भारतीय सूचना प्रौद्योगिकी, अभिकल्पना एवं विनिर्माण संस्थान, कर्नूल

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, DESIGN AND MANUFACTURING, KURNOOL

Jagannathagattu, Dinnedevarapadu, Kurnool - 518007, Andhra Pradesh, India

(An Institute of National Importance under MoE, Govt. of India)

Scheme and Syllabus

MASTER OF TECHNOLOGY

IN

Computer Science & Engineering

With Specialization in

Data Analytics and Decision Sciences

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,

DESIGN AND MANUFACTURING, KURNOOL

May, 2021

Scheme and Syllabus

Department of **Computer Science and Engineering**
M. Tech. in **Computer Science and Engineering**
with Specialization in **Data Analytics and Decision Sciences**

Sl. No.	Course Code	Course Name	Category Code	I	P	No. of Credits
Semester I						
1	CS501T	Mathematical Foundations for Data Science	PEC	3	0	3
2	CS502T	Advanced Data Structures and Algorithms	PEC	3	0	3
3	CS503T	Principles of Data Analytics	PEC	3	0	3
4	CS504T	Statistical Learning	PEC	3	0	3
5	CS505I	Decision Sciences	PEC	2	2	3
6	CS5XXT	Elective-I	PEC	3	0	3
7	CS506P	Decision Sciences Practice	PEC	1	3	3
8	CS601	Seminar	PCD	0	3	2
Total				**18**	**8**	**23**
Semester II						
1	CS507T	Marketing Science and Predictive Analytics	PEC	3	**0**	3
2	CS508T	Advanced Decision Modelling Techniques	PEC	3	**0**	3
3	CS509T	Machine Learning	PEC	3	**0**	3
4	CS510T	Decision Support Systems	PEC	2	**2**	3
5	CS5XXT	Elective-II	PEC	3	**0**	3
6	CS5XXT	Elective-III	PEC	3	**0**	3
7	CS507P	Predictive Analytics Practice	PEC	0	**3**	2
8	CS509P	Machine Learning Practice	PEC	0	**3**	2
9	CS602	Comprehensive Viva-Voce	PCD	0	**3**	2
Total				**17**	**11**	**24**
Semester III						
1	CS603	Dissertation Work-I	PCD	0	25	**10**
Semester IV						
1	CS604	Dissertation Work-II	PCD	0	25	**20**

PEC: Professional Engineering Course
PCD: Professional Career Development

	List of Electives					
Sl. No.	**Code**	**Course Name**	**Category**	**I**	**P**	**Credit**
1	CS511T	Data Mining	PEC	3	0	3
2	CS512T	Intelligent Systems	PEC	3	0	3
3	CS513T	Advanced Database Systems	PEC	3	0	3
4	CS514T	Artificial Intelligence	PEC	3	0	3
5	CS515T	Social Media and Networks Analytics	PEC	3	0	3
6	CS516T	Time Series Analysis	PEC	3	0	3
7	CS517T	Data Exploration and Visualization	PEC	3	0	3
8	CS518T	Game Theory	PEC	3	0	3
9	CS519T	Randomized and Approximation Algorithms	PEC	3	0	3
10	CS520T	Cloud Computing	PEC	3	0	3
11	CS521T	Big Data Processing	PEC	3	0	3
12	CS522T	Scalable Systems for Data Science	PEC	3	0	3
13	CS523T	Recommender Systems	PEC	3	0	3

PEC: Professional Engineering Course

Department of **Computer Science and Engineering**
M. Tech. in **Computer Science and Engineering**
with Specialization in **Data Analytics and Decision Sciences**

Semester I						
Sl. No.	**Code**	**Course Name**	**Category**	**I**	**p**	**Credit**
1	CS501T	Mathematical Foundations for Data Science	PEC	3	0	3
2	CS502T	Advanced Data Structures and Algorithms	PEC	3	0	3
3	CS503T	Principles of Data Analytics	PEC	3	0	3
4	CS504T	Statistical Learning	PEC	3	0	3
5	CS505T	Decision Sciences	PEC	2	2	3
6	CS5XXT	Elective-I	PEC	3	0	3
7	CS506P	Decision Sciences Practice	PEC	1	3	3
8	CS601	Seminar	PCD	0	3	2
Total				18	8	**23**

