Silicon Institute of Technology
| An Autonomous Institute |

Curriculum Structure and Detailed Syllabus

Master of Science (Data Science)
(Two-Year Post-Graduate Program)

Silicon
...beyond teaching

Department of Computer Science & Engineering
Silicon Institute of Technology
Silicon Hills, Patia, Bhubaneswar - 751024

Effective from Academic Year 2020-21
Build: 1.20 (18-10-2021)

Approval History

ACM#	Date	Resolutions
AC-4	18/08/2020	The curriculum structure & detailed syllabus of 1st Year, as proposed by the Board of Studies, is approved by the Academic Council.
AC-6	09/10/2021	The curriculum structure & detailed syllabus of 2nd Year, as proposed by the Board of Studies, is approved by the Academic Council.

Program Outcomes

Graduates Attributes (GAs) form a set of individually assessable outcomes that are the components indicative of the graduate's potential to acquire competence to practice at the appropriate level. The National Board of Accreditation (NBA) has defined Program Outcomes (POs) for UG Engineering programs only. Silicon Institute of Technology has defined the POs for the M.Sc. (Data Science) program in line with NBA, so that the outcomes can be assessed in a similar manner to UG programs. The Program Outcomes for M.Sc. (Data Science) program are given below:

PO1. Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.

PO2. Develop efficient applications to analyze data and make predictions for taking timely business decisions.

PO3. Design solutions for analysis of huge data with considerations towards societal and environmental aspects.

PO4. Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.

PO5. Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.

PO6. Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.

PO7. Work with a professional context pertaining to ethics, social, cultural and cyber regulations.

PO8. Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.

PO9. Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.

PO10. Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Program Specific Outcomes (PSOs)

PSO1. Understand & learn the concepts of Data Science in diverse fields dealing with large amount of stored or streaming data, and obtain useful inferences.

PSO2. Implement & apply appropriate algorithms, techniques, and strategies for developing applications and tools for analysis of large volume of data for decision support systems.

PSO3. Utilize cutting-edge technologies, programming languages, environments, tools & platforms leading to a rewarding career and a zest for entrepreneurship or higher studies.

Program Educational Objectives (PEOs)

PEO1. To build successful career based on concepts of programming, software and design principles using various methods of Data Science.

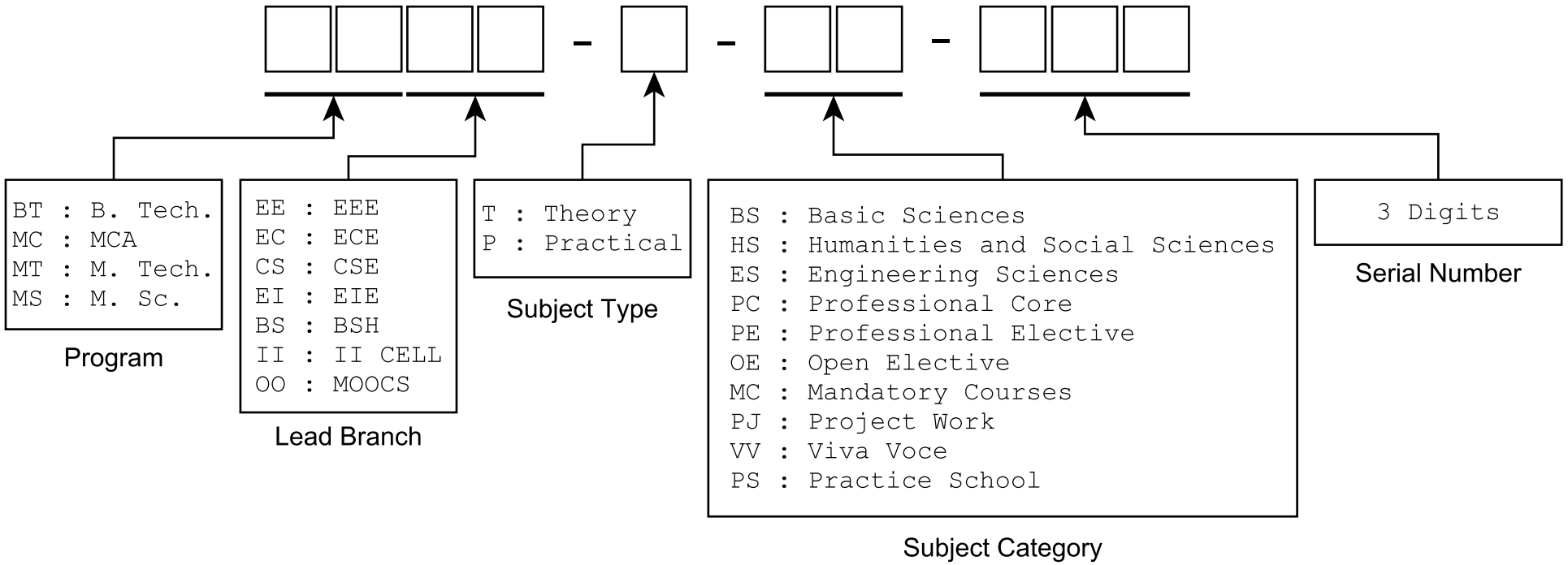
PEO2. To work independently or in a diverse team with effective communication in interdisciplinary environment, and demonstrate leadership in industry and academia.

PEO3. To engage in lifelong learning and career development through analysis, discussion, professional studies, literature study, and continued research.

Course Types & Definitions

L	Lecture
T	Tutorial
P	Practical / Sessional
WCH	Weekly Contact Hours
BS	Basic Sciences
HS	Humanities & Social Sciences (including Management)
ES	Engineering Sciences
PC	Professional Core
PE	Professional Elective
OE	Open Elective
MC	Mandatory Course
PJ	Internship / Project Work
VV	Viva Voce

Subject Code Format



Contents

Part I

1st Year M. Sc. (Data Science)

Curriculum Structure

Semester I									
Type	**Code**	**Course Title**	**WCH** L-T-P			**Credits** L-T-P		
THEORY									
PC	MSBS-T-PC-001	Inferential Statistics	3	0	0	3	0	0
PC	MSBS-T-PC-002	Computational Linear Algebra	3	0	0	3	0	0
PC	MSCS-T-PC-001	Data Structures & Algorithms	3	0	0	3	0	0
PC	MSCS-T-PC-002	Programming for Data Science	3	0	0	3	0	0
PC	MSCS-T-PC-003	Data Mining & Exploration	3	0	0	3	0	0
PRACTICAL									
PC	MSBS-P-PC-003	Inferential Statistics Lab	0	0	2	0	0	1
PC	MSCS-P-PC-004	Data Structures & Algorithms Lab	0	0	4	0	0	2
PC	MSCS-P-PC-005	Programming for Data Science Lab	0	0	4	0	0	2
		SUB-TOTAL	**15**	**0**	**10**	**15**	**0**	**5**
		TOTAL	**25**			**20**		

Semester II									
Type	**Code**	**Course Title**	**WCH** L-T-P			**Credits** L-T-P		
THEORY									
PC	MSBS-T-PC-004	Graph Theory & Stochastic Processes	3	0	0	3	0	0
PC	MSBS-T-PC-005	Optimization Techniques	3	0	0	3	0	0
PC	MSCS-T-PC-006	Scalable Database Systems	3	1	0	3	1	0
PC	MSCS-T-PC-007	Machine Learning	3	1	0	3	1	0
PC	MSCS-T-PC-008	Artificial Intelligence	3	0	0	3	0	0
PRACTICAL									
PC	MSCS-P-PC-009	Scalable Database Systems Lab	0	0	4	0	0	2
PC	MSCS-P-PC-010	Machine Learning Lab	0	0	4	0	0	2
PC	MSCS-P-PC-011	Artificial Intelligence Lab	0	0	2	0	0	1
		SUB-TOTAL	**15**	**2**	**10**	**15**	**2**	**5**
		TOTAL	**27**			**22**		

Type	Code	Inferential Statistics	L-T-P	Credits	Marks
PC	MSBS-T-PC-001		3-0-0	3	100

Objectives	The objective this course is exercise statistical thinking in designing data collection, derive insights from visualizing data, obtain supporting evidence for data-based decisions and construct models for predicting future trends from data. Additionally, this course prepares the foundation to recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
Pre-Requisites	Basic UG level knowledge of probability and statistics is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	**Axioms of Probability**: Sample space and events, axioms of probability, some simple proposition, sample spaces having equally likely outcomes. **Conditional Probability & Independence**: Conditional probabilities, Bayes' formula, independent events. **Random Variables**: Random variables, discrete random variables, expected value, expectation of function of random variable, variance, Bernoulli and binomial random variables, Poisson random variable, properties of cumulative distribution function.	**8 Hours**
Module-2	**Continuous Random Variables**: Expectation and variance of continuous random variables, uniform random variable, Normal random variables, exponential random variables, distribution of a function of a random variable. **Properties of Expectation**: Expectation of sums of random variables, covariance, variance of sums and correlations, conditional expectation, conditional expectation and prediction, Moment generating function. **Distributions Derived from the Normal Distribution**: χ^2, t, and F distributions, The sample mean and the sample variance.	**8 Hours**
Module-3	**Survey Sampling**: Population parameters, simple random sampling (The expectation and variance of the sample mean, estimation of the population variance, The normal approximation to the sampling distribution of $\bar{X}$), estimation of a ratio. **Estimation of Parameters & Fitting of Probability Distributions**: Fitting the Poisson distribution, parameter estimation, the method of moments, and maximum likelihood (Large sample theory for maximum likelihood estimates, confidence intervals from maximum likelihood estimates), the Bayesian approach to parameter estimation(large sample normal approximation to the posterior, computational aspects).	**9 Hours**

Cont'd...

Module-#	Topics	Hours
Module-4	**Testing Hypotheses & Assessing Goodness of Fit**: The Neyman-Person paradigm (specification of the significance level and the concept of a p-value, The null hypothesis, uniformly most powerful tests), the duality of confidence intervals and hypothesis tests, generalized likelihood ratio test, probability plots, tests for normality; Large scale hypothesis testing and false discovery rates. **Comparing Two Samples**: Comparing two independent sample (methods based on the normal distribution, power, a non-parametric method - the Mann-Whitney test, Bayesian approach), comparing paired samples (methods based on the normal distribution, The signed rank test).	**9 Hours**
Module-5	**The Analysis of Variance**: The one-way layout (normal theory, F test, problem of multiple comparisons, Kruskal Wallis test). **The Analysis of Categorical Data**: Fisher's exact test, the Chi-square test of homogeneity and independence, matched pairs designs, odds ratios.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. S. Ross, ***A First Course in Probability***, 8th Edition, Pearson Education, 2010.
T2. J. A. Rice, ***Mathematical Statistics and Data Analytics***, 3rd Edition, Cengage Learning, 2013.

Reference Books:

R1. L. Wasserman, ***All of Statistics : A Concise Course in Statistical Inference***, Springer, 2004.
R2. B. Efron and T. Hastie, ***Computer Age Statistical Inference : Algorithms, Evidence, and Data Science***, 1st Edition, Cambridge University Press, 2016.

Online Resources:

1. https://nptel.ac.in/courses/111/105/111105043/: By Prof. S. Kumar, IIT Kharagpur
2. https://nptel.ac.in/courses/111/102/111102112/: By Prof. N. Chaterjee, IIT Delhi

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Apply sampling distributions such as χ^2, t, and F distribution in real life problems.
CO2	Estimation the parameters and fitting of probability distributions.
CO3	Apply methods of tests of hypothesis and goodness of fit.
CO4	Conduct hypothesis tests, make decisions using p-value, and draw appropriate conclusions.
CO5	Analyze categorical data, formulate and use linear regression for the given data sets.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.

Cont'd...

PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	3		2	2				1	3	2	2
CO2	3	2	3		2	2			1	1	3	2	2
CO3	3	3	3	1	2	2			1	1	3	2	2
CO4	3	3	3	1	2	2			1	1	3	3	2
CO5	3	3	3	1	2	2			1	1	2	3	2

Type	Code	Computational Linear Algebra	L-T-P	Credits	Marks
PC	MSBS-T-PC-002		3-0-0	3	100

Objectives	The objective of this course is to study linear algebra along with different computational methods to handle large linear systems and large scale matrices.
Pre-Requisites	Basic concepts of system of linear equations and Matrix Algebra are required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Vector Space & Subspaces, Solving $AX = 0$ & $AX = b$, Linear Independence, Basis and Dimension, The Four Fundamental Subspaces, Linear Transformation, Orthogonal Vectors, Projections to a line, Projections and Least Squares, Orthogonal Bases & Gram-Schmidt.	**10 Hours**
Module-2	Eigenvalues and eigenvectors, Diagonalization of a Matrix, Complex Matrices, Similarity Transformation, Test for positive definiteness, Singular Value Decomposition.	**8 Hours**
Module-3	Errors in Computations, Computing Norm, Inner product and solution of Triangular System, Efficiency and stability of an Algorithm, Conditioning, Perturbation Analysis, Perturbation Analysis of linear system.	**7 Hours**
Module-4	LU Factorization Methods, Scaling, Effects of the condition number on accuracy, computing and estimating the condition number, Householder's matrices and QR factorization, Classical and Modified Gram-Schmidt Algorithm for QR factorization, Solution of AX = b using QR Factorization, Projections Using QR Factorization, SVD and its computation.	**9 Hours**
Module-5	Existence and uniqueness of least square solutions, Pseudoinverse and the least square problem, sensitivity of the least square problem, Computational Methods for Over determined Problems, Computing selected eigenvalues and eigenvectors, Jacobi, Gauss-Seidel and SOR methods.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. G. Strang, ***Linear Algebra and Its Applications***, 4th Edition, Cengage Learning, 2006.
T2. B. N. Datta, ***Numerical Linear Algebra and Applications***, 2nd Edition, PHI Learning, 2012.

Reference Books:

R1. J. W. Demmel, ***Applied Numerical Linear Algebra***, 1st Edition, University Press, 1997.
R2. G. H. Golub and C. F. Van Loan, ***Matrix Computations***, 4th Edition, Hindustan Book Agency, 2015.

Online Resources:

1. https://nptel.ac.in/courses/111/107/111107106/: By Prof. D. N Pandey & Prof. P. N. Agrawal, IIT Roorkee
2. https://nptel.ac.in/courses/111/108/111108066/: By Prof. V. Rao, IISc Bangalore

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Understand the geometry of spaces associated with a matrix and apply them in computing.
CO2	Use eigenvalues and eigenvectors of a matrix to factorize it.
CO3	Analyze the error and stability in matrix computations.
CO4	Apply different factorization techniques of matrices to solve linear systems.
CO5	Compute egen values and eigenvectors and solve over determined systems.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	1	1							3	3	2
CO2	3	2	1	1							3	3	2
CO3	3	3	3	2						1	2	2	2
CO4	3	3	3	3						1	2	2	2
CO5	3	3	3	3						1	2	2	2

Type	Code	Data Structures & Algorithms	L-T-P	Credits	Marks
PC	MSCS-T-PC-001		3-0-0	3	100

Objectives	The objective of this course is to introduce the abstract data types, classic algorithms in various domains, techniques for designing efficient algorithms, use various data structures and apply the algorithm design techniques to solve real life problems.
Pre-Requisites	Basic concepts and knowledge of a programming language are required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to data structures, classification of data structures, complexity of algorithms, growth of functions, asymptotic notations, Recurrences, Solving recurrences using Master Method, abstract data types, Arrays: introduction, representation of arrays (row and column major), basic operations on array (traverse, insert, delete, search), sparse matrix, representation of sparse matrix using triplet form, operation on sparse matrix (addition, transpose).	**9 Hours**
Module-2	Stack: stack model, representation using array, basic operations, and applications; Queue: queue model, representation using array, basic operations, circular queue; Linked List: introduction, types of linked list, representation in memory, operations on linked list (traverse, search, insert, delete), Representation of polynomial and its operations (addition, multiplication), implementation of stack and queue using linked list.	**9 Hours**
Module-3	Sorting Algorithms: Bubble sort, Selection sort, Insertion sort; Tree: terminology, representation, Binary tree: traversal algorithms, Binary search tree, Operations on Binary search tree, Height balanced tree; Divide and conquer strategy for designing algorithms, Merge Sort, Quick Sort; Heaps, Types of Heap, Maintaining the heap property, Building a Heap, The Heap-sort algorithm, Priority Queue.	**8 Hours**
Module-4	Dynamic Programming, Elements of dynamic programming, Longest Common Subsequence; Greedy algorithms, Elements of Greedy strategy, Fractional Knapsack problem, Huffman codes; String matching algorithms (Naive, Rabin-Karp).	**8 Hours**
Module-5	Graphs: terminology, representation, graph traversal (BFS, DFS), Minimum spanning trees, Kruskal's algorithm, Dijkstra's algorithm, Warshall's algorithm; Introduction to NP completeness (Polynomial time, Polynomial time verification, NP completeness & reducibility), Examples of NP complete problems (without proof); Introduction to Approximation algorithms.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. M. Weiss, ***Data Structures and Algorithm Analysis in C***, 2nd Edition, Pearson Education, 2002.
T2. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, ***Introduction to Algorithms***, 3rd Edition, PHI Learning, 2014.

