

3.1 Overview of Data Visualization

3.2 Tables and Pivot Table

3.3 Get & Transform Data from different sources using Power Query

3.4 Advanced Data Visualization

3.4.1 Advanced Charts

3.4.2 Geographic Information Systems Charts (Field Map)

3.5 Data Dashboards

3.5.1 Principles of Effective Data Dashboards

3.5.2 Applications of Data Dashboards

3.5.3 Design an interactive Business Dashboard

Unit IV: Persuasive Communication and Critical Communication

4.1 Communicating without intimidation

4.2 Art of Persuasion

4.2.1 Credibility

4.2.2 Common ground

4.2.3 Emotional Connection

4.3 How to deal with difficult people

4.4 Persuasion ethics

4.4.1 Beneficence

4.4.2 non-Maleficence

4.5 Public Speaking and Presentations

4.5.1 Voice and body language

4.5.2 Vocal 1projection

4.5.3 Pace and breathing

4.5.4 Stance and posture

4.5.5 Gesture and movement.

4.6 Content and structure

4.6.1 Preparation and organization

4.6.2 Narrative and messaging

4.6.3 Visual aids

4.7 Audience engagement

4.8 Quality of execution

Unit V: Spreadsheet Models

5.1 Building Good Spreadsheet Model

5.2 Data Tables

5.3 Goal Seek

5.4 Some Useful Spreadsheet Functions for Modeling

5.5 Auditing Spreadsheet Models

5.5.1 Trace Precedents and Dependents
5.5.2 Show and Evaluate Formulas
5.5.3 Error Checking and Watch Window

Unit VI: Monte Carlo Simulation

6.1 What-If Analysis

6.1.1 Base-Case Scenario

6.1.2 Worst-Case Scenario

6.1.3 Best-Case Scenario

6.2 Simulation Modeling with Native Spreadsheet Functions

6.2.1 Generating Values for Random Variables with Spreadsheet

6.2.2 Executing Simulation Trials with Spreadsheet

6.2.3 Measuring and Analysing Simulation Output

6.3 Simulation Modeling with Analytic Solver Platform

6.3.1 Spreadsheet Model for a real-time Problem

6.3.2 Generating Values for a real-time Problem's Random Variables

6.3.3. Tracking Output Measures of a real-time Problem

6.3.4 Executing Simulation Trials and Analysing Output for a real-time problem

List of Practical in Spreadsheet Package (Dataset Sources: nsb.gov.bt, Data.gov, Tableau.com, Contextures.com, Kaggle.com, Data.gov.sg, and archive.ics.uci.edu *(*UCI Machine Learning Repository): Data Sets

1. Frequency Distributions for Categorical Data and Quantitative Data
2. Sorting and Filtering of Data and Conditional Formatting of Data
3. Data wrangling using Text & Lookup Functions
4. Data Visualization and Cross tabulation of big data using Pivot table and Charts
5. Import .CSV & .TSV files into .xlsx/. gsheet/.ods
6. Using Power Query to analyze a real-time data
7. Determine the value of an input based on an output using Data Tables and Goal Seek
8. Solve Linear Programs using Solver Tool
9. Monte Carlo Simulation using Scenario Manager

List of functions to be delivered to carry out the above listed practical:

1. Financial Functions: PMT, IRR, IPMT, PPMT, NPV, FV, RRI & Yield.

2. Logical Functions: IF, IFERROR, AND, OR, XOR & LET.

3. Text Functions: PROPER, LOWER, UPPER, LEFT, RIGHT, MID, CONCAT, TEXTJOIN, TRIM, LEN & CLEAN.

4. Date & Time Functions: DATE, DAYS360, EOMONTH, MONTH, YEAR, DAY, WEEKDAY, WORKDAY, TODAY, & NOW.

5. Lookup & Reference Functions: CHOOSE, COLUMN, FILTER, FORMULATEXT, HLOOKUP, LOOKUP, VLOOKUP, XLOOKUP, MATCH, UNIQUE, INDIRECT & INDEX.

6. Mathematical Functions: EXP, FACT, POWER, PRODUCT, RAND, RANDBETWEEN, RANDARRAY, ROMAN, ROUND, ROUNDDOWN, ROUNDUP, SEQUENCE, SUM, SUMIF, SUMIFS, SUMPRODUCT, SUMSQ, SUBTOTAL, & SQRT.

7. Statistical Functions: AVERAGE, AVERAGEIF, AVERAGEIFS, COUNT, COUNTA, COUNTBLANK, COUNTIF, COUNTIFS, FREQUENCY, LARGE, SMALL, MAX, MAXA, MIN, MINA, RANK.AVG, RANK.EQ, STDEV.P, & STDEV.S

Reading List:

Essential Reading:

Winston, W. (2019). Microsoft Excel 2019 Data Analysis and Business Modeling: Data Analysis and Business Modeling. Pearson Education.

Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., & Anderson, D. R. (2014). Essentials of Business Analytics (Book Only). Nelson Education.

Lucas, S., & Suya, Y. (2004). The art of public speaking. New York: McGraw-Hill.

Stiff, J. B., & Mongeau, P. A. (2016). Persuasive communication. Guilford Publications.

Pochiraju, B., & Seshadri, S. (Eds.). (2018). Essentials of Business Analytics: An Introduction to the Methodology and Its Applications (Vol. 264). Springer.

Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. Communications of the Association for Information Systems, 37(1), 23.

Ragsdale, C. (2014). Spreadsheet modeling and decision analysis: a practical introduction to business analytics. Nelson Education.

Mithas, S., Lee, M. R., Earley, S., Murugesan, S., & Djavanshir, R. (2013). Leveraging big data and business analytics [Guest editors' introduction]. IT professional, 15(6), 18-20.

Additional Reading:

Albright, S. C., & Winston, W. L. (2019). Business analytics: Data analysis & decision making. Cengage Learning.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, *64*, 130-141.

Carnegie, D. (2017). *How to develop self-confidence and influence people by public speaking*. Simon and Schuster.

Hargie, O. (2018). *The Handbook of Communication Skills. (4th ed.)*. London:Routledge.

Seufert, A., & Schiefer, J. (2005, August). Enhanced business intelligence-supporting business processes with real-world business analytics. In *16th International Workshop on Database and Expert Systems Applications (DEXA'05)* (pp. 919-925). IEEE.

Gopalkrishnan, V., Steier, D., Lewis, H., & Guszcza, J. (2012, August). Big data, big business: bridging the gap. In *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications* (pp. 7-11).

Laursen, G. H., & Thorlund, J. (2016). *Business analytics for managers: Taking business intelligence beyond reporting*. John Wiley & Sons.

April 25, 2025

Year 4 Semester II

RSM413 Data Science Internship

Module Code & Title: RSM413 Data Science Internship
Programme: BSc in Data Science
Credit Value: 12
Module Tutor: Phub Namgay

General Objective

The internship programme is an opportunity to allow students to connect the knowledge (theories, principles, concepts), skills (soft and hard), and competence from the Data Science discourse in the college with a real-world work situation. Using the internship site as a platform to connect theory with practice, the programme will help students gain insights and hands-on experience into data management, stewardship, and governance by data practitioners in the real world and further hone their skills to become data science professionals. The experience of working as a subordinate with data practitioners or professionals on data workflows, such as planning, designing, testing, evaluating, computing organisational data, will be formative and prepare the student to transition from student life to career, thus enhancing their employability.

Learning Outcomes

On completion of the module, students will be able to:

1. Connect the academic knowledge and skills gained from data science discourse with data.
2. Effectively and efficiently assess data quantitatively and/or qualitatively and utilise appropriate methods to perform analysis.
3. Recognise data science tools and techniques that are current and central in the data ecosystem and implement them optimally on real-world problems.
4. Demonstrate basic proficiency in management and stewardship of data, namely developing metadata and data management plans (DMPs).
5. Develop soft and hard skills to organise, synthesise, and articulate ideas to handle data.
6. Perform essential data governance task through insights into current data principle and frameworks.
7. Interpret, critically evaluate, and effectively communicate data analysis procedure and results through a report, discussion, and presentation.
8. Demonstrate ability to work as a team, listen and respond to others, and use negation and conflict resolution skills.

Learning and Teaching Approach:

- **Student** – The student must undergo two and half months of internship in relevant government agencies or private organisations where data is central to their business operation. They must work closely with an immediate supervisor, referred to as a *mentor*, who is a data practitioner or have experience in dealing with data (refer to '*Responsible Conduct and Engagement'* below). The nature of the host organisation or data domain is irrelevant to choose an organisation or supervisor to intern. The student is required to familiarise themselves with the sociotechnical policies and guidelines of the host

organisation. Furthermore, s/he must abide by the policies on academic integrity and responsible conduct of the Royal University of Bhutan (RUB).