PEC: Professional Engineering Course
PCD: Professional Career Development

Course Title	Course Code	Structure (I-P-C)		
Mathematical Foundations for Data Science	CS501T	**3**	**0**	**3**

Prerequisite: Nil
Course outcomes:

CO1	Ability to use the mathematical concepts in the field of data science
CO2	Employ the techniques and methods related to the area of data science in variety of applications
CO3	Apply logical thinking to understand and solve the problem in context.

Syllabus:

Basics of Data Science: Introduction; Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems.

Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes.

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; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process.

Optimization: Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning. Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems.

References:

Textbooks:

1. G. Strang . Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA, 2016.
2. Bendat, J. S. and A. G. Piersol. Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA, 2010

Reference books:

3. Montgomery, D. C. and G. C. Runger. Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA, 2011.
4. David G. Luenberger . Optimization by Vector Space Methods, John Wiley & Sons (NY), 1969.
5. Cathy O'Neil and Rachel Schutt . Doing Data Science, O'Reilly Media, 2013.

Course Title	Course Code	Structure (I-P-C)		
Advanced Data Structures and Algorithms	CS502T	**3**	**0**	**3**

Prerequisite: Nil
Course outcomes:

CO1	Understand the implementation of symbol table using hashing techniques
CO2	Apply advanced abstract data type (ADT) and data structures in solving real world problem
CO3	Effectively combine the fundamental data structures and algorithmic techniques in building a solution to a given problem
CO4	Develop algorithms for text processing applications

Syllabus:

Dictionaries: Definition, Dictionary Abstract Data Type, Implementation of Dictionaries, Hashing: Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing.

Skip Lists: Need for Randomizing Data Structures and Algorithms, Search and Update Operations on Skip Lists, Probabilistic Analysis of Skip Lists, Deterministic Skip Lists, Trees: Binary Search Trees (BST), AVL Trees, Red Black Trees: Height of a Red Black Tree, Red Black Trees Bottom-Up Insertion, Top-Down Red Black Trees, Top-Down Deletion in Red Black Trees, Analysis of Operations.

2-3 Trees: Advantage of 2-3 trees over Binary Search Trees, Search and Update Operations on 2-3 Trees, Analysis of Operations, B-Trees: Advantage of B- trees over BSTs, Height of B-Tree, Search and Update Operations on 2-3 Trees, Analysis of Operations, Splay Trees: Splaying, Search and Update Operations on Splay Trees, Amortized Analysis of Splaying.

Text Processing: Sting Operations, Brute-Force Pattern Matching, The Boyer-Moore Algorithm, The Knuth-Morris-Pratt Algorithm, Standard Tries, Compressed Tries, Suffix Tries, The Huffman Coding Algorithm, The Longest Common Subsequence Problem (LCS), Applying Dynamic Programming to the LCS Problem, Computational Geometry: One Dimensional Range Searching, Two Dimensional Range Searching, Constructing a Priority Search Tree, Searching a Priority Search Tree, Priority Range Trees, Quadtrees, k-D Trees.

References:

Textbooks:

1. Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, second Edition, Pearson, 2004.
2. T.H. Cormen, C.E. Leiserson, R.L.Rivest, Introduction to Algorithms, Third Edition Prentice Hall, 2009.

Reference books:

Michael T. Goodrich, Roberto Tamassia, Algorithm Design, First Edition, Wiley, 2006.

.

