Nirma University

Institute of Technology

School of Technology

Department of Computer Science and Engineering

M Tech in Computer Science and Engineering (Data Science)

L	T	P	C
3	0	2	4

Course Code	3CS1112
Course Name	Advanced Database Systems

Course Learning Outcomes (CLO):

At the end of the course, students will be able to –

assess various storage and retrieval methods through appropriate indexing
design and analyze efficiency of algorithms for database operations
comprehend contemporary database architectures and its relevant issues

Syllabus:	**Teaching Hours:**
Unit I **Data storage:** Overview of RDBMS concepts**,** Basic File Structures, File Organization & Record formats, Heap sorted & Hashed Files, Buffer management, Disk Storage, Parallel Disk access with RAID, Modern Storage Architectures	**5**
Unit II **Indexing Structures:** Single level and Multilevel Indexes, B Tree and B+ Tree Indexes, Hash and bitmap based indexing, Index Structures for Single Dimensional and Multidimensional Databases	**8**
Unit III **Query Processing**: Query Execution, Algebra for Queries, Physical-Query-Plan-Operators, Algorithms for Database Operations, Algorithms for Joins andSorting, hash and index based algorithms, Buffer Management, Parallel Algorithms for Relational Operators	**9**
Unit IV **Query Optimization**: Algebraic Foundation for Improving Query Plans, Estimating Cost of Operations, Cost Based Plan Selection, Choosing Order of Joins, Optimization of Queries for Parallel, Distributed, Multidimensional and Text Database	**8**
Unit V **Transactions, Concurrency control and Recovery:** Transaction scheduling, serializability, Coping with System Failure, Concurrency Control techniques with locking, timestamp ordering and multiversion, Redo and Undo log based recovery, recovery in multi database systems	**7**
Unit VI **Advances in database systems:**Distributed database systems, fragmentation, replication and allocation techniques, NoSQL based systems: key-value based, document based, column based and Graph databases, Streaming SQL,Introduction to active, temporal, spatial, multimedia and deductive databases	**8**

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

RamezElmasri, Shamkant B Navathe, Fundamentals of Database System, Pearson Education
Garcia Molina, Ullman, Widom, Data Base System Implementation, Pearson education
Raghu Ramakrishnan& Johannes Gehrke, Database Management Systems, McGraw Hill
Silberschatz, Korth, Sudarshan, Database System Concepts, McGraw Hill
M.TamerOzsu, Patrick Valduriez, S.Sridhar, Principles of Distributed Database Systems, Pearson Education

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS1111
Course Name	Applied Machine Learning

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to –

comprehend statistical methods as basis of machine learning domain
apply and evaluate variety of machine learning algorithms
implement machine learning techniques to solve problems in interdisciplinary domains

Syllabus:

Teaching Hours

Unit I — 3

Introduction: Motivation and Applications, Basics of Supervised and Unsupervised Learning

Unit II — 13

Regression Techniques: Basic Concepts and applications of Regression, Simple Linear Regression – Gradient Descent and Normal Equation Method, Multiple Linear Regression, Non-Linear Regression, Linear Regression with Regularization, Hyper-parameters tuning, Loss Functions, Decision Tree Regression, Evaluation Measures for Regression Techniques

Unit III — 10

Classification Techniques: Naïve Bayes Classification: Fitting Multivariate Bernoulli Distribution, Gaussian Distribution and Multinomial Distribution, K-Nearest Neighbours, Classification Trees, Linear Discriminant Analysis, Support Vector Machines: Hard Margin and Soft Margin, Kernels and Kernel Trick, Evaluation Measures for Classification Techniques

Unit IV — 9

Artificial Neural Networks: Biological Neurons and Biological Neural Networks, Perceptron Learning, Activation Functions, Multilayer Perceptrons, Back-propagation Neural Networks, Learning with Momentum, Winner-take-all Learning, Competitive Neural Networks, Adaptive ANN

Unit V — 4

Clustering: Hierarchical Agglomerative Clustering, k-means Algorithm, Self-Organizing Maps

Unit VI — 6

Advances in Machine Learning: Basics of Semi-Supervised and Reinforcement Learning, Introduction to Deep Learning, Best Practices for Machine Learning, Case Studies in interdisciplinary domain