Reference Books:

R1. E. Horowitz, S. Sahni, and S. Anderson-Freed, ***Fundamentals of Data Structures in C***, 2nd Edition, Universities Press, 2008.
R2. A. Tenenbaum, ***Data Structures Using C***, 3rd Edition, Pearson Education, 2007.
R3. S. Lipchitz, ***Data Structures***, 1st Edition, Tata McGraw-Hill, 2005.
R4. J. Kleinberg and E. Tardos, ***Algorithm Design***, 1st Edition, Pearson Education, 2013.

Online Resources:

1. https://nptel.ac.in/courses/106/102/106102064/: By Prof. N. Garg, IIT Delhi
2. https://nptel.ac.in/courses/106/106/106106127/: By Prof. H. A. Murthy, Prof. S. Balachandran, and Dr. N. S. Narayanaswamy, IIT Madras
3. https://nptel.ac.in/courses/106/105/106105085/: By Dr. P. P. Chakraborty, IIT Kharagpur
4. https://nptel.ac.in/courses/106/106/106106131/: By Prof. M. Mukund, IIT Madras

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Analyze performance of algorithms and implement various operations on array and sparse matrix.
CO2	Apply the basic operations of stack,queue,and linked list to solve real world problems.
CO3	Compare various comparison based sorting algorithms and understand their advantages and limitations.
CO4	Develop solutions for a given optimization problem using dynamic programming and greedy algorithm.
CO5	Represent data using graphs to solve various real life problems and understand NP-Complete problems.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2		3					2	3	3	2
CO2	3	3	2		1					2	3	3	2
CO3	3	3	3		2					2	3	3	2
CO4	3	3	2		2					2	3	2	2
CO5	3	3	2		1					2	2	2	2

Type	Code	Programming for Data Science	L-T-P	Credits	Marks
PC	MSCS-T-PC-002		3-0-0	3	100

Objectives	The objective of this course is to develop programming skills in Python which rich in tools & libraries used for solving real-life Data Science problems.
Pre-Requisites	Basics of programming, algorithms and problem solving skills are required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned with focus on programming & problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Getting Started with Python: Introduction, Basic data types, Variables, Integers, Floating Points, Boolean types and Strings; Control Structures: if, if-elif-else, for, while, break, continue.	**9 Hours**
Module-2	Data Structures: Lists, Tuples, Sets, and Dictionaries; Functions: Defining functions, Calling functions, Passing arguments, Keyword arguments, Default arguments, Variable-length arguments, Anonymous functions, Function returning values, Scope of the variables in a function - global & local variables, User defined functions.	**8 Hours**
Module-3	Object Oriented Programming: Features, classes and objects, creating class and object, Using a class & its methods; Exception Handling: Errors, Types of exception, try, except and finally, assertion.	**7 Hours**
Module-4	Modules & Packages: Creating modules, Import statement, from ... import statement, name spacing; Creating user defined packages; Numpy: Introduction, Creating of arrays and matrices; File Handling: Handling of csv file.	**8 Hours**
Module-5	Introduction to Panda: Creating a data frame, Dealing with row & columns, Indexing & selection data, Working with missing data, Iterating over rows and columns; Merging and joining DataFrame objects, Concatenation, Reshaping DataFrame objects, Pivoting, Data transformation, permutation & sampling, Data aggregation and GroupBy operations; Creating data frame from CSV file; Introduction to scikit-learn: Fundamental of scikit-learn; Loading data set, Splitting of data set; Matplotlib: Creating effective visual representations of your data, Avoiding common pitfalls.	**10 Hours**
	Total	**42 Hours**

Text Books:

T1. R. N. Rao, ***Core Python Programming***, 2nd Edition, Dreamtech Press, 2018.
T2. J. V. Guttag, ***Introduction to Computation and Programming Using Python, with Application to Understanding Data***, 2nd Edition, PHI Learning, 2016.
T3. W. McKinney, ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython***, 2nd Edition, O'Reilly Media, 2017.

Reference Books:

R1. P. Barry, ***Head First Python***, 2nd Edition, O'Reilly Media, 2010.
R2. A. Downey, ***Think Python : How to Think Like a Computer Scientist***, 2nd Edition, Green Tea Press, 2015.
R3. J. Zelle, ***Python Programming : An Introduction To Computer Science***, 3rd Edition, Franklin, Beedle & Associates, 2016.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106182/: By Prof. S. Iyengar, IIT Ropar
2. https://nptel.ac.in/courses/106/106/106106145/: By Prof. M. Mukund, IIT Madras
3. https://nptel.ac.in/courses/106/106/106106212/: By Prof. R. Rengasamy, IIT Madras
4. https://nptel.ac.in/courses/106/107/106107220/: By Prof. A. Ramesh, IIT Roorkee

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Compile and debug basic python programs, and solve problems using control structures.
CO2	Apply the data structure for real life problems and design modular python programs.
CO3	Design object oriented programs and handle various types of run-time exceptions.
CO4	Create user-defined modules & packages and use the predefined modules appropriately.
CO5	Create DataFrame from CSV file and split the same into training & testing sets to prepare them for application of various data science techniques.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	2		1	1				1	2	3	1
CO2	3	3	3	1	2	2				1	2	3	2
CO3	3	3	3	1	2	2				1	2	3	1
CO4	3	3	3	1	2	2			1	1	2	3	1
CO5	3	3	3	1	2	2			1	1	2	3	2

Type	Code	Data Mining & Exploration	L-T-P	Credits	Marks
PC	MSCS-T-PC-003		3-0-0	3	100

Objectives	The objective of this course is to study the fundamentals of data mining, understand the need for analysis of large, complex, and information-rich data sets, analyse & use various data mining algorithms, and explore different graphical methods for data exploration.
Pre-Requisites	Knowledge of probability & statistics and algorithms is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Data Mining Basics: Introduction, Application areas in data mining, KDD process; Getting to Know Your Data: Data objects and attributes types; Data Pre-processing: Why pre-process data? Data cleaning, Data integration, Data transformation and reduction.	**8 Hours**
Module-2	Graphical Methods for Data Mining & Exploration: Histograms, Boxplots, Quantile plots, Bagplots, Glyph plots, Scatterplots, Dynamic graphics, Coplots, Dot charts, Plotting points as curves, Biplots.	**8 Hours**
Module-3	Mining Frequent Patterns: Introduction to Associations & Correlations, Market-basket analysis, Frequent item-set generation using Apriori algorithm, Rule generation; Alternative methods for Generating frequent item sets using FP-Growth algorithm, Evaluation of association patterns; From association analysis to correlation analysis.	**8 Hours**
Module-4	Classification: Introduction, Naïve Bayes Classifier, Decision Tree Induction, Nearest Neighbor Classifier; Classification model evaluation techniques, Techniques to improve classification accuracy: Bagging, Boosting, Handling the class imbalance problem.	**10 Hours**
Module-5	Clustering: Overview, K-Means, K Medoid, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-based Clustering, Graph-based Clustering, Scalable Clustering Algorithms; Visualizing Clusters: Dendogram, Treemaps, Rectangle Plots, Data image.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. J. Han, M. Kamber, and J. Pei, ***Data Mining Concepts and Techniques***, 3rd Edition, Elsevier, 2011.
T2. W. L. Martinez, A. R. Martinez, and J. L. Solka, ***Exploratory Data Analyis with Matlab***, 2nd Edition, CRC Press (Taylor & Francis Group), 2010.

Reference Books:

R1. C. Bishop, ***Pattern Recognition and Machine Learning***, 1st Edition, Springer, 2007.
R2. G. James, D. Witten, T. Hastie, and R. Tibshirani, ***An Introduction to Statistical Learning with Applications in R***, Springer 2013.

Online Resources:

1. https://nptel.ac.in/courses/106/105/106105174/: By Prof. P. Mitra, IIT Kharagpur
2. http://infolab.stanford.edu/~ullman/mining/2003.html: Lecture Notes and Resources by Prof. J. D. Ullman, Stanford University.

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the basic concepts & techniques of data mining.
CO2	Explore the different graphical methods of data mining & exploration.
CO3	Generate frequent patterns, derive association rules, and perform correlation analysis.
CO4	Analyze and apply different classification algorithms on real-life data.
CO5	Analyze and apply different clustering algorithms on real-life data.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	1							1	3	1	
CO2	3	3	1							1	3	2	1
CO3	3	3	1							1	3	3	
CO4	3	3	1							1	3	3	1
CO5	3	3	1							1	3	3	1

Type	Code	Inferential Statistics Lab	L-T-P	Credits	Marks
PC	MSBS-P-PC-003		0-0-2	1	100

Objectives	The objective of this course is to give the students hands-on exposure to statistical programming using R language and analyze the given sample data for inferring meaningful information for whole population.
Pre-Requisites	Knowledge of Statistics, Numerical analysis, and basic programming skills in a programming language like C/C++/Java/MATLAB etc., are required.
Teaching Scheme	Regular laboratory experiments executed by the students under supervision of the teacher. Experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1, 2	Introduction to R Programming and writing simple programs using R.
3, 4	Fitting data (one or more dimensions) using different distribution and visualizing.
5	Parameter estimation using maximum likelihood method.
6	Parameter estimation using Bayesian approach.
7, 8	Formulation of null hypothesis, computation of confidence interval and p-value.
9	Likelihood ratio test for multinomial distribution and obtaining probability plots.
10	Drawing boxplots, scatterplots, etc., and estimating measures of central tendency and dispersion.
11	Comparing two samples of data sets using various techniques.
12	Use of F-test and analysis of variance for multiple comparisons.
13	Analysis of categorical data using Fisher's test.
14	Fitting of data set to linear regression and interpret the coefficients.

Text Books:

T1. S. Ross, ***A First Course in Probability***, 8th Edition, Pearson Education, 2010.
T2. J. A. Rice, ***Mathematical Statistics and Data Analytics***, 3rd Edition, Cengage Learning, 2013.

Reference Books:

R1. N. Matloff, ***The Art of R Programming - A Tour of Statistical Software Design***, 1st Edition, No Starch Press, 2011.
R2. L. Wasserman, ***All of Statistics : A Concise Course in Statistical Inference***, Springer, 2004.
R3. B. Efron and T. Hastie, ***Computer Age Statistical Inference : Algorithms, Evidence, and Data Science***, 1st Edition, Cambridge University Press, 2016.

Online Resources:

1. https://www.coursera.org/learn/statistical-inference: Statistical inference course by Brian Caffo, PhD, Professor, Biostatistics, Johns Hopkins University.
2. https://daviddalpiaz.github.io/appliedstats/applied_statistics.pdf: Applied Statistics with R e-Book

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Study and use the normal distribution & distributions derived from it using R language.
CO2	Estimate the parameters of probability distributions using maximum likelihood and Bayesian approach.
CO3	Develop understanding of Testing Hypotheses and use it for examining the validity of inferences obtained from computational models.
CO4	Develop understanding of Comparing of Two Samples and analysis of variance and use them for computing the performance of different models.
CO5	Learn methodology to analyze categorical data and linear regression along with statistical signification of the coefficient in the regression equation.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	2	2	1					2	2	3	2
CO2	2	2	2	2	1					2	3	3	1
CO3	3	3	2	2	1					2	2	3	2
CO4	3	3	2	2	1					2	2	3	2
CO5	3	3	2	2	1					2	2	3	1

Type	Code	Data Structures & Algorithms Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-004		0-0-4	2	100

Objectives	The objective of this laboratory course is to provide practical exposure on how to use various data structures efficiently, with emphasis on design & implementation of efficient algorithms for specific real world applications.
Pre-Requisites	Knowledge of programming language is required. The experiments shall go along with the subjects taught in the theory class.
Teaching Scheme	Regular laboratory classes conducted under supervision of the teacher. The experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Operation on array – insert, delete, merge.
2	Linear Search and Binary search.
3, 4	Representation of sparse matrix, addition and transpose of sparse matrix.
5	Implementation of stack using array.
6	Conversion of infix to postfix expression.
7	Evaluation of postfix expression.
8	Operations of queue using array.
9	Operations of circular queue.
10	Single linked list operations.
11	Double linked list operations.
12	Stack using linked list.
13	Queue using linked list.
14	Selection Sort, Bubble sort.
15	Binary Search Tree operations.
16, 17	Quick Sort, Merge Sort, and Heap Sort.
18	Priority Queue using min-Heap
19	Longest Common Subsequence
20	Fractional Knapsack Problem
21	Rabin-Karp String matching algorithm
22, 23	Graph Traversal using BFS and DFS.
24	Kruskal's Algorithm for Minimum Spanning Tree.
25	Dijkstra's Single source shortest path algorithm.
26	Warshall's all pair shortest path algorithm.
27, 28	Mini Project

Text Books:

T1. M. Weiss, ***Data Structures and Algorithm Analysis in C***, 2nd Edition, Pearson Education, 2002.
T2. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, ***Introduction to Algorithms***, 3rd Edition, PHI Learning, 2014.

Reference Books:

R1. A. K. Rath and A. K. Jagadev, ***Data Structures Using C***, 2nd Edition, Scitech Publication, 2010.
R2. Y. Kanetkar, ***Data Structures Through C***, 2nd Edition, BPB Publication, 2010.

Online Resources:

1. https://nptel.ac.in/courses/106/102/106102064/: By Prof. N. Garg, IIT Delhi
2. https://nptel.ac.in/courses/106/106/106106127/: By Prof. H. A. Murthy, Prof. S. Balachandran, and Dr. N. S. Narayanaswamy, IIT Madras
3. https://nptel.ac.in/courses/106/105/106105085/: By Dr. P. P. Chakraborty, IIT Kharagpur
4. https://nptel.ac.in/courses/106/106/106106131/: By Prof. M. Mukund, IIT Madras
5. https://nptel.ac.in/courses/106/101/106101060/: By Prof. S. Viswanathan, Prof. A. A. Diwan, and Prof. A. G. Ranade, IIT Bombay

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Implement various operations on arrays and sparse matrices.
CO2	Design functions to implement basic operations on stack, queue, and linked list.
CO3	Construct binary search tree and perform traversal, insertion, deletion, and search operations on it.
CO4	Apply dynamic programming and greedy paradigms to solve real life problems.
CO5	Formulate engineering problems and solve them using graph algorithms.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	2		1					2	2	3	2
CO2	3	3	3		3					2	3	3	2
CO3	2	3	2		2					2	3	3	2
CO4	3	3	3		3					2	2	3	2
CO5	3	3	2		2					2	2	3	2

Type	Code	Programming for Data Science Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-005		0-0-4	2	100

Objectives	The objective of this laboratory course is to develop problem solving skills using python programming language to prepare the students solve data science problems using python.
Pre-Requisites	Knowledge of programming and basic problem solving skills are required.
Teaching Scheme	Regular laboratory classes conducted under supervision of the teacher. The experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Write, compile, test, and debug simple Python programs.
2	Write programs using control structures (if, if-elif-else).
3, 4	Write programs using loop control structure (while & for loops).
5	Solve mathematical problems (sin(x), cos(x) etc.) using Taylor's series expansion.
6	Write program based on the concept of lists and tuples.
7	Write program based on the concept of set and dictionaries.
8, 9	Develop the Python programs step-wise by defining functions and calling them.
10	Write simple program on user defined function.
11, 12	Write programs using built-in functions, control flow, and parameter passing.
13	Write programs using function with variable number of parameters.
14, 15	Write programs using object oriented programming and exception handling.
16	Write programs using predefined modules, create user defined module.
17	Write program using packages and user-defined package.
18	Write larger programs using files, exception, modules and packages.
19	Introduction to NumPy, solving problems using NumPy.
20	Program on CSV, file handling, solve problems on some real-life data sets.
21, 22	Introduction to the Panda module, creating data frame, data frame from CSV file, reshaping & filtering.
23	Develop programs on Strings and experiment with immutable nature of strings.
24	Introduction to scikit-learn module and simple programs using scikit-learn.
25	Plotting using Matplotlib.
26	Write programs various searching and sorting techniques using python.
27, 28	Mini Project.

Text Books:

T1. R. N. Rao, ***Core Python Programming***, 2nd Edition, Dreamtech Press, 2018.

T2. J. V. Guttag, ***Introduction to Computation and Programming Using Python, with Application to Understanding Data***, 2nd Edition, PHI Learning, 2016.
T3. W. McKinney, ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython***, 2nd Edition, O'Reilly Media, 2017.

