



HSNC UNIVERSITY, MUMBAI

HSNCU Syllabus

School of Applied Sciences

Syllabus of MSc Data Science and Business Analytics

Board of Faculty of Science & Technology

Board of Studies in the Subject of Data Science & Business
Analytics

1. Name of the Chair Person: Dr. Nandini Sengupta, Associate professor, Department of Economics
2. Name of the Co – Chairperson: Mrs. Shailaja. J. Rane, Co- Ordinator of Data Science and Business Analytics.

Teachers from the college:

3. Beenarani Karutharan, Assistant Professor, Department of Computer science
4. Mrs. Mrunal M Hardikar, Assistant Professor, Department of Mathematics
5. Mr. Sikandar Yadav, Assistant Professor, Department of Data Science and Business Analytics.

External Professor:

6. Rosemary Gosling, Director of External Studies (Retd) for the London School of Economics and Political Science (LSE)
7. Dr. Santosh Bothe, Founder and Director AiSense (Start-up funded by BIRAC, Govt. of India), Principal, Saraswati College, Shegaon, Affiliated to SGBU Amravati University.

External experts:

8. Prof. Parag Mahulikar is Ex- Dean and Senior Professor of Marketing at IES Management College and Research Centre, Bandra, India and Management Consultant.
9. Dr. Alok Deepak Dabade, Assistant Professor, Department of Statistics, University of Mumbai.
10. Dr. Sujata Suvarnapathki, Assistant Professor, Department of Statistics, Ramnarayan Ruia Autonomous College, Matunga, Mumbai.
11. Subhash Kumar, Assistant Professor, MCA, MPHIL IT department, St. Xavier's college, Mumbai.
12. Industry Expert: Vinayak Deshpande, Managing Director, Sankhya Analytical Research Pvt. Ltd.

13. Industry Expert: Nishad Kapadia, MCA, Technical trainer, Data and Solution Architect Project Manager, TeraData, Mumbai.

14. Industry Expert: Praveena Premanand Menon, MSc in Big Data Analytics.

15. Industry Expert: Mr. Awesh Bhornya, Infinity Learning (Founder)

Alumni: Proposed names Sara kale, Pranit Kadam

Part – I

Outline of Choice Based Credit System as outlined by University Grants Commission:

R. **: The Definitions of The Key Terms Used In The Choice Based Credit System And Grading System Introduced From The Academic Year 2023-2024 are as Under:**

1. **Core Course:** A course, which should compulsorily be studied by a candidate as a core requirement is termed as a core course.

2. **Elective Course:** Generally, a course which can be chosen from a pool of courses and which may be very specific or specialized or advanced or supportive to the discipline/subject of study or which provides an extended scope or which enables an exposure to some other discipline/subject/domain or nurtures the candidate's proficiency/skill is called an Elective Course.

2.1 **Discipline Specific Elective (DSE) Course:** Elective courses may be offered by the main discipline/subject of study is referred to as Discipline Specific Elective. The University/Institute may also offer discipline related Elective courses of interdisciplinary nature (to be offered by main discipline/subject of study).

2.2 **Dissertation/Project:** An elective course designed to acquire special/advanced knowledge, such as supplement study/support study to a project work, and a candidate studies such a course on his own with an advisory support by a teacher/faculty member is called dissertation/project. A Project/Dissertation work would be of 6 credits. A Project/Dissertation work may be given in lieu of a discipline specific elective paper.

2.3 **Generic Elective (GE) Course:** An elective course chosen generally from an unrelated discipline/subject, with an intention to seek exposure is called a Generic Elective.

P.S.: A core course offered in a discipline/subject may be treated as an elective by other discipline/subject and vice versa and such electives may also be referred to as Generic Elective.

3. **Choice Base Credit System:** CBCS allows students to choose inter- disciplinary, intra-disciplinary courses, skill-oriented papers (even from other disciplines according to their learning needs, interests and aptitude) and more flexibility for students.

4. **Honours Program:** To enhance employability and entrepreneurship abilities among the learners, through aligning Inter Disciplinary / Intra Disciplinary courses with Degree Program. Honours Program will have 40 additional credits to be undertaken by the learner across three years essentially in Inter / Intra Disciplinary course.

A learner who joins Regular Undergraduate Program will have to opt for Honours Program in the first year of the Program. However, the credits for honours, though divided across three years can be completed within three years to become eligible for award of honours Degree.

5. **Program:** A Program is a set of course that are linked together in an academically meaningful way and generally ends with the award of a Degree Certificate depending on the level of knowledge attained and the total duration of study, B.Sc. Programs.

6. **Course:** A 'course' is essentially a constituent of a 'program' and may be conceived of as a composite of several learning topics taken from a certain knowledge domain, at a certain level. All the learning topics included in a course must necessarily have academic coherence, i.e. there must be a common thread linking the various components of a course. A number of linked courses considered together are in practice, a 'program'.

7. **Bridge Course:** Bridge course is visualized as Pre-semester preparation by the learner before commencement of regular lectures. For each semester the topics, whose knowledge is considered as essential for effective and seamless learning of topics of the Semester, will be specified. The Bridge Course can be conducted in online mode. The Online content can be created for the Bridge Course Topics.

8. **Module and Unit:** A course which is generally an independent entity having its own separate identity, is also often referred to as a 'Module' in today's parlance, especially when we refer to a 'modular curricular structure'. A module may be studied in conjunction with other learning modules or studied independently. A topic within a

course is treated as a Unit. Each course should have exactly 3 Units.

9. Self-Learning: 20% of the topics will be marked for Self-Learning. Topics for Self-Learning are to be learned independently by the student, in a time-bound manner, using online and offline resources including online lectures, videos, library, discussion forums, fieldwork, internships etc.

Evaluative sessions (physical/online), equivalent to the credit allocation of the Self Learning topics, shall be conducted, preferably, every week for each course.

Learners are to be evaluated real time during evaluative sessions. The purpose of evaluative sessions is to assess the level of the students' learning achieved in the topics earmarked for Self-Learning.

The teacher's role in these evaluative sessions will be that of a Moderator and Mentor, who will guide and navigate the discussions in the sessions, and offer concluding remarks, with proper reasoning on the aspects which may have been missed by the students, in the course of the Self-Learning process.

The modes to evaluate self-learning can be a combination of the various methods such as written reports, handouts with gaps and MCQs, objective tests, case studies and Peer learning. Groups can be formed to present self-learning topics to peer groups, followed by Question and Answer sessions and open discussion. The marking scheme for Self-Learning will be defined under Examination and Teaching.

The topics stipulated for self-learning can be increased or reduced as per the recommendations of the Board of Studies and Academic Council from time to time. All decisions regarding evaluation need to be taken and communicated to the stakeholders preferably before the commencement of a semester. Some exceptions may be made in exigencies, like the current situation arising from the lockdown, but such ad hoc decisions are to be kept to the minimum possible.

10. Credit Point: Credit Point refers to the 'Workload' of a learner and is an index of the number of learning hours deemed for a certain segment of learning. These learning hours may include a variety of learning activities like reading, reflecting, discussing, attending lectures / counselling sessions, watching especially prepared videos, writing assignments, preparing for examinations, etc. Credits assigned for a single course

always pay attention to how many hours it would take for a learner to complete a single course successfully. A single course should have, by and large a course may be assigned anywhere between 2 to 8 credit points wherein 1 credit is construed as corresponding to approximately 30 to 40 learning hours.

11. Credit Completion and Credit Accumulation: Credit completion or Credit acquisition shall be considered to take place after the learner has successfully cleared all the evaluation criteria with respect to a single course. Thus, a learner who successfully completes a 4 CP (Credit Point) course may be considered to have collected or acquired 4 credits. Learner level of performance above the minimum prescribed level (viz. grades / marks obtained) has no bearing on the number of credits collected or acquired. A learner keeps on adding more and more credits as he completes successfully more and more courses. Thus, the learner 'accumulates' course wise credits.

12. Credit Bank: A Credit Bank in simple terms refers to stored and dynamically updated information regarding the number of Credits obtained by any given learner along with details regarding the course/s for which Credit has been given, the course-level, nature, etc. In addition, all the information regarding the number of Credits transferred to different programs or credit exemptions given may also be stored with the individual's history.

13. Credit Transfer: (performance transfer) When a learner successfully completes a program, he/she is allowed to transfer his/her past performance to another academic program having some common courses and Performance transfer is said to have taken place.

