IIT Palakkad

UG Curriculum 2022

BTech Data Science and Engineering

Applicable to all from 2022-23 Academic Year

21 April, 2022

The UG Curriculum and Sample Template for Data Science and Engineering

Program : **Bachelor of Technology**

Department : **Data Science**

Year : **2022 Onwards**

The BTech program in Data Science and Engineering aims to impart the principles of analysis and design of building large-scale data driven decision-making systems that involves humans, machines, and the environment at large. The core courses in the curriculum build a strong foundation in mathematics, probability and statistics, computational thinking, data management, modeling and assessment, and ethics. The elective courses cut across various facets and applications of data science and engineering. The courses are designed to impart both the theoretical and practical aspects of the discipline. The table below lists the program core courses

List of program core courses:

Level 1
Introduction to Data Science and Engineering, Discrete Mathematics
Level 2
Introduction to Optimization, Data Structures and Algorithms for Data Science, Computer Systems for Data Science, Introduction to Artificial Intelligence
Level 3
Database Systems, Data Analytics, Machine Learning, Deep Learning, AI Ethics
Note: Level 1 courses require no prerequisites while Level 2 and 3 courses may require the prerequisites from Level 1 and 2 courses.

The tentative list of program elective courses includes (but not limited to): Signal Processing for Data Science, AI of Things, Applied Accelerated AI, Big Data Lab, Information Theory and Statistics, Reinforcement Learning, Probabilistic Machine Learning, Computer Vision, AI for Cybersecurity, Econometrics, Natural Language Processing, Information Retrieval, Foundations of Data Science and Machine Learning, Bioinformatics, Responsible AI. The humanities and social science and open elective courses provide students the opportunity to learn interdisciplinary courses.

Total Credits	
Institute Core (IC)	42
Program Major Core (PMC)	43
Program Major Elective (PME)	20
Humanities and Social Sciences Elective (HSE)	9
Sciences and Mathematics Elective (SME)	6
Open Elective (OE)	15
Project	9

BTech Data Science and Engineering (Sample Template 2022 Batch Onwards)

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	I	PH1030	Physics	Institute Core	2-1-0-3
2		MA1011	Linear Algebra and Series	Institute Core	3-1-0-4
3		ME1130	Engineering Drawing	Institute Core	1-0-3-3
4		ID1010	Ecology and Environment	Institute Core	2-0-0-2
5		ID1050A	Engineering Design	Institute Core	1-0-3-3
6		ME1150	Mechanical Workshop	Institute Core	0-0-3-2
7		PH1130/ CY1140	Physics/Chemistry Lab	Institute Core	0-0-3-2
			Total		**19**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	II	MA1021	Multivariable Calculus	Institute Core	3-1-0-4
2		CY1040	Basic Chemistry for Engineers	Institute Core	2-1-0-3
3		HS1010	Technology and Society	Institute Core	2-0-0-2
4		CE1020	Engineering Mechanics	Institute Core	3-1-0-4
5		ID1110	Introduction to Programming	Institute Core	2-0-3-4
6		EE1110	Electrical Workshop	Institute Core	0-0-3-2
7		PH1130/ CY1140	Physics/Chemistry Lab	Institute Core	0-0-3-2
			Total		**21**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	III	DS1010	Introduction to Data Science and Engineering	PMC	1-0-0-1
2		CS2020A	Discrete Mathematics	PMC	3-1-0-4
3		DS2010	Introduction to Optimization	PMC	3-0-0-3
4		DS2030	Data Structures and Algorithms for Data Science	PMC	3-0-3-5
5		BT2010	Life Sciences	Institute Core	2-0-0-2
6			Science and Mathematics Elective 1	SME	3
7			Humanities and Social Sciences Elective 1	HSE	3
			Total		**21**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	IV	DS2020	Introduction to Artificial Intelligence	PMC	3-0-2-4
2		DS2040	Computer Systems for Data Science	PMC	3-0-3-5
3		DS3020	Database Systems	PMC	3-0-3-5
4			Science and Mathematics Elective 2	SME	3
5			Open Elective 1	OE	3
			Total		**20**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	V	DS3010	Machine Learning	PMC	3-0-3-5
2		DS3030	Data Analytics	PMC	2-0-3-4
3			Program Major Elective 1	PME	3
4			Open Elective 2	OE	3
5			Humanities and Social Sciences Elective 2	HSE	3
			Total		**18**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	VI	DS3040	Deep Learning	PMC	3-0-3-5
2		DS3060	AI Ethics	PMC	2-0-0-2
3			Program Major Elective 2	PME	3
4			Program Major Elective 3[1]	PME	2
5			Humanities and Social Sciences Elective 3	HSE	3
6			Program Major Elective* (for Honors)	PME*	3
			Total		**15/18**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	VII		Program Major Elective 4	PME	3
2			Program Major Elective 5	PME	3