Reference Books:

R1. P. Barry, ***Head First Python***, 2nd Edition, O'Reilly Media, 2010.
R2. A. Downey, ***Think Python : How to Think Like a Computer Scientist***, 2nd Edition, Green Tea Press, 2015.
R3. J. Zelle, ***Python Programming : An Introduction To Computer Science***, 3rd Edition, Franklin, Beedle & Associates, 2016.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106182/: By Prof. S. Iyengar, IIT Ropar
2. https://nptel.ac.in/courses/106/106/106106145/: By Prof. M. Mukund, IIT Madras
3. https://nptel.ac.in/courses/106/106/106106212/: By Prof. R. Rengasamy, IIT Madras
4. https://nptel.ac.in/courses/106/107/106107220/: By Prof. A. Ramesh, IIT Roorkee

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Develop programs using various features of python programming language.
CO2	Develop programs using built-in as well as user-defined functions in python.
CO3	Apply object oriented concepts, modules, packages, file & exception handling.
CO4	Explore NumPy and Panda modules of python for solving real-life problems.
CO5	Solve basic data science problems using scikit-learn and matplotlib libraries.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	1	2	2		1	1					2	3	1
CO2	3	3	3		1	1					2	3	1
CO3	3	3	3		2	1				1	2	2	1
CO4	3	3	3		3	2				1	2	2	2
CO5	3	3	3		3	2				1	2	2	2

Type	Code	Graph Theory & Stochastic Processes	L-T-P	Credits	Marks
PC	MSBS-T-PC-004		3-0-0	3	100

Objectives	The objective of this course is to study graph theory and stochastic processes and their applications to various real-life data science problems.
Pre-Requisites	Knowledge of elementary probability theory and matrix algebra is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Graphs, Degrees, Graph Isomorphism, Trees, Cut-Vertices.	**8 Hours**
Module-2	Eulerian graphs and Hamiltonian graphs, Matchings, Planar graphs, Vertex coloring.	**9 Hours**
Module-3	Joint distribution, Independent random variables, Covariance and Correlation coefficient, Variance-Covariance matrix, Conditional distribution and conditional expectation.	**9 Hours**
Module-4	Multivariate normal distribution, The weak law of large numbers, The strong law of large numbers, Central limit theorem, Stochastic processes - definitions and properties.	**7 Hours**
Module-5	Discrete-Time Markov Chain, Classification of states, Measure of stationary probability, Continuous-Time Markov Chains, Poisson Process.	**9 Hours**
	Total	**42 Hours**

Text Books:

T1. G. Chartrand and P. Zahang, ***Introduction to Graph Theory***, 1st Edition, Tata McGraw-Hill, 2006.
T2. L. B. Castañeda, V. Arunachalam, and S. Dharmaraja, ***Introduction to Probability and Stochastic Processes with Applications***, 1st Edition, Wiley-Blackwell, 2012.

Reference Books:

R1. D. B. West, ***Introduction to Graph Theory***, 2nd Edition, Pearson Education, 2002.
R2. S. M. Ross, ***Introduction to Probability Models***, 9th Edition, Academic Press, 2006.

Online Resources:

1. https://nptel.ac.in/courses/106/108/106108054/: By Dr. L. S. Chandran, IISc Bangalore
2. https://nptel.ac.in/courses/111/106/111106050/: By Prof. S. A. Choudum, IIT Madras
3. https://nptel.ac.in/courses/111/106/111106102/: By Dr. S. Maity, IISER Pune
4. https://nptel.ac.in/courses/111/102/111102014/: By Dr. S. Dharmaraja, IIT Delhi
5. https://nptel.ac.in/courses/110/104/110104024/: By Dr. R. N. Sengupta, IIT Kanpur
6. https://nptel.ac.in/courses/111/103/111103022/: By Dr. S. Dharmaraja and Dr. N. Selvaraju, IIT Guwahati

P.T.O

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Use graphs and trees to express and analyze analytical processes.
CO2	Use special types of graphs in computational and design activities.
CO3	Explore the relationship between multiple random variables.
CO4	Understand the laws of large numbers and their uses in real-life problems.
CO5	Model processes using Markov Models and apply them to data science.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	1	1						1	2	2	1
CO2	3	2	1	1						1	2	2	1
CO3	2	3	3	1						1	2	1	1
CO4	2	3	3	2						1	3	1	1
CO5	3	3	3	1						1	3	3	2

Type	Code	Optimization Techniques	L-T-P	Credits	Marks
PC	MSBS-T-PC-005		3-0-0	3	100

Objectives	The objective of this course is to learn different techniques for solving optimization problems using both linear and nonlinear programming and apply them in data science.
Pre-Requisites	Basic concepts of multivariable calculus and matrix algebra are required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Types of Optimization Problems, Linear Programming, Simplex Method, Artificial Variables, Matrix Form, Revised Simplex Method.	**8 Hours**
Module-2	Duality, The Duality Theorem, Sensitivity, Dual Simplex Method, Matrix Calculus, Conditions for Solution of an Unconstrained Problem.	**9 Hours**
Module-3	Fibonacci and Golden Section Search, Steepest Descent, Newton's Method, Conjugate Directions, Conjugate Direction Method, Conjugate Gradient Method, Modified Newton Method, Construction of Inverse, Davidon-Fletcher-Powel Method.	**10 Hours**
Module-4	Constrained Minimization Conditions, Gradient Projection Method, Penalty Methods, Barrier Methods, Sub-gradient Method.	**8 Hours**
Module-5	Primal-Dual Methods, Genetic Algorithm.	**7 Hours**
	Total	**42 Hours**

Text Books:

T1. D. G. Luenberger and Y. Ye, ***Linear and Nonlinear Programming***, 3rd Edition, Springer, 2008.
T2. M. C. Joshi and K. M. Moudgalya, ***Optimization Theory & Practice***, Narosa Publishing, 2013.

Reference Books:

R1. K. Dev, ***Optimization for Engineering Design : Algorithms and Examples***, 2nd Edition, PHI Learning, 2012.
R2. J. Nocedal and S. J. Wright, ***Numerical Optimization***, 2nd Edition, Springer, 2006.
R3. S. Boyd and L. Vandenberghe, ***Convex Optimization***, 1st Edition, Cambridge University Press, 2015.

Online Resources:

1. https://nptel.ac.in/courses/111/104/111104071/: By Dr. J. Dutta, IIT Kanpur
2. https://nptel.ac.in/courses/106/108/106108056/: By Dr. S. K. Shevade, IISc Bangalore
3. https://nptel.ac.in/courses/111/105/111105039/: By Prof. A. Goswami and Dr. D. Chakraborty, IIT Kharagpur
4. https://nptel.ac.in/courses/111/105/111105100/: By Dr. D. Chakraborty and Prof. A. Goswami, IIT Kharagpur
5. https://nptel.ac.in/courses/111/104/111104068/: By Dr. J. Dutta, IIT Kanpur

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Apply Simplex Method and its variations to solve linear programming problems.
CO2	Apply the concepts of duality in a linear programming problem and understand the conditions for an unconstrained optimization problem.
CO3	Apply various computational methods to solve unconstrained optimization problems.
CO4	Apply various computational methods to solve constrained optimization problems.
CO5	Apply Prime-dual methods and evolutionary algorithms to solve optimization problems.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	3		1					1	2	1	1
CO2	2	3	3		2					1	3	1	2
CO3	3	3	3		2					1	2	2	2
CO4	3	3	3		2					1	2	2	2
CO5	3	3	3		2					1	2	2	2

Type	Code	Scalable Database Systems	L-T-P	Credits	Marks
PC	MSCS-T-PC-006		3-1-0	4	100

Objectives	The objective of the course is to revisit relational databases and study advanced scalable database systems for managing large amounts of structured, semi-structured and complex data for various data science applications.
Pre-Requisites	Basic knowledge of data structures and algorithms is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required. Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Basic concepts & definitions, Three-schema architecture, Data independence, Data models, Database languages, ER model, Constraints & keys, Relational model, Mapping ER model to relational schema, Relational algebra, Basic operations, Joins operations, Grouping & aggregation, Modification of database.	**11 Hours**
Module-2	Database design, Functional dependencies, Armstrong axioms, Attribute closure, Normalization, Dependency & attribute preservation, Lossless join, Normal forms (1NF, 2NF, 3NF, BCNF), Storage strategies & architecture, File and record organization, Types of indexes, B-Tree, B+ Tree, Index files, Hashing.	**11 Hours**
Module-3	Query processing, Evaluation of relational algebra expressions, Query optimization; Transaction processing, ACID properties, Serializability, Concurrency control – Lock & Timestamp-based protocols, Deadlocks - prevention, detection & recovery, Database recovery, Types of failures, Log-based recovery, Checkpoints.	**11 Hours**
Module-4	Parallel Databases - Introduction, Parallelism in Databases, Distributed database systems, Reference architecture, Fragmentation, Allocation, Replication, Distribution transparency, Distributed database design, Distributed query processing, Distributed transactions, 2-Phase commit protocol, Distributed concurrency control & deadlock handling.	**11 Hours**
Module-5	Concepts of NoSQL, Why NoSQL, Aggregate data models (key-value & document data models, column-family stores), Data modeling details – Relationships, Graph databases, Schemaless databases, Materialized views, Modeling for data access, Distribution Models – single server, shrading, replication, Consistency, Relaxing consistency & durability, Version stamps, Map-Reduce.	**12 Hours**
	Total	**56 Hours**

Text Books:

T1. A. Silberschatz, H. F. Korth, and S. Sudarshan, ***Database System Concepts***, 6th Edition, McGraw-Hill Education, 2013.

T2. S. Ceri and G. Pellagatti, ***Distributed Databases: Principles and Systems***, 1st Edition, McGraw-Hill Education, 2017.
T3. P. J. Sadalage and M. Fowler, ***NoSQL Distilled***, 1st Edition, Pearson Education, 2012.

Reference Books:

R1. R. Elmasri and S. B. Navathe, ***Fundamentals of Database Systems***, 7th Edition, Pearson Education, 2016.
R2. R. P. Mahapatra and G. Verma, ***Database Management Systems***, 1st Edition, Khanna Publishing, 2016.
R3. M. T. Özsu and P. Valduriez, ***Principles of Distributed Database Systems***, 2nd Edition, Pearson Education, 2006.
R4. D. Sullivan, ***NoSQL for Mere Mortals***, 1st Edition, Addison Wesley, 2015.

Online Resources:

1. https://nptel.ac.in/courses/106104135/: By Prof. A. Bhattacharya, IIT Kanpur
2. https://nptel.ac.in/courses/106105175/: By Prof. P. P. Das, IIT Kharagpur
3. https://nosql-database.org/: Resources for NoSQL

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain concepts of various data models and write queries using relational algebra.
CO2	Design normalized relational databases and implement appropriate indexing.
CO3	Understand query optimization, transactions, concurrency, and recovery in RDBMS.
CO4	Visualize design & working principles of distributed databases for enterprise applications.
CO5	Explore NoSQL databases for storage, manipulation, and analysis of non-relational data.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2		2	1				1	2	3	2
CO2	3	3	3		2	1				1	3	3	3
CO3	3	3	3		2	1			1	1	3	3	3
CO4	3	3	3		3	2			2	2	3	3	3
CO5	3	3	3		3	2			2	3	3	3	3

Type	Code	Machine Learning	L-T-P	Credits	Marks
PC	MSCS-T-PC-007		3-1-0	4	100

Objectives	The objective of this course is to learn patterns and concepts from data using various machine learning techniques focusing on recent advances. Students will explore supervised and unsupervised learning paradigms, deep learning technique and various feature extraction strategies.
Pre-Requisites	Knowledge of algorithms, optimization, and matrix theory is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required. Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple linear and polynomial regression, Ridge regression, Least angle regression and Lasso, Elastic Net, Linear Discriminant Analysis, Logistic regression, Bayes decision theory and applications.	**11 Hours**
Module-2	Dimensionality Reduction: Principal Components, Kernel PCA, Independent Component Analysis, LLE, Feature Selection, Matrix factorization and Collaborative filtering.	**10 Hours**
Module-3	Model Assessment and Selection: Bias, Variance, and Model Complexity, Bias-variance trade-off, Optimism of the training error rate, Estimate of in-sample prediction error, Bayesian approach and BIC, Cross-validation, Bootstrap methods.	**11 Hours**
Module-4	Neural Networks: Model of a neuron, LMS, Perceptron and its learning algorithm, MLP and Back Propagation algorithm, Heuristics for improving performance of BPA, Higher order convergence methods for BPA (Newton method, Conjugradient method, LM, BFGS); Radial Basis Function Networks, Self-Organizing Maps.	**12 Hours**
Module-5	Support Vector Machines (SVM) and Others: SVM for classification and Novelty detection (1-class classification), Reproducing Kernels, SVM for Regression, Decision Tree for Regression, Random Forests, Ada-boost, Gradient boosting, EM algorithm and Gaussian Mixture model and application to clustering and outlier detection.	**12 Hours**
	Total	**56 Hours**

Text Books:

T1. T. Hastie, R. Tibshirani, and J. Friedman, ***The Elements of Statistical Learning : Data Mining, Inference and Prediction***, 2nd Edition, Springer Verlag, 2009.

T2. S. Haykin, ***Neural Networks : A Comprehensive Foundation***, 2nd Edition, Pearson Education, 1999.

Reference Books:

R1. C. Bishop, ***Pattern Recognition and Machine Learning***, 1st Edition, Springer, 2007.
R2. T. Mitchel, ***Machine Learning***, 1st Edition, McGraw-Hill Education, 1997.
R3. G. James, D. Witten, T. Hastie, and R. Tibshirani, ***An Introduction to Statistical Learning with Applications in R***, 7th Edition, Springer, 2013.
R4. K. P. Murphy, ***Machine learning : A Probabilistic Perspective***, 4th Edition, MIT Press, 2012.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106202/: Prof. C. G. Jansson, IIT Madras
2. https://nptel.ac.in/courses/106/105/106105152/: By Prof. S. Sarkar, IIT Kharagpur
3. https://github.com/josephmisiti/awesome-machine-learning: An exhaustive index of machine learning concepts and programming materials.
4. http://mlss.cc/: Machine Learning Summer School Study Material

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Formulate and solve machine learning problems using linear models of regression and classification.
CO2	Develop understanding of unsupervised learning models of dimensionality reduction and factor analysis.
CO3	Analyze the building blocks of probabilistic model assessment and selection.
CO4	Understand neural networks and their applications to real-world problems.
CO5	Apply the tools in cluster analysis, support vector machines and K-nearest neighbors.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2		1	2			1	1	3	3	2
CO2	3	3	2		1	2			1	1	3	3	2
CO3	3	2	2		1	2			1	1	3	2	2
CO4	3	2	2		2	2			1	1	3	3	2
CO5	3	2	2		2	2			1	1	3	3	2

Type	Code	Artificial Intelligence	L-T-P	Credits	Marks
PC	MSCS-T-PC-008		3-0-0	3	100

Objectives	The objective of the course is to study the basics of Artificial Intelligence (AI), problem solving techniques, methods of knowledge representation and applications of AI in various information processing applications.
Pre-Requisites	Knowledge of algorithms and data structures is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Artificial Intelligence: Introduction to AI; Intelligent Agents: Agents and Environment, Good Behavior, The nature of Environments, The structure of agents; Problem-solving: Solving Problems by Searching: Problem-Solving Agents, Example Problems, Searching for Solutions, Uninformed search strategies, Searching with Partial Information.	**8 Hours**
Module-2	Informed Search and Exploration: Informed (Heuristic) Search Strategies, Heuristic Functions, Local Search Algorithms and Optimization Problems; Constraint Satisfaction Problems: Constraint Satisfaction Problems, Backtracking search for CSPs, Local Search for Constraint Satisfaction Problems; Adversarial Search: Games, Optimal Decisions in Games, Alpha-Beta Pruning; Knowledge and Reasoning: Knowledge-Based Agents, The Wumpus World.	**10 Hours**
Module-3	Knowledge and Reasoning: Logic, Propositional Logic, Reasoning Patterns in Propositional Logic; First-Order Logic: Syntax and Semantics of First-Order Logic, Using FOL, Knowledge Engineering in FOL; Inference in FOL: Propositional vs. First-Order Logic, Unification and Lifting, Forward Chaining, Backward Chaining, Resolution.	**8 Hours**
Module-4	Planning: The Planning Problem,Planning with State-Space Search, Partial-Order Planning, Planning Graphs; Uncertain Knowledge and Reasoning: Acting under Uncertainty, Bayes Rule and its use; Probabilistic Reasoning: Representing Knowledge in an Uncertain Domain. The Semantics of Bayesian Networks.	**8 Hours**
Module-5	Learning: Learning from Observations: Forms of Learning, Inductive Learning, Learning Decision Trees; Statistical Learning, Instance Based Learning, Neural Networks; Reinforcement Learning: Passive and Active Reinforcement Learning.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. S. J. Russell and P. Norvig, ***Artificial Intelligence - A Modern Approach***, 3rd Edition, Pearson Education, 2016.