14. Course Exemption: Occasionally, when two academic programs offered by a single university or by more than one university, may have some common or equivalent course-content, the learner who has already completed one of these academic programs is allowed to skip these 'equivalent' courses while registering for the new program. The Learner is 'exempted' from 'relearning' the common or equivalent content area and from re-appearing for the concerned examinations. It is thus taken for granted that the learner has already collected in the past the credits corresponding to the exempted courses.

O*** The fees for transfer of credits or performance will be based on the number of credits that a learner has to complete for award of the degree.**

The Scheme of Teaching and Examination:

The performance of the learners shall be evaluated in two components: Internal Assessment with 25% marks by way of continuous evaluation and by Semester End Examination with 50% marks by conducting the theory examination.

INTERNAL ASSESSMENT: - It is defined as the assessment of the learners on the basis of continuous evaluation as envisaged in the credit-based system by way of participation of learners in various academic and correlated activities in the given semester of the programme.

A). Internal Assessment – 25% 25 marks Practical' s (internal Components of the Practical Course)

Sr. No.	Particulars	Marks
1	ONE class test / online examination to be conducted in the given semester	10 Marks
2	Self-Learning Evaluation	10 Marks
3	Active participation in routine class instructional deliveries	05 Marks

1. For Theory Courses

2. For Courses with Practicals

Each practical course can be conducted out of 25 marks with 10 marks for internal and 15 marks for external

Practical' s (Internal component of the Practical Course)

Sr. No	Evaluation type	Marks
1	Two Best Practical' s/Assignments/Presentation /Preparation of models/ One Assignment/ project/presentation to be assessed by teacher concerned	10
2	Journal	05
3	Viva	05

The semester end examination (external component) of 60% for each course will be as follows:

i) **Duration – 2 Hours**

ii) **Theory Question**

Paper Pattern: -

1. There shall be five questions each of 10 marks. On each unit there will be one question and the fourth one will be based on entire syllabus.
2. All questions shall be compulsory with internal choice within the questions. (Each question will be of 15 marks with options.)
3. Question may be subdivided into sub-questions a, b, c... and the allocation of marks depend on the weightage of the topic.

The marks will be given for all examinations and they will be converted into grade (quality) points. The semester-end, final grade sheets and transcripts will have only credits, grades, grade points, SGPA and CGPA.

3. Project:

- Project which can in the following forms
 - Case Studies
 - Videos
 - Blogs
 - Research paper (Presented in Seminar/Conference)
 - Field Visit Report

- Presentations related to the subject (Moot Court, Youth Parliament, etc.)
- Internships (Exposition of theory into practice)
- Open Book Test
- any other innovative methods adopted with the prior approval of Director Board of Examination and Evaluation.

4. Self-Learning Evaluation:

- **20% OF THE TOPICS OF CURRICULUM ARE LEARNED BY THE STUDENT THROUGH SELF LEARNING USING ONLINE / OFFLINE ACADEMIC RESOURCE SPECIFIED IN THE CURRICULUM.**
- **HENCE 20% OF THE LECTURES SHALL BE ALLOCATED FOR EVALUATION OF STUDENTS ON SELF LEARNING TOPICS**
- **The identified topics in the syllabus shall be learnt independently by the students in a time bound manner preferably from online resources. Evaluative sessions shall be conducted by the teachers and will carry 10 Marks.**

CLUB the self-learning topics into 3-4 GROUPS OF TOPICS ONLY FOR EVALUATION.

- PRESCRIBE TIME DURATION (IN DAYS) FOR COMPLETION OF EACH GROUP OF TOPIC AND EARMARK SELF LEARNING EVALUATION LECTURES IN THE IMETABLE. HENCE EACH GROUP OF TOPIC CAN BE ASSIGNED 3 REGULAR LECTURES FOR THIS EVALUATION FOR THE ENTIRE CLASS

3 Sub Topics

Each evaluative session shall carry 3 Marks (3 x 3 Units = 9 Marks). Students who Participate in all evaluative sessions shall be awarded 1 additional Mark.

4 Sub Topics

Each evaluative session shall carry 2.5 Marks (2.5 x 4 Units = 10 Marks)

-EVALUATION OF SELF LEARNING TOPICS CAN COMMENCE IN REGULAR LECTURES ASSIGNED FOR SELF LEARNING EVALUATION IN THE TIMETABLE

3 Evaluative sessions

Each evaluative session shall carry 3 Marks (3 x 3 = 9 Marks). Students who participate in all evaluative sessions shall be awarded 1 additional Mark.

4 Evaluative sessions

Each evaluative session shall carry 2.5 Marks (2.5 x 4 = 10 Marks).

Methods for Evaluation of Self-learning topics:

- Seminars/presentation (PPT or poster), followed by Q&A – Objective questions /Quiz / Framing of MCQ questions.
- Debates
- Group discussion
- You-Tube videos (Marks shall be based on the quality and viewership)
- Improvisation of videos
- Role Play followed by question-answers
- Viva Voce
- Any other innovative method

TEACHERS CAN FRAME OTHER METHODS OF EVALUATION ALSO PROVIDED THAT THE METHOD, DULY APPROVED BY THE COLLEGE EXAMINATION COMMITTEE, IS NOTIFIED TO THE STUDENTS AT LEAST 7 DAYS BEFORE THE COMMENCEMENT OF THE EVALUATION SESSION AND IS FORWARDED FOR INFORMATION AND NECESSARY ACTION AT LEAST 3 DAYS BEFORE THE COMMENCEMENT OF THE EVALUATION SESSION.

SEMESTER END EXAMINATION: -

It is defined as the examination of the learners on the basis of performance in the semester end theory / written examinations.

B. Semester End Examination- 60 % of 75 Marks or 50 Marks Examination.

- 1) Duration – These examinations shall be of 2 Hours duration.
- 2) Question Paper Pattern: -
 - i. There shall be five questions each of 10 marks.
 - ii. All questions shall be compulsory with internal choice within the questions.
 - iii. Question may be sub- divided into sub- questions a, b, c, d & e only and the allocation of marks depends on the weightage of the topic.

THE MARKS OF THE INTERNAL ASSESSMENT SHOULD NOT BE
DISCLOSED TO THE STUDENTS TILL THE RESULTS OF THE
CORRESPONDING SEMESTER IS DECLARED.

HSNC University, Mumbai

(2023-2024)

Ordinances and Regulations

With Respect to

Choice Based Credit System (CBCS)

For the Programmes Under

The Faculty of Science and Technology

For the Course

Data Science & Business Analytics

Curriculum – First Year Postgraduate Programmes

Semester-I and Semester -II

2023-2024

Data Science & Business Analytics

Part 1- Preamble

M. Sc. Data Science and Business Analytics program is of minimum 88 credits cover four semesters. Data is the new oil. The analytics may be input for human decisions or may drive fully automated decisions. It helps decision maker in building strategies to perform deep-dive understanding and provide descriptive, predictive, and prescriptive analytics. It is used to run the business effectively and is instrumental in growing the business. It is the area for huge potential for corporate investments. Business Analytics include identifying KPIs,

measurement strategy, data analysis, complex statistical model and analysis, data mining and deep understanding of cause-and-effect models. Business analytics can drive key decision making in the organization and help executive decision makers in building strategy, predictive analysis, forecasting, risk analysis, identify and prevent fraud, market analysis, etc. Data Scientists use these skills can provide insights into discrete data sets, build complex model and present them in Scorecard format and use the same in executive reviews to lead data-driven discussion and decisions. Some of the impactful use of this is in the areas of Management Information Systems, Financial Service, Marketing Research, Process Improvements, Six Sigma, Process Excellence, Scorecard, Dashboard, End-to-End Product Management, etc.

The program emphasizes both theory and modern applications of Data Science and Business analytics and is structured to provide knowledge and skills in depth necessary for the employability of students in industry, in academics and other government and non-government organizations. The program has some unique features like independent projects, number of elective courses and extensive computer training of statistical computations including standard software packages like SQL, SPSS AMOS, MINITAB, R and PYTHON etc. Due to Cluster University, the department got the academic autonomy and it has been utilized to add the new and need based elective courses. The independent project work is one among the important components of this program. The syllabus has been framed to possess a decent balance of theory, methods, and applications of statistics. It is possible for the students to study basic courses from other disciplines like economics, life sciences, computer science and Information Technology in place of optional/electives. The thrust of the course is to prepare students to enter a promising career after graduation, as also provide them a platform for pursuing higher studies resulting in postgraduation degrees.

