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: 1.10 (20-03-2021)

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.

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.

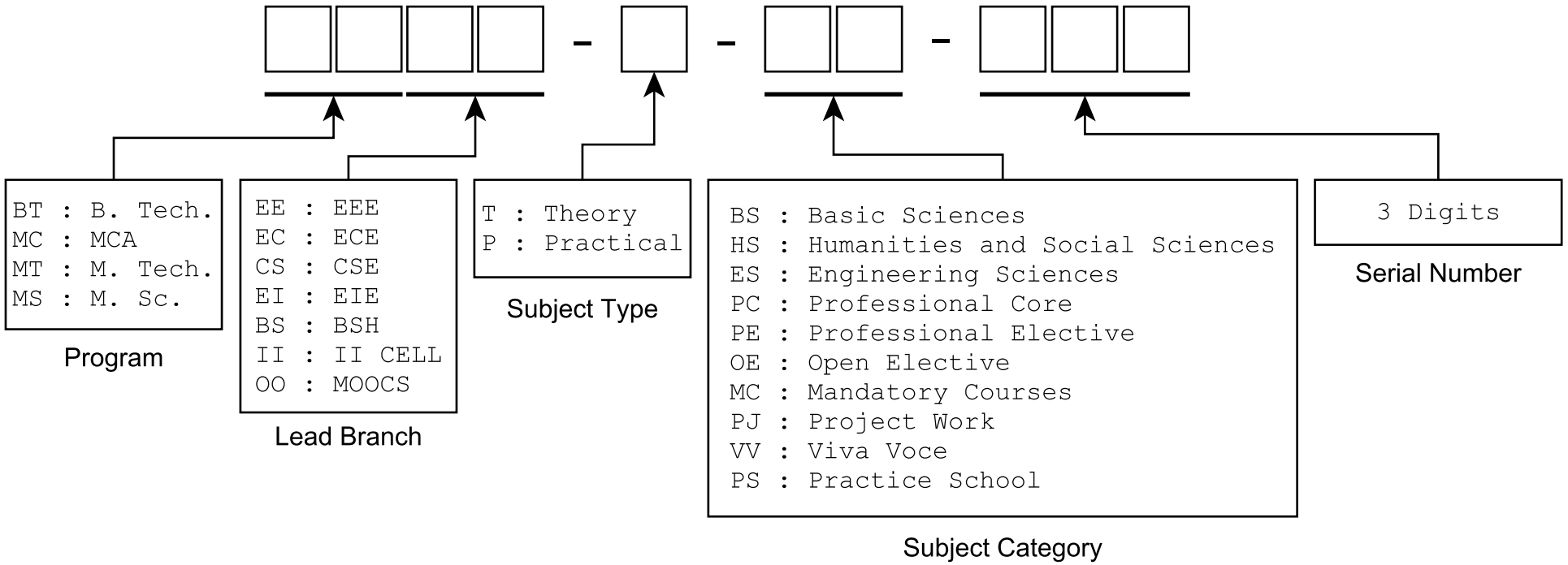
PEO2. To work independently or in a diverse team with effective communication in interdisciplinary environment, and demonstrate leadership in industry and academia.

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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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 (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	3	0	0	3	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	4	0	0	2
PJ	MSCS-P-PJ-032	Capstone Project - I	0	0	4	0	0	2
PJ	MSII-P-PJ-002	Summer Internship	0	0	0	0	0	1
		SUB-TOTAL	**14**	**0**	**12**	**14**	**0**	**7**
		TOTAL	**26**			**21**		

Semester IV									
Type	**Code**	**Course Title**	**WCH** L-T-P			**Credits** L-T-P		
THEORY									
PC	MSCS-T-PC-024	Data Visualization & Reporting	3	0	0	3	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	4	0	0	2
PJ	MSCS-P-PJ-033	Capstone Project - II	0	0	16	0	0	8
		SUB-TOTAL	**9**	**0**	**20**	**9**	**0**	**10**
		TOTAL	**29**			**19**		

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	Bioinformatic Algorithms
MSCS-T-PE-022	Bio-Medical Imaging
MSCS-T-PE-023	Healthcare Analytics
Professional Elective-III	
MSCS-T-PE-027	Social Media Analysis
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	Business Data Analytics
MSCS-T-PE-031	Realtime Analytics