T2. D. W. Patterson, ***Introduction to Artificial Intelligence & Expert Systems***, Pearson Education, 2015.

Reference Books:

R1. E. Rich, K. Knight, and S. B. Nair, ***Artificial Intelligence***, 3rd Edition, McGraw Hill, 2017.
R2. G. F. Luger, ***Artificial Intelligence***, 5th Edition, Pearson Education, 2009.
R3. M. Negnevitsky, ***Artificial Intelligence: A Guide to Intelligent Systems***, 2nd Edition, Pearson Education, 2008.
R4. N. J. Nilson, ***Principles of Artificial Intelligence***, 1st Edition, Narosa, 2002.
R5. E. Charniak and D. McDermott, ***Introduction to Artificial Intelligence***, 1st Edition, Addison-Wesley, 1985.

Online Resources:

1. https://nptel.ac.in/courses/106105077/: by Prof. S. Sarkar & Prof. A. Basu, IIT Kharagpur
2. https://nptel.ac.in/courses/106105079/: by Prof. P. Mitra, IIT Kharagpur
3. https://nptel.ac.in/courses/106106140/: by Prof. D. Khemani, IIT Madras

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explore agents and working environments with utilization of uninformed techniques in state space search.
CO2	Apply search techniques for Game playing and solving CSP problems.
CO3	Interpret Logic, Inference rules for decision making and represent knowledge using semantic nets, frames.
CO4	Apply Planning and Reasoning to handle uncertainty in real life.
CO5	Use Learning to solve complex real life problems in science, engineering and business.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	3		1				1	1	2	1	1
CO2	3	3	2		1				2	1	3	3	1
CO3	3	2	2		1				2	1	3	2	1
CO4	3	2	2		1				2	1	3	1	1
CO5	3	2	2		1				2	1	3	1	1

Type	Code	Scalable Database Systems Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-009		0-0-4	2	100

Objectives	The objective of this course is to provide hands-on practice on storage, retrieval and manipulation of relational data using SQL, along with other data models & query languages on some of the popular NoSQL databases.
Pre-Requisites	Knowledge of databases and programming skills is required.
Teaching Scheme	Regular laboratory classes conducted under supervision of the teacher. The experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Introduction to relational databases, data types, and syntax of SQL
2	Data retrieval using simple queries based on conditions and sorting the results.
2	Single-row functions, grouping and aggregate functions in SQL queries.
3	Writing complex queries using sub-queries and co-related sub-queries.
4	Create database, create tables, alter and manipulate structure of tables.
5	Imposing various constraints on tables for maintaining data integrity.
6	Insert, Update, and Delete data in the tables (DML statements).
7	Retrieve data from multiple tables using various types of Join operations.
8	Create, alter, and manage Views from single & multiple base tables.
9	Create and use other data base objects like sequence, indexes, and synonyms.
10	Performing set operations on tables, advanced operations like rollup and cube.
11	Introduction to PL/SQL, identifiers, literals, and keywords.
12	Write PL/SQL block by using conditional statements and expressions.
13	Using different types of Loops in a PL/SQL block and Exception handling.
14	Write PL/SQL block by using numeric, string, and other miscellaneous data types.
14	Write PL/SQL block to retrieve data using cursors.
16	Introduction to Stored procedures, Write PL/SQL block using procedures.
17	Develop functions with in/out parameters and using them in a PL/SQL block.
18	Write PL/SQL block using package and trigger.
19	Introduction to NoSQL databases, Document & Graph data models.
20	MongoDB - Introduction to MQL, Data Definition - Create, Alter, Drop, Truncate.
21	MongoDB - Data Manipulation - Select, Insert, Update, Delete, Batch.
22	MongoDB - Aggregate Framework, executing advanced queries.
23	Cassandra - Introduction to CQL, Create database, Create tables, Insert data.
24	Cassandra - Data retrieval and manipulation using CQL.

Cont'd...

Experiment-#	Assignment/Experiment
25	Cassandra - Indexes and Materialized Views.
26	Neo4j - Introduction to GQL (Cypher), Design & implement graph database.
27	Neo4j - Executing simple queries on graph databases.
28	Neo4j - Executing complex pattern queries on graph databases.

Text Books:

T1. K. Loney, ***Oracle Database 11g : The Complete Reference***, 1st Edition, McGraw-Hill, 2009.
T2. S. Bradshaw, E. Brazil, and K. Chodorow, ***MongoDB: The Definitive Guide***, 3rd Edition, O'Reilly Media, 2019.
T3. E. Hewitt, ***Cassandra: The Definitive Guide***, 1st Edition, O'Reilly Media, 2010.
T4. R. V. Bruggen, ***Learning Neo4j 3.x***, 1st Edition, Packt Publishing, 2014.

Reference Books:

R1. I. Bayross, ***Teach Yourself SQL/PLSQL Using Oracle 8i and 9i with SQLJ***, 1st Edition, BPB Publications, 2003.
R2. S. Feuerstein, ***Oracle PL/SQL Programming***, 6th Edition, O'Reilly, 2014.
R3. S. Tiwari, ***Professional NoSQL***, 1st Edition, Willey, 2011.
R4. D. Bechberger and J. Perryman, ***Graph Databases in Action***, 1st Edition, Manning Publications, 2020.

Online Resources:

1. https://docs.oracle.com/cd/E11882_01/server.112/e40402/toc.htm
2. https://docs.mongodb.com/
3. https://cassandra.apache.org/doc/latest/
4. https://neo4j.com/docs/

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Construct queries using SQL and retrieve data from a database using single/multi-row functions, and sub-queries.
CO2	Design relational tables imposing integrity constraints, operate on and manipulate database tables using DDL/DML statements.
CO3	Create other database objects like views, sequences and indices.
CO4	Develop complex PL/SQL programs including control structures, procedures, functions and triggers for real life applications.
CO5	Implement different types of NoSQL databases for unstructured data as per real-world requirements and analyze the data using NoSQL query languages.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.

Cont'd...

PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	1	1									3	2	2
CO2	2	2				1					3	3	2
CO3	2	2	1		1	1				1	3	3	3
CO4	1	3	2		2	1			1	2	3	3	3
CO5	2	3	3		3	2			2	3	3	3	3

Type	Code	Machine Learning Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-010		0-0-4	2	100

Objectives	The objective of this laboratory course is to provide practical exposure on implementing various machine learning techniques, extract features from data sets, and compare the results thereby realizing appropriate use of the machine learning techniques to specific real-world problems.
Pre-Requisites	Knowledge of optimization, and matrix theory is required. The experiments shall go along with the subjects taught in the theory class.
Teaching Scheme	Regular laboratory classes conducted under supervision of the teacher. The experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1, 2	Introduction and overview of scikit-Learn & its features.
3, 4	Implement simple and multiple linear, polynomial, and ridge regression.
5, 6	Implement binary classification using Logistic regression.
7, 8	Implement Linear Discriminant Analysis for dimensionality reduction and classification.
9, 10	Implementation of Lasso and Elastic net.
11, 12	Experiments on Cross-validation and bootstrap methods.
13, 14	Implementation of random forest, and gradient boost.
15, 16	Implementation of collaborative filtering.
17, 18	Experiments on Principal Component Analysis and Kernel PCA.
19, 20	Implementation of Perceptron Learning technique.
21, 22	Implementation of Backpropagation algorithm.
23, 24	Implementation of SVM for classification and regression.
25, 26	Novelty and outlier detection using SVM and GMM respectively.
27, 28	Mini Project

Text Books:

T1. T. Hastie, R. Tibshirani, and J. Friedman, ***The Elements of Statistical Learning : Data Mining, Inference and Prediction***, 2nd Edition, Springer Verlag, 2009.
T2. C. Bishop, ***Pattern Recognition and Machine Learning***, 1st Edition, Springer, 2007.

Reference Books:

R1. K. P. Murphy, ***Machine learning : A Probabilistic Perspective***, 4th Edition, MIT Press, 2012.
R2. H. Daumé III, ***A Course in Machine Learning***, Unpublished, Online: http://ciml.info/dl/v0_9/ciml-v0_9-all.pdf.
R3. T. Mitchel, ***Machine Learning***, 1st Edition, McGraw-Hill Education, 1997.
R4. S. Shalev-Shwartz and S. Ben-David, ***Understanding Machine Learning : From Theory to Algorithms***, 1st Edition, Cambridge University Press, 2014.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106202/: Prof. C. G. Jansson, IIT Madras
2. https://nptel.ac.in/courses/106/105/106105152/: By Prof. S. Sarkar, IIT Kharagpur
3. https://github.com/josephmisiti/awesome-machine-learning: An exhaustive index of machine learning concepts and programming materials.
4. http://mlss.cc/: Machine Learning Summer School Study Material

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Develop expertise with different Machine Learning toolkits.
CO2	Apply basic machine learning algorithms for predictive modeling.
CO3	Compare and contrast pros and cons of various machine learning techniques.
CO4	Extract meaningful information using non-statistical modeling on real world applications.
CO5	Evaluate recent advances & latest research in the field of machine learning.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	1		3	1			1	2	3	3	2
CO2	3	2	1		2	1			1	2	3	3	2
CO3	2	2	2		2	1			1	2	3	3	2
CO4	2	2	2		2	1			1	2	3	3	2
CO5	2	2	1		2	1			1	2	3	3	2

Type	Code	Artificial Intelligence Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-011		0-0-2	1	100

Objectives	The objective of this course is to motivate and prepare the students to appreciate and implement intelligent systems and incorporate artificial intelligence in data science applications.
Pre-Requisites	Knowledge of data structure, database management systems, and strong logical ability, and proficiency in programming are required.
Teaching Scheme	Regular laboratory experiments executed by the students under supervision of the teacher. Experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Heuristic search: Tic-Tac-Toe
2	Heuristic search: Water Jug Problem
3	Heuristic search: Missionaries and Cannibals
4	Heuristic search: Blocks World Problem
5	8 Queens Problem
6	Depth First Search (DFS)
7	Breadth First Search (BFS)
8	Best First Search
9	A* Algorithm
10	Adversarial Search: Optimal decision in games
11	Constraint satisfaction problems: Backtracking
12	Reasoning Systems for categories
13	Bayesian Networks: Exact inference
14	Bayesian Networks: Approximate inference

Text Books:

T1. I. Bratko, ***Prolog Programming***, Prentice Hall India, 2002.
T2. M. T. Jones, ***Artificial Intelligence Application Programming***, 2nd Edition, DreamTech, 2006.
T3. E. Charniak, C. K. Riesbeck, D. V. McDermott, and J. R. Meehan, ***Artificial Intelligence Programming***, 2nd Edition, Amazon Kindle, 2014.

Reference Books:

R1. P. Joshi, ***Artificial Intelligence with Python***, Packt Publishing, 2017.
R2. S. J. Russell and P. Norvig, ***Artificial Intelligence: A Modern Approach***, 3rd Edition, Prentice-Hall, 2010.

Online Resources:

1. https://www.csail.mit.edu/: MIT Computer Science and Artificial Intelligence Lab

2. https://www.expertsystem.com: Multi disciplinary industry solutions, World
3. https://www.eecs.umich.edu/eecs: University of Michigan AI Lab
4. https://onlinelibrary.wiley.com/journal/14680394: Expert System Wiley online library

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Solve problems by applying AI techniques to different complex problems using LISP, PROLOG.
CO2	Represent difficult real life problems in a state space and solve those using AI techniques.
CO3	Apply various AI methods like searching and game playing to solve real world applications.
CO4	Build inference engines by applying knowledge representation and Logic.
CO5	Obtain understanding of planning, Bayes networks, NLP and concepts of cognitive computing.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	3	1	2		2					3	3	1
CO2	3	3	1	2		2					3	3	1
CO3	3	3	2	3		2					3	3	1
CO4	2	3	2	2		2					3	3	1
CO5	2	3	2	2		2					3	3	1

Part II

2nd Year M. Sc. (Data Science)

Curriculum Structure

Semester III									
Type	**Code**	**Course Title**	**WCH** L-T-P			**Credits** L-T-P		
THEORY									
PC	MSCS-T-PC-012	Advanced Machine Learning	4	0	0	4	0	0
PC	MSCS-T-PC-013	Big Data Analytics	3	0	0	3	0	0
PC	MSCS-T-PC-014	Data Security & Privacy	2	0	0	2	0	0
PE		Professional Elective - I	3	0	0	3	0	0
PE		Professional Elective - II	3	0	0	3	0	0
PRACTICAL									
PC	MSCS-P-PC-015	Advanced Machine Learning Lab	0	0	4	0	0	2
PC	MSCS-P-PC-016	Big Data Analytics Lab	0	0	2	0	0	1
PJ	MSCS-P-PJ-032	Capstone Project - I	0	0	8	0	0	4
PJ	MSII-P-PJ-002	Summer Internship	0	0	0	0	0	1
		SUB-TOTAL	**15**	**0**	**14**	**15**	**0**	**8**
		TOTAL	**29**			**23**		

Semester IV									
Type	**Code**	**Course Title**	**WCH** L-T-P			**Credits** L-T-P		
THEORY									
PC	MSCS-T-PC-024	Data Visualization & Reporting	2	0	0	2	0	0
PE		Professional Elective - III	3	0	0	3	0	0
PE		Professional Elective - IV	3	0	0	3	0	0
PRACTICAL									
PC	MSCS-P-PC-025	Data Visualization & Reporting Lab	0	0	2	0	0	1
PJ	MSCS-P-PJ-033	Capstone Project - II	0	0	16	0	0	8
		SUB-TOTAL	**8**	**0**	**18**	**8**	**0**	**9**
		TOTAL	**26**			**17**		

List of Electives

Code	Elective # and Subjects
Professional Elective-I	
MSCS-T-PE-018	Time Series Analysis
MSCS-T-PE-019	Investment Analysis
MSCS-T-PE-020	Computational Finance
Professional Elective-II	
MSCS-T-PE-021	Bioinformatics Algorithms
MSCS-T-PE-022	Biomedical Image Analysis
MSCS-T-PE-023	Healthcare Analytics
Professional Elective-III	
MSCS-T-PE-027	Social Media Mining
MSCS-T-PE-028	Natural Language Processing
MSBS-T-PE-006	Probabilistic Graphical Models
Professional Elective-IV	
MSCS-T-PE-029	Multimedia Database Systems
MSCS-T-PE-030	Computer Vision
MSCS-T-PE-031	Realtime Analytics

Type	Code	Advanced Machine Learning	L-T-P	Credits	Marks
PC	MSCS-T-PC-012		4-0-0	4	100

Objectives	The objectives of this course is to learn the concepts behind regularization of parameters, deep neural networks, probabilistic graphical models, reinforcement learning etc., and use them to solve various machine learning problems.
Pre-Requisites	The first course in machine learning taught in previous semester, along with knowledge of probability, optimization, and linear algebra are required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	**Regularization for Deep Learning**: Parameter Norm Penalties, Norm Penalities as Constrained Optimization, Regularization and Under-Constrained Problems, Dataset Augmentation, Noise Robustness, Early Stopping, Parameter Tying and Parameter Sharing, Sparse Representations, Bagging and Other Ensemble Methods, Dropout. **Optimization for Training Deep Models**: How learning differs from optimization. Challenges in Neural Network Optimization, Basic Algorithms (SGD, momentum, Nesterov momentum), Parameter Initialization Strategies, Algorithms with Adaptive Learning Rates (AdaGrad, RMSProp, Adam), Approximate Second-order Methods (Newton, Congugradient, BFGS), Optimization Strategies.	**10 Hours**
Module-2	**Convolutional Networks**: The Convolution Operation, Convolution ideas (sparse interaction, parameter sharing, and equivalent representation), Pooling, Variants of the Basic Convolution Function, Structured Outputs, Data Types, Efficient Convolution Algorithms, Random or Unsupervised Features, The Neuroscientific Basis for Convolutional Networks, Applications.	**8 Hours**
Module-3	**Sequence Modeling**: Recurrent and Recursive Nets – Unfolding Computational Graphs, Recurrent Neural Networks, Bidirectional RNNs, Encoder-Decoder Sequence-to-Sequence Architecture, Deep Recurrent Networks, Recursive Neural Networks, The Challenge of Long-Term Dependencies, Echo State Networks, Leaky Units and Other Strategies for Multiple Time Scales, The Long Short-Term Memory and Other Gated RNNs, Optimization for Long-Term Dependencies, Practical methodologies of Deep Learning, Autoencoders, Applications.	**12 Hours**
Module-4	**Directed Graphical Models**: Conditional independence, representing joint, marginal, and conditional distributions, Bayesian Networks, D-separation, inference, learning, Hidden Markov Models; Undirected Graphical Models – Markov Random Fields, Spectral Clustering; Advanced SVM – Multiclass SVM, Multiple Kernels, Kernels for texts and strings.	**14 Hours**

Cont'd...