[1] The 3rd PME course of 2-credit is mentioned in the course plan assuming that all other PME courses are 3 credits each. In practice, as PME courses may have varied credits, the number of PME courses is not important and it is only required that a student obtains at least 20 credits in total.

3			Open Elective 3	OE	3
4			Open Elective 4	OE	3
5			Project I	Project	3
6			Program Major Elective* (for Honors)	PME*	3
			Total		**15/18**

Sl No.	Semester	Course Code	Course Title	Category	Credits
1	VIII		Program Major Elective 6	PME	3
			Program Major Elective 7	PME	3
2			Open Elective-5	OE	3
3			Project	Project	6
4			Program Major Elective* (for Honors)	PME*	6
			Total		**15/21**

Data Science and Engineering

Core course Syllabi

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	: Introduction to Data Science and Engineering
Course Code	: DS1010
Credit	: 1-0-0-1
Category	: Program Core
Target Program	: Undergraduate
Target Discipline	: Data Science and Engineering
Prerequisites	:
Data of proposal	: 9/02/2022
Data of approval	:
Proposing faculty	: All faculty of Department of Data Science

Course Content:

A walk through different real world applications drawn from the fields of computer vision, speech processing, natural language processing, bioinformatics, AI systems, etc., connecting it to the theory and engineering aspects of data science.

Learning Outcomes:

- Students will get familiar with real-world applications along with the data science and engineering principles and theory driving these successful applications.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New)

Course Title : Discrete Mathematics[1]

Course Code : CS2020A

Credit : 3-1-0-4 (L-T-P-C)

Category : Core

Target Programme : UG

Target Discipline : CSE

Prerequisite (if any) :

Date of proposal :

Date of approval :

Proposing faculty : Krishnamoorthy Dinesh & Deepak Rajendraprasad

Course Content:

Logic and set theory. Introduction to proofs, axiomatic method, proof patterns. Well ordering principle. Logical formulas - propositional and predicate. Sets, relations, equivalences and partial orders. Induction. Recursive definitions. Infinite Sets: Cantor's theorem, diagonalization argument, halting problem. Logic of sets. [*12 lectures + 4 tutorial sessions*]

Graph Theory. Graphs, degree. Common graphs. Walks, paths, connectivity, cycles, trees, forests. Cliques, Independent sets. Graph Isomorphism, bipartite graphs and matchings. Colouring. Planar graphs: Euler's formula, 6-colouring of planar graphs. Regular polyhedra. [*8 lectures + 3 tutorial sessions*]

Combinatorics. Product rule, division rule, counting subsets, sequences with repetitions - binomial and multinomial theorems. Pigeonhole principle. Inclusion-exclusion. Combinatorial proofs. Twelvefold way. Recurrence relations - Fibonacci numbers and Towers of Hanoi puzzle. [*8 lectures + 3 tutorial sessions*]

Discrete Probability. Events and probability spaces, The four step method, birthday principle. Set theory and probability. Conditional probability, law of total probability. Independence. Probability versus confidence. Random variables: independence. distribution functions, expectations. Linearity of Expectation [*14 lectures + 4 tutorial sessions*]

---

[1] Revision of CS2020 Discrete Mathematics for Computer Science and CS2010 Logic for Computing

Learning Outcomes:

1. Use logical notation to define and reason about fundamental mathematical concepts such as sets, relations, functions, and integers.
2. Evaluate elementary mathematical arguments and identify fallacious reasoning (not just fallacious conclusions).
3. Synthesize new proofs based on standard proof patterns.
4. Apply graph-theoretic models to solve problems like job allocation, scheduling, connectivity etc.
5. Calculate numbers of possible outcomes of elementary combinatorial processes such as permutations and combinations.
6. Calculate probabilities and discrete distributions for simple combinatorial processes; calculate expectations.