Course Title	Course Code	Structure (I-P-C)		
Principles of Data Analytics	CS503T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Understand the ideas of statistical approaches to learning
CO2	Work with big data platform and explore the data analysis techniques for business applications
CO3	Design efficient algorithms for mining the data from large volumes
CO4	Perform appropriate statistical tests using R and visualize the outcome
CO5	Understand the significance of exploratory data analysis (EDA) in data science and apply basic tools (plots, graphs, summary statistics) to perform EDA
CO6	Apply basic machine learning algorithms (Linear Regression, k-Nearest Neighbors (k-NN), k-means, Naive Bayes) for predictive modeling. Explore the merits of Naive Bayes technique
CO7	Recognize the characteristics of machine learning techniques that are useful to solve real-world problems
CO8	Identify basic approaches used for feature generation and feature selection algorithms (Filters, Wrappers, Decision Trees, and Random Forests) and to apply the techniques in applications
CO9	Identify and explain fundamental mathematical and algorithm paradigms that constitute a recommendation engine. Build their own recommendation system using existing components

Syllabus:

Introduction: What is Data Science? Big Data and Data Science hype and getting past the hype, Why now?, Datafication, Current landscape of perspectives, Skill sets, Life cycle of Data Science, Different phases.

Exploratory Data Analysis and the Data Science Process: Basic tools (plots, graphs and summary statistics) of EDA, Philosophy of EDA, The Data Science Process, Case Study: RealDirect (online real estate firm), Three Basic Machine Learning Algorithms: Linear Regression, k-Nearest Neighbors (k-NN), k-means.

One More Machine Learning Algorithm and Usage in Applications: Motivating application: Filtering Spam, Why Linear Regression and k-NN are poor choices for Filtering Spam, Naive Bayes and why it works for Filtering Spam, Data Wrangling: APIs and other tools for scrapping the Web, Feature Generation and Feature Selection (Extracting Meaning From Data), Motivating application: user (customer) retention, Feature Generation (brainstorming, role of domain expertise, and place for imagination), Feature Selection algorithms: Filters; Wrappers; Decision Trees; Random Forests, Recommendation Systems: Building a User-Facing Data Product: Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis, Exercise: build your own recommendation system.

Data Visualization: Basic principles, ideas and tools for data visualization, Case study on industry projects, Exercise: create your own visualization of a complex dataset, Data Science and Ethical Issues: Discussions on privacy, security, ethics, A look back at Data Science, Next-generation data scientists.

References:

Textbooks:

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O'Reilly, 2014.
2. Jure Leskovek, Anand Rajaraman and Jerey Ullman. Mining of Massive Datasets, Cambridge University Press, 2014.

Reference books:

3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2013.
4. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. O′Reilly, 2013.
5. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. Springer, 2009.
6. Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science.2018.
7. Mohammed J. Zaki and Wagner Miera Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, 2014.
8. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, Third Edition. Morgan Kaufmann, 2011.

Course Title	Course Code	Structure (I-P-C)		
Statistical Learning	CS504T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Familiarize with a broad range of approaches and techniques in machine learning
CO2	Choose effective methods to solve various learning problem
CO3	Explore statistical learning methods and their application to modern problems in science, industry, and society
CO4	Build analytics pipelines for regression problems and classification problems
CO5	Build analytics pipelines for recommendation problems

Syllabus:

Probabilistic formulations of prediction problems: Plug-in estimators, empirical risk minimization, Linear threshold functions, perceptron algorithm, Risk bounds, Concentration inequalities, Uniform convergence, Rademacher averages; combinatorial dimensions, Convex surrogate losses for classification, Linear regression, Regularization and linear model selection, Feature Selection Methods, Cross Validation methods.

Game-theoretic formulations of prediction problems, High Dimensional methods, Lasso, Ridge Regression, Dimensionality Reduction, Minimax strategies for log loss, linear loss, and quadratic loss, Universal portfolios, Online convex optimization.

Neural networks: Stochastic gradient methods, Combinatorial dimensions and Rademacher averages, Hardness results for learning, Efficient learning algorithms.

Kernel methods: Reproducing kernel Hilbert spaces, Mercer's theorem, Convex optimization for kernel methods, Representer theorem, Ensemble methods: AdaBoost, AdaBoost as I-projection, Convergence and consistency of AdaBoost.