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

C. Bishop, Pattern Recognition and Machine Learning, Springer
R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification and Scene Analysis, Wiley
KishanMehrotra, Chilukuri Mohan and Sanjay Ranka, Elements of Artificial Neural Networks, Penram International
Tom Mitchell, Machine Learning, TMH
RajjanShinghal, Pattern Recognition, Techniques and Applications, OXFORD
AthemEalpaydin, Introduction to Machine Learning, PHI
Andries P. Engelbrecht, Computational Intelligence - An Introduction, Wiley Publication
Andrew Kelleher, Adam Kelleher, Applied Machine Learning for Data Scientist and Software engineers, Addison-Wesley Professional

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	0	3

Course Code	3CS1113
Course Name	Applied Mathematics for Computer Science

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to –

comprehend the mathematical fundamentals related to sets, probability, statistics, linear algebra and mathematical optimization
apply the mathematical principles to solve wide range of problems in computer science
use the mathematical concepts as per the need of the application

Syllabus:	**Teaching Hours**
Unit I **Review of Linear Algebra**: Matrices, Vectors properties, Eigenvalues and eigenvectors, Matrix factorizations, Distance measures, Projections, Notion of hyperplanes, Half-planes, Application for Linear Algebra in Computer Science	**8**
Unit II **Probability, Statistics and Random Processes**: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate), Expectations and moments, Covariance and correlation, Confidence intervals, Correlation functions, Random walks, Markov chains, Statistical inference, Applications in Regression and Classifications.	**16**
Unit III **Optimization**: Basic Concepts, Linear Programming, Duality, Constrained and unconstrained optimization, gradient decent and non-gradient techniques, Introduction to least squares optimization, optimization in Practice.	**12**
Unit IV **Advanced topics:** Nonlinear dimensionality reduction methods, PCA in high dimensions and random matrix theory (Marcenko-Pastur), Linear Discriminant Analysis, Non-Negative Matrix Factorization, Hypothesis testing, Proof Techniques, Random Graphs	**9**

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Suggested Readings^:

Gilbert Strang, Introduction to Linear Algebra, Cambridge Press.
Gilbert Strang, Linear Algebra and its applications, Harcourt, Brace, Jovanovich Publishers
Douglas C. Montgomery, George C. Runger, Applied Probability and Statistics for Engineers, Wiley
M. Mitzenmacher and E. Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press
Sheldon Ross, A first course in Probability, Pearson
Cathy O'Neil and Rachel Schutt, Doing Data Science, O'Reilly Media
Avrim Blum, John Hopcroft, and RavindranKannan, Foundations of Data Science, e-book, Cornell University
Afonso S. Bandeira, Ten Lectures and Forty-Two Open Problems in the Mathematics of Data Science, e-book, MIT OCW
Jeff M. Phillips, Mathematical Foundations for Data Analysis, e-book, University of Utah
O. Paneerselvam, Operational Research, PHI

L=Lecture, T=Tutorial, P=Practical, C=Credit

---

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS1109
Course Title	Complexity Theory and Algorithms

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to -
comprehend time & space complexity and formal aspects of algorithms
identify appropriate data structures and methodologies for efficient algorithm design
design and implement efficient algorithms using various approaches

Syllabus:	**Teaching Hours:**
Unit I	**6**
Mathematical Preliminaries of computational complexity: Asymptotic Notations, Proof of correctness, Performance analysis, Recursive Algorithms and Recurrences	
Unit II	**8**
Complexity Theory: Various complexity classes, linear reductions. Probabilistic algorithms, Approximation algorithms and complexity classes relating to Parallel algorithms	
Unit III	**6**
Data Structures: Hash tables, Binomial heaps , Fibonacci heaps, Disjoint set structures	
Unit IV	**12**
Greedy Algorithms: Making change, graphs and minimum spanning tree, Shortest path, Knapsack problem, Scheduling, etc. **Divide and Conquer**: General Template, Various algorithm implementation like Binary search, Merge Sort, Quick Sort, Convex Hull, Matrix multiplication, etc.	
Unit V	**6**
Dynamic Programming: Introduction of Dynamic Programming, Principle of Optimality, Examples like Single source shortest paths, Knapsack problem, Chained matrix multiplication, Longest Common Subsequence, etc.	
Unit VI	**7**
Graph Algorithms: Elementary algorithms, DFS, BFS, Backtracking, and Branch & Bound techniques with related examples	

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 10 experiments to be incorporated.