- **Mentor** – The mentor is a specific individual at the host organisation who will supervise the intern (mentee). S/he will also serve as a point of contact with the internship coordinator at the college. The mentor is also responsible for students' activities and learning; observe and advise the student routinely; periodically consult with the coordinator on the students' progress; and complete the post-internship evaluation.
- **Examiner**– Apart from the mentor, the college's visiting examiner will visit all the host organisations to carry out a mid-term review of the student's progress. Students must produce an original work plan and other tasks executed to the visiting examiner. All the documents should be submitted in hardcopy along with the final report after the internship to the coordinator of the internship.
- **Internship Coordinator**– A faculty member from the department will be nominated to coordinate the internship and liaise with organisations or agencies for internship opportunities, networking, and other relevant matters. The coordinator will also maintain communication with students and mentors during the internship period. Additional responsibilities include compiling assessment marks, preparing a work plan for an internship, developing report templates and rubrics. Furthermore, a mechanism for quality control through the regular meetings (at least twice during the internship period) among faculty members and the host organisation is also the coordinator's responsibility.

Assessment approach:

The Programme Team will objectively assess the fruitfulness of the internship experience through the following formative assessments:

a. **IAP report** – Internship Activity Plan (IAP) aims to outline what the student wants to realise while on internship. The document should be developed in consultation with the mentor. It should identify and include core tasks and a description of each task the intern will perform during the internship period. The student should submit an IAP report (300 – 500 words) to the internship coordinator in the programme and mentor in the host organisation. IAP reports will serve as a guideline for the student to accomplish the task assigned to him or her by the host organisation. Furthermore, the document will serve as a mechanism to track student's work for the internship coordinator and mentor.
b. **Progress report** –The student must submit a progress report of 2-3 pages (1000 – 1500 words) with status at the end of each month. The report should be approved and sealed by the student's mentor.
c. **Final report** – The student is supposed to submit a final report of 8-10 pages (4000 – 5000 words APA format) detailing internship organisation, work domain, activities in the IAP executed, knowledge and skills obtained, and overall internship experience.
d. **Presentation –** The student must present their internship report (max.15 slides for 10-15 minutes) during an internship report session to the programme staff and other students. The presentation modalities, such as template and deadlines for report and presentation and student assessment form, will be taken care of by the semester guide.
e. **Post-internship mentor evaluation** – After the internship period, the assigned mentor will objectively assess their mentee's performance. The assessment will be part of the formative assessment of the student.
f. **Internship coordinator evaluation** – After the internship, the internship coordinator will assess the students' performance based on the documents they submitted, such as IAP, final report, presentation, and overall enthusiasm in the internship programme.

An overview of the assessment approaches and weighting:

Areas of assessment	Quantity	Weighting
Progress report	1	10%
Final report	1	35%
Final presentation	1	25%
Post-internship mentor evaluation	1	10%
Examiner evaluation	1	20%

Detailed assessment

Sl. No	Assessment	No	Mark	Weightage
i.	Continuous Assessments			
1	**Progress Report** a. Introduction (5) b. Major activities accomplished (10) c. Review of IAP (10)	1	25	10%
2	**Examiner Evaluation** a. IAP fulfilment and report verification (10) b. Data analytics performed demo (10) c. Workplace assessment (5) d. Question & Answer (5)	1	30	20%
3	**Post-internship Mentor Evaluation** a. Domain knowledge (10) b. IAP accomplishment (5) c. Worth ethics and responsibilities (5) Punctuality (5)	1	25	10%
ii.	**Semester-end Assessment**			
4	**Final report** a. Executive summary (5) b. Introduction (10) c. About host organisation (20) d. Domain of work (20) e. Critical appraisal of activity (10) f. Contribution to knowledge and skills (15) g. Internship experience and relevance (10) h. Conclusion (5) i. Format and referencing (5)	1	100	35%
5	**Final Presentation** a. Introduction (5) b. Content (20) c. Presentation skills and techniques (5) d. Q & A (10) e. Demo of data analytics (10)	1	50	25%

	Note– students must present a hardcopy report and all the other necessary reports to the examiners.			
			Total	**100**

A group of examiners comprising faculty members from the Data Science programme will evaluate the student's overall internship performance.

Responsible Conduct and Engagement

A. **Conduct** –Students represent the programme and college; they will take on the responsibilities assigned to them with true professionalism at all times. The student must ensure that their effort and contribution meet or exceed the expectations of the host organisation. Furthermore, they must show respect for and follow the workplace's policies and routines in the host organisation.
B. **Training** – During the internship training, the student is expected to actively engage in data-related organisational activities to accomplish the learning goals outlined in the IAP report. The essential activities include (but is not limited to) the following:
    i. Execute non-critical (or critical if one is capable of handling) data analytics task in the organisation;
    ii. Design and develop flowcharts, pseudocodes, and algorithms that are efficient and effective in time and space complexity, to solve organisational data-related problems;
    iii. Assist the host organisation in developing data management and stewardship plans, such as DMP and metadata to handle heterogeneous data;
    iv. Participate in data-related internal meetings and/or workshops to gain insights into data; and
    v. Work on miscellaneous data related task or activities that require the student's knowledge and expertise
C. **Critical appraisal** – The student must submit a critical appraisal report of any data-intensive task they carried out. The report should incorporate the theories, principles, and concepts learned in the classroom with practice. It should be supplemented with facts and figures. However, the student should seek permission from the host organisations to use any organisational-related facts and figures to use in the report. The report will be a component of the final report.

Date: April, 2021

RSM414 Capstone Project

Module Code & Title: RSM414 Capstone Project
Programme: BSc in Data Science
Credit Value: 48
Module Tutor: Phub Namgay

General Objective

Students are expected to carry out a capstone project under a supervisor, which culminates in a final project report. The time frame of the project is 12 weeks. The project aims to provide a platform for students to apply knowledge and skills gained throughout the data science academic discourse on a real-world project. Moreover, it provides an avenue for students to use the theoretical foundations of the programme in a real-world case, thereby better preparing them to handle data science undertakings in a real work environment. It will also help students gather, analyse, model, interpret, and apply data analytics tools and techniques on small or big data from varied sources. The successful submission of a project report and presentation of their project is a strong indicator of data science repertoire in a student. The semester-long project is equivalent to three modules of 48 credits.

Learning Outcomes

On completion of the capstone project, students will be able to:

1. Identify a suitable topic for a project, formulate and articulate questions, hypotheses, and objectives
2. Carry out a literature search and summarise the state of the art
3. Articulate suitable methodology and professionally engage with primary or secondary data sources
4. Handle real-world data science problems and apply expertise and skills to solve a problem
5. Design and develop effective and efficient resources and methods to solve a data science problem optimally
6. Implement tools for activity in data science projects such as programming, statistics, database, and project management
7. Demonstrate proficiency in management and stewardship of data, namely developing metadata, ontologies, and data management plans (DMPs)
8. Interpret, evaluate, and communicate project results through a report, discussion, and presentation
9. Develop portable skills, especially in writing project proposals, status reports, documentation, and presentation slides

Learning and Teaching Approach:

- **Project coordinator** – Capstone Project Committee (CPC) consists of the head of department (HoD), programme leader (PL), and supervisors. The Programme's PL will serve as a capstone project coordinator and is responsible for administrative aspects of the project, such as creating a project work plan, designing templates and rubrics for proposals, and chairing presentation seminars. Towards the end of year 4, semester I, the coordinator shall convene a meeting with the CPC members to discuss the modalities of the capstone project, such as finding projects, nominating a supervisor, and other project-related activities. With support from the HoD, the coordinator will liaise with other organisations and agencies for

data-science-related projects that students can take up as capstone projects. Students will also be encouraged to find projects worth working on as capstone projects. However, the caveat is that a project should be relevant to the data science programme or related to some aspects of data. The coordinator shall present a working report to the CPC members during the autumn semester's closing meeting to ensure that the number of projects is adequate and aspects of the capstone project are in place.

- **Supervisor** – Student(s) will work under the supervision of a faculty (supervisor) nominated by the CPC. The supervisor guides the student's project and provides timely feedback on all graded and ungraded components. It is worth noting that the supervisor need not have a background or subject knowledge of the project undertaken by a student. Furthermore, the supervisor will report students' progress or related matters to the coordinator and the CPC. Students must meet their supervisor at least once a week to update work progress and resolve issues that impede project progress. The supervisor and students can internally decide on modalities to assess the project's status, such as weekly meetings and demo presentations. However, it should be consistent with other students.

Table 1. Credit breakdown for the capstone project

Type	Teaching and Learning Approaches	Hours per week	Total credit hours
Contact	Lecture	2	30
	Individual tutorial and guidance	5	75
Independent study	Writing of reports, including seminar	25	375
Total			**480**

Assessment approach

A Panel with at least four members that includes an external evaluator, preferably with a technical background, invited from another college will assess the capstone project. The student's evaluation will be 60% (240 marks) by the Panel and 40 % (160 marks) by the supervisor for a total mark of 400. As for assessment, the Panel will be involved in activities C and D of **Table *2***, whereas the supervisor will be involved in activities A, B, and E. The rationale for not involving the supervisor in activities C and D is for fresh insights and perspectives by an external evaluator to improve the project. Moreover, involving the supervisor is redundant as they are already involved across the capstone project lifecycle. The page limit and word count (including footnotes, excluding title, abstract, bibliography, and appendices) and assessment weighting for each task are given in each section. **Table *2*** encapsulates the quantity of assessment and its weight for the capstone project.