References:

Textbooks:

1. James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013.
2. Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer, 2009

Course Title	Course Code	Structure (I-P-C)		
Social Media and Network Analysis	CS515T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Formalize different types of entities and relationships as nodes and edges and represent this information as relational data
CO2	Plan and execute network analytical computations
CO3	Use advanced network analysis software to generate visualizations and perform empirical investigations of network data
CO4	Interpret and synthesize the meaning of the results with respect to a question, goal, or task
CO5	Collect network data in different ways and from different sources while adhering to legal standards and ethics standards

Syllabus:

Class logistics, Overview on Network Analysis, The Network Analysis Process and Methodology, Network Visualization, Sociometric Analysis, Preliminaries on Network Structures, Models and simulation on Network evolution, Sociological direction.

Subgroups and Cliques, Clustering of graphs, Organization of relational data, Mining Social-network graphs, Direct discovery of communities in graph, Neighborhood properties in graphs, Ego Networks, Structural Holes, Cognitive Social Structures.

Introduction: Integration of Text and Network Analysis, Types of Networks Extracted From Texts Across Disciplines, Local Centrality, Global Centrality, Bank centrality in corporate network.

Introduction: Multi-Agent Models for Representing Networks, Social Network Based Multi-Agent systems, Trends on social networks, Case Studies.

References:

Textbooks:

1. Scott, J. Social network analysis: A handbook, Second Edition, Newbury Park. Sage, 2008
2. Knoke. Social Network Analysis, Second Edition, Sage, 2008.

Reference books:

Hanneman, Robert A. and Mark Riddle, Introduction to social network methods, Riverside, CA, 2005.

Course Title	Course Code	Structure (I-P-C)		
Decision Sciences	CS505T	**2**	**2**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Familiarize with the usage of Microsoft Excel for business analysis
CO2	Understand the role of quantitative techniques for managerial decision making
CO3	Ability to structure problems and to perform logical analyses
CO4	Explore the functionalities from various models and to use those insights to communicate, persuade and motivate change

Syllabus:

Decision making process, Introduction to spreadsheet modeling and problem solving, Spreadsheet engineering, Spreadsheet analysis, Modeling and Prototyping,

Measures of Central Tendency and Measures of Dispersion, Data analysis, Modeling in Practice, Enterprise Resource Planning, Introduction to Optimization and Solver, Case study.

Creating optimization models, Optimization models and sensitivity analysis, Production planning, Revenue management, Sales force sizing and allocation, Case Study.

Introduction to simulation, Simulation techniques and examples, Simulation modeling and analysis, Managing risk with insurance, Optimization in simulation, Assessing acquisition value with simulation.

References:

Textbooks:
1. Francis J. Clauss, Applied Management Science and Spreadsheet Modeling, Duxbury.
2. Jeffrey D. Camm and James R. Evans, Management Science: Modeling, Analysis, and Interpretation, South-Western.

Reference books:

3. Cliff Ragsdale, Spreadsheet Modeling and Decision Analysis, South-Western.
4. Wayne L. Winston and S. Christian Albright, Practical Management Science: Spreadsheet Modeling and Applications, Duxbury.
5. Jeffrey Moore et al., Introductory Management Science, Prentice-Hall.

Course Title	Course Code	Structure (I-P-C)		
Decision Sciences Practice	CS506P	**1**	**3**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Familiarize with various features of R language
CO2	Understand the role of quantitative techniques for managerial decision making
CO3	Ability to perform data analysis with popular datasets
CO4	Explore and familiarize with experiments in linear model selection, cross validation and other statistical learning techniques

Syllabus:

1. Designing a spreadsheet that will allow an analyst to predict the monthly expenses in a company.
2. Importing and exporting spreadsheet data.
3. Regression: linear regression, test of significance, residual analysis, polynomial regression using R tool.
4. Experiments on cross validation, and linear model selection.
5. Experiments on advanced linear regression methods such as lasso and ridge regression.
6. Exploratory data analysis: Charts and Plots, Data Visualization, combining visualization and data transformation that allows to efficiently explore the data.
7. Information Extraction from given Textual data.
8. Apply the techniques of Linear Algebra using R tool on given dataset.
9. Optimization in R: Common R Packages for Linear, Quadratic and Non-linear optimization and sensitivity analysis.
10. Linear Programming in R: lpsolve.