Program Outcomes:

1. Develop in depth understanding of the key technologies in data science and business analytics: data mining, machine learning, visualization techniques, predictive modelling, and statistics.
2. Practice problem analysis and decision-making.
3. Gain practical, hands-on experience with statistics programming languages and

big data tools through coursework and applied research experiences.

4. Explore text mining analysis/techniques to understand the influence of social media applications.
5. Apply Principles and skills of Finance, Marketing and Decision making in Data Science Contexts and Environments.
6. Encourage an aptitude for business improvement, innovation and entrepreneurial action.
7. Analyze legal and Ethical Principles in Data Science and Decision making contexts.
8. Enable all participants to recognise, understand and apply the language, theory and models of the field of business analytics.
9. Employ cutting edge tools and technologies to analyze Big Data.
10. Create and configure virtual machines, storage, networking, and other resources in a cloud environment.
11. Apply advanced data analysis techniques to financial data and explore emerging technologies and trends in FinTech.
12. Demonstrate use of teamwork, leadership skills, decision making and organization theory.

1. Course objective: The main course objectives are:

- Provide hands-on training to students to develop and enhance the strong analytical, quantitative modelling skills and business skills for solving team-based, real-world business problems and to make students ready for the role of Data Scientist.
- Provide opportunity to work on some real-life data/problems or simulated data through building the business Intelligence reports, scorecard and dashboard.

Pre requisites:

1. Measures of central tendency:

- 1.1 Concept of central tendency of data. Requirements of good measure
- 1.2 Locational averages: Median, Mode, and Partition Values: Quartiles, Deciles, and Percentiles.
- 1.3 Mathematical averages Arithmetic mean (Simple, weighted mean, combined mean), Geometric mean, Harmonic mean,
- 1.4 Empirical relation between mean, median and mode
- 1.5 Merits and demerits of using different measures & their applicability

2. Probability Basic definitions of probability, Events, Properties of probability

SEMESTER-I

MS-DS-101 Foundations of Data Science

Objective: Students will learn the ethics and privacy concerned with data while collecting data, dealing with the data through different case studies.

The objective of the course is to make the students conversant with various techniques used in summarization, certainty involved in uncertainty in happening of events, understanding the data and the distributions, estimation, testing of hypothesis and analysis of data using statistical techniques. The focus will be both on theoretical as well as practical approach using commonly used Statistical Software.

C01:Understand the Foundations of Data Ethics

C02:Analyze Ethical Dilemmas in Data Collection and Usage

C03:Apply probability distribution functions to solve real-world data science problems.

C04:Apply sampling methods to estimate population parameters and assess the accuracy of estimates.

C05:Recognize the role of statistics in making data-driven decisions and solving complex business and research challenges.

MS-DS-102 Computational Mathematics

Objective: The objective of the course is to provide students with a comprehensive understanding of mathematical concepts and techniques essential for solving problems in data science through computational methods. Students will learn essential mathematical concepts and their practical applications in the context of data analysis, machine learning, and statistical modeling.

C01: Solve systems of linear equations using techniques like Gaussian elimination and matrix factorization.

C02: Perform matrix operations, including addition, multiplication, transpose, and inversion.

C03: Apply eigen-decomposition for dimensionality reduction and feature extraction.

C04: Utilize multivariate calculus to optimize algorithms, such as gradient descent for machine learning optimization.

C05: Solving mathematical equations, integrating functions with numerical methods.

C06:Formulate real-world problems as mathematical models and select appropriate techniques for their solutions.

MS-DS-103 Python Programming

Objective: This course aims to provide students with advanced skills and knowledge in data analysis and data visualization using the Python programming language. The objective of the course is to equip students with the skills necessary to effectively gather, clean, analyze, and interpret data using Python. Students will also learn to create data visualizations to communicate insights.

C01: Apply functional programming concepts in Python, including lambda functions, map, filter, and reduce.

C02: Analyze data and perform scientific computations using Numpy

C03: Cleaning and managing data using pandas.

C04: Visualize and interpret data using matplotlib.

MS-DS-104A Data Analysis using R

Objective: The objective of this course is to equip students with a comprehensive understanding of data analysis techniques using the R programming language. Students will develop the skills required to effectively analyze and interpret datasets, make data-driven decisions through hands-on practical experience.

C01: Develop R Programming Skills

C02: Demonstrate proficiency in R syntax, data structures, and functions.

C03: Utilize R packages and libraries for various data analysis, visualization, and statistical tasks.

C04: Profile and optimize R code for improved performance and memory usage.

C05: Generate informative and visually appealing data visualizations using R's visualization libraries.

C06: Create well-structured data analysis workflows using R's functional programming capabilities.

MS-DS-104B Cyber Security

Objective: Objective: To prepare students with the technical knowledge and skills needed to protect and defend computer systems and networks. To develop graduates that can identify, analyze, and remediate computer **security** breaches.

C01: Understand the Foundations of Cyber Security.

C02: Apply techniques to limit the visibility of personal information and enhance online privacy.

C03: Design and deliver awareness campaigns to promote safe online behavior and skepticism.

C04: Analyze Cyber Threats and Attacks.

C05: Demonstrate safe practices for sharing links, files, and multimedia to prevent spreading malware.

C06: Design and practice effective incident response procedures to minimize damage and recover from attacks.

Mi 20 Research Methodology

Objective:

SEMESTER-II

MS-DS-201 DBMS with SQL

Objective: The objective of this fundamental course for data handling and analysis is to provide students with a strong foundation in working with databases and using SQL for data manipulation. Students will learn the practical skills related to MySQL for data manipulation and management.

C01:Understand Fundamental Database Concepts

C02:Identify the components of a DBMS architecture, including data storage, query processing, and transaction management.

C03:Create and Manage Databases and Tables

C04:Perform Basic Data Manipulation.

C05:Write Complex SQL Queries

C06:Apply normalization techniques to design and optimize database schemas

MS-DS-202 Machine Learning

Objective: The objective of the course is to provide students with the knowledge and skills to effectively understand, apply, and evaluate various supervised machine learning techniques for solving real-world problems. Students will learn the theory and practical implementation of supervised machine learning algorithms and make informed decisions regarding model selection, hyperparameter tuning, and evaluation.

C01: Differentiate between supervised, unsupervised, and reinforcement learning.

C02: Preprocess and clean data to prepare it for machine learning tasks.

C03: Implement various regression techniques, including linear regression, polynomial regression, and regularization.

C04: Implement classification algorithms, such as logistic regression, decision trees, support vector machines, and k-nearest neighbors.

C05: Learn methods to evaluate model performance, including accuracy, precision, recall, F1-score, and ROC curves.

C06: Implement ensemble methods such as random forests, boosting, and bagging.

MS-FDS-203 Advanced Statistical methods

Objective: The objective of this course is to make students understand advanced statistical methods and their applications in data science. The students will learn to make informed decisions based on rigorous statistical reasoning.

C01:Understand Hypothesis Testing Concepts.

C02:Explain the importance of significance levels, p-values, and critical regions in hypothesis testing.

C03:Interpret and draw conclusions from multiple comparison results.

C04:Assess model fit and perform variable selection techniques.

C05:Apply GLMMs to account for clustered or correlated data structures.

C06:Interpret results and handle overdispersion in count data analysis.

MS-DS-204A Cloud Computing

Objective: The objective of this course is to make students understand the fundamentals of cloud computing, including service models (IaaS, PaaS, SaaS) and deployment models. Students will learn about cloud platforms such as AWS, Azure, and Google Cloud and manage virtual machines, containers, and serverless architectures.

C01:Implement load balancing techniques to distribute network traffic effectively.

C02:Analyze challenges and benefits of designing and managing distributed systems.

C03:Identify cloud security risks and threats, including data breaches and unauthorized access.

C04:Develop strategies for data anonymization, retention, and compliance in cloud environments.

C05:Evaluate cost-effectiveness, scalability, and agility benefits of cloud adoption.

C06:Implement cloud governance frameworks to monitor and control cloud resources.

MS-DS-204B Financial technology

Objective:

C01:Analyse and model financial data.

C02:Construct and optimize asset portfolios.

C03:Evaluate and model Risk on various financial assets.