Suggested Readings^:

Gilles Brassard and Paul Bratley, Fundamentals of Algorithmics, PHI Publication.
Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest& Clifford Stein, Introduction to Algorithms, PHI Publication.
Ellis Horowitz, SartajSahni,SanguthevarRajasekaran, Fundamentals of Computer Algorithms, University Press
Jean-Paul Tremblay and Paul G. Sorenson, An Introduction to Data Structures with Applications, Tata McGraw Hill
Robert L. Kruse, Data Structures and Program Design in C, PHI

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS4101
Course Title	Introduction to Scalable Systems

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to –
comprehend the distributed computing models for scalable systems
analyse the scalable systems in the context of various performance parameters
apply concepts of scalable systems in designing data intensive applications

Syllabus:	**Teaching hours:**
Unit I	**5**
Introduction and Architectures: Features and types of distributed systems, Distributed Models of Computation, Architectures: Regular graphs, random graphs, power-law, and small-world networks, middleware, and self-management in distributed systems, causality & logical time	
Unit II	**5**
Distributed System Models and Enabling Technologies: Scalable Computing over the Internet, Technologies for Network-Based Systems, System Models for Distributed and Cloud Computing, Software Environments for Distributed Systems and Clouds, Performance, Security, and Energy Efficiency	
Unit III	**5**
Consistency and Replication: Consistency models: strong and weak, Scalable Causal Consistency, Highly reliable distributed coordination with Zookeeper, Replication management, Distributed Replication, PAXOS and RAFT Algorithms	
Unit IV	**6**
Computer Clusters for Scalable Parallel Computing: Clustering for Massive Parallelism, Computer Clusters and MPP Architectures, Design Principles of Computer Clusters, Cluster Job and Resource Management, Case Studies of Top Supercomputer Systems	
Unit V	**7**
Cloud Platform Architecture over Virtualized Data Centers: Cloud Computing and Service Models, Data-Center Design and Interconnection Networks, Architectural Design of Compute and Storage Clouds, Public Cloud Platforms: GAE, AWS, Azure, Inter-Cloud Resource Management, Cloud Security and Trust Management	
Unit VI	**6**

Cloud Programming and Software Environments: Features of Cloud and Grid Platforms, Parallel and Distributed Programming Paradigms, Programming Support of Google App Engine, Programming on Amazon AWS and Microsoft Azure, Emerging Cloud Software Environments

Unit VII **6**

Fault Tolerance and Security: Failure models, failure detection, algorithms for fault tolerance, and recovery from failure in distributed systems, Authentication in Distributed Systems, Distribution of security mechanisms, access control, and security management

Unit VIII **5**

File Systems and Distributed Storage Systems: Network File System, Andrew File system, Google File System, Hadoop Distributed File System

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Kai Hwang, Geoffrey C. Fox and Jack J. Dongarra, Distributed and Cloud Computing from Parallel Processing to the Internet of Things, Elsevier

Andrew S. Tanenbaum and Maarten van Steen, Distributed Systems: Principles and Paradigms, Createspace

George Coulouris, Jean Dollimore, Tim Kindberg, Gordon Blair, Distributed Systems: Concepts and Design, Addison Wesley

Ajay D. Kshemkalyani and MukeshSinghal, Distributed Computing: Principles, Algorithms, and Systems, Cambridge University Press

Kenneth P Birman, Guide to Reliable Distributed Systems: Building High-Assurance Applications and Cloud-Hosted Services, Springer

Relevant research papers in the area of distributed systems

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
2	0	2	3

Course Code	3CS42D201
Course Name	Analytics for the IoT

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

implement the architectural components and protocols for application development
identify data analytics and data visualization tools as per the problem characteristics
collect, store and analyse IoT data

Syllabus: Teaching Hours

Unit I **2**

Introduction to IoT, applications, IoT architectures, introduction to analytics, IoT analytics challenges

Unit II **7**

IoT devices, Networking basics, IoT networking connectivity protocols, IoT networking data messaging protocols, Analyzing data to infer protocol and device characteristics

Unit III **5**

IoT Analytics for the Cloud: Introduction to elastic analytics, Decouple key components, Cloud security and analytics, Designing data processing for analytics, Applying big data technology to storage