Capstone Project.

References:

Textbooks:

1. Francis J. Clauss, Applied Management Science and Spreadsheet Modeling, Duxbury.
2. Jeffrey D. Camm and James R. Evans, Management Science: Modeling, Analysis, andInterpretation, South-Western.

Reference books:

3. Cliff Ragsdale, Spreadsheet Modeling and Decision Analysis, South-Western.
4. Wayne L. Winston and S. Christian Albright, Practical Management Science: SpreadsheetModeling and Applications, Duxbury.

Jeffrey Moore et al., Introductory Management Science, Prentice-Hall.

Department of **Computer Science and Engineering**
M. Tech. in **Computer Science and Engineering**
with Specialization in **Data Analytics and Decision Sciences**

Semester II						
Sl. No.	**Code**	**Course Name**	**Category**	**I**	**p**	**Credit**
1	CS507T	Marketing Science and Predictive Analytics	PEC	3	0	3
2	CS508T	Advanced Decision Modelling Techniques	PEC	3	0	3
3	CS509T	Machine Learning	PEC	3	0	3
4	CS510T	Decision Support Systems	PEC	2	2	3
5	CS5XXT	Elective-II	PEC	3	0	3
6	CS5XXT	Elective-III	PEC	3	0	3
7	CS507P	Predictive Analytics Practice	PEC	0	3	2
8	CS509P	Machine Learning Practice	PEC	0	3	2
9	CS602	Comprehensive Viva-Voce	PCD	0	3	2
Total				17	11	**24**

PEC: Professional Engineering Course
PCD: Professional Career Development

Course Title	Course Code	Structure (I-P-C)		
Market Science and Predictive Analysis	CS507T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Understand the different strategic tools used for predictive analysis
CO2	Recognize the quantitative methods used to apply in predictive analysis
CO3	Apply various tools and techniques which help in predicting the future decision making. Examine the predictive analysis by investigating through decision trees or unstructured data analytics
CO4	Justify the importance of predictive analysis through forecasting techniques like time series analysis. To measure the ability of students to apply the predictive analytics tools in different business scenarios through example projects

Syllabus:

Introduction to Predictive Analysis: Introduction to Analytics, Analytics in Decision Making, Game changers & Innovators, Experts view on Predictive Analysis. Simple Linear Regression (SLR): Case-let Overview Introduction to Regression, Model Development, Model Validation, Demo using Excel & R Programming

Multiple Linear Regression (MLR): Multiple Linear Regression, Estimation of Regression Parameters, Model Diagnostics, Dummy, Derived & Interaction Variables, Multi-co linearity, Model Deployment, Implementation using R Programming

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, Implementation using R Programming.

Forecasting and Time series Analysis: Forecasting Time Series Analysis, Additive & Multiplicative models, Exponential smoothing techniques, Forecasting Accuracy, Autoregressive and Moving average models, Implementation using R Programming.

References:

Textbooks:

1. Applied Predictive Modeling by Max Kuhn and Kjell Johnson.
2. Predictive Analytics For Dummies by Anasse Bari, Mohamed Chaouchi and Tommy Jung

Reference books:

3. Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science (FT Press Analytics) by Thomas W. Miller

Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die by Eric Siegel

Course Title	Course Code	Structure (I-P-C)		
Advanced Decision Modelling Techniques	CS508T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Conduct probabilistic sensitivity analysis, including the identification and appropriate parameterization of parameter distributions, correct use and interpretation of the net-benefit framework using ICER planes, CEACs, and confidence intervals
CO2	Identify and understand when other types of advanced decision models are more appropriate than standard decision analysis techniques, and to be able to design, conduct analysis, and interpret results from these approaches.
CO3	To conduct of probabilistic sensitivity analysis and development of sophisticated decision models

Syllabus:

Introductions and overview/discussion of why and how we model, Infectious disease modeling, Introduction to Agent-based modeling and AnyLogic, More with AnyLogic,