Textbook

1. *Mathematics for Computer Science*. Eric Lehman, F Thomson Leighton, Albert R Meyer. This book is available online under the terms of the Creative Commons Attribution ShareAlike 3.0 license. URL: https://courses.csail.mit.edu/6.042/spring18/mcs.pdf Printed version: 12th Media Services. ISBN-13: 978-1680921229.

Reference

1. *Discrete Mathematics and Applications*. Kenneth Rosen (7th Edition, 2012), McGraw-Hill Education (ISBN-13: 978-0073383095)
2. *Invitation to Discrete Mathematics*. Jiří Matoušek and Jaroslav Nešetřil (2nd Edition) Oxford University Press (ISBN-13: 978-0198570424)
3. *Discrete Mathematics: Elementary and Beyond*. László Lovász, József Pelikán, Katalin Vesztergombi, Springer 2003, ISBN-13: 978-0387955858.

Popular Readings

1. *Logicomix: An epic search for truth*. Apostolos Doxiadis and Christos Papadimitriou. Bloomsbury USA. ISBN-13: 978-1596914520
2. *What is the Name of This Book?: The Riddle of Dracula and Other Logical Puzzles*. Raymond M. Smullyan. Dover Publications. ISBN-13: 978-0486481982
3. *Proofs from THE BOOK*. Martin Aigner and Günter M. Ziegler. Springer. ISBN-13 : 978-3642008559
4. *Combinatorics: A Very Short Introduction*. Robin Wilson. Oxford University Press. ISBN-13 : 978-0198723493

Notes

1. This course is modelled along the Mathematics for Computer Science course (6.042J, Spring 2015) at MIT.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	:	Introduction to Optimization
Course Code	:	DS2010
Credit	:	3-0-0-3
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Introduction to Programming Python, Mathematics I - Linear Algebra, Mathematics II - Multivariate Calculus
Date of proposal	:	
Date of approval	:	
Proposing faculty	:	Dr. Sahely Bhadra

Course Content

1. Revision of prerequisite : Linear algebra and multivariate derivative [3 lectures]
2. Characterization of maxima and minima: Conditions of maxima or Minima for constraint and unconstrained problem [3 lectures]
3. Application of Linear programming and Simplex [3 lectures]
4. Major algorithms in unconstrained optimization [6 lectures]
    a. Newton, quasi-Newton,
    b. steepest descent, line search methods]
5. Application of constraint optimization [6 lectures]
6. Major algorithms in constrained optimization with application [e.g.,, barrier, penalty, augmented Lagrangian, interior point methods][6 lectures]
7. Large scale optimization [3 lectures]
8. Discrete Optimization (e.g. Combinatorial and integer programming)[3 lectures]
9. Major algorithms in derivative-free methods (e.g., simulated annealing, Bayesian optimization, Surrogate-assisted optimization)[6 lectures]

Learning Outcomes:

1. Students will be able to state the conditions of optimality and their meaning.
2. Students will have familiarity with some of the standard optimization packages.
3. Students will be able to implement important types of unconstrained and constrained optimization algorithms.

4. Students will also be able to apply appropriate optimization methods given a practical problem.

Textbooks:

1. **Practical Optimization** by Philip E. Gill, Walter Murray, Margaret H. Wright, *Emerald Group Publishing Limited* (1982). ISBN-13: 978-0122839528, ISBN-10: 0122839528
2. **Numerical Optimization**, J. Nocedal and S. Wright, *Springer Series in Operations Research and Financial Engineering, 2006*. ISBN-10 : 0387303030, ISBN-13 : 978-0387303031

Reference Books:

1. **Practical Methods of Optimization** by R. Fletcher 2nd edition, *Wiley, 1987*.ISBN-10 : 0471494631, ISBN-13 : 978-0471494638.
2. **Introduction to Linear Optimization** by Bertsimas, Tsitsiklis. *MIT Press (1997)*. ISBN-10 : 1886529191, ISBN-13 : 978-1886529199