As for ethics in the project, ethical considerations will be taken seriously across all phases of the capstone project. For example, any sensitive data will be handled with the utmost care, such as using de-identification or anonymisation techniques, depending on the needs and requirements of the project. The capstone project must be conducted in line with the code of ethics and intellectual property policies of the Royal University of Bhutan. Where necessary, capstone project will consider the ethical permit and approval requirement. The College Research Committee (CRC) will be responsible for giving the ethical clearance for the project. Moreover, the project coordinator will orient students on project ethics and personal conduct through a week-long seminar (Unit I: Introduction to Capstone Project and Ethics) at the beginning of the semester. The project coordinator and supervisor are responsible for ensuring that activities are conducted following good project practices. Supervisor consent is necessary if students want to write a

conference or journal paper based on the project. Moreover, students are encouraged to explore such avenues to disseminate knowledge.

A: Project proposal report: 20%

Student(s) will prepare a preliminary proposal (page limit is 10–15 pages and approximately 3500 words) and submit it to their supervisor. The supervisor will be responsible for supervisory support and assistance while planning, designing, and drafting a project proposal. It is worth noting that the project proposal is permissible for change only before acceptance, and the change request will not be considered afterwards. As for assessment, the project proposal will be assessed for 80 marks (converted into 20% of the final mark) using the following criteria:

- Introduction: 5
- Problem description:15
- Purpose of the project: 10
- Overview of literature: 15
- Methodology and methods: 20
- Evaluation techniques: 15

Regarding the evaluation of the capstone project, evaluation techniques will constitute modalities for evaluating the project deliverables, such as algorithms, programs, and software. Students must draft a preliminary framework for assessing deliverables, such as test cases to evaluate an algorithm and program. If there is a well-established framework, students are advised to use it to avoid rework. Moreover, it enhances the validity of project deliverables.

B: Progress review: 10%

The review of project progress is crucial for timely constructive feedback and comments, thereby contributing to the project's milestones. Moreover, it assures the progression of a project according to the goals and objectives outlined in the proposal. The project will be reviewed twice—before and after the mid-semester examination. The supervisor will assess their respective student's progress reports (2–3 pages, approximately 750 words). The report will be evaluated for 40 marks (10% of the final mark) using the following criteria:

- Project status: 15
- Deliverables: 20
- Timeline and the way forward: 5

C: Project report: 40%

Student(s) must submit the final project report (two copies) to the coordinator two weeks prior to the presentation. Subsequently, the coordinator will hand over a report to the external evaluator to assess the student through a presentation and question and answer session. The page and approximate word limits are 15–25 pages and 10000-12000 words, respectively. Likewise, the project report will be assessed for 160 marks, which is equivalent to 40% of the final mark, using the following criteria (detailed rubrics in **Appendix A**):

- Introduction: 20
- Overview of literature: 25
- Methodology and methods: 40
- Results, findings, and interpretation: 40
- Metadata and DMPS: 20
- Implications of the project: 15

The report should align with the initial project proposal—that is, the introduction, literature, methodology and methods should simply be an augmented version of things mentioned in the proposal. The emphasis will be more on the last three criteria in the list. Students should comprehensively provide a detail of the analysis and synthesis of the results and findings of their project. As for interpretation, students should discuss their project's implications with industry practitioners. They also have to highlight the limitations of their projects. In order to ensure good data governance and management practices, students have to use some metadata schema, such as Dublin Core, to document and describe their project data set. Likewise, they are also advised to maintain DMPS for all the data sets throughout the project. Finally, regarding implication, students should discuss the wider implication of their project for practitioners. The elements above will be the basis for evaluating the final project.

D: Project seminar: 20%

Student(s) will present their project to the panel. Student(s) will be given 60 minutes for presentation; the time breakdown will depend on the rubrics. The assessment rubrics will be made available to the student and supervisor prior to the presentation. The coordinator and examiner will compile the marks and be certified by the respective supervisors. The presentation will be assessed for 80 marks, which is equivalent to 20% of the final mark, using the following criteria (detailed rubrics in **Appendix A**):

- Organisation and coherence: 20
- Understanding and clarity of the subject matter: 25
- Central idea of the capstone project: 20
- Question and answer sessions: 15

E: Monthly progress meeting: 10%

A student should maintain a project diary that records each activity from inception to completion. Additionally, the project diary should be submitted separately to their supervisor. Supervisors could use the diary as a reference to assess individual students' learning and progress. The meeting will be evaluated for 40 marks, which is 10% of the final mark, using the following criteria:

- Project diary: 15
- Question and answer session: 15
- Way forward of the project: 10

Table 2. An overview of the assessment approaches and weighting

Areas of assessment	Quantity	Weighting	Assessor	
			Panel	Supervisor
A. Project proposal report	1	20%		✓
B. Progress review	2	10%		✓
C. Project report	1	40%	✓	
D. Project seminar	1	20%	✓	
E. Monthly progress meeting	4	10%		✓

Pre-requisites: RSM301 Research Methods

Subject Matter:

Unit 1: Introduction to Capstone Project and Ethics

- 1.1 Capstone project
  - 1.1.1 Writing proposal and final report
- 1.2 Literature review
  - 1.2.1 Planning and organising literature review
- 1.3 Project scheduling
  - 1.3.1 Project activity schedule using project management tools
- 1.4 Metadata and Data Management Plans (DMPs)
- 1.5 Data collection approaches and techniques
  - 1.5.1 Data privacy
  - 1.5.2 Data protection acts
  - 1.5.3 General data protection regulation (GDPR)
  - 1.5.4 Commercial confidentiality
- 1.6 Ethics in the capstone project

Unit 2: Project proposal report and References

- 2.1 Introduction
  - 2.1.1 Brief overview of the project topic
  - 2.1.2 Goals of the project
  - 2.1.3 Scope of the project
- 2.2 Purpose
  - 2.2.1 Statement of the purpose
- 2.3 Problem description
  - 2.3.1 Elaborate on the problem statement
  - 2.3.2 Nature of the problem
    - 2.3.2.1 Rationale for believing the problem exists
  - 2.3.3 Describe how the project will address the problem
  - 2.3.4 Limitations of the project
- 2.4 Literature review
  - 2.4.1 Theoretical justification of the project
  - 2.4.2 Process, validity, and value of the project
- 2.5 Methodology
  - 2.5.1 Potential methods and techniques
  - 2.5.2 Gathering data/materials
  - 2.5.3 Ethical consideration
  - 2.5.4 Project activity timeline and tools
    - 2.5.4.1 Gantt chart
- 2.6 Evaluation
  - 2.6.1 Specific expected outcomes
  - 2.6.2 Measurable expected outcomes

Unit III: Project Report

- 3.1 Introduction
  - 3.1.1 Problem statement
  - 3.1.2 Significance of the study
  - 3.1.3 Scope statement
  - 3.1.4 Project timeline
    - 3.1.4.1 Tools to render project timeline visually

3.1.5 Feasibility study
3.1.5.1 Assessment of the practicality or viability of the project
3.2 Literature review
3.2.1.1 Strengths, weakness, and opportunities
3.3 Methodology
3.3.1 Methods used to provide structure and cohesiveness
3.3.1.1 Workflow from inception to completion
3.3.2 Subject or participants of the project
3.3.3 Data collection approaches and strategies
3.3.3.1 Advantage, limitation, and ethical issues
3.3.4 Data analysis approaches and software
3.3.4.1 Coding method, analysis of interviews/recordings
3.3.4.2 Statistical analysis
3.4 Results, findings, and discussions
3.5 Metadata and Data Management Plans (DMPs)
3.5.1 Information concerning how data and materials are collected, organised, and stored
3.5.2 Findability, accessibility, interoperability, and reusability (FAIR) of the project data
3.6 Implications, recommendations, and applications
Refer to Appendix A for the final report structure.

Unit IV: Project Seminar

4.1. Oral presentation of the capstone project
4.1.1. Aims and objectives of the project
4.1.2. Organisation and coherence of the presentation
4.1.3. Understanding and clarity of the subject matter
4.1.4. Central idea of the capstone project

Reading List:

Essential Reading List

Hauhart, R. C., & Grahe, J. E. (2015). *Designing and teaching undergraduate capstone courses*. San Francisco: Jossey-Bass.

Hoffman, H. F. (2014). *Engineering Capstone Course: fundamentals for students and instructors.* Cham: Springer.

Additional Reading List

Grix, J. (2004). *The foundations of research. New York:* Palgrave Macmillan.

Strunk, W., & White, E. B. 1. (2000). *The elements of style*. 4th ed. New York: Longman.