Module-#	Topics	Hours
Module-5	**Reinforcement Learning and Control**: Elements of Reinforcement Learning, Model based Decision Making – Markov processes, Markov reward process (MRP), Computing value function of a MRP, Markov decision process (MDP), Value function, Bellman equations, MDP control, Optimal value function, Policy iteration, Value iteration; Model Free Prediction and Control – SARSA, Q-Learning.	**12 Hours**
	Total	**56 Hours**

Text Books:

T1. I. Goodfellow, Y. Bengio, and A. Courville, ***Deep Learning***, The MIT Press, 2016.
T2. K. P. Murphy, ***Machine Learning : A Probabilistic Perspective***, The MIT Press, 2012.
T3. R. S. Sutton and A. G. Barto, ***Reinforcement Learning : An Introduction***, 2nd Edition, The MIT Press, 2018.
T4. J. S. Taylor, M. Mitchell, and N. Cristianini, ***Kernel Methods for Pattern Analysis***, Cambridge University Press, 2004.

Reference Books:

R1. D. Koller and N. Friedman, ***Probabilistic Graphical Models : Principles and Techniques***, The MIT Press, 2009.
R2. D. Barber, ***Bayesian Reasoning and Machine Learning***, Cambridge University Press, 2012.
R3. C. M. Bishop, ***Pattern Recognition and Machine Learning***, Springer, 2006.
R4. T. M. Mitchell, ***Machine Learning***, McGraw Hill Education, 2013.

Online Resources:

1. https://nptel.ac.in/courses/106/105/106105215/: by Prof. P. K. Biswas, IIT Kharagpur
2. https://nptel.ac.in/courses/106/106/106106184/: by Prof. S. Iyengar and Prof. Padmavati, IIT Ropar
3. https://nptel.ac.in/courses/106/106/106106201/: by Prof. M. Khapra, IIT Madras
4. https://cedar.buffalo.edu/~srihari/CSE676

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain and apply knowledge of regularization and efficient optimization techniques to improve the performance of deep learning methods.
CO2	Explore the concepts of CNN and apply this to solve related problems.
CO3	Apply RNN and Autoencoder techniques to solve real world machine learning problems.
CO4	Explore the representation, learning and inference of some graphical models.
CO5	Aquire the concepts & methods of reinforcement learning and apply them in decision making.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.

Cont'd...

PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	2	2	3				2	2	2	2	1
CO2	3	3	3	2	3				3	3	3	3	2
CO3	3	3	3	2	3				3	3	3	3	3
CO4	3	3	3	2	3				3	3	3	3	3
CO5	3	3	3	2	3				3	3	3	3	3

Type	Code	Big Data Analytics	L-T-P	Credits	Marks
PC	MSCS-T-PC-013		3-0-0	3	100

Objectives	The objective of the course is to study different techniques to find similar items, mining data streams, link analysis, clustering techniques, recommendation systems, and collaborative filtering for Big Data, along with the concepts of batch processing, Hadoop, MapReduce & Spark.
Pre-Requisites	Knowledge of basics of data mining & algorithm design is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Identifying Big Data, Understanding the Challenges of Big Data, Building Blocks of Effective Data Management; MapReduce and the New Software Stack – Distributed File Systems, MapReduce, Algorithms Using MapReduce, Extensions to MapReduce, The Communication Cost Model, Complexity Theory for MapReduce; Hadoop, Spark.	**9 Hours**
Module-2	Big Data Algorithms-I: Finding Similar Items - Nearest Neighbor Search, Shingling of Documents, Similarity Preserving Summaries of Sets, Locality Sensitive Hashing for Documents, Distance Measures, Theory of Locality Sensitive Functions, LSH Families for High Degree of Similarities.	**9 Hours**
Module-3	Big Data Algorithms-II: Mining Data Streams - Streaming Data Models, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments, Counting Ones in Window. Link Analysis-Page Rank, Efficient Computation of Page Rank, Topic Sensitive PageRank.	**10 Hours**
Module-4	Big Data Algorithms-III: Clustering Techniques - BFR Algorithm, CURE Algorithm, Clustering in Non-Euclidean Space, Clustering for Streams and Parallelism.	**7 Hours**
Module-5	Big Data Algorithms-IV: Recommendation Systems - A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering, Dimensionality Reduction, The NetFlix Challenge.	**7 Hours**
	Total	**42 Hours**

Text Books:

T1. J. Leskovec, A. Rajaraman, and J. D. Ullman, ***Mining of Massive Datasets***, 2nd Edition, Cambridge University Press, 2014.
T2. M. Wessler, ***Big Data Management For Dummies***, 3rd Informatica Special Edition, John Wiley & Sons.

Reference Books:

R1. J. Han, M. Kamber, and J. Pei, ***Data Mining Concepts and Techniques***, 3rd Edition, Morgan Kaufman Publications, 2011.
R2. T. M. Mitchell, ***Machine Learning***, 1st Edition, McGraw-Hill Education, 2017.

R3. J. Bell, ***Machine Learning for Big Data: Hands-On for Developers and Technical Professionals***, Wiley Publications, 2014.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106142/: by Prof. J. Augustine, IIT Madras
2. https://nptel.ac.in/courses/106/104/106104189/: by Dr. R. Misra, IIT Patna
3. http://www.mmds.org: Study material on Mining of Massive Data Sets
4. http://lintool.github.com/MapReduceAlgorithms/index.html

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the basic concepts of Big Data and MapReduce techniques.
CO2	Apply different tools and techniques used for finding similar items.
CO3	Demonstrate application of algorithms for analysis of streaming data and link analysis.
CO4	Apply different techniques for recommendation systems & collaborative filtering and compare different clustering techniques to apply them on large data sets.
CO5	Implement big data algorithms using the concepts of Hadoop, MapReduce, and Spark.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	2	2	3					2	2	2	2
CO2	3	3	3	2	3					2	2	2	2
CO3	3	3	3	2	3					2	2	2	2
CO4	3	3	3	2	3					2	2	2	2
CO5	1	1	1	2	3					2	2	2	2

Type	Code	Data Security & Privacy	L-T-P	Credits	Marks
PC	MSCS-T-PC-014		2-0-0	2	100

Objectives	The objective of this course is to study the security goals, services and mechanisms from a data science perspective with focus on cryptography techniques on user data stored and communicated through unsecured channels.
Pre-Requisites	Knowledge on computer networks and algorithms is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to Computer Security Concepts, Security Attacks, Security Services and Mechanisms, Symmetric Cipher model, Cryptography & Cryptanalysis, Substitution Techniques: Caesar cipher, Monoalphabetic cipher, Playfair cipher, Hill Cipher, Polyalphabetic ciphers: Vignere cipher, Vernam cipher, Transposition cipher.	**7 Hours**
Module-2	Symmetric key Cryptography: Data Encryption Standard (DES), Advanced Encryption Standard (AES); Public Key Cryptography: RSA, El-Gamal, Elliptic Curve Cryptography (ECC): Introduction to elliptic curve, arithmetic, and applications.	**7 Hours**
Module-3	Message Integrity and Authentication, Cryptographic Hash Functions: MD5, SHA, Digital Signature algorithms using RSA, and ECC (ECDSA), Key Distribution, Certificate Authority, X.509.	**7 Hours**
Module-4	Typical attacks on network and web and their security; Security Protocols: SSL/TLS; Role of Firewall and IDS in data security & privacy.	**7 Hours**
	Total	**28 Hours**

Text Books:

T1. W. Stallings, ***Cryptography and Network Security: Principle and Practice***, 7th Edition, Pearson Education, 2017.
T2. C. P. Pfleeger, S. L. Pfleeger, and J. Margulies, ***Security in Computing***, 5th Edition, Prentice-Hall India, 2015.

Reference Books:

R1. B. A. Forouzan, D. Mukhopadhaya, ***Cryptography and Network Security***, 2nd Edition, McGraw-Hill Education, 2010.
R2. C. Kaufman, R. Perlman, and M. Speciner, ***Network Security: Private Communication in a Public World***, 2nd Edition, Prentice-Hall India, 2002.
R3. A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, ***Handbook of Applied Cryptography***, CRC Press, 1996.

P.T.O

Online Resources:

1. https://nptel.ac.in/courses/106105031/
2. https://csrc.nist.gov/
3. https://www.sans.org/
4. https://www.cryptool.org/en/
5. https://www.rsa.com/en-us

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Analyze the performance of traditional symmetric key cryptography techniques.
CO2	Explain modern symmetric key cryptography and asymmetric key cryptography techniques and compare their strengths and weaknesses.
CO3	Apply public key cryptography and Hash algorithms for data integrity, authentication, and digital signatures.
CO4	Explore security threats on computer networks, web, and role of Firewalls/IDS to preserve data security and privacy.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO7	Work with a professional context pertaining to ethics, social, cultural and cyber regulations.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2										1	1	1
CO2	3		2	3	1	1	1				1	1	1
CO3	3		2	3	2	1	1				1	1	1
CO4	1		3	3	2	2	3				1	1	1

Type	Code	Time Series Analysis	L-T-P	Credits	Marks
PE	MSCS-T-PE-018		3-0-0	3	100

Objectives	The objective of the course is to study the basics and analysis techniques of various time series data from a data science perspective and utilize R/Python for computation, analysis, and visualization of the same.
Pre-Requisites	Fundamental knowledge of probability & statistics and proficiency in Python and R programming is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	**Characteristics of Time Series**: Nature of Time Series Data, Time Series Statistical Models (White Noise, Moving signal in Noise); Measures of Dependence: Auto-correlation and Cross-Correlation, Stationary Time Series, Estimation of Correlation, Vector-Valued and Multidimensional Series.	**10 Hours**
Module-2	**Regression and ARMA Models**: Classical Regression in the Time Series Context, Exploratory Data Analysis, Smoothing in the Time Series Context, Auto-regressive Moving Average Models (Auto-regressive Models, Moving Average Models, ARMA Model), Difference Equations (ACF of an AR(2) Process, Sample Path of an AR(2)), Auto-correlation and Partial Auto-correlation Functions, Forecasting – Prediction for an AR(2), PACF of an AR(2), The Innovations Algorithm, Prediction for an MA(1), Forecasting ARMA Processes, Backcasting an ARMA(1; 1).	**9 Hours**
Module-3	**ARIMA & GARCH Model**: Integrated Models for Non stationary Data, Building ARIMA Models, Multiplicative Seasonal ARIMA Models, GARCH Model.	**8 Hours**
Module-4	**Spectral Analysis and Filtering**: Cyclical Behavior and Periodicity, The Spectral Density, Periodogram and Discrete Fourier Transform, Parametric & Non-parametric Spectral Estimation, Linear Filters.	**8 Hours**
Module-5	**State-Space Models**: Introduction, Filtering, Smoothing, and Forecasting, Maximum Likelihood Estimation; Structural Models: Signal Extraction and Forecasting, Dynamic Linear Models.	**7 Hours**
	Total	**42 Hours**

Note: The time series models studied above will be implemented using R/Python in class.

Text Books:

T1. R. H. Shumway and D. S. Stoffer, ***Time Series Analysis and Its Applications with R Examples***, 4th Edition, Springer Verlag, 2017.

Reference Books:

R1. J. D. Hamilton, ***Time Series Analysis***, Princeton University Press, 2012.
R2. R. S. Tsay, ***Analysis of Financial Time Series***, 3rd Edition, Wiley Publications, 2010.
R3. P. J. Brockwell and R. A. Davis, ***Time Series Theory and Methods***, 2nd Edition, Springer Verlag, 1991.
R4. D. R. Brillinger, ***Time Series: Data Analysis and Theory***, Society for Industrial and Applied Mathematics (SIAM), 2001.

Online Resources:

1. https://nptel.ac.in/courses/103/106/103106123/
2. https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm
3. https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/
4. https://www.stat.tamu.edu/~suhasini/teaching673/time_series.pdf
5. https://people.duke.edu/~rnau/411home.htm

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the concepts of statistical models. Stationary time series, and estimating auto-correlation on an observed time-series data.
CO2	Estimate parameters of auto regressive moving average models (ARMA) and to forecast ARMA process.
CO3	Apply ARIMA and GARCH models to make predictions on an observed time-series data.
CO4	Explain the foundations of spectral theory and to estimate spectral density.
CO5	Evaluate results from statistical tools for time-series models such as ARIMA, GARCH, spectral, and state space models.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO8	Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2		2	1		2	3	3	1		1
CO2	3	2	1		2	1		2	2	1	1		1
CO3	3	2	1		2	1		2	2	2	2		1
CO4	3	2	2		2	2		2	2	2	1		1
CO5	3	3	3		2	3		3	3	2	2		1

Type	Code	Investment Analysis	L-T-P	Credits	Marks
PE	MSCS-T-PE-019		3-0-0	3	100

Objectives	The objective of the course is to study the theory and empirical evidences relevant for investing, particularly in the context of portfolio management, optimal portfolio selection, the relation between risk and return market efficiency.
Pre-Requisites	Knowledge of statistical methods and machine learning is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction: Investment and markets, Typical investment problems; Mean-Variance Portfolio Theory: Asset return, random variables, random returns, portfolio mean and variance, the feasible set, the Markowitz model, the two-fund theorem, inclusion of risk-free asset, the one fund theorem.	**9 Hours**
Module-2	Capital Asset Pricing Model: Market equilibrium, the capital market line, Pricing Model, Security Market Line, Divestment implications, Performance evaluation, CAPM pricing formula; Models and Data: Factor model, CAPM as a Factor Model, Arbitrage Pricing theory, data and statistics.	**9 Hours**
Module-3	Forwards and Futures: Forward contracts, Forward prices, Value of froward contract, Basics of futures contracts, Futures prices, Relation to expected spot price, Perfect hedge, Minimum variance hedge.	**8 Hours**
Module-4	Basic Option Theory: Option concept, Nature of option values, Option combinations and put-call parity, Single period binomial options theory, Multi-period options, More general binomial problems, Evaluating real investment opportunities, Efficient portfolio construction model using stock price predicted by support vector regression.	**8 Hours**
Module-5	Additional Options: The Black-scholes equation, Call option formula, Risk natural valuation, Delta hedging; Hybrid SVM-TLBO forecasting model incorporating dimensionality reduction techniques.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. D. G. Luenberger, ***Investment Science***, 2nd Edition, Oxford University Press, 2014.

Reference Books:

R1. E. J. Elton, M. J. Gruber, S. J. Brown, and W. N. Goetzmann, ***Modern Portfolio Theory and Investment Analysis***, 7th Edition, John Wiley & Sons, 2007.
R2. Z. Bodie, A. Kane, A. J. Marcus, and P. Mohanty, ***Investments***, 11th Edition, McGraw-Hill Education, 2019.
R3. J. C. Hull, ***Options Futures & Other Derivatives***, 4th Edition, Pearson Education, 1999.

Online Resources:

1. https://www.sciencedirect.com/science/article/abs/pii/S1062940818302481
2. https://link.springer.com/article/10.1007/s10489-016-0801-3

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain Markovich model and mean-variance portfolio theory and use them to construct efficient portfolios.
CO2	Use capital asset pricing model and arbitrage pricing theory to design efficient portfolio and obtain inferences about returns.
CO3	Apply forward and futures contracts and use them for better investment decisions.
CO4	Analyze the basic option theory to evaluate real investment opportunities.
CO5	Explore current research works on real world problem for forecasting option price.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	3		2	2					3	2	2
CO2	3	3	3		2	2			1	1	3	1	2
CO3	3	3	3	1	2	2			1	1	2	2	1
CO4	3	3	3	1	2	2			2	1	3	1	1
CO5	3	3	3	1	2	2			3	1	3	1	1

Type	Code	Computational Finance	L-T-P	Credits	Marks
PE	MSCS-T-PE-020		3-0-0	3	100

Objectives	The objective of this course is to study and apply mathematical methods to the pricing and hedging of financial derivative, securities, risk management etc., and to formulate mathematical models for quantifiable systems.
Pre-Requisites	Knowledge of probability theory, statistical inference, linear algebra, and differential equations is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Fixed-Income Securities: Valuations, Interest Rate Sensitivity, Portfolio Management, Portfolio Optimization (Basic Mean-Variance Portfolio Optimization), Option Pricing and Binomial Methods: Options (The No-Arbitrage Principle), The Binomial Model (Pricing American options by Binomial Methods), Case study of Portfolio construction using Machine Learning methods.	**9 Hours**
Module-2	Stochastic Differential Equations: Stochastic Ito Processes, Ito Lemma, And Applications in Stock Market, The Black-Scholes Equation: Derivation of the Black-Scholes Equation, Solution of the Black-Scholes Equation, Closed-Form Solutions of European Call and Put Options, Hedging Portfolios: The Greeks, Implied Volatility.	**8 Hours**
Module-3	Random Numbers and Monte Carlo Simulation: Pseudo-Random Numbers, Transformation of Random Variables, Inverse Transform Method, Acceptance-Rejection Method, Generating Normal Variates (Box-Muller Method, The Polar Method of Marsaglia, Multivariate Normal Variables), Monte Carlo Integration, Option Pricing by Monte Carlo Simulation, Variance Reduction Techniques (Antithetic Variates, Control Variates)	**9 Hours**
Module-4	Option Pricing by Partial Differential Equations: Classification of PDEs, Finite Difference Methods for Parabolic Equations (An Explicit Method, An Implicit method, Crank-Nicolson Method), Option Pricing by the Heat Equation, Pricing American Options (Projected SOR Method for American Options).	**8 Hours**
Module-5	Derivatives Pricing via the FFT: Call option pricing, Put option pricing via Fourier Transform, Evaluating the pricing integral, Implementation of FFT, Damping factor α, Derivatives Pricing via Fourier Cosine method: COS method, Cosine Series Expansion of arbitrary functions, Cosine Series coefficients, COS option pricing for Different Payoffs, Numerical Results for COS method: Geometrical Brownian Motion, Heston Stochastic Volatility Model, Case study: Machine learning methods for derivative pricing.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. O. Ugur, ***An Introduction to Computational Finance***, 1st Edition, Imperial College Press, 2009.
T2. A. Hirsa, ***Computational Methods in Finance***, Chapman & Hall/CRC Press, 2013.