Unit IV **6**

Exploring IoT Data: Exploring and visualizing data, Techniques to understand data quality, Basic time series analysis, Statistical analysis

Unit V **5**

Data Science for IoT Analytics: Introduction to Machine Learning, Feature engineering with IoT data, Validation methods, Understanding the bias–variance tradeoff, Use cases for deep learning with IoT data

Unit VI **5**

Strategies to Organize Data for Analytics: Linked Analytical Datasets, Managing data lakes, data retention strategy

Self Study:

The self study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Minteer, Andrew, Analytics for the Internet of Things (IoT), Packt Publishing Ltd.
Kai Hwang, Min Chen, Big-Data Analytics for Cloud, IoT and Cognitive Computing, Wiley
Hwaiyu Geng, Internet of Things and Data Analytics Handbook, Wiley
John Soldatos, Building Blocks for IoT Analytics Internet-of-Things Analytics, RiverPublishers
Gerardus Blokdyk, IoT Analytics A Complete Guide, 5starcooks

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS12D301
Course Name	Big Data Systems

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

analyse the big data analytic techniques for business applications.
manage big data using different tools and frameworks.
design efficient algorithms for mining the data from large volumes.
implement the HADOOP and MapReduce technologies associated with big data analytics

Syllabus — Teaching Hours

Unit I — **5**

Introduction to Big Data: Introduction to Big Data Platform, Challenges of Conventional Systems, Intelligent Data Analysis, Nature of Data, Analytic Processes and Tools, Analysis vs Reporting, Modern Data Analytic Tools, Statistical Concepts: Sampling Distributions, Re-Sampling, Statistical Inference - Prediction Error

Unit II — **10**

The Big data technology landscape : NoSQL, Types of No SQL databases, SQL Vs No SQL, why No SQL, Introduction to MongoDB, Data Types in MongoDB, CRUD, Practice examples, Apache Cassandra, Features of Cassandra, CRUD operations

Unit III — **10**

Hadoop: History of Hadoop, The Hadoop Distributed File System, Components of Hadoop, Analysing the Data with Hadoop, Scaling Out, Hadoop Streaming, Design of HDFS, Java Interfaces to DFS Basics, Developing a Map Reduce Application, How Map Reduce Works, Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features, Hadoop ecosystem.

Unit IV **10**

Hadoop Environment: Setting up a Hadoop Cluster, Cluster Specification, Cluster Setup and Installation, Hadoop Configuration, Security in Hadoop, Administering Hadoop, HDFS, Monitoring, Maintenance, Hadoop benchmarks, Hadoop in the Cloud.

Unit IV **10**

Frameworks: Applications on Big Data Using Pig and Hive, Data Processing Operators in Pig, Hive Services, HiveQL, Querying Data in Hive, Fundamentals of HBase and ZooKeeper, IBM Info Sphere Big Insights and Streams, Visualizations, Visual Data Analysis Techniques, Interaction Techniques, Systems and Applications

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 6 experiments to be incorporated.

Suggested Readings^:

Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
Tom White, Hadoop: The Definitive Guide, Third Edition, O'reilly Media
Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Publishing
Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press
Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & sons
Glenn J. Myatt, Making Sense of Data, John Wiley & Sons
Pete Warden, Big Data Glossary, O'Reilly
Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, Second Edition, Elsevier
Da Ruan, Guoquing Chen, Etienne E. Kerre, Geert Wets, Intelligent Data Mining, Springer
Paul Zikopoulos, Dirk deRoos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corrigan, Harness the Power of Big Data: The IBM Big Data Platform, Tata McGraw Hill Publications
Michael Minelli, Michele Chambers, Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley Publications
Zikopoulos, Paul, Chris Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Tata McGraw Hill Publications
Seema Acharya and Subhashini C, Big Data and Analytics, Wiley India

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

Course Code	3CS12D201
Course Name	Blockchain Technology

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

comprehend the structure of a Blockchain networks
evaluate security issues relating to Blockchain and cryptocurrency
design and analyze the applications based on Blockchain technology