C04: Use the most powerful and sophisticated routines in R for analytical finance.

2. Salient features, how it has been made more relevant.

- Syllabus is prepared by top most industry experts, eminent statisticians, Economists, IT faculty, Computer Science faculty, Data Scientists, and many reputed faculties of HSNC University.
- Skills covered includes Statistics, Mathematics, Economics Data Analytics, Business Analytics, Data Visualization, machine learning algorithm, Artificial Intelligence, Text Mining, Forecasting/ Predictive Analytics, Hypothesis Testing, Survival Analysis etc.
- Tools and Technology covered includes R, Python, My SQL, NO SQL, Tableau, Power BI, Machine Learning, Artificial Intelligence, SPSS, Minitab, Excel, and many more.
- Course includes several Assignments and Case Studies.
- Getting industrial experience by working on industry relevant live projects and internship.
- University has tie up with 100+ companies to provide job to many students.
- University has a dedicated placement cell for the participants who will complete course.
- Teaching Faculties will include good blend of Academicians and Industry Experts.
- Focus is to prepare Participants with clear, concise concept to experts in data Science field to add quality and value to institute they join.
- University/ College has well placed Alumni working at top position.

3. Learning Outcomes.

Students will learn Analytics from basics concepts to creating basic models for predictions. Students will learn how analytics is actually used large corporations like Microsoft, Jio, Amazon and other top companies globally. In the hands-on session, Students will gain skills on identifying opportunities for Analytics, Machine Learning, IoT, AI, Blockchain, coming up with right set of metrics/KPIs, use cases, defining the metrics, measuring, and implementing it.

Students will be able to come up with specific Analytics Opportunities, Define Metrics and KPIs successfully for any business.

Part 2- The Scheme of Teaching and Examination is as under:

Semester – I

Summary

Sr. N o.	Choice Based Credit System			Subject Code	Remarks
1	Core Course (Analytics – I, Foundation of Data Science, Computational Mathematics, Python Programming), (Practical' s of MS-DS-101, MS-DS-102, MS-DS-103)			MS-DS-101 MS-DS-102, MS-DS-103, MS-DS-P1	
2	Elective Course	Discipline Specific Elective (DSE) Course			
		2.1	Interdisciplinary Specific Elective (IDSE) Course : 1.Data Analysis Using R 2.Cyber security	MS-DS-104	
		2.2	Dissertation/Project		
		2.3	Generic Elective (GE) Course		
3	Minor course: Research Methodology			Mi-20(RM)	
4	Ability Enhancement Courses (AEC)				

First Year Semester I Internal and External Detailed Evaluation Scheme

Major Papers (03)

Sr. N o.	Subject Code	Subject Title	Periods Per Week						Evaluation Scheme				Mark s
			Units	S. L.	L	T	P	Cred it	S. L. E	CT	T A	SE E	
1	MS-DS-101	Foundations of data	3	20%	3	0	0	3	10	10	5	50	75

		science												
2	MS-DS-102	Computational Mathematics	3	20%	3	0	0	3	10	10	5	50	75	
3	MS-DS-103	Python Programming	3	20%	3	0	0	3	10	10	5	50	75	
4	MS-DS-104A	Data Analysis Using R	3	20%	3	0	0	3	10	10	5	50	75	
5	MS-DS-104B	Cyber security	3	20%	3	0	0	3	10	10	5	50	75	
6	MI20(RM)	Research Methodology	4	20%	4	0	0	4	20	15	5	60	100	
7	MS-DS-P1	Practicals of MS-DS-101, MS-DS-102, MS-DS-103 MS-DS-104 A/ MS-DS-104 B	5 practicals per batch	0	0	1	4	4						100
	Total Hours / Credit							20	Total Marks					500

* Three hours per week to be taken for CONTINUOUS self– learning Evaluation.

First Year Semester I - Units – Topics – Teaching Hours

S. N	Subject Code	Subject Unit Title	Hours	Total No. of Hours	Credit	Total Marks
------	--------------	--------------------	-------	--------------------	--------	-------------

1	MS-DS-101	I	Data Ethics and Privacy	15	45	3	75 (50+25)
		II	Probability Discrete random variable and standard distributions	15			
		III	Standard distributions. Estimation, Bayes Estimation	15			
2	MS-DS-102	I	Linear Algebra	15	45	3	75 (50+25)
		II	Differentiation	15			
		III	Optimization Technique	15			
3	MS-DS-103	I	Introduction to Numpy and Pandas	15	45	3	75 (50+25)
		II	Data Wrangling using Pandas	15			
		III	Visualization with Matplotlib	15			
4	MS-DS-104A	I	Fundamentals of R	15	45	3	75 (50+25)
		II	Data Handling	15			
		III	Statistical Computing	15			
5	MS-DS-104B	I	Social Media Security	15	45	3	75 (50+25)
		II	Security Guidelines	15			
		III	Social Engineering and IT Security	15			
6	Mi20(RM)	I	Research methodology	15	60	4	100
		II		15			

		III		15			
		IV		15			
7	MS-DS-P 1	Practicals of MS-DS-101, MS-DS-102, MS-DS-103 MS-DS-104A / MS-DS-104B			60	4	100

- **Lecture Duration =1 Hour.**

- **One Credit =15 hours**

L: Lecture: Tutorials P: Practical Ct-Core Theory, Cp-Core Practical, SLE- Self learning evaluation CT- Commutative Test, SEE- Semester End Examination, PA-Project Assessment, AT- Attendance

Part -3 Detailed Scheme Theory

Curriculum Topics along with Self-Learning topics

Teaching of the unit will be done through teaching mode and through self-learning mode. Evaluation of self-learning topics to be undertaken before the concluding lecture instructions of the respective UNIT.

MS-DS-101 Foundations of Data Science Lectures)

(Total Hours : 45

Unit	Content	No. of Lectures
1	Data Ethics and Privacy: 1.1 Disinformation 1.2 Bias & Fairness 1.3 Ethical Foundations & Practical Tools 1.4 Privacy and surveillance 1.5 How did we get here? Our Ecosystem 1.6 Algorithmic Colonialism, and Next Steps	15

2	<p>Measures of Dispersion, Skewness & Kurtosis:</p> <p>2.1 Concept of dispersion. Requirements of good measure.</p> <p>2.2 Absolute and Relative measures of dispersion: Range, Quartile Deviation, Mean absolute deviation, Standard deviation, Variance and Combined variance.</p> <p>Raw moments and central moments and relations between them.</p> <p>2.3 Concept of Skewness and Kurtosis: Measures of Skewness: Karl Pearson's, Bowley's and Coefficient of skewness based on moments, Box Plot.</p> <p>2.4 Conditional probability - Multiplication rule, Independence, Law of total probability, Bayes' theorem</p> <p>2.5 Random Variables - Random variable, Discrete random variable, Probability mass function, Cumulative density function</p> <p>2.6 Expectation and Variance - Expectation of a discrete random variable, Variance, and standard deviation of a discrete random variable. Properties of Expectation and variance. Correlation coefficient.</p> <p>2.7 Binomial and Poisson random variables - Bernoulli trials, Binomial distribution, Expectation and variance of a binomial random variable, Poisson distribution.</p> <p>2.8 Continuous random variable, Expectation and variance.</p>	15
3	<p>3.1 Continuous Probability distributions: Normal Distribution and Exponential distribution.</p> <p>3.2 Central Limit theorem (statement only).</p> <p>3.3 Sampling distribution of sample means and sample proportion (For large sample only).</p> <p>Estimation:</p> <p>3.4 Concept of Parameter, statistic, estimator, and estimate.</p> <p>3.5 Properties of good estimator (Only names), unbiasedness and standard error of an estimator. (Development of critical region is not expected.)</p> <p>3.6 Point and Interval estimate of single proportion, difference of two proportions.</p> <p>3.7 Confidence intervals for (i) Mean of Normal population, (ii) difference between means of two independent Normal populations having the same variance.</p> <p>3.8 Introduction of t distribution, Chi square distribution, F distribution</p> <p>3.9 Confidence intervals using t distribution, Chi square distribution, F distribution.</p>	15

--	--	--

Self-LLearning topics (Unit wise)

Sub Unit	Topics
2.2	Absolute and Relative measures of dispersion: Range, Quartile Deviation, Mean absolute deviation, Standard deviation, Variance and Combined variance. Raw moments and central moments and relations between them.
2.3	Bowley' s and Coefficient of skewness based on moments,
2.6	Expectation and Variance - Expectation of a discrete random variable, Variance, and standard deviation of a discrete random variable.
2.7	Bernoulli trials, Binomial distribution, Expectation and variance of a binomial random variable, Poisson distribution.
3.4	Concept of Parameter, statistic, estimator, and estimate.
3.6	Difference of two proportions.