Date: April 2021

Appendix A: General Guideline and Template for Capstone Project

1. Introduction

The following guideline is designed to provide a framework for the capstone project of students in data science. The guideline outlines the clear roles and responsibilities of project coordinator, supervisors, students, and external evaluators. The guideline will be used to ensure that all aspects of the capstone project are completed to a high standard, with clear expectations for each category and criteria for assessment. By adhering to the guideline, students can ensure that their capstone project is a comprehensive and insightful piece of work, while supervisors and external evaluators can use the criteria to assess and provide feedback on the project in a consistent and fair manner. Overall, this guideline aims to provide a clear and effective framework for the successful completion and assessment of the capstone project in data science.

2. Roles and Responsibilities

The following are the roles and responsibilities of the students, supervisors, external evaluators, the CPC, the Project Coordinator.

a. CPC:

i. Nominate supervisors for projects.
ii. Identify members for the Panel to evaluate the projects, including an external evaluator.
iii. Ensure that ethical consideration are taken seriously across all phases of the capstone project.
iv. Chiefly responsible for all logistical arrangements of capstone project.

b. Project Coordinator:

i. Responsible for administrative aspects of the project, such as creating a project work plan, designing templates and rubrics for proposals, and chairing presentation seminars.
ii. Convene a meeting with the CPC members to discuss the modalities of the capstone project, such as finding projects, nominating a supervisor, and other project-related activities.
iii. Present a working report to the CPC members during the autumn semester's closing meeting to ensure that the number of projects is adequate and aspects of the capstone project are in place.
iv. Make arrangement for orientation of students on project ethics and personal conduct through a week-long seminar at the beginning of the semester.
v. Hand over final reports to the external evaluator(s).

c. Supervisor:

i. Guide the student's project and provide timely feedback on all graded and ungraded components.
ii. Report students' progress or related matters to the coordinator and the CPC.

iii. Meet with their student at least once a week to review work progress and address any issues.
iv. Collaborate with the student to determine modalities to assess the project's status, such as weekly meetings and demo presentations.
v. Ensure that the student's work is consistent with other students in the program, even if the supervisor does not have a background or subject knowledge of the project undertaken by the student.
vi. Ensure that students have access to necessary resources and tools.

d. Student:

i. Identify a real-world problem and develop a proposal for the project.
ii. Get ethical clearance for the project from the CRC.
iii. Conduct research and collect relevant data.
iv. Work on the capstone project under the supervision of a faculty (supervisor) nominated by the CPC.
v. Meet with their supervisor at least once a week to update work progress and resolve issues that impede project progress.
vi. Submit graded and ungraded components of the project to their supervisor in a timely manner.
vii. Collaborate with the supervisor to determine modalities to assess the project's status, such as weekly meetings and demo presentations.
viii. Analyze the results and draw conclusions.
ix. Prepare a final report and submit final report (two copies) to the PC at least two weeks before the presentation.
x. Present their project to the Panel.
xi. Explore avenues to disseminate knowledge by writing a conference or journal paper based on the project.

e. External Evaluator(s):

i. Review the final report and presentation.
ii. Provide feedback and suggestions to students.
iii. Evaluate the project based on the criteria provided.

3. Template for Project Diary

The following template could be used by the student for maintaining project diary for monthly progress meeting with the supervisor.

Date	Activity	Time Spent	Progress Made	Challenges Encountered	Solution/Next Steps

4. Marking Criteria for Project Report

The following marking rubrics will be used for evaluating the project report.

Criteria	Outstanding (5)	Very Good (4)	Good (3)	Satisfactory (2)	Fail (1)
Introduction (20)	Clearly states the problem and its significance, with a comprehensive explanation of the project's goals and objectives.	Clearly states the problem and its significance, with a well-developed explanation of the project's goals and objectives.	Clearly states the problem and its significance, with a satisfactory explanation of the project's goals and objectives.	States the problem and its significance, but with some lack of clarity or focus on the project's goals and objectives.	Fails to state the problem and its significance, or with significant deficiencies in explanation of the project's goals and objectives.
Score: Comments:					
Overview of Literature (25)	Presents a thorough and insightful review of relevant literature, with strong evidence of critical analysis and synthesis.	Presents a comprehensive review of relevant literature, with good evidence of critical analysis and synthesis.	Presents a satisfactory review of relevant literature, with reasonable evidence of critical analysis and synthesis.	Presents a satisfactory review of relevant literature, with reasonable evidence of critical analysis and synthesis.	Presents a basic review of relevant literature, but with some gaps in critical analysis and synthesis.
Score: Comments:					
Methodology and Methods (40)	Describes the methodology and methods with exceptional clarity and precision, demonstrating a high	Describes the methodology and methods with good clarity and precision, demonstrating a strong level of	Describes the methodology and methods with acceptable clarity and precision, demonstrating a reasonable level of	Describes the methodology and methods with some lack of clarity or precision, demonstrating a basic level of expertise and mastery.	Fails to describe the methodology and methods, or presents them with significant deficiencies in

	level of expertise and mastery.	expertise and mastery.	expertise and mastery.		clarity, precision, or expertise.
Score: Comments:					
Results, Findings, and Interpretation (40)	Presents the results and findings with exceptional clarity and coherence, with insightful and comprehensive interpretation.	Presents the results and findings with good clarity and coherence, with insightful interpretation.	Presents the results and findings with acceptable clarity and coherence, with reasonable interpretation.	Presents the results and findings with some lack of clarity or coherence, with basic interpretation.	Presents the results and findings with significant deficiencies in clarity, coherence, or interpretation.
Score: Comments:					
Metadata and Data Management Plans (20)	Demonstrates an exceptional level of expertise and care in metadata and data management, with comprehensive and effective plans.	Demonstrates a strong level of expertise and care in metadata and data management, with effective plans.	Demonstrates a reasonable level of expertise and care in metadata and data management, with satisfactory plans.	Demonstrates a basic level of expertise and care in metadata and data management, with some gaps in plans.	Demonstrates significant deficiencies in expertise and care in metadata and data management, or presents plans that are inadequate or non-existent.
Score: Comments:					
Implications of the Project (15)	Presents comprehensive and insightful implications of the project, with clear	Presents strong implications of the project, with well-supported recommendations for	Presents reasonable implications of the project, with some support for recommendations for	Presents basic implications of the project, but with some gaps in support for recommendations for	Presents inadequate or non-existent implications of the project, or with

	and compelling recommendations for future research or practical applications.	future research or practical applications.	future research or practical applications.	future research or practical applications.	significant deficiencies in support for recommendations for future research or practical applications.
Score:					
Comments:					

To calculate the score for a given criterion, the formula can be used:

Score = (Weightage for the criterion X Category Achieved) / 5

For instance, if a student's **Introduction** is rated as "**Very Good**", their score is: *(20 x 4) / 5 = 16*

This formula is also applicable for determining the scores for Project Seminar.

5. Marking Criteria for Project Seminar

The following marking rubrics will be used for assessing presentation during Project Seminar.

Criteria	Outstanding (5)	Very Good (4)	Good (3)	Satisfactory (2)	Fail (1)
Organization and coherence (20)	The presentation is exceptionally well-organized, coherent, and easy to follow. Transitions between sections are seamless.	The presentation is well-organized and coherent. Transitions between sections are generally smooth.	The presentation is mostly well-organized and coherent, but some sections may lack clarity or coherence. Transitions between sections are sometimes abrupt.	The presentation is somewhat disorganized and lacks coherence. Transitions between sections are frequently abrupt.	The presentation is poorly organized and lacks coherence. Transitions between sections are often unclear.
Score: Comments:					
Understanding and clarity of the subject matter (25)	The presenter demonstrates an exceptional understanding of the subject matter and presents it with exceptional clarity.	The presenter demonstrates a very good understanding of the subject matter and presents it with very good clarity.	The presenter demonstrates a good understanding of the subject matter and presents it with good clarity.	The presenter demonstrates a satisfactory understanding of the subject matter and presents it with satisfactory clarity.	The presenter demonstrates a poor understanding of the subject matter and presents it with poor clarity.
Score: Comments:					
Central idea of the capstone project (20)	The central idea of the capstone project is exceptionally clear and is effectively conveyed	The central idea of the capstone project is very clear and is effectively conveyed throughout the presentation.	The central idea of the capstone project is mostly clear and is conveyed throughout the presentation.	The central idea of the capstone project is somewhat unclear and is conveyed inconsistently throughout the presentation.	The central idea of the capstone project is unclear and is not effectively conveyed

	throughout the presentation.				throughout the presentation.
Score: Comments:					
Question and answer sessions (15)	The presenter handles questions and answers exceptionally well, providing detailed and insightful responses.	The presenter handles questions and answers very well, providing detailed and insightful responses.	The presenter handles questions and answers generally well, but some responses may lack detail or insight.	The presenter handles questions and answers somewhat poorly, providing incomplete or unclear responses.	The presenter handles questions and answers poorly, providing incomplete or unclear responses.
Score: Comments:					