Reference Books:

R1. J. C. Hull and S. Basu, ***Options, Future and Other Derivatives***, 10th Edition, Pearson, 2018.
R2. P. Wilmott, ***Option Pricing : Mathematical Models and Computation***, 1st Edition, Oxford Financial Press, 1993.
R3. L. Clewlow and C. Strickland, ***Implementing Derivative Models***, John Wiley & Sons, 1998.
R4. D. G. Luenberger, ***Investment Science***, 2nd Edition, Oxford University Press, 2013.

Online Resources:

1. https://nptel.ac.in/courses/111/104/111104089/: by Dr. J. Dutta, IIT Kanpur
2. https://nptel.ac.in/courses/109/104/109104104/: by Dr. J. Dutta, IIT Kanpur
3. https://www.youtube.com/watch?v=Qq6v27t2-c8: Lecture Series by Leipzig University

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Manage and Optimize Portfolio Fixed-Income using Binomial and Machine Learning methods.
CO2	Explain Stochastic Ito Processes and illustrate its application in stock market. Apply the Black-Scholes methodology in deriving option prices.
CO3	Apply various methods to generate Normal variates and design Option Pricing by Monte Carlo Simulation.
CO4	Explain Finite Difference Methods for Parabolic Equations and use Partial Differential Equations for Option Pricing and Heat equation.
CO5	Apply Transform techniques such as Fast Fourier Transform, Fourier Cosine method and analyze Numerical Results for COS method using various statistical models.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

P.T.O

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2	2	2	1			1	1	2	2	1
CO2	3	3	2	2	2	1			1	1	2	2	1
CO3	3	3	2	2	2	1			1	1	2	2	1
CO4	3	2	2	2	2	1			1	1	2	1	1
CO5	3	2	2	2	2	1			1	1	2	1	1

Type	Code	Bioinformatics Algorithms	L-T-P	Credits	Marks
PE	MSCS-T-PE-021		3-0-0	3	100

Objectives	The objectives of this course is to understand biological data and databases, map biological problems as computational problems and solve them using various statistical and computational tools and techniques.
Pre-Requisites	Knowledge of design and analysis of algorithms is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Basic Concepts of Molecular Biology: Cellular Architecture, Nucleic Acid (RNA & DNA), DNA Replication, Repair and Recombination, Transcription, Translation, Genetic Code, Gene Expression, Protein structure and function, Molecular biology tools; Suffix Trees: Definition and examples, Ukkonen's linear-time suffix tree algorithm, Applications – Exact string matching, LCS of two strings, Pair-wise Sequence Alignment (Global, Local), Dynamic Programming approach, Edit distance, Dynamic programming calculation of Edit distance, Gaps.	**12 Hours**
Module-2	HMM for Pair-wise & Multiple Sequence Alignment: Need of MSA, Family and Super Family representation, Multiple sequence comparisons for structural inferences, Multiple alignments with sum-of-pairs, Consensus objective functions, Database searching for similar sequences (FASTA, BLAST), PAM, BLOSOM substitution matrices.	**8 Hours**
Module-3	Sequencing: Sequencing by Hybridization, Shortest common super string, Algorithms for overlap multi-graph, Fragment Assembly, Protein Sequencing and identification, The peptide sequencing problem, spectrum graph.	**6 Hours**
Module-4	Motif Prediction: Regulatory motifs in DNA sequences, The motif finding problem, Finding motifs and median string, Greedy approach to motif finding; Statistical and similarity based approach to Gene prediction, Spliced alignment.	**8 Hours**
Module-5	Evolutionary trees, distance and character based methods for tree reconstruction. Gene expression clustering tumor classification with SVM, Reconstruction of biological network by supervised machine learning approaches, short review on deep learning for computational biology.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. N. C. Jones and P. A. Pevzner, ***An Introduction to Bioinformatics Algorithms***, MIT Press, 2005.
T2. D. Gusfields, ***Algorithms on Strings, Trees and Sequences : Computer Science and Computational Biology***, Cambridge University Press, 1997.

T3. J. C. Setubal and J. Meidanis, ***Introduction to Computational Molecular Biology***, 1st Edition, PWS Publishing Co., 1997.

Reference Books:

R1. W. J. Ewens and G. R. Grant, ***Statistical Methods in Bioinformatics : An Introduction***, 2nd Edition, Springer Science & Business Media, 2006.
R2. R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, ***Biological Sequence Analysis : Probabilistic Models of Proteins and Nucleic Acid***, Cambridge University Press, 1999.
R3. D. E. Krane and M. L. Raymer, ***Fundamental Concepts of Bioinformatics***, 1st Edition, Pearson Education, 2003.
R4. H. M. Lodhi and S. H. Muggleton, ***Elements of Computational Systems Biology (Vol. 08)***, 1st Edition, John Willey & Sons, 2010.

Online Resources:

1. https://nptel.ac.in/courses/102/106/102106065/: By Prof. M. M. Gromiha, IIT Madras
2. https://ocw.mit.edu/courses/biology/7-91j-foundations-of-computational-and-systems-biology-spring-2014/index.htm
3. https://dspace.mit.edu/bitstream/handle/1721.1/103560/6-047-fall-2008/contents/lecture-notes/index.htm
4. https://www.embopress.org/doi/pdf/10.15252/msb.20156651

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain basics of molecular biology, map biological problems into computational problems and solve string matching & sequence alignment problems using appropriate techniques.
CO2	Apply different algorithms to solve pair-wise and multiple sequence alignment problems.
CO3	Solve DNA/Protein sequencing problems to identify genetic disorders.
CO4	Apply Motif/Gene prediction algorithms for some real life problems and drug design.
CO5	Use the power of machine learning to solve real life problems in computational biology.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	1	2		3				2	2	2	2	1
CO2	3	2	2		3				2	2	3	2	1
CO3	3	3	3		3				2	2	2	3	1
CO4	3	2	3		3				3	2	3	3	1
CO5	3	2	3						3	2	3	3	1

Type	Code	Biomedical Image Analysis	L-T-P	Credits	Marks
PE	MSCS-T-PE-022		3-0-0	3	100

Objectives	The objectives of this course is to study state-of-the-art medical image analysis techniques and apply them for computer aided diagnostics.
Pre-Requisites	Basic knowledge of algebra, statistics, signal and image processing is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to Bio medical image, Nature of Biomedical Images, Body Temperature as an Image, X-Ray Imaging, Objectives of Biomedical Image Analysis, Computer aided Diagnosis, Image Quality and Information Content: Difficulties in Image Acquisition and Analysis, Characterization of Image Quality, Digitization of Images, Histogram, Entropy, Fourier Transform and Spectral Content, Signal-to-Noise Ratio.	**9 Hours**
Module-2	Removal of Artifacts: Characterization, Matrix Representation of Image Processing, Optimal Filtering, Image Enhancement: Dual energy and Energy subtraction X-Ray Imaging, Grayscale Transforms, Histogram Transformation, Convolution Mask Operators; Detection of Regions of Interest: Thresholding and Binarization, Detection of Isolated Points and Lines, Edge Detection, Segmentation and Region Growing.	**9 Hours**
Module-3	Analysis of Shape: Representation of Shapes and Contours, Shape Factors, Application - Shape Analysis of Classifications, Analysis of Texture: Texture in Biomedical Images, Models for the Generation of Texture, Statistical Analysis of Texture, Fractal Analysis, Segmentation and Structural Analysis of Texture; Overview of Fingerprints – Sensing, Representation, Analysis, Matching, Classification and Indexing.	**9 Hours**
Module-4	Physics of Magnetic Resonance Imaging (MRI) - Data acquisition, Image reconstruction, Advanced MRI - Image quality, Deep learning of brain images and its application to multiple sclerosis; Machine learning and its application in microscopic image analysis. Physics of Ultrasound - Ultrasound imaging principles, Beam pattern formation and focusing, reconstruction, and segmentation, Multi-template-based multiview learning for Alzheimer's disease diagnosis.	**9 Hours**
Module-5	Image Coding and Data Compression: Considerations Based on Information Theory,Direct Source Coding, Image Coding and Compression Standards, Lower limit Analysis of Lossless Data Compression, Pattern Classification and Diagnostic Decision:Supervised and unsupervised pattern classification, Probabilistic Models and Statistical Decision, Measures of Diagnostic Accuracy, Case Studies - Breast Cancer.	**6 Hours**
	Total	**42 Hours**

Text Books:

T1. R. M. Rangayyan, ***Biomedical Image Analysis***, CRC Press, 2004.
T2. J. L. Prince and J. M. Links, ***Medical Imaging : Signals and Systems***, 2nd Edition, Prentice Hall, 2015.

Reference Books:

R1. J. S. Suri, D. Wilson, and S. Laxminarayan (Eds.), ***Handbook of Biomedical Image Analysis (Vol. 2)***, Springer Science & Business Media, 2005.
R2. R. Salzer (Ed.), ***Biomedical Imaging : Principles and Applications***, John Wiley & Sons, 2012.
R3. A. G. Webb, ***Introduction to Biomedical Imaging***, John Wiley & Sons, 2017.

Online Resources:

1. https://nptel.ac.in/courses/108/105/108105091/: by Prof. D. Sheet, IIT Kharagpur
2. https://blog.tensorflow.org/2018/07/an-introduction-to-biomedical-image-analysis-tensorflow-dltk.html
3. https://eng.ox.ac.uk/biomedical-image-analysis/
4. https://www.kdnuggets.com/2017/03/medical-image-analysis-deep-learning.html

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain basics of biomedical imaging and describe various aspects of digital imaging.
CO2	Remove artifacts from the images and detect the regions of interest for analysis.
CO3	Analyze the shape & texture of biomedical images and classify fingerprints.
CO4	Analyze MRI & Ultrasound images by applying machine learning techniques.
CO5	Compare between coding and compression techniques and classify patterns for diagnostics.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO8	Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	1	1	1	2			2	2	1	2	1	2
CO2	2	2	1	2	2			2	2	2	1	1	1
CO3	3	3	2	3	2			2	3	3	2	2	2
CO4	3	3	2	2	2			2	3	3	2	2	2
CO5	3	3	2	2	2			2	3	3	3	3	3

Type	Code	Healthcare Analytics	L-T-P	Credits	Marks
PE	MSCS-T-PE-023		3-0-0	3	100

Objectives	The objective of this course is to study electronic healthcare information systems and analysis of healthcare data with machine learning including risk stratification, disease progression modeling, diagnosis and precision medicine.
Pre-Requisites	Knowledge of mathematics, probability & statistics, algorithms and machine learning techniques is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction: Brief history of AI & ML applied to healthcare, Uniqueness of ML in healthcare, Goals of Health Care - Mortality, Disability, Morbidity and Tasks of healthcare (Diagnosis, Prognosis, Treatment, Prevention); Healthcare Data: Deep Dive into Clinical Data - Understanding of Clinical Data, Types of Data, Risk Stratification, Machine Learning Formulation, Model Evaluation.	**8 Hours**
Module-2	NLP & ML in Healthcare: Value of the data in clinical text, NLP for Healthcare data, Foundations of NLP, Goals of NLP, Term spotting and handling negation, Uncertainty, Unified Medical Language System (UMLS), Language models and Neural methods; Translating Technology into the Clinic: Existing healthcare technology, Diffusion of new medical technologies and biomedical research, Application of ML to Cardiac Imaging - Cardiac structure and functions, Major types of cardiac diagnostics, Medical imaging data, Image classification and semantic segmentation.	**8 Hours**
Module-3	ML for Differential Diagnosis - Models for diagnostic reasoning; ML for Pathology - Background, Pathology, Computational Pathology and building of model; ML for Mammography - Background, Deep learning models for mammogram interpretation; Causal Inference - Background, Potential outcomes framework, Covariate adjustment and Propensity score re-weighting.	**8 Hours**
Module-4	Reinforcement Learning - Overview of treatment policies and potential outcomes, Reinforcement Learning - Applications, Patient Management, Decision Processes, Value Maximization; Reinforcement Learning Paradigms, Dynamic programming, Learning from Off-Policy Data, Causal Inference vs. Reinforcement Learning and Evaluating dynamic treatment strategies; Disease Progression Modeling and Sub-typing – Disease Progression Modeling, Staging, Sub-typing, Multi-task Learning, Unsupervised Learning, Stage vs. Subtype, Data points in different dimensions, COPD Diagnosis and Progression, Pseudo-time Methods from Computational Biology.	**10 Hours**

Cont'd...

Module-#	Topics	Hours
Module-5	Precision Medicine - Collecting Genome Data, Precision Medicine Modality Space (PMMS) and Genotypes and Phenotypes; Automating Clinical Workflows: Improving Medical Care - Expert Systems, Protocol Systems, Narrowing Performance Distribution, Decision support from local data, Change-Point Detection to Monitor Rule Firings and Workflow Issues; Miscellaneous - Fairness, Robustness to dataset shift, Building a robust model and interpretability.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. C. K. Reddy and C. C. Aggarwal, ***Healthcare Data Analytics***, 1st Edition, Taylor & Francis, 2015..

Reference Books:

R1. H. Yan and E. K. Lee, ***Healthcare Analytics : From Data to Knowledge to Healthcare Improvement***, 1st Edition, Wiley Publications, 2016..

Online Resources:

1. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-s897-machine-learning-for-healthcare-spring-2019/: Primary source of study material
2. https://nptel.ac.in/courses/110/104/110104095/: by Prof. A. Sengupta, IIT Kanpur
3. https://www.coursera.org/lecture/hi-five-clinical/introduction-to-healthcare-data-analytics-overview-NAv03: by Columbia University
4. http://healthcareanalytics.info
5. https://www.edx.org/course/data-analytics-in-health-from-basics-to-business

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the basic concepts of healthcare data analytics and application of AI & ML techniques on healthcare data.
CO2	Explore application of NLP & ML in healthcare and apply it to Cardiac Imaging
CO3	Implement ML for differential diagnosis, pathology and mammography and explain the concept of Causal Inference.
CO4	Evaluate the use of Reinforcement Learning in healthcare and understand Disease Progression Modeling along with Sub-typing.
CO5	Explain the details of Precision Medicine and the process of automating clinical workflows.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO8	Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.

Cont'd...

PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	1	1	1			1	1	2	3	1	1
CO2	3	2	2	1	1			1	1	2	2	1	2
CO3	3	2	2	1	2			1	1	1	2	3	3
CO4	3	2	2	1	2			1	1	2	2	2	2
CO5	2	1	2	1	1			1	1	1	1	1	1

Type	Code	Advanced Machine Learning Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-015		0-0-4	2	100

Objectives	The objectives of this laboratory course is to implement different algorithms for CNN, RNN, Probabilistic graphical models, Reinforcement learning etc., for getting hands-on exposure on application of these techniques to solve related machine learning problems in the real world.
Pre-Requisites	Knowledge of topics taught in the theory class and programming is required.
Teaching Scheme	Regular laboratory experiments executed by the students under supervision of the teacher. Experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
	Part-A: Implementation using Python
1, 2	Back-propagation algorithm in vector-matrix form using SGD with L2-regularization.
3, 4	Comparison of performance of algorithms with adaptive learning rates (AdaGrad, RMSprop, Adam)
5, 6	Convolutional Neural Network (CNN) algorithms
7, 8	Recurrent Neural Network (RNN) algorithm
9	Long Short-Term Memory (LSTM) network algorithm
10, 11	Auto-encoder algorithms
12	Bayesian network algorithm
13, 14	Hidden Markov Model (HMM)
15	Spectral Clustering
16	Multi-class SVM algorithms (one vs. one)
17	SVM for text classification
18, 19	Markov reward process (MRP), Markov decision process (MDP): computation of optimal value function and optimal policy
20	Implementation of Q-learning
	Part-B: Implementation using Keras and TensorFlow
21-23	Quick introduction to TensorFlow and keras: customizing model and training algorithms
24	Loading and preprocessing data using TensorFlow
25	CNN using TensorFlow and keras
26	Processing sequences using RNN
27	Various types of auto-encoders using TensorFlow
28	Generative Adversarial Networks (GANs)

Cont'd...