Online Resources

<p>‘ Introduction to Probability and Statistics’ by Prof. G. Srinivasan from IIT Madras available on the Swayam portal</p> <p>https://nptel.ac.in/courses/111/106/111106112/</p>
<p>‘ Introduction to Probability Theory and Stochastic Processes’ by Prof. S Dhramaraja from IIT Delhi available on the Swayam portal</p> <p>https://nptel.ac.in/courses/111/102/111102111/ for unit II</p>

Statistics for Business Economics’ by Dr. Patel from University School of Sciences available on the Swayam portal http://ugcmoocs.inflibnet.ac.in/ugcmoocs/view_module_ug.php/227
‘ Business Statistics’ by Dr Mukesh Kumar Barua from IIT Roorkee available on the Swayam portal https://nptel.ac.in/courses/110/107/110107114/

MS-DS-102 Computational Mathematics

(Total Hours : 45 Lectures)

Unit	Content	No. of Lectures
1	Linear Algebra System of linear equations, Matrix operations, Computation of determinant and inverse of a matrix, Generalized inverse, matrix equations, Solution of system of linear equations by Gaussian elimination, parametric form, inverse of matrix, LU factorization Linear transformation, eigenvalues and eigenvectors, matrices diagonalization, orthogonality, orthogonal complements, orthogonal Diagonalization, Positive semi definite and position definite, QR-Factorization, Singular value decomposition, applications to principal component analysis.	15
2	Differentiation Differentiation of Univariate Functions, maxima and minima, Taylor series, Differentiation Rules. Partial Differentiation and Gradients, Basic Rules of Partial Differentiation, Chain rule, Jacobian matrix, Gradient of a vector-valued functions, Gradients of Matrices, Higher-Order Derivatives, Hessian Matrix, multivariate Taylor series, multivariate Taylor polynomial, local and global maxima & minima	15
3	Optimization Techniques	15

	<p>Necessary and sufficient conditions for the existence of an extremal point, Newton's method, Lagrange multipliers, gradient and conjugate gradient methods.</p> <p>Roots of Non-linear equations: Fixed point iteration, bisection and Newton-Raphson method,</p> <p>Integral Function, Riemann integration, improper integrals, Application of Integral.</p>	
--	---	--

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	System of linear equations, Matrix operations, Computation of determinant and inverse of a matrix,
2	Differentiation of Univariate Functions, maxima and minima, Differentiation Rules.
3	Bisection and Newton-Raphson method

Online Resources

<p>Essential Mathematics for Machine Learning by Prof. S.K. Gupta, Dr. Sanjeev Kumar from IIT Roorkee available on the Swayam portal https://nptel.ac.in/courses/111107137</p>
<p>Calculus of Several Real Variables by Dr. Joydeep Dutta from IIT Kanpur available on the Swayam portal https://nptel.ac.in/courses/111104125</p>
<p>Applied Linear Algebra in AI and ML by Prof. Swanand Khare from IIT Kharagpur available on the Swayam portal https://nptel.ac.in/courses/111105165</p>

MS-DS-103 Python Programming

(Total Hours : 45 Lectures)

Unit	Content	No. of Lectures
1	<p>Introduction to NumPy and Pandas</p> <p>Basics of NumPy arrays: indexing, slicing, reshaping, concatenation, splitting, Computation on NumPy arrays: Universal</p>	15

	<p>Functions, Aggregations, Broadcasting, Comparison, Sorting arrays, Structured arrays, Expressing Conditional Logic as Array Operations, Random number generation.</p> <p>Basics of Pandas: Pandas Objects- Series, Data Frame, Index, Data indexing and selection and filtering, dropping entries from an axis, arithmetic and data alignment, function application and mapping, Summarizing and computing descriptive statistics.</p> <p>Working with data in text format, Delimited formats, JSON Data, XML, HTML: Web Scrapping, Interacting with Databases, Combining Data sets: Concat, Append, Merge and Join.</p>	
2	<p>Data Wrangling using Pandas</p> <p>Data Cleaning and transformation: Reshaping, handling missing values, replacing values, removing duplicates, transforming data using a function or mapping, Discretization and Binning, Filtering outliers, permutation and random sampling, computing indicators/Dummy variables, group-wise operations and transformations.</p> <p>Data Manipulation: Aggregation and Grouping – simple aggregation in Pandas, Group BY: Split, Apply, Combine, Pivot Tables and cross-tabulation, String Manipulation: String Object Methods, Regular expressions, Vectorized string functions in pandas, high-performance Pandas: eval() and query()</p>	15
3	<p>Visualization with Matplotlib</p> <p>Line plots, Scatter plots, visualizing errors, density and contour plots, Histograms, Binnings nad density, customizing plot legend and colorbars, Multiple subplots, Text and Annotations, Customizing ticks, customizing matplotlib configurations and stylesheets, Three-Dimensional plotting in Matplotlib, geographic map projections, plotting data maps.</p> <p>Alternatives to Matplotlib: Exploring seaborn plots, Plotly plots, ggplot, PyViz, Bokeh and Panel, Yellow brick, etc.</p>	15

Sub Unit	Topics
1	Numpy Aggregations functions, dropping entries from an axis, arithmetic functions, computing descriptive statistics.
2	Random sampling, Aggregation and Grouping – simple aggregation in Pandas, Group BY: Split, Apply,
3	Plotly plots, ggplot, PyViz, Bokeh and Panel, Yellow brick, etc.

Online Resources

Python for Data Science by Prof. Ragunathan Rengasamy from IIT Madras available on the Swayam portal https://nptel.ac.in/courses/106106212
Business Analytics & Text Mining Modeling Using Python by Dr. Gaurav Dixit from IIT Roorkee available on the Swayam portal https://nptel.ac.in/courses/110107129

MS-DS-104A Data Analysis Using R

(Total Hours : 45 Lectures)

Unit	Content	No. of Lectures
1	<p>Fundamentals of R</p> <p>Introduction to R features of R, installation of R, Starting and ending R session, getting help in R ,</p> <p>Value assigning to variables.</p> <p>Basic Operations: +, -, *, ÷ , ^, sqrt.</p> <p>Numerical functions: log 10, log , sort, max, unique, range, length, var, prod, sum, summary, dim, sort, five num etc.</p> <p>Reading and writing data: From and to CSV files and HTML.</p> <p>Data Type: Vector, list, matrices, array and data frame 1.7 Variable Type: logical, numeric, integer, complex, character and factor.</p> <p>Operations on matrices.</p> <p>Control statements: if, if-else, if-else-if, while loop, for loop. 1.10</p> <p>Defining functions and Printing outputs.</p>	15

2	<p>Data Handling</p> <p>Data Manipulation: Selecting random N rows, removing duplicate row(s), dropping a variable(s), Renaming variable(s), sub setting data, creating a new variable(s), selecting of random fraction of row(s), appending of row(s) and column(s), simulation of variables.</p> <p>Data Processing: Data import and export, setting working directory, checking structure of Data :Str(), Class(), Changing type of variable (for eg as.factor, as.numeric)</p> <p>Introduction to dplyr and data.table packages</p> <p>Data Visualisation using simple functions and ggplot: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot eg plot(), correlation plot.</p>	15
3	<p>Statistical Computing:</p> <p>Descriptive Statistics: Averages, Positional Averages, Dispersion, Skewness , Kurtosis, Correlation Curve Fitting and Regression</p> <p>Some Probability Distributions: Plotting of pmf/pdf, cdf, computation of probabilities of Binomial, Poisson, Normal, Exponential.</p> <p>Statistical Tests for t, Chi-square, F and ANOVA</p> <p>Operations Research Techniques:</p> <p>A Concept and Mathematical Formulation of Linear Programming Problem, Transportation Problems and Assignment Problems.</p> <p>Numerical problems of all above using lpSolve.</p>	15