General template for the Capstone Project

1. Title page - required, not numbered
   1.1. Title of the project
   1.2. Name of the candidates (alphabetical order)
   1.3. Name and designation of the supervisor
   1.4. Degree for which the project is submitted
   1.5. Name of the college
   1.6. Month and year the project is presented
   1.7. Declaration of the student and supervisor
2. Preliminary pages
   2.1. Signature page - not numbered. Pages following it are numbered using lowercase Roman numerals
   2.2. Table of Contents - required
   2.3. List of tables - if needed
   2.4. List of figures - if needed
   2.5. Glossary: if needed
   2.6. List of abbreviations - if needed
   2.7. Acknowledgement
3. Abstract - required, 300 words
4. Main text
   4.1. Introduction
   4.2. Purpose statement
   4.3. Description of problem
   4.4. Literature review
   4.5. Methodology
   4.6. Data collection, analysis and results interpretation
   4.7. Conclusions, recommendations and summary
   4.8. References - Required (15 sources of different types, preferably authoritative documents)
5. Appendices: if needed. Pagination follows the last References page
   5.1. Questionnaire and interview schedule (optional)
   5.2. Observation schedule (optional)
   5.3. Coding frame (optional)
   5.4. Letters sent to sample members (optional)
   5.5. Any other
6. Length of the Project Report
   6.1. Report - 15–25 pages
   6.2. Text style - APA format
   6.3. Alignment - Justify
   6.4. Font - Times New Roman
   6.5. Font size - 12
   6.6. Line spacing - 1.5

Essential Reading List

Hauhart, R. C., & Grahe, J. E. (2015). Designing and teaching undergraduate capstone courses. San Francisco: Jossey-Bass.

Hoffman, Harvey. F. (2014). Engineering Capstone Course: fundamentals for students and instructors. Cham: Springer.

Suggested Reading List

Grix, J. (2004). *The foundations of research*. New York: Palgrave Macmillan.

Strunk, W., & White, E. B. 1. (2000). The elements of style. 4th ed. New York: Longman.

Appendix B: Library Resources

Appendix B: Library Resources								
Sl.No.	Module Code and Title	Title	Price	In Stock	Addl. Qty Needed	Total addl. Cost	Essential (E)/Additional(A)	Acquire before:
	Year I Semester I							
1	PLT102 Programming in Python	Python Data Science Handbook	3500	0	4	14000	E	Spring 2021
2		Python Programming: An Introduction to Computer Science	4200	0	4	16800	E	Spring 2021
3		Introduction to Programming in Python	5200	0	1	5200	A	Spring 2021
4		*Deitel, P., Deitel, J. (2019). Introduction to Python for Computer Science and Data Science Learning to Program with AI, Big Data and the Cloud. Pearson.*	6100	0	4	24400	E	Spring 2021
1	DST101: Descriptive Statistics	Freund, J. E. (2008). Mathematical Statistics, Prentice Hall, New Delhi.		10	0	0	E	Spring 2021
2		Goon, A. M., Gupta, M. K., and Dasgupta, B. (1991): Fundamentals of Statistics, Vol. I and II, World Press, Calcutta.		18	0	0	E	Spring 2021
3		Gupta, S. P. (2002): Statistical Methods, Sultan Chand and Sons, New Delhi.		8	0	0	E	Spring 2021
4		Gupta, S. C., and Kapoor, V. K. (2007). Applied Statistics, Sultan Chand and Sons, New Delhi.		5	0	0	A	Spring 2021

5		Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of mathematical statistics. Sultan Chand & Sons.		5	0	0	A	Spring 2021
6		Hogg, R. V., and Criag, A. T. (2013). Introduction to Mathematical Statistics, Macmillan Publishing, New York.		5	0	0	A	Spring 2021
7		Mood, A. M., Graybill, F. A., and Bose D. C. (2001). Introduction to the theory of Statistics, Tata McGraw Hill, New Delhi.		5	0	0	A	Spring 2021
8		Sheldon, R. (2007). Introduction to Probability Models, Academic Press, New Delhi.		5	0	0	A	Spring 2021
9		Walpole, R. E., Myers, R. H., Myers S. L., and Ye, K. (2001). Probability and Statistics for Engineers and Computer Scientists, Prentice Hall, New Delhi.		3	0	0	A	Spring 2021
1	CAE102 Computer Organization and Programming	*Yadin, A. (2016). Computer systems architecture. CRC Press.*	7879	0	4	31516	E	Spring 2021
2		*Mano, M. M. (2005). Computer system architecture. Prentice-Hall of India.*	6099	1	3	18297	E	Spring 2021
3		*Chaudhuri, P. P. (2008). Computer organization and design. PHI Learning Pvt. Ltd.*	2608	0	4	10432	E	Spring 2021
4		*Kumar, N. S., Saravanan, M., & Jeevananthan, S. (2011). Microprocessors and microcontrollers. Oxford University Press, Inc.*	4398	0	1	4398	A	Spring 2021
5		*Stallings, W. (2003). Computer organization and architecture: designing for performance. Pearson Education India.*	6373	5	0	0	A	Spring 2021

6		*Hamacher, C., Vranesic, Z., & Zaky, S. (2002). Computer organization. McGraw-Hill*	3815	0	1	3815	A	Spring 2021
1	CAL111 Calculus	Thomas, G.B and Finney, R.L.(2012). Calculus and Analytic Geometry (12th ed.)	4536		4	18144	E	Spring 2021
2		Kreyszig, E. (2006). Advanced Engineering Mathematics. (8thedition)		24	0	0	E	Spring 2021
3		*Salas, S.L., Hille, E. & Etgen, G.J. (2003). Calculus: One and Several Variables. (9th edition*	0	5	0	0	E	Spring 2021
4		*Apostal, T.M. (2005). Calculus: Volume I*	0	4	0	0	A	Spring 2021
5		*Spiegel, M.R. (1981). Advanced Calculus. (Schaum's Outline Series, Asian Student Edition)*	0	5	0	0	A	Spring 2021
6		*Frank, A. Jr., & Mendelson, E. (1992). Differential and Integral Calculus. (Schaum's Outline series, 3rd edition).*	0	2	0	0	A	Spring 2021
7		*Steward, J. (1999). Calculus (6th ed.)*	6552	0	1	6552	A	Spring 2021
	Year I Semester II							
1	AMT114 Discrete Structures	*Johnsonbaugh, R. (2009). Discrete Mathematics (6thed.). New York: Pearson Education Inc.*		5	0		E	Autumn 2021
2		*Liu, C.L. (2000). Elements of Discrete Mathematics (2nded.). New Delhi: McGraw-Hill.*		12	0		E	Autumn 2021

3		*Tremblay, J.P., & Manohar, R. (2007). Discrete Mathematical Structures with Applications to Computer Science. New Delhi: Tata McGraw Hill.*		11	0		E	Autumn 2021
1	DST106 Probability and Distributions	A.M.Mood, F.A. Graybill and D.C. Boes (2017): Introduction to the theory of statistics, International student ed. McGraw Hill.		13	0	0	E	Autumn 2021
2		Hogg, R.V. and Craig, A.T. (2020): Introduction to Mathematical Statistics, 4th ed. Acadmic Press.		9	0	0	E	Autumn 2021
3		A.M.Goon, M.K.Gupta & B. Dasgupta (1998): An outline of Statistical theory, Vol.I, 6th revised ed., World Press	400	0	4	1600	E	Autumn 2021
4		Rohatgi, V.K. (2015): An introduction to probability theory and mathematical statistics.	11,163	2	2	22326	E	Autumn 2021
5		P. G.Hoel (2005): Introduction to Mathematical Statistics, Asia publishing house.	5,449	0	1	5449	A	Autumn 2021
6		Murray R. Spiegal (1997): Theory and problems of Probability and Statistics, Schaum's outline series, McGraw Hill.	5,919	0	1	5919	A	Autumn 2021
7		Marek Fisz (2015): Probability theory and Mathematical Statistics, John Wiley.	4,097	0	1	4097	A	Autumn 2021
1	ADS104 Data Structures	Goodrich, T. M., Tamassia, R., & Goldwasher, H, M. (2016). Data Structures and Algorithms in Python. Wiley.	2590	0	4	10360	E	Autumn 2021

2		Bradley N. Miller, David L. Ranum.(2011) Problem Solving with Algorithms and DataStructures Using Python [2nd ed]. Franklin, Beedle & Associate.	2800	0	4	11200	E	Autumn 2021
3		Tremblay J.P and Sorenson P.G. (2002). An Introduction to Data Structures with Applications 2nd edition. TMH.	3150	0	1	3150	A	Autumn 2021
4		Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2017). Introduction to algorithms. MIT Press.	4900	1	0	0	A	Autumn 2021
1	ALG108: Matrix Analysis and Vector Spaces	Anton. H. & Rorres. C. (2005). Elementary Linear Algebra (9 ed.). USA: John Wiley & Sons.	599	0	4	2396	E	Autumn 2021
2		Biswas, S. (2012). Textbook of Matrix Algebra (3 ed.). New Delhi: PHI Learning Pvt. Ltd.	615	0	4	2460	E	Autumn 2021
3		Lang, S. (1988). Introduction to Linear Algebra (2 ed.). New York: Springer Verlag Inc.		8	0	0	E	Autumn 2021
4		Cohen, A.M, Cuypers, H., & Sterk, H. (1999). *Algebra Interactive*. New Yok: Springer Verlag Inc.	680	0	1	680	A	Autumn 2021
5		Hoffman, K., & Kunze, R. (2006). *Linear Algebra* (2 ed.) New Delhi: Dorly Kindersly (India) Pvt. Ltd.		10	0	0	A	Autumn 2021
6		Kolman, B., & Hill, D.R. (2006). *Introduction to Linear Algebra with applications* (7 ed.) New Delhi: Dorly Kindersly (India) Pvt. Ltd.		9	0	0	A	Autumn 2021