Syllabus:	**Teaching Hours**
Unit I	
Introduction to Blockchain: History, Digital Money to Distributed Ledgers, Design Primitives, Protocols, Security, Consensus, Permissions, Privacy	**3**
Unit II	
Blockchain Architecture, Design and Consensus: Basic crypto primitives: Hash, Signature, Hashchain to Blockchain, Basic consensus mechanisms, Requirements for the consensus protocols, PoW and PoS, Scalability aspects of Blockchain consensus protocols	**8**
Unit III	**9**
Permissioned and Public Blockchains: Design goals, Consensus protocols for Permissioned Blockchains, Hyperledger Fabric, Decomposing the consensus process, Hyperledger fabric components, Smart Contracts, Chain code design, Hybrid models (PoS and PoW)	
Unit IV	**6**
Blockchain cryptography: Different techniques for Blockchain cryptography, privacy and security of Blockchain, multi-sig concept	
Unit V	

Recent trends and research issues in Blockchain: Scalability, secure cryptographic protocols on Blockchain, multiparty communication, FinTech and Blockchain applicabilities **4**

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Narayanan, Arvind,. et al, Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton University Press.
Wattenhofer, Roger, The science of the blockchain, CreateSpace Independent Publishing Platform
Bahga, Arshdeep, and Vijay Madisetti,. Blockchain Applications: A Hands-on Approach, VPT
Nakamoto, Satoshi, Bitcoin: A peer-to-peer electronic cash system, Research Paper
Antonopoulos, Andreas M, Mastering Bitcoin: Programming the open blockchain, O'Reilly Media, Inc
Diedrich, Henning, Ethereum: Blockchains, digital assets, smart contracts, decentralized autonomous organizations, Wildfire Publishing (Sydney)

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS42D105
Course Name	Data Mining and Visualization

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

identify a number of common data domains and corresponding analysis tasks, including multivariate data, networks, text and cartography
comprehend the key processes of data mining, data warehousing and knowledge discovery process
implement data mining techniques to solve problems in other disciplines in a mathematical way
exercise building and evaluating visualization systems

Syllabus: Teaching Hours

Unit I 10

Data Understanding: types of data, information and uncertainty, classes and attributes, interactions among attributes, relative distributions, summary statistics.

Data Visualization: using different tools - refine data and create, edit, alter, and display their visualizations (x-y graph, bar chart, pie chart, cube etc)

Data Quality: inaccurate data, sparse data, missing data, insufficient data, imbalanced data

Social Challenges: data ownership, data security, ethics and privacy

Unit II 15

Data Reduction and Feature Enhancement: standardizing data, sampling data, using principal components to eliminate attributes, limitations and pitfalls of principal component analysis (PCA), curse of dimensionality

Clustering: dissimilarity and scatter, categorization, k-means clustering, hierarchical clustering, distance measures, shape of clusters, determining the number of clusters, evaluating clusters

Association Analysis: association rule learning, the Apriori algorithm, FP-Growth, market basket analysis

15

Unit III

Machine Learning Algorithms for Data Mining: Regression, review of linear regression, assumptions underlying linear regression. Classification, supervised categorization, linear classifiers, logistic regression, regression trees, classification trees, Bayes' Theorem, Na¨ive Bayes, support vector machines, confusion matrices, receiver operating characteristic (ROC) curves, precision and recall, lift curves, cost curves

Model Selection and Validation: training error and optimism, the Bayes error rate, inductive bias, the bias-variance trade off, overfitting, Occam's Razor, minimum description length (MDL), sampling bias, the validation set approach, leave-one-out cross-validation, k-fold cross-validation, boot strapping, jack knifing, data snooping

Ensemble Learning: bootstrap aggregating (bagging), boosting, stacking/blending, random subspaces, random forests.

Unit IV **5**

Recommender Systems, Reinforcement Learning, Active Learning, Semi-supervised Learning, Transfer Learning, Deep Learning, Data Stream Mining

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Jiahei Han & Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kaufmann
Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson
Wes McKinney, Python for Data Analysis, Oreilly
S. Nagabhushana, Data Warehousing OLAP and Data Mining, New Age publishers

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS12D302
Course Name	Deep Learning and Applications

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

comprehend the strengths and weaknesses of deep networks

analyze suitability of different deep networks for variety of problems

design and implement deep networks for solving problems pertaining to computer science and interdisciplinary research

Syllabus: Teaching Hours

Unit I

Applications of Convolutional Neural Networks: Understanding CNN, Image Classification with Localization, Object Detection, Semantic Segmentation, Instance Segmentation — **10**