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	Introduction to R features of R, installation of R, Starting and ending R session, getting help in R, Operations on matrices.
2	Data Visualisation using simple functions and ggplot: Simple bar diagram, subdivided bar diagram, multiple bar diagram, pie diagram, Box plot for one and more variables, histogram, frequency polygon, scatter plot eg plot(), correlation plot.
3	Descriptive Statistics: Averages, Positional Averages, Dispersion, Skewness , Kurtosis, Correlation Curve Fitting and Regression

Online Resources

‘ Introduction to R Software’ by Prof. Shalabh from IIT Kanpur available on the Swayam portal

https://nptel.ac.in/courses/111/104/111104100/
Business analytics and data mining Modeling using R by Dr. Gaurav Dixit from IIT Roorkee available on the Swayam portal https://nptel.ac.in/courses/110107092
Data Analysis with R Programming available on coursera https://www.coursera.org/learn/data-analysis-r
Data Analysis with R by IBM available on coursera https://www.coursera.org/learn/data-analysis-with-r

MS-DS-104B Cyber Security

(Total Hours : 45 Lectures)

Unit	Content	No. of Lectures
1	Social Media Security 1.1 Introduction to Cyber Space: History of Internet, Cyber Crime, Information Security, Computer Ethics and Security Policies 1.2 Choosing the Best Browser according to the requirement and email security: Guidelines to choose web browsers, Securing web browser, Antivirus, Email security 1.3 Guidelines for secure password and wi-fi security 1.4 Guidelines for setting up a Secure password 1.5 Two-steps authentication 1.6 Password Manager 1.7 Wi-Fi Security 1.8 Guidelines for social media and basic Windows security 1.9 Guidelines for social media security 1.10 Tips and best practices for safer Social Networking 1.11 Basic Security for Windows 1.12 User Account Password	15
2	Security Guidelines Smartphone security guidelines: Introduction to mobile phones, Smartphone Security, Android Security, IOS Security 2 Cyber Security Initiatives in India: Counter Cyber Security Initiatives in India, Cyber Security Exercise, Cyber Security Incident Handling, Cyber Security Assurance 3 Online Banking, Credit Card and UPI Security: Online Banking	15

	Security, Mobile Banking Security, Security of Debit and Credit Card, UPI Security, Micro ATM, e-wallet and POS Security, Security of Micro ATMs. 2.4 e-wallet Security Guidelines 2.5 Security Guidelines for Point of Sales(POS)	
3	Social Engineering and IT Security Social Engineering: Social Engineering, Types of Social Engineering, How Cyber Criminal Works, How to prevent for being a victim of Cyber Crime. 2 Cyber Security Threat Landscape and Techniques: Cyber Security Threat Landscape, Emerging Cyber Security Threats, Cyber Security Techniques. 3 IT Security Act and Misc. Topics: IT Act, Hackers-Attacker Countermeasures, Web Application Security, Digital Infrastructure Security, Defensive Programming 3.4 Information Destroying and Recovery Tools 3.5 Recovering from Information Loss 3.6 Destroying Sensitive Information 3.7 CCleaner for Windows	15

Self-Learning topics (Unit wise)

Sub Unit	Topics
2	Online Banking, Credit Card and UPI Security: Online Banking Security, Mobile Banking Security, Security of Debit and Credit Card, UPI Security, Micro ATM, e-wallet and POS Security, Security of Micro ATMs. e-wallet Security Guidelines, Security Guidelines for Point of Sales(POS)
3	Engineering: Social Engineering, Types of Social Engineering, How Cyber Criminal Works, How to prevent for being a victim of Cyber Crime.

Online Resources

Cyber Security and Privacy from IIT Madras available on the Swayam portal https://nptel.ac.in/courses/106106248
Privacy and Security in Online Social Media by Prof. Ponnurangam Kumaraguru from IIIT Hyderabad available on the Swayam portal https://nptel.ac.in/courses/106106146

Uni t	Content	No. of Lectures
1		15
2		15
3		15
4		15

Self-Learning topics (Unit wise)

Sub Unit	Topics

Online Resources

Detailed Scheme Practicals

Code Practical: MS-DS-P1

Title of Paper: Practicals of MS-DS-101, US-FDS-102, US-FDS-103

Title of Paper	Content	No. of Lectures
MS-DS-101	Practicals based on MS-DS-101(Using Excel)	2 hours per Batch per Practical
MS-DS-102	Practicals based on MS-DS-102	2 hours per Batch per Practical
MS-DS-103	Practicals based on MS-DS-103	2 hours per Batch per Practical
MS-DS-104A	Practicals based on MS-DS-104A	2 hours per Batch per Practical
MS-DS-104B	Practicals based on MS-DS-104B	2 hours per

4B		Batch per Practic al
----	--	-------------------------------

MS-DS-101 References:

Foundations of data science

1. Excel Statistics-A Quick Guide, Nel J. Salkind, Sage Publications.
2. Zeynep Tufekci, How social media took us from Tahrir Square to Donald Trump.
3. Timnit Gebru et al, Datasheets for Datasets.
4. Rachele Hampton, The Black Feminists Who Saw the Alt-Right Threat Coming.

Statistical Methods:

1. Medhi J.: Statistical Methods, An Introductory Text, Second Edition, New Age International Ltd.
2. Agarwal B.L.: Basic Statistics, New Age International Ltd.
3. Spiegel M.R.: Theory and Problems of Statistics, Schaum's Publications series. Tata McGraw Hill.
4. Kothari C.R.: Research Methodology, Wiley Eastern Limited.
5. David S.: Elementary Probability, Cambridge University Press.
6. Hoel P.G.: Introduction to Mathematical Statistics, Asia Publishing House.
7. Hogg R.V. and Tannis E.P.: Probability and Statistical Inference. McMillan Publishing Co. Inc.
8. Pitan Jim: Probability, Narosa Publishing House.
9. Goon A.M., Gupta M.K., Dasgupta B.: Fundamentals of Statistics, Volume II: The World Press Private Limited, Calcutta.
10. Gupta and Kapoor: Fundamentals of Applied Statistics, S. Chand
11. Gupta and Kapoor: Fundamentals of Mathematical Statistics, S. Chand
12. Sharma S. D.: Operations Research, Kedar Nath Ram Nath
13. Taha Hamdy A.: Operations Research-An Introduction, Tenth Edition, Pearson

MS-DS-102 Computational Mathematics References

1. GGG_Nicholson,_W_Keith_Linear_algebra_with_applications_McGraw_Hill.
2. Mathematics for Machine Learning Marc Peter Deisenroth, A. Aldo Faisal and Cheng Soon Ong, Cambridge University Press .
3. Higher Engineering Mathematics- H.K. Dass, Er. Rajnish Verm-S.chand

MS-DS-103 Python Programming References

1. "Python for Data Analysis" by Wes McKinney.
2. " Python Data Science Handbook" by Jake VanderPlas.

MS-DS-104A data Analysis Using R References

Excel Statistics-A Quick Guide, Nel J. Salkind, Sage Publications.

MS-DS-104B Cyber Security References

1. Introduction to Cyber Security available at <http://uou.ac.in/foundation-course>

2. Fundamentals of Information Security <http://uou.ac.in/progdetail?pid=CEGCS-17>
3. Cyber Security Techniques <http://uou.ac.in/progdetail?pid=CEGCS-17>
4. Cyber Attacks and Counter Measures: User Perspective
<http://uou.ac.in/progdetail?pid=CEGCS-17>
5. Information System <http://uou.ac.in/progdetail?pid=CEGCS-17>

Mi 20 (RM) Research Methodology References

Part 5- The Scheme of Teaching and Examination is as under: First Year Semester- II

Summary

Sr. No.	Choice Based Credit System		Subject Code	Remarks
1	Core Course (DBMS with SQL, Machine Learning, Advanced Statistical Methods)		MS-DS-201, MS-DS-202, MS-DS-203 MS-DS-P2	
2	Elective Course	Discipline Specific Elective (DSE) Course		
			1 Interdisciplinary Specific Elective (IDSE) Course (Cloud Computing, Financial Technology)	MS-DS-204A MS-DS-204B
			2 Dissertation/Project	
			3 Generic Elective (GE) Course	

First Year Semester -II Internal and External Detailed Evaluation Scheme

* Three hours per week to be taken for CONTINUOUS self -learning Evaluation.