7		Kreyszig, E. (2005). *Advanced Engineering Mathematics* (8 ed.) Singapore: John Whiley & Sons (Asia) Pvt. Ltd.		10	0	0	A	Autumn 2021
	Year II Semester I							
1	PLT203: R Programming for Data Analysis	Crawley, M.J. (2012). The R book. John Whiley & Sons.	7756	4	0	31024	E	Spring 2022
2		Matloff, N. (2011). The Art of R programming: A tour of statistical software design. No Starch Press.	1655	4	0	6620	E	Spring 2022
3		Braun, W. J., Murdoch, D. J. (2017). A First Course in Statistical Programming with R. Cambridge University Press. New York.	2899	0	1	2899	A	Spring 2022
4		Gardener, M. (2012). Beginning R: The Statistical Programming Language. Wiley Publications.	2688	0	1	2688	A	Spring 2022
1	CAE201 Database Systems	*Elmasri, R., & Navathe, S. B. (2016). Fundamentals of Database Systems (7 ed.). Pearson.*	5095	5	0	0	E	Spring 2022
2		Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019). Database System Concepts (7 ed.). McGraw-Hil	3577	1	3	10731	E	Spring 2022
3		*Chaudhri, A. B., Rashid, A., & Zicari, R. (2003). XML data management: Native XMLand XML-enabled database systems. Addison-Wesley Professional.*	4499	0	4	17996	E	Spring 2022
4		*Ramakrishnan, R., & Gehrke, J. (2000). Database management systems. McGraw Hill.*	7606	5	0	0	A	Spring 2022

5		Deitel, P., & Deitel, H. (2007). Internet & world wide web: how to program. Prentice Hall Press	3243	6	0	0	A	Spring 2022
6		*Williams, K., Brundage, M., Dengler, P., Gabriel, J., Hoskinson, A., Kay, M. R. & Vanmane, M. (2000). Professional XML databases. Birmingham, UK: Wrox press.*	5878	0	1	5878	A	Spring 2022
1	IST303: Statistical Inference	Casela, G., and Berger, R. L. (2017). Statistical Inference, Duxbury Thompson Learning, New Delhi.	9897.7	0	4	39590.8	E	Spring 2022
2		Christensen, R. (2013). Log Linear Models, Springer, New Delhi.	5648.8	0	4	22595.2	E	Spring 2022
3		Conover, W. J. (2007). Practical Non-parametric Statistics, John Wiley, New Delhi.	5485	0	4	21940	E	Spring 2022
4		Hogg, R. V., and Craig, T. T. (2019). Introduction to Mathematical Statistics, Prentice-Hall, New Delhi.		10	0	0	E	Spring 2022
5		Mood, A. M. Graybill, F. A., and Boes, D. C. (2017). Introduction to Theory of Statistics, Cambridge University Press, New Delhi.		23	0	0	E	Spring 2022
6		Box, G. E. P., and Tiao, G. C. (2011). Bayesian Inference in Statistical Analysis, Addison Wesely, New Delhi.	980	0	1	980	A	Spring 2022
7		Dudewicz, E. J., and Mishra, S. N. (1988). Modern Mathematical Statistics. John Wiley, New Delhi.	4892	0	1	4892	A	Spring 2022
8		Rao, C. R. (2001). Linear Statistical Inference and its Applications, John Wiley, New Delhi.	23607	0	1	23607	A	Spring 2022

9		Rohatgi, V. K., and Saleh, A. K. Md. E. (2005). An Introduction to Probability and Statistics, John Wiley, New Delhi.		3	0	0	A	Spring 2022
10		Wald, A. (2013). Sequential Analysis, Dover Publications, New Delhi.	1241	0	1	1241	A	Spring 2022
1	IST302: Regression Analysis	Kleinbaum, D. G., and Kupper, L. L. (2014). Applied Regression analysis and other Multivariate Methods, Duxbury Press, Massachusetts, USA.	20480	4	0	81920	E	Spring 2022
2		Draper, N. R., and Smith, H. (1998). Applied Regression Analysis, John Wiley, New Delhi.	11822	0	4	47288	E	Spring 2022
3		Bapat, R. B. (2012). Linear Algebra and Linear Models, Springer-Verlag, New Delhi.	3869	0	4	15476	E	Spring 2022
4		Barnett, V., and Lewis, T. (1998). Outliers in Statistical Data, John Wiley, New Delhi.	11822	0	4	47288	E	Spring 2022
5		Belsley, D. A., Kuh, E., and Welsch, R. E. (2004). Regression Diagnostics-Identifying Influential Data and Sources of Collinearity, John Wiley, New Delhi.	8311.4	0	4	33245.6	E	Spring 2022
6		Graybill, F. A. (2000). Theory and Application of the Linear Model, Duxbury, North Scituate, USA.		23	0	0	E	Spring 2022
7		Joshi, D. D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern, New Delhi.	310	0	1	310	A	Spring 2022
8		Montgomery, D. C., Peck, E., and Vining, G. (2012). Introduction to Linear Regression Analysis, John Wiley, New Delhi.	290	0	2	580	A	Spring 2022

9		McCullagh, P., and Nelder, J.A. (1989). Generalized Linear Models, Chapman and Hall, New Delhi.	2150	0	1	2150	A	Spring 2022
10		Rao, C. R. (2001). Linear Statistical Inference and Its Applications, Wiley Eastern, New Delhi.	23607	0	1	23607	A	Spring 2022
11		Searle, S. R., and Gruber, H.J. (2016). Linear Models, John Wiley, New Delhi.	8485	0	1	8485	A	Spring 2022
12		Searle, S. R. Casella, G. McCulloch, C.E. (1998). Variance Components, John Wiley, New Delhi.	3421	0	1	3421	A	Spring 2022
13		Scheffé, H. (1999). The Analysis of Variance, John Wiley, New Delhi.	10297	0	1	10297	A	Spring 2022
	ECO101 Fundamentals of Economics	Begg, D., Fischer, S., & Dornbusch, R. (2008). Economics (9th edition). McGraw-Hill Contemporary.	2378	0	4	9512		Spring 2022
1		Bodman, D. (2002). Macroeconomics. McGraw-Hill Europe.	5450	0	4	21800	E	Spring 2022
2		Charlé, E. (1983). Macroeconomics of developing countries. Tata McGraw-Hill.		NA	NA	#VALUE!	E	Spring 2022
3		Mankiw, N. G. (2015). Macroeconomics. New York: Worth Publishers.	4653	0	4	18612	E	Spring 2022
4		NSB (2017). Consumer price Index bulletin January 2017. National Accounts and Price Division, NSB	0	NA	NA	#VALUE!	E	Spring 2022
5		Katz M. L., & Rosen, H. S. (1991). Microeconomics.-Richard D. McGraw Hill. Koutsoyiannis, A. (2001). Modern Micro Economics. ELBS	0	NA	NA	#VALUE!	A	Spring 2022

6		Pindyck, R. S., & Rubinfeld, D. (1999). Micro Economics. McGraw Hill.	6143	0	1	6143	A	Spring 2022
7		Varian, H. (1998). Intermediate Micro Economics. WW Norton	2300	0	1	2300	A	Spring 2022
	Year II Semester II							
1	AMT210 Numerical Methods	*Burden, R. L. & Faires, J. D. (2011). Numerical Analysis*	8390.91	0	4	33563.64	E	Autumn 2022
2		*Gerald, C. F. & Wheatley, P. O. (2013). Applied Numerical Analysis*	0	12	0	0	E	Autumn 2022
3		*Greenbaum A. & Chartier T. P. (2012). Numerical Methods: Design, Analysis and Computer Implementation of algorithms*	5567.41	0	4	22269.64	E	Autumn 2022
4		*Atkinson, K. E. (1989). An Introduction to Numerical Analysis*	0	5	0	0	A	Autumn 2022
5		*Prasad, D. (2012). An Introduction to Numerical Analysis*	0	4	0	0	A	Autumn 2022
6		*Sastry, S. S. (2012). Introductory Methods of Numerical Analysis*	0	3	0	0	A	Autumn 2022
1	ADS205 Algorithm Analysis	Horowitz E, Sahni S, Rajasekaran S. (2010). Fundamentals of Computer Algorithms (2 ed.). Universities Press Pvt Ltd.	1501	0	4	6004	E	Autumn 2022
2		Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2017). Introduction to algorithms. MIT Press.	4900	1	3	14700	E	Autumn 2022
3		Donald E. Knuth. (1997). The Art of Computer Programming Volume 3, Sorting and Searching, (2 ed.). Pearson Education, Addison-Wesley.	4340	0	1	4340	A	Autumn 2022