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	: Data Structures and Algorithms for Data Science
Course Code	: DS2030
Credit	: 3-0-3-5
Category	: Program Core
Target Program	: Undergraduate
Target Discipline	: Data Science
Prerequisites	: Introduction to Programming
Data of proposal	: 10/01/2022
Data of approval	:
Proposing faculty	: Narayanan C Krishnan

Course Content:

1. Introduction to Object Oriented Programming, Arrays, and Linked Lists [3 lectures]
    a. Introduction to OOPS programming language, basic array and list manipulations- images as 3D arrays [6 lab hours]
2. Analysis Tools, Recursion [3 lectures]
    a. Recursion [3 lab hours]
3. Stacks and Queues [4 lectures]
    a. Breadth and depth first AI search [3 lab hours]
4. Tree Structures, Heaps, Sorting, Priority Queues, Divide and conquer strategy [6 lectures]
    a. Decision tree construction [6 lab hours]
5. Hashing functions, Bloom Filters [5 lectures]
    a. Data deduplication [3 lab hours]
6. Search Tree Structures [6 lectures]
    a. Mini-database using search tree structures [6 lab hours]
7. String processing, Dynamic Programming [6 lectures]
    a. Genome sequence matching using KMP Algorithm [6 lab hours]
8. Graphs, greedy algorithms (Graph Traversal Algorithms) [8 lectures]
    a. Social network analysis (node and network level summary statistics) [6 lab hours]

Learning Outcomes:

- Learn the fundamentals of data structures and algorithms- the building blocks of managing data efficiently.
- Gain hands-on programming experience through the implementation of various data structures for data science applications.

Textbooks:

Data Structures and Algorithms in Java, by Michael T Goodrich, Roberto Tamassia, and Michael H Goldwasser, 6th edition, Wiley 2010, ISBN: 978-1-118-77133

Reference Books:

Introduction to Algorithms by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, MIT Press, 2009, ISBN: 978-0262533058

Data Structures and Algorithms in C++, by Michael T Goodrich, Roberto Tamassia, and Michael H Goldwasser, edition and year, ISBN:xxx

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	: Introduction to Artificial Intelligence
Course Code	: DS2020
Credit	: 3-0-2-4
Category	: Program Core
Target Program	: UG
Target Discipline	: Data Science and Engineering
Prerequisites	: Introduction to Programming Discrete Mathematics
Data of proposal	: 10/01/2022
Data of approval	:
Proposing faculty	: Narayanan C Krishnan

Course Content:

1. Introduction to AI - Rational Agents - [2 lectures]
2. Search-(BFS, UCS, A* Search, Heuristics, Local Search) - [4 lectures, 6 lab hours]
3. Adversarial Search [3 lectures, 6 lab hours]
4. Knowledge Representation and logical inference (propositional logic, resolution, predicate logic, ontologies)[9 lectures]
5. Planning [4 lectures]
6. Probabilistic reasoning (probabilistic graphical model, exact and approximate inference)[5 lectures, 4 lab hours]
7. Markov Decision Processes (introduction to mdp) [6 lectures, 6 lab hours]
8. Decision making under uncertainty (introduction to PoMDPS, game theory, mechanism design, [3 lectures]
9. Reinforcement learning (introduction to rl) [6 lectures, 6 lab hours]

Learning Outcomes:

- Learn the fundamentals of field artificial intelligence. The students will be familiar with a broad range of topics in AI, ready to undertake specialized courses.
- Gain hands-on programming experience through the implementation of the AI techniques for various synthetic and real-world applications.