Sr. N o.	Subject Code	Subjec t Title	Periods Per Week						Evaluation Scheme				Mark s
			Units	S. L.	L	T	P	Cred it	S. L. E	CT	T A	SE E	
1	MS-DS-201	DBMS with SQL	3	20 %	3	0	0	3	10	10	5	50	75
2	MS-DS-202	Machine Learning-I	3	20 %	3	0	0	3	10	10	5	50	75
3	MS-DS-203	Advanced Statistical Methods	3	20 %	3	0	0	3	10	10	5	50	75
4	MS-DS-204A	Cloud Computing	3	20%	3	0	0	3	10	10	5	50	75
5	MS-DS-204B	Financial Technolog y	3	20%	3	0	0	3	10	10	5	50	75
6	MS-DS-P1	Practicals of MS-DS-201, MS-DS-202, MS-DS-203 MS-DS-204 A/ MS-DS-204 B	5 practical s per batch	0	0	1	4	4					100
	Total Hours / Credit							20	Total Marks				500

S. N	Subject Code	Subject Unit Title		Hou rs	Total No. of Hours	Cre dit	Tot al Marks
1	MS-DS-2	I	Introduction & DBMS	15	45	3	75

	01		Architecture				(50+25)
		II	Basics of MySQL	15			
		III	Advanced SQL	15			
2	MS-DS-202	I	Regression	15	45	3	75 (50+25)
		II	Classifications	15			
		III	Model Selection and improvement	15			
3	MS-DS-203	I	Testing Hypothesis	15	45	3	75 (50+25)
		II	Linear Models	15			
		III	Generalized Linear Models	15			
4	MS-DS-204A	I	Networking Concepts & Distributed Systems	15	45	3	75 (50+25)
		II	Cloud Computing and Security	15			
		III	Business demands and needs	15			
5	MS-DS-204B	I	Introduction to Fintech	15	45	3	75 (50+25)
		II	Working of Fintech	15			
		III	Sectors in Fintech	15			
7	MS-DS-P2	Practicals of MS-DS-201, MS-DS-202, MS-DS-203 MS-DS-204A			60	4	100

		/					
		MS-DS-204B					

First Year Semester – II Units – Topics – Teaching Hours

- **Lecture Duration =1 Hour.**
- **One Credit =15 hours**

L: Lecture: Tutorials P: Practical Ct-Core Theory, Cp-Core Practical, SLE- Self learning evaluation
 CT-Commutative Test, SEE- Semester End Examination, PA- Project Assessment, AT- Attendance

Part -6 - Detailed Scheme Theory

Curriculum Topics along with Self-Learning topics

Teaching of the unit will be done through teaching mode and through self-learning mode. Evaluation of self-learning topics to be undertaken before the concluding lecture instructions of the respective UNIT.

MS-DS-201 DBMS with SQL (Total Hours: 45 Lectures)

Unit	Content	No. of Lectures
1	Introduction & DBMS Architecture 1.1 Data, Database, Database management system 1.2 Characteristics of the Database Approach 1.3 Advantages and Disadvantages of DBMS 4 Data Models, Categories of Data models, Schemas, Instance and Database states 1.5 Data Independence 1.6 The Three schema architecture 1.7 DBMS language and interface	15

	1.8 Classifications of Database Management Systems	
--	--	--

2	Basics of MySQL 2.1 Introduction of MySQL as a DBMS 2.2 Relational Database 2.3 Creating a Database, CREATE Table statement 2.4 SELECT Statement 2.5 COUNT, DISTINCT, LIMIT Statements 2.6 INSERT Statement 2.7 UPDATE and DELETE Statements 2.8 DROP Statement	15
3	Advanced SQL 3.1 Using String Patterns, Ranges 3.2 Sorting and Grouping 3.3 Built-in Database Functions 3.4 Date and Time Built-in Functions 3.5 Numeric Built-in Functions 3.6 String Built-in Functions 3.7 Sub-Queries and Nested Selects 3.8 Working with Multiple Tables 3.9 WHERE and HAVING Clause 3.10 LIKE Clause 3.11 Transactions: ROLLBACK and COMMIT Case study to explore 3.1 to 3.11	15

Sub Unit	Topics
----------	--------

1	1.1 Data, Database, Database management system 1.2 Characteristics of the Database Approach 1.4 Data Models, Categories of Data models, Schemas, Instance and Database states
2	2.1 Introduction of MySQL as a DBMS 2.3 Creating a Database, CREATE Table statement 2.4 SELECT Statement
3	3.6 String Built-in Functions Case study to explore 3.1 to 3.11

Online Resources

SQL for Data Science available on coursera https://www.coursera.org/learn/sql-for-data-science
SQL: A Practical Introduction for Querying Databases by IBM available on coursera https://www.coursera.org/learn/sql-practical-introduction-for-querying-databases
Advanced MySQL Topics available on coursera https://www.coursera.org/learn/advanced-mysql-topics

MS-DS-202 Machine Learning:

Unit	Content	No. of Lectures
1	Regression Introduction: Goals and applications of machine learning, the concept of learning, types of machine learning: supervised, unsupervised and reinforcement learning, framework for developing machine learning models, issues in machine learning. Classical Regression models: Identifying regression problems, prediction using regression model – simple and multivariate, model diagnostics – coefficient of determination, adjusted R-squared, hypothesis testing for the regression coefficient, ANOVA in regression analysis, residual analysis, Identifying significant features, multi- collinearity, Non-normality and	15

	Heteroscedasticity, Polynomial regression. Checking for linear regression assumptions. Case study regression.	
--	--	--

2	<p>Classifications</p> <p>Identifying classification problems, Logistic regression in classification, Logit function and interpretation, model evaluation metrics for classification: confusion matrix, accuracy, sensitivity, specificity, f1-score, ROC, AUC, dealing with imbalanced data, decision trees: ID4, C4.5, CART, Naïve Bayes classifier, k-nearest neighbour algorithm, support vector machine for classification and regression, multi-class classifications</p> <p>Case study Classifications.</p>	15
3	<p>Selection and improvement</p> <p>Concept of overfitting and underfitting, bias-variance trade off, applying regularizations, Cross-Validations, Grid search for optimal parameters, Feature Reduction using AIC, BIC, Dimensionality reduction, Principal components analysis.</p> <p>Ensemble learning: bagging and boosting, random forest, adaboost, gradient boosting, XGBoost.</p> <p>Case study Random forest.</p>	15

Sub Unit	Topics
1	Regression model – simple, Checking for linear regression assumptions. Case study regression.

2	Logit function and interpretation, Case study Classifications.
3	Case study Random forest.

Online Resources

Introduction to Machine Learning by Prof. S. Sarkar from IIT Kharagpur available on the Swayam portal https://nptel.ac.in/courses/106105152
Introduction to Machine Learning(Course sponsored by Aricent) by Dr. Balaraman Ravindran from IIT Madras available on the Swayam portal https://nptel.ac.in/courses/106106139
Introduction to Machine Learning: Supervised Learning available on coursera https://www.coursera.org/learn/introduction-to-machine-learning-supervised-learning

MS-DS-203 Advanced Statistical methods

Unit	Content	No. of Lectures
1	<p>Testing of hypothesis: Concept of hypothesis, Simple Hypothesis and composite hypothesis, Null and alternate hypothesis, Types of errors, Critical region, Level of significance.</p> <p>Large sample tests: Test of significance for specified value of mean of Normal population. Test of significance for difference between means of two independent Normal populations with equal variances and unequal variances.</p> <p>Exact tests using t distribution: Independents and Dependent samples (Paired t test)</p> <p>Applications of Chi-Square:</p>	15

	<p>Testing for association, Testing for variance, Testing for goodness of fit.</p> <p>Applications of F:</p> <p>Test procedure for testing equality of variances of two independent Normal populations</p> <p>i. Mean is known</p> <p>ii. Mean is unknown.</p>	
2	<p>Linear models: Linear parametric function and its estimability, Gauss Markoff theorem, Interval estimates and test of hypothesis, fundamental theorems on conditional error ss, Test of $\beta=d$, generalized least squares</p> <p>Analysis of variance, fixed effect models: (i) One – way classification model. (ii) Checking assumptions of ANOVA Model. (iii) Simultaneous Confidence Intervals: Scheffe’ s, Bonferroni and Turkey’ s interval.</p> <p>Two – way classification model with and without interaction effect, one observation per cell .Tukey’ s test for non-additivity.</p>	15
3	<p>Generalized Linear models:</p> <p>GLM for Binary data: Linear probability model, Logistic regression model and Probit regression model.</p> <p>GLM for Count data: Poissons regression, Negative Binomial regression.</p> <p>Model with constant coefficient of variation: Gamma Regression, Variance function, Canonical link, Multiplicative model- Log link and Linear model- Identity link.</p>	15

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	<p>Test of significance for difference between means of two independent Normal populations with equal variances and unequal variances.</p> <p>Test procedure for testing equality of variances of two independent Normal</p> <p>ii. Mean is unknown.</p>

2	Confidence Intervals: Scheffe' s, Bonferroni and Turkey' s interval.
3	Probit regression model, Negative Binomial regression.