4		Pai G.A.V. (2008). Data structures and Algorithms, Tata McGraw Hill.	1890	0	1	1890	A	Autumn 2022
1	IST306: Multivariate Analysis	Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis, John Wiley, New Delhi.	1609	0	4	6436	E	Autumn 2022
2		Rao, C. R. (2001). Linear Statistical Inference and its Applications, John Wiley, New Delhi	46329	0	4	185316	E	Autumn 2022
3		Muirhead, R. J. (2009). Aspects of Multivariate Statistical Theory, John Wiley, New Delhi.	12736	0	4	50944	E	Autumn 2022
4		Johnson, R. A., and Wichern, D. W. (2008). Applied Multivariate Statistical Analysis, Prentice Hall, New Delhi.	3865	0	4	15460	E	Autumn 2022
5		Rencher, A. C. (2002). Methods of Multivariate Analysis, John Wiley, New Delhi.	1458	0	4	5832	E	Autumn 2022
6		Srivastava, M. S., & Khatri, C. G. (1979). An Introduction to Multivariate Statistics, North Holland, New Delhi.	1065	0	1	1065	A	Autumn 2022
7		Chatfield, C., & Collins, A. J. (1982). Introduction to Multivariate Analysis, Prentice Hall, New Delhi.	995	0	1	995	A	Autumn 2022
1	DSC201: Machine Learning I	Coelho, L. P., & Richert, W. (2015). Building machine learning systems with Python. Packt Publishing Ltd.	3264	0	4	13056	E	Autumn 2022
2		Albon, C. (2018). Machine learning with python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.".	900	0	4	3600	E	Autumn 2022

3		Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2005). Data mining: practical machine learning tools and techniques (Morgan-Kaufman Series of Data Management Systems).	465	0	4	1860	E	Autumn 2022
4		Garreta, R., & Moncecchi, G. (2013). Learning scikit-learn: machine learning in python. Packt Publishing Ltd.	1943	0	1	1934	A	Autumn 2022
5		Bowles, M. (2015). Machine learning in Python: essential techniques for predictive analysis. John Wiley & Sons.	3749	0	1	3749	A	Autumn 2022
	GSE101 Analytical Skills	Bano, E. d. (2000). *Six Thinking Hats* (2nd ed.). New Delhi, India: Penguin India.	2510	2	2	5020	E	Autumn 2022
		Michalko, M. (2006). *Thinkertoys: A handbook of creative-thinking techniques* (2nd ed.). Ten Speed Press.	1869	0	4	7476	E	
1		Puccio, G.J., Mance, M. & Switalski, L.B. (2017). *Creativity Rising Creative Thinking and Creative Problem Solving in the 21st Century.* ICSC Press, International Center for Creativity, US	5571	0	4	22284	E	
2		Treffinger, D. J. (2006). *Creative Problem Solving: An introduction* (4th ed.). Prufrock.	5570.79	0	4	22283.16	E	
3		Bano, E. d. (2009). *Lateral Thinking.* e-Penguin	776	0	1	776	A	
4		Bono, E. d. (2005). *Thinking course (Revised Edition).* Bernes and Nobel	4938	0	1	4938	A	

5		Chopra, R. (n.d.). *Logical Critical Analytical Reasoning.* Galgoba Publications Pvt Ltd	199	0	1	199	A	
6		Kahneman, D. (2015). *Thinking fast and slow.* New York: Farrar, Straus and Giroux.	1800	0	1	1800	A	
7		Scott, J. W. (2016). *Critical Thinking: Proven strategies for improving your decision making skills, retaining information longer and analyzing situations with simple logic ---- Logical thinking and critical thinking skills.* New Familiar Publishing.	240	0	1	240	A	
	Year III Semester I							
1	RMS301: Research Methods	Bhattacherjee, A. (2012) Social science research: Principles, methods, and practices. Published under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Open access book available at: https://open.umn.edu/opentextbooks/textbooks?term=research&commit=Go		20	0	0	E	Spring 2023
2		Bryman, A. (2016). Social research methods. Oxford university press.		10	0	0	E	Spring 2023

3		Jhangiani,R. S., I-Chant A. Chiang, I. A., Cuttler, C & Leighton, D. C (2019). Research methods in psychology. Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Open access book available at: https://open.umn.edu/opentextbooks/textbooks?term=research&commit=Go		15	0	0	E	Spring 2023
4		Kothari, C. R. (2004). Research methodology: Methods and techniques. New Age International.		10	0	0	E	Spring 2023
5		Kumar, R. (2019). Research methodology: A step-by-step guide for beginners. Sage Publications Limited.		10	0	0	E	Spring 2023
6		Creswell, J. W., & Creswell, J. D. (2017). Research design: Qualitative, quantitative, and mixed methods approaches. Sage publications.		5	0	0	A	Spring 2023
7		Ghosh, B. N. (2007). Scientific Method and Social Research (Revised edition). New Delhi: Sterling Publishers Pvt. Ltd.		5	0	0	A	Spring 2023
8		Grix, J. (2002). The foundations of Research. Palgrave Foundations. New York. (accession no.001.4 GRI)		5	0	0	A	Spring 2023
9		Patten, M. L., & Newhart, M. (2017). Understanding research methods: An overview of the essentials. Taylor & Francis.		5	0	0	A	Spring 2023
10		Wolf, H. K. and Pant, P. R. (2002). Social Science Research and Thesis Writing. Kathmandu: Buddha Academic		6	0	0	A	Spring 2023

11		Punch, K. F. (2005). Introduction to Social research: Qualitative and Quantitative Approaches. London: Sage Publication.		6	0	0	A	Spring 2023
1	APC317 Data Warehousing and Data Mining	*Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.*	4612	0	4	18448	E	Spring 2023
2		*Bhatia, P. (2019). Data mining and data warehousing: principles and practical techniques. Cambridge University Press.*	3282	0	4	13128	E	Spring 2023
3		*Inmon, W. H. (2005). Building the data warehouse. John wiley & sons*	3577	0	4	14308	E	Spring 2023
4		*Dunham, M. H. (2006). Data mining: Introductory and advanced topics. Pearson Education India.*	2378	0	1	2378	A	Spring 2023
5		*Mourya, S. K., & Gupta, S. (2012). Data mining and data warehousing. Alpha Science International, Ltd.*	6185	0	1	6185	A	Spring 2023
1	IST201: Sampling Theory	Cassel, C. M., Sarndal, C. E., and Wretman, J. H. (1992). Foundations of Inference in Survey Sampling, John Wiley, New Delhi.	2768.21	0	4	11072.8	E	Spring 2023
2		Chaudhari, A., and Stenger, H. (2005). Survey Sampling Theory and Methods, Chapman and Hall, New Delhi.	275	0	4	1100	E	Spring 2023
3		Chaudhari, A., and Voss, J. W. E. (1988). Unified Theory and Strategies of Survey Sampling, North Holland, New Delhi.	275	0	4	1100	E	Spring 2023
4		Cochran, W. G. (2005). Sampling Techniques, John Wiley, New Delhi.	1153.56	0	4	4614.26	E	Spring 2023

5		Murthy, M. N. (1977). Sampling Theory and Methods, Statistical Publishing Society, Kolkatta.	3276.55	0	4	13106.2	E	Spring 2023
6		Hansen, M. H., Hurwitz, W. H., and Madow, W. G. (1993). Sample Survey Methods and Theory, Vol. I and Vol. II., John Wiley, New Delhi.	71	0	1	71	A	Spring 2023
7		Hedayat, A. S., and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, John Wiley, New Delhi.	560	0	1	560	A	Spring 2023
8		Singh, D., and Chaudhary, F. S. (2020). Theory and Analysis of Sample Survey Designs, New Age International Pvt. Ltd., New Delhi.	448	0	1	448	A	Spring 2023
9		Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984). Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.	361	0	1	361	A	Spring 2023
10		Thompson, S. K. (2012). Sampling, John Wiley, New Delhi.	308	0	1	308	A	Spring 2023
1	DSC302 Unsupervised Learning	Albon, C. (2018). Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning. " O'Reilly Media, Inc.".	4200	0	4	16800	E	Spring 2023
2		Hinton, G. E., & Sejnowski, T. J. (Eds.). (1999). Unsupervised learning: foundations of neural computation. MIT press.	3800	0	4	15200	E	Spring 2023