Experiment-#	Assignment/Experiment
Part-C: Homework for Implementation	
1	Back-propagation algorithm in vector-matrix form using SGD with: (i) Momentum, (ii) Nesterov momentum
2	Gated-RNN algorithm
3	Markov Random fields
4	Multi-class SVM algorithms: (i) One vs. rest, (ii) Class labels as binary sequence
5	SARSA algorithm

Text Books:

T1. I. Goodfellow, Y. Bengio, and A. Courville, ***Deep Learning***, The MIT Press, 2016.
T2. K. P. Murphy, ***Machine Learning : A Probabilistic Perspective***, The MIT Press, 2012.
T3. R. S. Sutton and A. G. Barto, ***Reinforcement Learning : An Introduction***, 2nd Edition, The MIT Press, 2018.
T4. J. S. Taylor, M. Mitchell, and N. Cristianini, ***Kernel Methods for Pattern Analysis***, Cambridge University Press, 2004.
T5. A. Géron, ***Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow : Concepts, Tools, and Techniques to Build Intelligent Systems***, 2nd Edition , O'Reilly Media, 2019.

Reference Books:

R1. D. Koller, and N. Friedman, ***Probabilistic Graphical Models : Principles and Techniques***, The MIT press, 2009.
R2. D. Barber, ***Bayesian Reasoning and Machine Learning***, Cambridge University Press, 2012.
R3. C. M. Bishop, ***Pattern Recognition and Machine Learning***, Springer, 2006.
R4. T. M. Mitchell, ***Machine Learning***, 1st Edition, McGraw Hill Education, 2017.

Online Resources:

1. https://cedar.buffalo.edu/~srihari/CSE676

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Implement regularization and efficient optimization techniques to improve the performance of deep learning methods.
CO2	Implement Convolutional Neural Network (CNN) algorithms.
CO3	Implement Recurrent Neural Network (RNN) algorithms and autoencoder.
CO4	Implement various graphical models.
CO5	Implement reinforcement learning techniques.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	2		2					2	3	3	2
CO2	3	3	2		2					2	3	3	2
CO3	3	3	2		2					2	3	3	2
CO4	3	3	2		2					2	3	3	2
CO5	3	3	2		2					2	3	3	2

Type	Code	Big Data Analytics Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-016		0-0-2	1	100

Objectives	The objective of this laboratory course is to provide hands on exposure on implementation of different big data analytic algorithms using Hadoop, MapReduce and Spark.
Pre-Requisites	Knowledge of data mining & algorithm design is required. The experiments shall go along with the subjects taught in the theory class.
Teaching Scheme	Regular laboratory experiments executed by the students under supervision of the teacher. Experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Setup and install Hadoop
2	Adding Files and Directories, Retrieving Files, Deleting Files and Exploring various shell commands in Hadoop
3	Find the number of occurrences of each word appearing in an input file using basic Word Count Map Reduce program
4	Look for specific keywords in a file using MapReduce Job for word search count.
5	Eliminate stop word (given in a file) from a large text file.
6	Write a Map Reduce program to find the average, max and min temperature for each year in NCDC weather data set (Data is available at: https://github.com/tomwhite/hadoop-book/tree/master/input/ncdc/al)
7	Install and run Apache Pig in Ubuntu so as to work with Hadoop.
8	Explore basic Data Processing operators in Pig.
9	Write Pig Latin program to sort, group, join, project, and filter data (SalesJan2009.csv). Find out Number of Products Sold in Each Country.
10	Write Pig Latin script to count the number of occurrences of each word in an input text file. (Word Count Problem using Pig Script)
11	Install and run HIVE in Ubuntu to work with Hadoop.
12	Explore Hive with its basic commands: create, alter, and drop databases, tables, views, functions and indexes.
13	Install, Deploy and configure Apache Spark.
14	Data analytics using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together and count the frequencies.

Text Books:

T1. J. Leskovec, A. Rajaraman, and J. D.Ullman, ***Mining of Massive Datasets***, 2nd Edition, Cambridge University Press, 2014.
T2. J. Bell, ***Machine Learning for Big Data: Hands-On for Developers and Technical Professionals***, Wiley Publications, 2014.

Reference Books:

R1. J. Han, M. Kamber, and J. Pei, ***Data Mining Concepts and Techniques***, 3rd Edition, Morgan Kaufman Publications, 2011.
R2. T. M. Mitchell, ***Machine Learning***, 1st Edition, McGraw-Hill Education, 2017.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106142/: by Prof. J. Augustine, IIT Madras
2. https://nptel.ac.in/courses/106/104/106104189/: by Dr. R. Misra, IIT Patna
3. http://www.mmds.org: Study material on Mining of Massive Data Sets
4. http://lintool.github.com/MapReduceAlgorithms/index.html

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Install, set up and configure Hadoop.
CO2	Implement problems relating to Big Data using Map Reduce program.
CO3	Analyze bulk data sets using Apache Pig and sort, group, join, project, and filter big data.
CO4	Create, alter, and drop databases, tables, views, functions and indexes using HIVE.
CO5	Install, deploy & configure Apache Spark and perform Data Analytics tasks with it.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	3	3	2	3					2	2	2	2
CO2	3	3	3	2	3					2	2	2	2
CO3	3	3	3	2	3					2	2	2	2
CO4	3	3	3	2	3					2	2	2	2
CO5	3	3	3	2	3					2	2	2	2

Type	Code	Data Visualization & Reporting	L-T-P	Credits	Marks
PC	MSCS-T-PC-024		2-0-0	2	100

Objectives	The objective of this course is to introduce the design principles and techniques for interactively visualizing data and its analysis in graphical manner for proper understanding. This course involves theories, techniques, strategies, and tools for constructing information visually.
Pre-Requisites	Basic knowledge of graphs & charts, and programming in Python/R is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with programming and problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	What is Visualization? The visualization process. Seven stages of data visualization, Types of data, Perception, Eight visual variables: position, shape, size, brightness, color, orientation, texture, motion.	**6 Hours**
Module-2	Visualization Techniques for Multivariate data, Tree, Graph, Networks, Text, Documents.	**5 Hours**
Module-3	Visualization Techniques for Spatial data, Geospatial data, Time-oriented data, Evaluating Visualizations.	**5 Hours**
Module-4	Introduction to D3, Working with data, Data-binding, Data-driven design and interaction, General charting principles, creating an axis, line charts and interpolations, Layouts (Histograms, Pie charts, Stack layout), Visualization with Scalable Vector Graphics (SVG), Drawing, Transformations, Building Charts with SVG.	**6 Hours**
Module-5	Visualization libraries in R/Python: Matplotlib (Histograms, Bar Charts, Line plots, Pie Charts, Box Plots, Scatter Plots), Seaborn (Box, Violin Plots, Regression Plots, Heatmaps), Bokeh, ggplot2, Creating Dashboards with Plotly and Dash.	**6 Hours**
	Total	**28 Hours**

Text Books:

T1. M. Ward, G. Grinstein, and D. Keim, ***Interactive Data Visualization : Foundations, Techniques, and Applications***, 2nd Edition, CRC Press, 2015.
T2. E. Meeks, ***D3.js in Action : Data Visualization with JavaScript***, 2nd Edition, Manning Publications, 2018.
T3. A. C. Telea, ***Data Visualization Principles and Practice***, 2nd Edition, CRC Press, 2015.

Reference Books:

R1. B. Fry, ***Visualizing Data***, O'Reilly Media, 2007.
R2. S. Murray, ***Interactive Data Visualization for the Web***, 2nd Edition, O'Reilly Media, 2017.
R3. K. Sosulski, ***Data Visualization Made Simple : Insights Into Becoming Visual***, Routledge, 2018.
R4. K. Healy, ***Data Visualization : A Practical Introduction***, Princeton University Press, 2019.

Online Resources:

1. https://1lib.in/book/2551564/b19e0b
2. https://1lib.in/book/5216746/8f8e9b
3. https://matplotlib.org/stable/tutorials/index.html
4. https://seaborn.pydata.org/tutorial.html
5. https://docs.bokeh.org/en/latest/docs/gallery.html
6. https://www.r-graph-gallery.com/ggplot2-package.html

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explore the use of data visualization and perceptions.
CO2	Determine appropriate visualization techniques for one, two and multi-dimensional data.
CO3	Create visualizations for complex data and evaluate the visualization techniques.
CO4	Build interactive graphs and charts with D3.js and Scalable Vector Graphics.
CO5	Develop programs using Python/R libraries for interactive data visualization.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO8	Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	1	2	2	1	1			3	2	1	2	1	2
CO2	1	2	3	1	1			3	2	1	1	1	1
CO3	1	3	3	1	1			3	3	1	1	1	1
CO4	1	2	3	1	1			3	2	1	1	1	1
CO5	2	2	2	1	1			2	2	1	2	1	1

Type	Code	Social Media Mining	L-T-P	Credits	Marks
PE	MSCS-T-PE-027		3-0-0	3	100

Objectives	The objective of the course is to study the techniques to model, analyze, and understand large-scale social media along with dynamic processes over social and information networks, and understand the link between qualitative and quantitative methods of social media mining.
Pre-Requisites	Knowledge of algorithms, graph theory and AI is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required, sessions are planned to be interactive with problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	**Introduction**: Social media mining, New challenges for mining, Social networks as graphs, Graph basics, Graph representation, Types of graphs, Connectivity in graphs, Special graphs, Graph algorithms.	**8 Hours**
Module-2	**Network Level Measures and Models**: Centrality, Directed networks and centrality measures, Transitivity and Reciprocity, Balance and Status, Similarity, Properties of Real-world Networks, Random graphs, Small-World Model, Preferential Attachment Model.	**8 Hours**
Module-3	**Data Mining Essentials**: Data, Data preprocessing, Data mining algorithms, Supervised learning, Unsupervised learning; Social Theory – Individual Actors, Social Exchange Theory, Social Forces, Hierarchy of Social Network Motivation.	**8 Hours**
Module-4	**Community Analysis**: Community Detection, Community Evolution, Community Evaluation; Information Diffusion in Social Media – Herd Behavior, Information Cascades, Diffusion of Innovations, Epidemics.	**8 Hours**
Module-5	**Applications**: Measuring Assortative, Influence, Homophily, Distinguishing Influence and Homophily; Recommendation in Social Media – Challenges, Classical Recommendation Algorithms, Recommendation Using Social Context, Evaluating Recommendations; Behavior Analytics – Individual and Collective Behavior.	**10 Hours**
	Total	**42 Hours**

Text Books:

T1. R. Zafarani, M. A. Abbasi, and H. Liu, ***Social Media Mining : An Introduction***, Cambridge University Press, 2014.

T2. I. McCulloh, H. Armstrong, and A. Johnson, ***Social Network Analysis with Applications***, John Willey & Sons, 2013.

Reference Books:

R1. P. Mika, ***Social Networks and the Semantic Web***, Springer, 2007.

R2. G. Xu, Y. Zhang, and L. Li, ***Web Mining and Social Networking : Techniques and Applications***,

Springer, 2011.
R3. M. A. Russell, ***Mining the Social Web : Data Mining Facebook, Twitter, Linkedin, Google+, Github, and More***, 2nd Edition, O'Reilly Media, 2013.
R4. J. Golbeck, ***Analyzing the Social Web***, Morgan Kaufmann, 2013.
R5. S. K. Ravindran, V. Garg, ***Mastering Social Media Mining with R***, Packt Publishing, 2015.

Online Resources:

1. https://nptel.ac.in/courses/106/106/106106169/: by Prof. S. Iyengar and Prof. P. Saini, IIT Ropar
2. https://archive.siam.org/meetings/sdm08/TS1.pdf
3. https://www.coursera.org/learn/social-network-analysis
4. https://www.javatpoint.com/social-media-data-mining

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the basic concepts of social media mining and model social networks as graphs.
CO2	Apply network level measures to appropriately model real-world social networks.
CO3	Analyze large-scale data derived from social media using data mining techniques.
CO4	Apply community analysis and information diffusion in social media.
CO5	Measure influence and homophily in social media, evaluate recommendations, and analyze individual vs. collective behaviour or users in social media context.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	1				2				1	2	2	1	1
CO2	3	2	1	2	3				1	2	2	1	2
CO3	1	1	3	2	2				2	2	1	2	1
CO4	2	2				3			2	3	1	2	2
CO5			3	3	2	1			2	2	1	2	2

Type	Code	Natural Language Processing	L-T-P	Credits	Marks
PE	MSCS-T-PE-028		3-0-0	3	100

Objectives	The objective of this course is to study the fundamentals, algorithms, and techniques to enable processing of natural languages by computers in order to design different human-computer interactive systems.
Pre-Requisites	Knowledge on grammar rules, statistics, automata theory and machine learning techniques is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required. Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to Natural Language Processing, various applications, issues and processing complexities, Regular Expressions, Text Normalization, Edit Distance, N-gram Language Models, Smoothing techniques.	**8 Hours**
Module-2	Naive Bayes and Sentiment Classification, Other text classification task and logistic regression, Vector Semantics and Embeddings, Sequence Labeling for Parts of Speech and Named Entities.	**8 Hours**
Module-3	Deep Learning Architectures for Sequence Processing, Recurrent neural networks applied to language problem, Contextual Embeddings, Machine Translation and Encoder-Decoder Models.	**8 Hours**
Module-4	Constituency Grammars, Constituency Parsing, Dependency Parsing, Logical Representations of Sentence Meaning, Computational Semantics and, Semantic Parsing, Information Extraction.	**8 Hours**
Module-5	Word Senses and WordNet, Coreference Resolution, Discourse Coherence, Question Answering, Chatbots & Dialogue Systems, Automatic Speech Recognition, Text-to-Speech.	**10 Hours**
	Total	**42 Hours**

Text Books:

T1. D. Jurafsky and J. H. Martin, ***Speech and Language Processing – An Introduction to Language Processing, Computational Linguistics, and Speech Recognition***, 3rd Edition, Pearson Education, 2020.
T2. C. D. Manning and H. Schütze, ***Foundations of Statistical Natural Language Processing***, 2nd Edition, MIT Press, 2000.

Reference Books:

R1. T. Siddiqui and U. S. Tiwary, ***Natural Language Processing and Information Retrieval***, 1st Edition, Oxford University Press, 2008.
R2. C. C. Aggarwal, ***Machine Learning for Text***, 1st Edition, Springer, 2018.
R3. J. Allen, ***Natural Language Understanding***, 2nd Edition, Pearson Education, 2008.

Online Resources:

1. https://nptel.ac.in/courses/106/101/106101007/: by Prof. P. Bhattacharyya, IIT Bombay
2. https://nptel.ac.in/courses/106/105/106105158/: by Prof. P. Goyal, IIT Kharagpur
3. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-863j-naturallanguage-and-the-computer-representation-of-knowledge-spring-2003/lecture-notes/

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain the text pre-processing techniques for natural language processing.
CO2	Apply machine learning techniques to text classification task and sequence labeling.
CO3	Apply deep learning for sequence processing and other language processing tasks.
CO4	Perform semantic level analysis on natural language processing applications.
CO5	Perform discourse level analysis and appreciate advanced applications of NLP with applied machine learning techniques.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	2	2	3				2	2	2	2	1
CO2	3	3	3	2	3				3	3	3	3	2
CO3	3	3	3	2	3				3	3	3	3	3
CO4	3	3	3	2	3				3	3	3	3	3
CO5	3	3	3	2	3				3	3	3	3	3

Type	Code	Probabilistic Graphical Models	L-T-P	Credits	Marks
PE	MSBS-T-PE-006		3-0-0	3	100

Objectives	The objective this course is to study representation of complex domains using probability distributions and graph theory concepts that is necessary to solve real world problems in computer vision, NLP, computational biology, finance, etc.
Pre-Requisites	Knowledge of probability & statistics, graph theory, and ML is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; sessions are planned to be interactive with focus on problem solving activities.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	**Representation and Exact Inference**: Introduction to graphical models, Directed graphical models (Bayesian network), Undirected graphical models (Markov Random Fields), Exact inference (variable elimination, sum-product message passing).	**10 Hours**
Module-2	**Learning**: Generalized linear models, Maximum likelihood estimation, Sufficient statistics, Learning in fully observed Bayesian networks, Learning in fully observed Markov network, The expectation maximum algorithm.	**8 Hours**
Module-3	**Popular Graphical Models and Applications**: Modeling networks, Ising models, Gaussian graphical models, Factor analysis, State space models and applications, Case study of conditional random field (Image segmentation of computer vision).	**8 Hours**
Module-4	**Approximate Inference**: Variational inference (loopy belief propagation, mean field approximation), Monte-Carlo methods, Markov Chain Monte-Carlo (MCMC).	**8 Hours**
Module-5	**Non-Parametric Bayesian Models**: Dirichlet processes, Hierarchhial Dirichlet processes, Kernel graphical models, Graph-induced structured input-output methods, Case study – Disease association analysis, Causal inference.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. D. Koller and N. Friedman, ***Probabilistic Graphical Models : Principles and Techniques***, The MIT Press, 2009.
T2. K. P. Murphy, ***Machine Learning : A Probabilistic Prospective***, The MIT Press, 2012.
T3. M. I. Jordan, ***An Introduction to Probabilistic Graphical Models***, Center for Biological and Computational Learning (MIT), 2003.