Textbooks:

Artificial Intelligence: A Modern Approach, Stuart Russell and Peter Norvig, Pearson, 2020. ISBN 978-0134610993

Reference Books:

Artificial intelligence: Kevin Knight, Elaine Rich, Shivashankar Nair, McGraw Hill, 3rd Edition 2017 ISBN 978-0070087705

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/Revised/MOOC)

Course Title	:	Computer Systems for Data Science
Course Code	:	DS2040
Credit	:	3-0-3-5
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Data Structures and Algorithms for Data Science
Date of proposal	:	11-Nov-21
Date of approval	:	
Proposing faculty	:	Dr. Satyajit Das

Course Content

1. ORGANIZATION OF COMPUTING SYSTEMS: Introduction, Instruction and Data representation, processing unit, memory and I/O subsystem (15 Lecture + 10 Labs)
2. SYSTEMS PROGRAMMING: Kernel, virtual memory, exceptions, processes, files, threads, scheduling, List potential threats to operating systems (e.g., software vulnerabilities, authentication, issues, malware) and the types of security features designed to guard against them (10 Lectures + 8 Labs)
3. PROGRAM EXECUTION: How programs are executed on a machine (with a particular focus on Linux-based operating systems); Program segments/sections; The ELF Format; Linking and loading; Linux dynamic libraries (shared objects); Multitasking and paging; Address translation; Memory Protection. (5 Lectures + 4 Labs)
4. ADDITIONAL TOPICS: Threads, virtualization, cloud computing, CUDA, security vulnerabilities and hardening. (10 Lectures + 8 Labs)

Learning Outcomes:

1. Develop a detailed understanding of computer systems from a programmer's point of view
2. Proficiency in writing cache friendly programs
3. Understand object and executable file formats, linking, system calls and memory protection schemes
4. Identify and describe a couple of the most basic security vulnerabilities

Textbooks:

1. Computer Systems: A Programmer's Perspective, Randal E. Bryant and David R. O'Hallaron

Reference Books:

1. Computer Organization and Design The Hardware/Software Interface: RISC-V Edition David A. Patterson, John L. Hennessy
2. The Elements of Computing Systems-Building a Modern Computer from First Principles: Noam Nisan and Shimon Schocken
3. C Programming Language, 2nd Edition, Brian W. Kernighan and Dennis M. Ritchie

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/Revised/MOOC)

Course Title	:	Database Systems
Course Code	:	DS3020
Credit	:	3-0-3-5 (L-T-P-C)
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Data Structures and Algorithms for Data Science
Date of proposal	:	13-Jan-22
Date of approval	:	
Proposing faculty	:	Dr. Koninika Pal

Course Content

1. Introduction: Database applications and purpose, Data abstraction and manipulation, Relational Databases, Database schema, Keys, Relational query languages, algebra, tuple and domain calculus. [6 lecture hours]
    - [Lab] Creating Databases schema, defining data and relations. [1 Week]
2. SQL: Data definition, basic SQL query structure, set operations, nested subqueries, aggregation, null values, database modification, join expressions, views, Integrity constraints in SQL, triggers. [5 lecture hours]
    - [Lab] Implementation of different manipulations of data using different SQL operators, introducing events in databases [3 Weeks]
3. Database Design: Entity-relationship model, reduction to relational schema, E-R design issues, Relational Database Design: Features of good design, Functional dependency theory, Decomposition using functional dependency, Normal forms (1NF, 2NF, 3NF, BCNF), Algorithms for decomposition. [5 lecture hours]
4. Storage Management and Indexing: Overview of secondary storage, Storing tables, Index concept, clustered and nonclustered indices, B+-tree indices, hash indices, bitmap indices. [4 lecture hours]
    - [Lab] Comparing effects of different in-built index methods in DBMS on benchmark datasets. [1 Week]
5. Query processing and optimization: Evaluation of relational algebra expressions, query equivalence, join strategies, query optimization algorithms. [4 lecture hours]
    - [Lab] Comparing effects of different join strategies in DBMS on benchmark datasets. [2 Weeks]

6. Transactions: Concept, ACID properties, Transaction as SQL statements. [4 lecture hours]
    - [Lab] Implementing transactions with DBMS. [1 Week]
7. Concurrency control: Lock-based protocols, 2-phase locking, Deadlock handling, Multiple granularities, Timestamp-based protocols, Multi-version protocols, Concurrency control for index structures. [4 lecture hours]
8. Recovery: Failure Classification, Recovery and atomicity, Recovery algorithms, Buffer management. [4 lecture hours]
9. Introduction to modern databases: Different types of NoSQL databases, their advantages, and disadvantages. [6 lecture hours]
    - [Lab] Introducing MongoDB and its basic features. [1 Week]
10. [Lab] A small DBMS project– sketching user requirements, database design and implementation, testing different DBMS functionalities, web interface. [3 Weeks]

Learning Outcomes:

1. Understanding how to organize data in a structured format using data models that provide easy and efficient accessibility of the data.
2. Acquiring the fundamental concepts of designing and maintaining a database.
3. Learning how to define and optimize a database query language.
4. To be able to create a database according to the requirement of an application.
5. Can access and manipulate the data in databases using a query language.