3		Bowles, M. (2015). Machine learning in Python: essential techniques for predictive analysis. John Wiley & Sons.	6800	0	1	6800	A	Spring 2023
4		Garreta, R., & Moncecchi, G. (2013). Learning scikit-learn: machine learning in python. Packt Publishing Ltd.	3100	0	1	3100	A	Spring 2023
		Charvat, J. (2003). *Project management methodologies: selecting, implementing, and supporting methodologies and processes for projects*. John Wiley & Sons.	2700	0	4	10800	E	Spring 2023
		Duncan, W. R. (1996). A guide to the project management body of knowledge.	2000	0	4	8000	E	Spring 2023
		Horine, G. (2009). *Absolute beginner's guide to project management*. Pearson Education.	1200	0	4	4800	E	Spring 2023
		Kerzner, H. R. (2013). *Project management: a systems approach to planning,scheduling, and controlling*. John Wiley & Sons.	1100	0	4	4400	E	Spring 2023
		Khatua, S. (2011*). Project management and appraisal*. Oxford University Press.	1500	0	4	6000	E	Spring 2023

		Meredith, J. R., Mantel, S. J., & Shafer, S. M. (2016). *Project management: a managerial approach*. Singapore: Wiley.	4400	0	4	17600	E	Spring 2023
		Potts, D. (2002). *Project planning and analysis for development*. Boulder: LynneRienner	3000	0	4	12000	E	
Year III Semester II								
1	AMT311: Introduction to Operations Research	Taha, H. A. (2013). Operations research: an introduction. (9th ed.). Pearson Education India.		10	0	0	E	Autumn 2023
2		Hillier, F. S. (2021). Introduction to operations research. (11th ed.). Tata McGraw-Hill Education.		8	0	0	E	Autumn 2023
3		Vanderbei, R. J. (2020). Linear programming: foundations and extensions (Vol. 285). Springer Nature.	8115	4	0	32460	E	Autumn 2023
4		Bernard W. Taylor III. (1993). Introduction to Management Sciences. (4th ed.). New York: Allyn and Bacon.	236	0	1	236	A	Autumn 2023
5		Kanti Swarup, Manmohan, Gupta. (2018). Operations Research. (10th ed.). New Delhi: Sultan Chand and Sons Publishing Co.	1572	0	1	1572	A	Autumn 2023

6		Manmohan & Gupta P.K. (1987). Operations Research and Statistical Analysis. (3rd ed.). New Delhi: Sultan Chand and Sons Publishers.	1330	0	1	1330	A	Autumn 2023
1	DSC303 Ethics and Security	Ladley, J. (2019). Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program. Academic Press.	6000	0	4	24000	E	Autumn 2023
2		Davis, K. (2012). Ethics of Big Data. OReilly	2300	0	4	9200	E	Autumn 2023
3		Bunnik, A.,Cawley, A.,Mulqueen, M.,Zwitter, A. (2016). Big Data Challenges. Palgrave Macmillan	6100	0	1	6100	A	Autumn 2023
4		Freeman, L.A, Peace, A.G. (2005). Information Ethics: Privacy and Intellectual Property. Information Science Publishing.	6200	0	1	6200		Autumn 2023
1	AMT312: Stochastic Processes	Bhat, B. R. (2000). Stochastic Models: Analysis and Applications, New Age International, New Delhi.	374	0	4	1496	E	Autumn 2023
2		Medhi, J. (2001). Stochastic Processes, Wiley Eastern Ltd., New Delhi.	4241	0	4	16964	E	Autumn 2023
3		Ross, S. (1996). Stochastic Processes (2 ed.). Wiley.	850	0	4	3400	E	Autumn 2023
4		Adke, S. R., and Manjunath, S. M. (1984). An Introduction to Finite Markov Processes, John Wiley, New Delhi.	1308	0	1	1308	A	Autumn 2023
5		Bailey, N. T. J. (1990). Elements of Stochastic Processes with Applications to the Natural Sciences, Wiley Eastern Ltd., New Delhi.	422	0	1	422	A	Autumn 2023

1	APC318 Deep Learning	*Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press*	3508	0	4	14032	E	Autumn 2023
2		*Nelli, F. (2018). Python data analytics: with pandas, numpy, and matplotlib. Apress.*	3123	0	4	12492	E	Autumn 2023
3		*Subramanian, V. (2018). Deep Learning with PyTorch: A practical approach to building neural network models using PyTorch. Packt Publishing Ltd*	3812	0	4	15248	E	Autumn 2023
4		*Chollet, F. (2018). Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG.*	4176	0	1	4176	A	Autumn 2023
1	EDP101 Entrepreneurship	Hisrish, R., Peters, M., & Shepherd, D. (2016). *Entrepreneurship* (10th ed.). McGraw Hill Education.	1200	0	4	4800	E	Autumn 2023
2		Hsieh,T. (2014). *Delivering Happiness: A path to profits, passion and purpose*. Grand Central Publishing	1100	0	4	4400	E	Autumn 2023
3		Kuratko, D. F. (2016). *Entrepreneurship: Theory, process and practice* (10th ed.). South Western College Publication.	4600	0	4	18400	E	Autumn 2023

	Year IV Semester I							
1	APC419 Computer Vision	Bovik, A. (2005). Handbook of Image and Video Processing (2 ed.). Academic Press.	8120	0	4	32480	E	Spring 2024
2		Gonzalez, R.C., Woods, R.E. (2018). Digital Image Processing (4 ed.). Pearson Education.	2380	0	4	9520	E	Spring 2024
3		Szelisk, R. (2011). Computer Vision – Algorithms and Applications, Springer.	8000	0	4	32000	E	Spring 2024
4		Jain, A.K. (2011). Fundamentals of Digital Image Processing, PHI.	400	0	1	400	A	Spring 2024
5		Marques, O. (2011). Practical Image and Video Processing Using MatLab,Wiley.	15000	0	1	15000	A	Spring 2024
6		Woods, J.W. (2006). Multidimensional Signal, Image, Video Processing and Coding,	5331	0	1	5331	A	Spring 2024
1	DSC404 Big Data Technologies and Applications	*White, Tom. 2015. Hadoop the Definitive Guide. O'reilly Media.*	2644.4	0	4	10577.6	E	Spring 2024
2		*George, Lars, Jan Kunigk, Paul Wilkinson, and Ian Buss. 2018. Architecting Modern Data Platforms. California.*	5528.3	0	4	22113.2	E	Spring 2024
3		*Karanth, Sandeep. 2014. Mastering Hadoop. Birmingham: Packt Publishing.*	3776.96	0	1	3776.96	A	Spring 2024
1	APC420 Natural Language Processing	*Bird, Steven, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python. O'reilly Media.*	3552	0	4	14208	E	Spring 2024
2		Eisenstein, J. 2019. Introduction to Natural Language Processing. MIT Press.	5649.95	0	4	22599.8	E	Spring 2024

3		*Sarkar, Dipanjan. 2019. Text Analytics with Python. Second. Springer.*	2516.49	0	1	2516.49	A	Spring 2024
1	DSC405 Geospatial Data Analysis	*Zandbergen, P. A. (2015). Python scripting for ArcGIS. Esri press.*	11440	0	4	45760	E	Spring 2024
2		*Lawhead, J. (2013). Learning geospatial analysis with Python. Packt Publishing Ltd.*	3264	0	4	13056	E	Spring 2024
3		*Westra, E. (2013). Python geospatial development. Packt Publishing Ltd.*	3264	0	4	13056	E	Spring 2024
4		*Pimpler, E. (2013). Programming ArcGIS 10.1 with Python Cookbook. Packt Publishing Ltd.*	2979	0	1	2979	A	Spring 2024
5		*Tateosian, L. (2015). Python For ArcGIS. Springer.*	10457	0	1	10457	A	Spring 2024
6		*Kropla, B. (2006). Beginning MapServer: open source GIS development. Apress.*	3181	0	1	3181	A	Spring 2024
1	DSC406 Communications for Data Science	*Winston, W. (2019). Microsoft Excel 2019 Data Analysis and Business Modeling: Data Analysis and Business Modeling. Pearson Education.*	2700	0	4	10800	E	Spring 2024
2		Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., & Anderson, D. R. (2014). *Essentials of Business Analytics (Book Only)*. Nelson Education.	3000	0	4	12000	E	Spring 2024
3		Stiff, J. B., & Mongeau, P. A. (2016). *Persuasive communication*. Guilford Publications.	5100	0	4	20400	E	Spring 2024

4		Pochiraju, B., & Seshadri, S. (Eds.). (2018). *Essentials of Business Analytics: An Introduction to the Methodology and Its Applications* (Vol. 264). Springer.	8500	0	4	34000	E	Spring 2024
5		Ragsdale, C. (2014). *Spreadsheet modeling and decision analysis: a practical introduction to business analytics*. Nelson Education.	3200	0	4	12800	E	Spring 2024
6		Albright, S. C., & Winston, W. L. (2019). Business analytics: Data analysis & decision making. Cengage Learning.	7500	0	1	7500	A	Spring 2024
	Year IV Semester II							
1	RSM414 Capstone/Research Project	Hauhart, R. C., & Grahe, J. E. (2015). Designing and teaching undergraduate capstone courses. San Francisco: Jossey-Bass.	3109	0	2	3109	E	Autumn 2024
2		Hoffman, Harvey. F. (2014). Engineering Capstone Course: fundamentals for students and instructors. Cham: Springer.	4893	0	1	4893	E	Autumn 2024