Reference Books:

R1. M. Wainwright and M. Jordan, ***Graphical Models, Exponential Families, and Variational Inference***, Now Publishers, 2008.
R2. J. Pearl, M. Glymour, and N. P. Jewell, ***Causal Inference in Statistics : A Primer***, John Wiley, 2016.

Online Resources:

1. http://www.cs.cmu.edu/~epxing/Class/10708-14/lecture.html: By Prof. E. Xing, Carnegie Mellon University
2. https://cedar.buffalo.edu/~srihari/CSE674/: By Prof. S. Srihari, University at Buffalo
3. https://ermongroup.github.io/cs228-notes/

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Understand the basic, directed, and undirected graphical models and exact inference.
CO2	Apply learning algorithms of graphical models for solving real life problems.
CO3	Explore popular graphical models and use them to solve practical problems.
CO4	Apply variational inference and monte-carlo methods for solving real life methods.
CO5	Apply non-parametric Bayesian models and casuality to design graphical models.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	2	2	1				2	1	2	3	3
CO2	3	3	3	2	1				2	1	2	3	3
CO3	3	3	3	2	1				2	1	2	3	3
CO4	3	3	3	2	1				2	1	2	3	3
CO5	3	3	3	2	1				2	1	2	3	3

Type	Code	Multimedia Database Systems	L-T-P	Credits	Marks
PE	MSCS-T-PE-029		3-0-0	3	100

Objectives	The objective of the course is to study the fundamental concepts of database management systems designed specifically for storage, retrieval, and manipulation of multimedia data such as images, audio, video etc.
Pre-Requisites	Knowledge of relational database systems is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Review of relational databases, Relational algebra, Relational calculus; Object oriented databases: Objects and values, Types and classes, Methods, Object definition and Query languages, Object-relational systems; Multimedia data structures: k-d trees, Point quadtrees, The MX-quadtree, R-Trees, Comparison of data structures.	**8 Hours**
Module-2	Metadata for multimedia, Classification of metadata, Metadata for text, Speech, Images, Video, Multimedia information modeling: Object oriented modeling, Temporal models, Spatial models, Multimedia authoring.	**6 Hours**
Module-3	Image Databases: Raw images, Compressed image representations, Image segmentation, Similarity-based retrieval, Alternative image DB paradigms, Representing image DBs with Relations and R-Trees, Retrieving images by Spatial Layout; Text/Document databases: Precision and recall, Stop lists, Word stems, Frequency tables, Latent semantic indexing, TV-Trees, Other retrieval techniques.	**10 Hours**
Module-4	Video Databases: Organizing content of a video, Querying content of video libraries, Video segmentation, Video standards; Audio Databases: General model of audio data, Capturing and indexing audio data; Multimedia Databases: Design & architecture of MMDB, Organizing multimedia data, Media abstractions, Query languages for retrieval, Indexing, Query relaxation/expansion.	**10 Hours**
Module-5	Physical Storage & Retrieval: Retrieving multimedia data from disks, CD ROMs, Tape drives, Creating distributed multimedia presentations, Distributed media servers, Optimal distributed retrieval plans, Algorithms to compute optimal retrieval plans.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. V. S. Subrahmanian, ***Principles of Multimedia Database Systems***, 1st Edition, Morgan Kaufmann, 2014.
T2. B. Prabhakaran, ***Multimedia Database Management Systems***, 1st Edition, Springer, 2009.

Reference Books:

R1. L. Dunckley, ***Multimedia Databases: An Object Relational Approach***, 1st Edition, Addison-Wesley, 2002.
R2. V. Castelli and L. D. Bergman (Ed), ***Image Databases: Search and Retrieval of Digital Imagery***, 1st Edition, Wiley-Interscience, 2002.

Online Resources:

1. http://www.ifis.cs.tu-bs.de/teaching/ss-16/mmdb
2. https://docs.oracle.com/cd/B28359_01/appdev.111/b28414/ch_intro.htm

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain and compare different multimedia data structures.
CO2	Understand multimedia models and metadata for multimedia data.
CO3	Analyze, store, index, retrieve and manipulate image and text data.
CO4	Organize, index and query video and audio content from multimedia databases.
CO5	Retrieve multimedia data from various media like CD ROM, Tape drives etc.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	2	1	2				1		2	3	2
CO2	3	2	2	1	2				1		2	3	2
CO3	3	2	2	1	3	1			2	1	3	3	3
CO4	2	2	2	1	3	1			2	1	3	3	3
CO5	2	2	2	1	2	1			2	1	3	3	3

Type	Code	Computer Vision	L-T-P	Credits	Marks
PE	MSCS-T-PE-030		3-0-0	3	100

Objectives	The objective of the course is to introduce algorithms & applications in the field of computer vision including image processing, detection & recognition, geometry-based and physics-based vision and video analysis.
Pre-Requisites	Knowledge of linear algebra, calculus, probability, and programming is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to Computer Vision, Human vision, Color spaces, Transformations, Image formation – Geometric primitives, 2D transformations, 3D transformations, 3D rotations, 3D to 2D projections, Lens distortions, Parametric transformations, Mesh-based warping, Image processing – Point operators, Linear filtering, More neighborhood operators, Fourier transformations.	**9 Hours**
Module-2	Interpolation and Optimization: Pyramids and wavelets, Image blending, Model fitting and optimization – Scattered data interpolation, Variational methods and regularization, Markov random field and application; Convolutional neural networks – Application to digit classification, Network architectures, model zoos, Visualizing weights and activations, Adversarial examples, self-supervised learning; Recognition – Object detection, Semantic segmentation.	**9 Hours**
Module-3	Feature Detection and Matching: Points and patches, Edges and contours with applications, Lines and vanishing points – Hough transformation, Segmentation – Graph-based segmentation, Mean shift, Image alignment and stitching – Pairwise alignment and application, Image stitching and applications.	**8 Hours**
Module-4	Motion Estimation: Translational alignment, Parametric motion and application, Spline-based motion and application, Optical flow – deep learning approaches and application, Layered motion and application; Depth Estimation – Epipolar geometry, Sparse correspondence, Dense correspondence, Local methods, Global optimization, Deep neural networks, Multi-view stereo.	**7 Hours**
Module-5	Structure from Motion and SLAM: Geometric intrinsic calibration, Pose estimation, Two-frame structure from motion, Multi-frame structure from motion, Simultaneous localization and mapping (SLAM), 3D Reconstruction – Shape from X, 3D scanning, Computational photography – Photometric calibration, High dynamic range imaging, Super-resolution, Denoising & blur removal; Image matting and composing, Hole filling and in-painting.	**9 Hours**
	Total	**42 Hours**

Text Books:

T1. R. Szeliski, ***Computer Vision : Algorithms and Applications***, Springer Science & Business Media, 2010.

Reference Books:

R1. D. A. Forsyth, J. Ponce, ***Computer Vision : A Modern Approach***, 2nd Edition, Pearson Education, 2012.
R2. R. Hartley and A. Zisserman, ***Multiple View Geometry in Computer Vision***, 2nd Edition, Cambridge University Press, 2005.
R3. R. C. Gonzalez and R.E. Woods, ***Digital Image Processing***, 4th Edition, Pearson Education, 2017.

Online Resources:

1. https://nptel.ac.in/courses/106/105/106105216/: by Prof. J. Mukhopadhyay, IIT Kharagpur
2. https://nptel.ac.in/courses/108/103/108103174/: by Prof. M. K. Bhuyan, IIT Guwahati
3. https://nptel.ac.in/courses/106/106/106106224/: by Prof. V. N. Balasubramanian, IIT Madras
4. https://www.cs.cornell.edu/courses/cs6670/2011sp/lectures/lectures.html
5. http://web.stanford.edu/class/cs223b/syllabus.html
6. https://www.cs.cmu.edu/~16385/

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explain color spaces, image transformation, wrapping, and basics of image processing.
CO2	Explore machine learning and deep neural networks for solving computer vision problems.
CO3	Apply different techniques used for object detection, image segmentation and stitching.
CO4	Develop deep learning methods for motion estimation, depth estimation, and stereo vision.
CO5	Design methods for SLAM, 3D image reconstruction and computational photography.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	3	2	1	2	2					1	2	3	1
CO2	3	3	3	3	2					1	2	3	1
CO3	2	2	3	2	2					1	2	3	1
CO4	2	2	3	3	1					1	2	3	1
CO5	2	2	2	2	1					2	2	3	1

Type	Code	Realtime Analytics	L-T-P	Credits	Marks
PE	MSCS-T-PE-031		3-0-0	3	100

Objectives	The objective of the course is to study the techniques, tools, and platforms to analyze and visualize real-time streaming data for data science applications.
Pre-Requisites	Basic knowledge of data processing, database systems and big data is required.
Teaching Scheme	Regular classroom lectures with use of ICT as and when required; Sessions shall be interactive with focus on problem solving and real-life examples.

Evaluation Scheme

Teacher's Assessment			Written Assessment		Total
Quiz	Surprise Test(s)	Assignment(s)	Mid-Term	End-Term	
05	05	05	25	60	100

Detailed Syllabus

Module-#	Topics	Hours
Module-1	Introduction to streaming data, Sources, Why streaming data is different, Components and features of real-time streaming architectures, Languages for real-time programming, Real-time architecture checklist.	**8 Hours**
Module-2	Service Configuration and Coordination: Motivation for configuration and coordination systems, Maintaining distributed state, Apache ZooKeeper, Data-flow Management in Streaming Analysis: Distributed data flows, Apache Kafka.	**8 Hours**
Module-3	Apache Flume, Distributed streaming data processing, Processing data with Storm and Samza.	**8 Hours**
Module-4	Storing Streaming Data: Consistent Hashing, NoSQL storage systems, Other storage technologies, Warehousing.	**10 Hours**
Module-5	Delivering Streaming Metrics: Streaming Web Applications, Visualizing data, Mobile streaming applications.	**8 Hours**
	Total	**42 Hours**

Text Books:

T1. B. Ellis, ***Real-Time Analytics : Techniques to Analyze and Visualize Streaming Data***, 1st Edition, John Wiley & Sons, 2014.

Reference Books:

R1. S. Gupta, ***Real-Time Big Data Analytics***, Packt Publishing, 2016.
R2. S. Saxena and S. Gupta, ***Practical Real-time Data Processing and Analytics***, Packt Publishing, 2017.

Online Resources:

1. https://zookeeper.apache.org/doc/r3.7.0/index.html
2. https://kafka.apache.org/documentation/
3. https://flume.apache.org/documentation.html
4. https://storm.apache.org/releases/2.2.0/index.html
5. http://samza.apache.org/learn/documentation/1.6.0/core-concepts/core-concepts.html

P.T.O

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Justify infrastructural decisions, computational approaches, software, frameworks and methods for dealing with real-time streaming architectures.
CO2	Explain systems and techniques used in the management of shared state and coordination between distributed processes.
CO3	Describe the frameworks like Storm and Samza for implementing stream processing systems.
CO4	Determine appropriate storage options for processing live streaming data analysis systems.
CO5	Develop web-based application for delivering, analyzing and visualizing real-time data.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO6	Function effectively both as a leader and team member on multi disciplinary projects to demonstrate computing and management skills.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	2	2	1	2	2					2	3	2	1
CO2	2	2	2	3						1	2	2	2
CO3	3	3	1	1	1	1				1	2	2	2
CO4	1	1	1	2	2					1	2	2	2
CO5	2	2	2	1	1	1				2	1	2	2

Type	Code	Data Visualization & Reporting Lab	L-T-P	Credits	Marks
PC	MSCS-P-PC-025		0-0-2	1	100

Objectives	The objective of this laboratory course is to provide hands on exposure on implementation of different data visualization tools & platforms and create interactive plots for better storytelling with data.
Pre-Requisites	Knowledge of MS Excel, JavaScript, Python, and R programming is required.
Teaching Scheme	Regular laboratory experiments executed by the students under supervision of the teacher. Experiments shall comprise of programming assignments.

Evaluation Scheme

Attendance	Daily Performance	Lab Record	Lab Test/ Mini Project	Viva-voce	Total
10	30	15	30	15	100

Detailed Syllabus

Experiment-#	Assignment/Experiment
1	Excel: Use Excel charts for presentation of quantitative data.
2	Excel: Presentation of qualitative data.
3	Tableau: Import and prepare data, Perform data cleansing and harmonization.
4	Tableau: Create data visualizations.
5	Tableau: Use advanced data visualization to discover trends in data sets.
6	D3: Creating lines and circles with select and append.
7	D3: Loading data, displaying it as a bar chart and Creating a scatterplot.
8	D3: Draw histograms, violin plots, pie charts and ring charts.
9	Matplotlib: Basic plotting with matplotlib library.
10	Seaborn: Createregression plots and heatmap using seaborn library.
11	Bokeh: Create few standalone interactive plots.
12	ggplot: Data visualization with ggplot2.
13	Plotly: Build a dashboard with Plotly and Dash.
14	Mini Project.

Text Books:

T1. H. Guerrero, ***Excel Data Analysis: Modeling and Simulation***, Springer, 2010.
T2. B. Jones, ***Communicating Data with Tableau***, O'Reilly Media, 2014.
T3. E. Meeks, ***D3.js in Action : Vata visualization with JavaScript***, Manning Publications, 2017.
T4. T. Rahlf, ***Data Visualization with R – 111 Examples***, 2nd Edition, Springer, 2019.

Reference Books:

R1. A. Pajankar, ***Practical Python Data Visualization: A Fast Track Approach to Learning Data Visualization with Python***, 1st Edition, Apress, 2020.
R2. L. Ryan, ***Visual Data Storytelling with Tableau***, 1st Edition, Addison-Wesley, 2018.
R3. K. Healy, ***Data Visualization: A Practical Introduction***, Princeton University Press, 2018.
R4. E. Pimpler, ***Introduction to Data Visualization and Exploration with R***, GeoSpatial Training Services, 2017.

Online Resources:

1. https://realpython.com/tutorials/data-viz/
2. http://www.datavisualisation-r.com/
3. https://matplotlib.org/stable/tutorials/index.html
4. https://www.tableau.com/learn/training/20212
5. https://seaborn.pydata.org/
6. https://docs.bokeh.org/en/latest/docs/gallery.html
7. https://www.r-graph-gallery.com/ggplot2-package.html

Course Outcomes: *At the end of this course, the students will be able to:*

CO1	Explore and use Excel charts for data representation.
CO2	Discover trends in data sets by using advanced data visualization.
CO3	Design and draw interactive plots using popular libraries.
CO4	Integrate graphs and charts with the web frameworks.
CO5	Create dashboards for better storytelling with data.

Program Outcomes Relevant to the Course:

PO1	Apply knowledge of mathematics, computing, analytics, and domain knowledge appropriate for creating computational models for defined problems and requirements.
PO2	Develop efficient applications to analyze data and make predictions for taking timely business decisions.
PO3	Design solutions for analysis of huge data with considerations towards societal and environmental aspects.
PO4	Assess the security & privacy aspects in storage, transmission, and analysis of large amounts of critical business information.
PO5	Integrate diverse IT tools and apply them efficiently for real-world data analysis applications.
PO8	Communicate effectively and present technical information in oral and written reports supported by graphs & charts for easy visualization.
PO9	Review research literature and conduct independent research in data science to develop advanced algorithms, techniques and tools.
PO10	Recognize the need for and engage in continuous lifelong learning to enhance the knowledge & skills in data science.

Mapping of COs to POs and PSOs (1: Low, 2: Medium, 3: High)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1	1	2	1	1	1			3	2	1	1	2	2
CO2	1	2	2	1	1			3	2	1	1	2	3
CO3	1	2	2	1	1			3	1	1	1	2	2
CO4	1	3	2	1	3			2	2	1	1	2	2
CO5	1	2	2	1	2			2	2	1	1	2	2

Department of Computer Science & Engineering
Silicon Institute of Technology
Silicon Hills, Patia, Bhubaneswar - 751024