Textbooks:

1. A Silberschatz, H Korth, and S Sudarshan, Database System Concepts, 7th Ed., McGraw-Hill, 2020.
2. H Garcia-Molina, JD Ullman and Widom, Database Systems: The Complete Book, 2nd Ed., Prentice-Hall, 2008.

Reference Books:

1. R Elmasri, S Navathe, Fundamentals of Database Systems, 7th edition, Pearson, 2016. online
2. R Ramakrishnan, J Gehrke, Database Management Systems, 3rd Ed., McGraw-Hill, 2003. online

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	:	Machine Learning
Course Code	:	DS3010
Credit	:	3-0-3-5 (L-T-P-C)
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Introduction to Optimization Probability and Statistics
Date of proposal	:	
Date of approval	:	
Proposing faculty	:	Dr. Sahely Bhadra

Course Content

1. Introduction to the course, recap of linear algebra(vector derivative) and probability theory (Bayes Rule) basics. [3 lecture hours]
2. Regression: linear regression, ridge regression [3 lectures hours, 3 lab hours]
3. Classifier[9 lab hours]:
    a. Linear classification (e.g. perceptron, maximum margin, logistic regression)[4 lecture hours]
    b. Non linear classification(e.g. KNN,use of kernel in SVM)[5 lecture hours]
4. Evaluation and Model Selection: ROC Curves, Evaluation Measures, Cross validation, Significance tests. [6 lecture hours, 3 lab hours]
5. Ensemble Methods: Boosting, Bagging, Decision Trees, Random Forests.[6 lectures, 6 lab hours]
6. Feature extraction(Principal Component Analysis, Canonical Correlation Analysis)[3 lectures, 3 lab hours]
7. Clustering: K-mean, Gaussian Mixture Model, Expectation Maximization, density based clustering. [6 lectures hours, 3 lab hours]
8. Sequential Learning(HMM) [6 lectures, 3 lab hours]

Learning Outcomes:

1. Upon successful completion, the students will be able to implement machine learning models
2. Students will be able to learn models from data and evaluate models on test data for

various tasks.

Textbooks:

1. Machine Learning.Tom Mitchell.*Mcgraw-Hill,1997* ISBN-10: 0070428077,ISBN-13: 978-0070428072.

2. Introduction to Machine Learning, Alpaydin Ethem PHI Learning Pvt. Ltd.; Third edition 2015, ISBN-10 : 8120350782 ISBN-13 : 978-8120350786.

3. An Introduction to Statistical Learning: with Applications in R Gareth James , Daniela Witten, Trevor Hastie, Robert Tibshirani Springer Texts in Statistics. ISBN-10 : 1461471370, ISBN-13 : 978-1461471370.

Reference Books:

1. Pattern Classification. Richard Duda, Peter Hart, David Stork, 2nd Ed., *John Wiley & Sons, 2001*. ISBN-10 : 9814126020 ISBN-13 : 978-9814126021.
2. **Understanding Machine Learning: From Theory to Algorithms**, Shai Shalev-Shwartz, and Shai Ben-David, *Cambridge University Press, 2014*. ISBN-10 : 1107057132, ISBN-13 : 978-1107057135.
3. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics)**. Trevor Hastie, Robert Tibshirani, Jerome Friedman. *Published by Springer, 2003*. ISBN 10: 0387952845ISBN 13: 9780387952840.
4. Sam Roweis's probability review
5. Sam Roweis's linear algebra review