Online Resources

1. ' Probability and Statistics' by Prof. Somesh Kumar from IIT Kharagpur available on the Swayam portal https://nptel.ac.in/courses/111/105/111105090/
2. ' Statistics for Business Economics' by Dr. Patel from University School of Sciences available on the Swayam portal http://ugcmoocs.inflibnet.ac.in/ugcmoocs/view_module_ug.php/227
3. ' Business Statistics' by Dr Mukesh Kumar Barua from IIT Roorkee available on the Swayam portal https://nptel.ac.in/courses/110/107/110107114/ for unit I and unit II.

MS-DS-204A Cloud Computing

(Total Hours: 45 Lectures)

Unit	Content	No. of Lectures
1	Networking Concepts & Distributed Systems 1.1 Introduction to networking 1.2 Classification of networks 1.3 Cloud Architecture 1.4 Cloud DevOps 1.5 Virtualization 1.6 Types of Operating systems 1.7 Hypervisors & Virtual Machines 1.8 Parallel Computing	15
2	Cloud Computing and Security 2.1 Cloud Service Models	15

	2.2 Cloud Deployment Models 2.3 Identity and Access Management	
3	Business demands and needs 3.1 Migrating to Cloud 3.2 Deployment of cloud based tools 3.3 Ease of cloud resource management 3.4 Financial projections w.r.t cloud 3.5 Comparison between CSPs (AWS, Azure, GCP Etc.)	15

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	1.2 Classification of networks 1.7 Hypervisors & Virtual Machines
3	3.5 Comparison between CSPs (AWS, Azure, GCP Etc.)

Online Resources

Cloud computing by Prof. Soumya Kanti Ghosh from IIT Kharagpur available on the Swayam portal https://nptel.ac.in/courses/106105167
Cloud Computing and Distributed Systems by Dr.Rajiv Misra from IIT Patna https://nptel.ac.in/courses/106104182

MS-DS-204B Financial technology

(Total Hours : 45 Lectures)

Unit	Content	No. of Lectures
1	Introduction to Fintech	15

	<p>Basic of Fintech</p> <p>Categories & Types of Fintech</p> <p>Fintech v/s Traditional Banking & Finance</p> <p>Technology Used in Fintech</p> <p>Rise of Fintech</p> <p>Emerging Sectors in Fintech</p> <p>Impact of Fintech</p> <p>Fintech Ecosystems</p> <p>Fintech Segmentation</p>	
2	<p>Working of Fintech</p> <p>AI and Machine Learning</p> <p>Blockchain Technology</p> <p>Introduction to Crypto Currency, Crypto Wallets</p> <p>Payment Gateways</p> <p>Algorithm Trading</p> <p>Data Analytics application in Fintech</p> <p>Valuation of a fintech organization at different stages</p>	15
3	<p>Sectors in Fintech</p> <p>Digital Lending</p> <p>Digital Banking</p> <p>Digital Identity</p> <p>Digital Wallet</p> <p>Asset Management</p> <p>Alternative Insurance Underwriting</p> <p>Alternative Credit Score</p> <p>Regulations in Fintech & Case Studies</p> <p>Regulations</p> <p>FinTech Regulations</p> <p>Global FinTech-enabling Regulations Database</p> <p>Evolving Regulations</p> <p>Regulatory and policy developments in the use and governance of blockchain and cryptocurrencies</p> <p>Case Studies</p> <p>M-Pesa</p>	15

	Goldman Sachs' Digital Journey The NEAT Account : Fintech Innovation in Hongkong Paytm : Building a Payment Network Binance Fidelity Investments Analysis : How Existing products can benefit from Fintech Trends	
--	---	--

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	Rise of Fintech Emerging Sectors in Fintech
2	Data Analytics application in Fintech Valuation of a fintech organization at different stages
3	Case Studies M-Pesa Goldman Sachs' Digital Journey

Online Resources

Introduction to Blockchain Technology and Applications by Prof. Sandeep Shukla from IIT Kanpur available on the Swayam portal https://nptel.ac.in/courses/106104220
Fintech: Foundations & Applications of Financial Technology Specialization available on coursera https://www.coursera.org/specializations/wharton-fintech

Detailed Scheme Practicals

Code Practical: MS-DS-P2

Title of Paper: Practicals of US-FDS-201, US-FDS-202, US-FDS-201-i

Title	Content	No. of
-------	---------	--------

of Pape r		Lectures
MS-D S 201	Practicals based on MS-DS-201	2 hours per Batch per Practical
MS-D S 202	Practicals based on MS-DS-202	2 hours per Batch per Practical
MS-D S 203	Practicals based on MS-DS-203	2 hours per Batch per Practica l
MS-D S-20 4A/ MS-D S-20 4B	Practical based on Ms-DS-204A or MS-DS-204B	2 hours per Batch per Practica l

MS-DS-201 References:

1. Raghu Ramakrishnan, J. G. (2003). Database Management Systems, . McGraw-Hill.
2. Ramez Elmasri, S. N. (2013, Sixth Edition). Database Systems. Pearson.
3. Tahaghogh, S. M. (2006). Learning MySQL: Get a Handle on Your Data. O'

Reilly.

MS-DS-202 References:

1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition
2. by Aurélien Géron.
3. MACHINE LEARNING FOR BUSINESS ANALYTICS, Concepts, Techniques, and Applications Second Edition, GALIT SHMUELI, PETER C. BRUCE, MIA L. STEPHENS, MURALIDHARA ANANDAMURTHY, NITIN R. PATEL. Wiley Publication.
4. An Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani.
5. Machine Learning using Python, By U Dinesh Kumar, Manaranjan Pradhan
6. Machine Learning and Deep Learning Using Python and TensorFlow, Venkata Reddy Konasani, Shailendra Kadre, McGraw Hill
7. Data Mining and Data Warehousing Principals and Practical techniques, Prateek Bhatia, Cambridge University Press.

MS-DS-203 References:

1. Medhi J. : Statistical Methods, An Introductory Text, Second Edition, New Age International Ltd.
2. Agarwal B.L. : Basic Statistics, New Age International Ltd.
3. Spiegel M.R. : Theory and Problems of Statistics, Schaum's Publications series. Tata McGraw Hill.
4. Kothari C.R. : Research Methodology, Wiley Eastern Limited.
5. David S. : Elementary Probability, Cambridge University Press.
6. Hoel P.G. : Introduction to Mathematical Statistics, Asia Publishing House.
7. Hogg R.V. and Tannis E.P. : Probability and Statistical Inference. McMillan Publishing Co. Inc.
8. Pitan Jim : Probability, Narosa Publishing House.
9. Goon A.M., Gupta M.K., Dasgupta B. : Fundamentals of Statistics, Volume II : The World Press Private Limited, Calcutta.
10. Gupta and Kapoor: Fundamentals of Applied Statistics, S. Chand
11. Gupta and Kapoor: Fundamentals of Mathematical Statistics, S. Chand

MS-DS-204A References:

1. Cloud Computing: Concepts, Technology & Architecture by Thomas Erl, Ricardo Puttini, Zaigham Mahmood (The Pearson Service Technology Series from Thomas Erl)
2. Infrastructure as Code: Dynamic Systems for the Cloud Age, by Kief Morris, Second Edition
3. Multi-Cloud Architecture and Governance: Leverage Azure, AWS, GCP, and VMware vSphere to build effective multi-cloud solutions by Jeroen Mulder
4. Distributed and Cloud Computing: From Parallel Processing to the Internet of

Things by Hwang

MS-DS-204B References:

1. Fintech and the Future of Finance: Market and Policy Implications by Erik Feyen, Harish Natarajan, Matthew Saal
2. Foundations in Fintech and Cryptocurrency by Shoba Premkumar