Unit II — **13**

Applications of Recurrent and Recursive Neural Networks: Understanding Recurrent and Recursive Neural Networks, Word Embedding, Language Models, Named-Entity Recognition, Machine Translation, Parsing, Sentiment Analysis, Speech Recognition

Unit III — **13**

Specific Applications: Image Captioning, Video Captioning, Document Classification, Prediction in Stock Markets, Recommender Systems, Dimensionality Reduction, Image Denoising, Anomaly Detection from Video, Face and Facial Expression Recognition

Unit IV **9**

Generative Applications: Understanding Generative Adversarial Networks, Image Inpainting, Image Super Resolution, Colorization of Black and White Images, Human Face Generation, Text2Image, Music Generation

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press

Adam Gibson, Josh Patterson, Deep Learning, O'Reilly Media, Inc.

Francois Chollet, Deep Learning with Python, Manning Publications

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
2	0	2	3

Course Code	3CS4201
Course Name	Exploratory Data Analysis

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

comprehend the basic concepts of probability and statistics and their need in engineering
apply concepts and methods of probability and statistics in simulation and modeling of various computer science problems
perform probabilistic and statistical analysis of data related to computer science research and projects

Syllabus: Teaching Hours

Unit I **3**

The Role of Statistics in Engineering: The Engineering Method and Statistical Thinking, Collecting Engineering Data, Mechanistic and Empirical Models, Probability and Probability Models

Unit II **10**

Probability and Probability Distributions: Independence, Bayes Theorem, Discrete& Continuous Random Variables, Probability Mass and Density Functions, Cumulative Distribution Functions, Mean and Variance of a Random Variable, Discrete& Continuous Uniform Distribution, Binomial Distribution, Geometric and Negative Binomial Distributions, Hypergeometric Distributions, Poisson Distributions, Normal Distribution, Normal Approximation to the Binomial and Poisson Distributions, Exponential Distributions, Erlang and Gamma Distributions, Lognormal Distribution, Two or more Random Variables, Covariance and Correlation, Multinomial and Bivariate Normal Distributions

Unit III **5**

Descriptive Statistics and Point Estimation of Parameters: Numerical Summaries of Data, Frequency Distributions and Histograms, Box and Probability Plots, Point Estimation, Sampling Distributions and the Central Limit Theorem, Methods of Point Estimation

Unit IV **6**

Statistical Intervals for a Single Sample: Confidence Interval on the mean of a Normal Distribution, Confidence Interval on the Variance and Standard Deviation of a Normal Distribution, Large-Sample Confidence Interval for a Population Proportion, Guidelines for Constructing Confidence Intervals, Tolerance and Prediction Intervals

Unit V **6**

Hypothesis Testing: Statistical Hypothesis, P-Values in Hypothesis Test, Tests on the Mean of a Normal Distribution, Tests on the Variance and Standard Deviation of a Normal Distribution, Tests on a Population Proportion, Testing for Goodness of Fit, Chi Square Test for Nominal Values.

Self Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Applied Statistics and Probability for Engineers, Douglas C. Montgomery and George C. Runger, Wiley
Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, J. Susan Milton and Jesse Arnold, McGraw Hill Education
Statistics in Plain English, Timothy C. Urdan, Routledge
Introduction to Probability, Bertsekas, Dimitri and John Tsitsiklis, Athena Scientific
Fundamentals of Applied Probability Theory, Alvin Drake. McGraw-Hill
A First Course in Probability, Sheldon Ross, Prentice Hall
Introductory Statistics with Randomization and Simulation, David M Diez, Christopher D Barr, Mine C etinkaya-Rundel, Openintro

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
-	-	10	5

Course Code	**3CS4202**
Course Title	**Minor Project**

Course Outcomes (COs):

At the end of the course, students will be able to -

identify the issues related with the recent trends in the field of computer science and its applications

formulate the problem definition, analyze and do functional simulation of the same

design, implement, test and verify the proposed solution related to problem definition

compile, comprehend and present the work carried out

A student is required to carry out project work in the relevant area of post-graduate study. The project may include design / simulation / synthesis / development of a system, etc. At the end of the semester, a student has to submit a detailed report incorporating literature survey, problem formulation, clear problem statement, research methods, result analysis, conclusion, etc. It is expected that the student should defend his/her work before the jury / panel of examiners.