Note: Changes from Previous version DS5001

1. After running this course as a PG core instructor realised that more lecture hours are required for Ensemble Method and HMM, hence lecture hours for these topics have been increased in this modified syllabu.
2. To support the above, the neural network has been removed. Please note that this topic is also covered in other core courses DS5002 Deep Learning.
3. Also few sub topics from Clustering have been removed to accommodate extra lectures. Removed portion also covered in Data Engineering(PG) and Data analytics(UG)

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/Revised/MOOC)

Course Title	:	Data Analytics
Course Code	:	DS3030
Credit	:	2-0-3-4 (L-T-P-C)
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Data structures and algorithms for Data Science, Linear algebra, Probability and Statistics, Artificial Intelligence
Date of proposal	:	13-Jan-22
Date of approval	:	
Proposing faculty	:	Mrinal Das

Course Content

1. Introduction: data definition, different types of data, data collection, data storage, data management, data driven decision making and system development [2L]
    a. [lab] Introduction to Python programming for data management, pandas, sciPy, numPy [1Week]
    b. [lab] Introduction to R, R Studio [2Week]
2. Descriptive and Inferential statistics : Measures of central tendency, Measures of location of dispersions, analysis of variance, Correlation analysis [2L]
    a. [lab] above topics [1Week]
3. Data cleaning: noise removal, outlier detection, missing value handling, feature engineering with cases of graph and texts, feature selection, normalization, standardization [4L]
    a. [lab] above topics [1Week]
4. Data transformation, dimensionality reduction (PCA, tSNE, auto-encoder), regression [4L]
    a. [lab] above topics [1Week]
5. Association rule mining, Pattern recognition using k-means clustering, hierarchical clustering, bi-clustering, density based clustering [4L]
    a. [lab] above topics [2Week]

6. Data visualization: table, graph, histogram, pie-chart, area-plot, box-plot, scatter-plot, bubble-plot, waffle charts, word clouds [2L]
    a. [lab] matplotlib [2Week]
    b. [lab] R, GGPLOT [2Week]

7. Introduction to Experimental Design, Basic Analysis Techniques: Statistical hypothesis generation and testing, Chi-Square test, t-test, Maximum likelihood test. [4L]

8. Case study: IPL/Covid/Traffic [4L] [1Week]

Learning Outcomes:

1. Understanding methodologies for finding meaningful data patterns.
2. Getting hands on experience to analyse the data and interpret graphically.
3. Developing data driven decision making systems

Textbooks:

An Introduction to Statistical Learning: with Applications in R. G James, D Witten, R Tibshirani, T Hastie. Springer, ISBN: 1461471370

The Art of Statistics: Learning from Data. David Spiegelhalter. Pelican books. ISBN: 0241258766

Reference Books:

1. R for Data Science. Garrett Grolemund and Hadley Wickham. O'Reilly.
2. Exploratory data analysis. J W Tukey. ADDISON-WESLEY. ISBN 0201076160.
3. Hastie, Trevor, et al. The elements of statistical learning. Vol. 2. No. 1. New York: springer, 2009.
4. Pattern Classification. Richard Duda, Peter Hart, David Stork, 2nd Ed., *John Wiley & Sons, 2001*. ISBN-10 : 9814126020 ISBN-13 : 978-9814126021.
5. Montgomery, Douglas C., and George C. Runger. Applied statistics and probability for engineers. John Wiley & Sons, 2010

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	:	Deep Learning
Course Code	:	DS3040
Credit	:	3-0-3-5 (L-T-P-C)
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Introduction to Optimization Machine Learning
Date of proposal	:	30-Jan-22
Date of approval	:	
Proposing faculty	:	Mrinal Kanti Das

Course content:

1. Revision of Linear Algebra, Probability, Optimization [2L]
    a. [lab] introduction to computation aspects like GPU, Colab, Tensorflow [1Week]
2. Basics of learning: model, weight, data, loss functions, learning, prediction, output functions, connection with gradients and optimization, linear and non-linear classifiers [4L]
    a. [lab] classification using logistic regression, implementation of cross validation, experimentation for classification, preprocessing and visualization [1Week]
3. Perceptron: model, algorithm, basic neuron, XOR [3L]
    a. [lab] implementation of perceptron from scratch [1Week]
4. Feed forward neural network (FNN)
    a. Multi-layer perceptrons, weights, learning problem [1L]
        i. [lab] implementation of MLP/FNN using tensorflow [1Week]
    b. Forward propagation [1L]
    c. Backpropagation [3L]
        i. [lab] implementation of FNN using PyTorch [1Week]
    d. Regularization methods, dropout, batchnorm, data augmentation [3L]
        i. [lab] implementation of dropout, batchnorm and verification [1Week]
5. Convolutional neural network (CNN)
    a. Convolution, filters, padding, pooling,[2L]
        i. [lab] image classification [1Week]
    b. CNN for image classification [2L]

    c. Example architectures, various applications [2L]
        i. [lab] segmentation, object detection [1Week]
6. Recurrent neural network (RNN)
    a. Recurrence, various types of RNNs for different applications [3L]
        i. [lab] implementation of RNN, sequence/text classification [1Week]
    b. Drawbacks of RNNs, Long Short Term Memory networks (LSTM), GRU [2L]
        i. [lab] implementation of LSTM [1Week]
    c. Attention and memory networks [1L]
7. Auto-encoders
    a. Encoder-decoder, seq2seq case study, dimensionality reduction, denoising [3L]
        i. [lab] implement auto-encoders apply on dimensionality reduction [1Week]
    b. Representation learning [2L]
        i. [lab] word2vec, LSTM representations for text classification or information retrieval [1Week]
8. Generative adversarial network (GAN), WGAN, conditional-GAN [3L]
    a. [lab] GAN on MNIST, FMNIST
9. Transformers [3L]

Learning Outcomes:

Students will be able to i) design and code programs for solving simple computational tasks involving deep learning, ii) use tools and libraries to develop deep learning models, iii) analyse and present outcomes related to applications.

Textbooks:

Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville. MIT Press. 2016.ISBN-13**:** 978-0262035613.

Dive into Deep Learning. https://d2l.ai/

Reference Books:

Pattern Recognition and Machine Learning. Christopher Bishop. Springer. 2006. ISBN-13 978-0-387-31073-2.

Deep Learning with Python. Francois Chollet. Publisher: Manning Publications; 1 edition. ISBN-13: 978-1617294433

Hands-On Machine Learning with Scikit-Learn and TensorFlow. Aurélien Géron. Publisher: O'Reilly Media; 1 edition. ISBN-13: 978-1491962299.

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD

Proforma for proposing course (New/~~Revised/MOOC~~)

Course Title	:	AI Ethics
Course Code	:	DS3060
Credit	:	2-0-0-2 (L-T-P-C)
Category	:	Program Core
Target Programme	:	UG
Target Discipline	:	Data Science and Engineering
Prerequisites (if any)	:	Introduction to Artificial Intelligence Introduction to Machine Learning
Date of proposal	:	
Date of approval	:	
Proposing faculty	:	Dr. Sahely Bhadra

Course Content

1. What is AI Ethics [2 lectures]
2. Potential Harms Caused by AI Systems [3 lectures]
3. Individual and social impact of ML and AI [2 lectures]
4. Support, Underwrite, and Motivate(SUM Values): Respect, Connect, Care, and Protect [3 lectures]
5. The FAST Track Principles:
    a. Fairness : data, design, outcome implementation [3 lectures]
    b. Accountability: answerability and auditability [3 lectures]
    c. Sustainability:
        i. Stakeholder Impact Assessment, accuracy, [2 lecture]
        ii. reliability, security, and robustness, [2 lecture]
        iii. Privacy [2 lecture]
        iv. Data poisoning, adversarial attack [3 lecture]
    d. Transparency: interpretable AI [3 lecture]

Learning Outcomes:

1. Upon successful completion, the students will be able to judge Ethical aspect of AI model
2. Students will be able to state an Ethical AI model.
3. Student will be aware of AI Ethics related regulations of some countries.

Textbooks:

1. **AI Ethics** Mark Coeckelbergh. *The MIT Press Essential Knowledge series 2020*.ISBN-10 : 0262538199, ISBN-13 : 978-0262538190

Reference Books:

1. **Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector**, David Leslie, SSRN Electronic Journal 2019. DOI:10.2139/ssrn.3403301