L = Lecture, T = Tutorial, P = Practical, C = Credit

L	T	P	C
3	0	2	4

Course Code	3CS42D101
Course Name	Natural Language Computing

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

comprehend the key concepts of NLP which are used to describe and analyse language
perform POS tagging and generate context free grammar for English language
realize semantics and pragmatics of English language for processing
implement natural language processing task

Syllabus:	**Teaching Hours**
Unit I **Introduction:** Knowledge in speech and language processing, Ambiguity, Models and algorithms, Language, Thought, and Understanding, State-of-the-art, History.	**2**
Unit II **Words:** Regular expressions and automata, Morphology and Finite-State Transducers, Computational Phonology and Text-to-Speech, Probabilistic Models of Pronunciation and Spelling, N-grams, HMMs and Speech Recognition	**8**
Unit III **Syntax:** Word classes and Part-of-speech tagging, Context-free grammars for English, Parsing with context-free grammars, Features and unification, Lexicalized and probabilistic parsing	**9**
Unit IV **Semantics:** Representing meaning, Semantic analysis, Lexical semantics, Word Sense Disambiguation and Information Retrieval	**9**
Unit V **Pragmatics:** Discourse, Dialogue and Conversational Agents, Generation,	**9**

Machine Translation

Unit VI **8**

Natural Language Processing: using Deep Learning: Data representation in NLP, Distributed representations, Word2Vec, applying work embeddings, sequence to sequence learning, Application of Deep learning in NLP for author attribution, text classification, word generation

Self Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

Jurafsky, D., & Martin, J. H., Speech and language processing (Vol. 3). London: Pearson.
Allen, J., Natural language understanding. Pearson.
Charniak, E., Statistical language learning. MIT press.
Manning, C. D., Manning, C. D., & Schütze, H., Foundations of statistical natural language processing. MIT press.
Dale, R., Moisl, H., & Somers, H., Handbook of natural language processing. CRC Press.
Radford, A., Atkinson, M., & Britain, D., Linguistics: an introduction. Cambridge University Press.

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list

L	T	P	C
3	0	2	4

Course Code	3CS42D303
Course Name	Predictive Analytics

Course Learning Outcomes (CLOs):

At the end of the course, students will be able to

apply statistical and regression analysis methods to identify new trends and patterns, uncover relationships, create forecasts, predict likelihoods, and test predictive hypotheses
compare the underlying predictive modeling techniques
develop the modeling skills from an industry perspective
select appropriate predictive modeling approaches suitable to various tasks

Syllabus: Teaching Hours

Unit I **4**

Introduction: Introducing CRISP-DM methodology, Need for it in business scenario

Unit II **8**

Modeling Techniques: Overview of Modeling Techniques, Unsupervised learning - Clustering, Supervised Learning - Classification, Linear Discriminant Analysis, Ensemble Learning, Random Forest models.

Unit III **10**

Regression: Concept of Regression, Covariance, Correlation and ANOVA Review, Simple linear regression, multiple linear regression, parameter estimation, logistic regression, Maximum Likelihood Estimation (MLE) of parameters

Unit IV **5**

Model Evaluation: Metrics for Performance Evaluation, Accuracy, confusion matrix, Precision, Recall, ROC Curves

Unit V 10

Decision Trees and Unstructured data analysis: Introduction to Decision Trees, CHI-Square Automatic Interaction Detectors (CHAID), Classification and Regression Tree (CART), Analysis of Unstructured data, Naive Bayes Classification

Unit VI 8

Advanced Topics: Forecasting and time series analysis, auto-regressive and moving average models, applications in stock market prediction weather prediction.

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings^:

James, Witten, Hastie and Tibshirani, An Introduction to Statistical Learning: with Applications in R, Springer
Dinov, Ivo D., Data Science and Predictive Analytics, Springer
Hastie, Trevor, et al., The elements of statistical learning, Vol. 2. No. 1. New York: Springer
Montgomery, Douglas C., and George C. Runger., Applied statistics and probability for engineers, John Wiley & Sons
Hann, J. and Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann.

L=Lecture, T=Tutorial, P=Practical, C=Credit

^this is not an exhaustive list