Discrete Event Simulation, Decision Trees and Evaluating the use of Imperfect Information, Advanced Sensitivity/Uncertainty Analysis, Advanced Sensitivity Analysis – Methods from Engineering

Advanced Preparation of Sensitivity and Uncertainty Analysis – ICER Planes, CEACs and the Net Benefit Framework, Model Calibration, Building Confidence in Models

Uncertainty Analysis, Value of Information and Design of Simulation

References:

Textbooks:

1. Advances in Decision Analysis: 4 (Mathematical Modelling: Theory and Applications) by Nadine Meskens, M.R. Roubens
2. Mathematical Modelling: Theory and Application, by X. Liao and P. Yu

Course Title	Course Code	Structure (I-P-C)		
Machine Learning	CS509T	**3**	**0**	**3**

Prerequisite: Nil

Course outcomes:

CO1	Familiarity with traditional and modern learning paradigms with their applications in the real-world systems
CO2	The learners can adapt human training for development of intelligent machines
CO3	Ability to model any real-world practical problem in a machine learning domain
CO4	Thorough grasp on the artificial neural networks with an understanding of the modern deep learning techniques

Syllabus:

Introduction to machine learning: learning systems, classification, clustering, regression, separability of problems; introduction to learning paradigms: supervised, unsupervised, semi-supervised, active, reinforcement with examples; cross-validation; performance evaluation metrics for classification and clustering; curse of dimensionality, feature selection, reduction and expansion, computation of Eigen co-ordinates and principle component analysis

Recognition systems and design cycle, Non-linearly separable problems: solutions through Cover's theorem with examples, parametric learning mechanisms like Maximum likelihood, expectation maximisation, aposteriori probabilities, Instance-based learning, Lazy learning with K-nearest neighbour, Eager learning with basis functions, non-parametric learning using support vector machines (SVMs)

Artificial neural networks: Analogy of biological neural network with artificial neural network; Perceptron learning; gradient descent algorithm; multi-layer perceptrons; back-propagation algorithm; activation functions, delta rule, learning curves: overfitting and underfitting of models; Hebbian learning, self-organising feature map, radial basis function neural networks

Deep neural networks: Introduction and advent of deep learning paradigm, solutions to vanishing and exploding gradient problems, regularisation, activation functions for deep learning, deep feed forward network, convolutional neural network (CNN), pretrained CNN models, attention network, generative models like auto-encoders and adversarial learning, recurrent neural networks, problem solving through deep learning and open areas of research

References:

Textbooks:

1. T. M. Mitchell, Machine Learning, McGraw-Hill, 1997.

2. S. Haykin, Neural Networks: A Comprehensive Foundation. Prentice-Hall of India, 2007.

Reference books:

3. R. O. Duda, P.E. Hart, D. G. Stork, Pattern Classification, John Wiley, 2001

4. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016

Course Title	Course Code	Structure (I-P-C)		
Decision Support Systems	CS510T	**2**	**2**	**3**

Prerequisite: Nil
Course outcomes:

CO1	To provide students with the basic and necessary knowledge, in order that they could identify when a given domain is really a complex one
CO2	To identify how many and of which nature are the decisions involved in complex domains management
CO3	To know how to analyse, to design, to implement and to validate an Intelligent Decision Support Systems (IDSS), emphasising the integration of Artificial Intelligence models and Statistical/Numerical models, and the knowledge discovery from data

Syllabus:

Introduction, Complexity of real-world systems or domains, The need of decision support tools, Decisions Theory, Modelling of Decision Process

Evolution of Decision Support Systems, Historical perspective of Management Information Systems, Decision Support Systems (DSS), Advanced Decision Support Systems (ADSS), Intelligent Decision Support Systems (IDSS), Intelligent Decision Support Systems (IDSS), IDSS Architecture, IDSS Analysis and Design, Requirements, advantages and drawbacks of IDSS, IDSS Validation, Implementation of an IDSS in a computer

Knowledge Discovery in a IDSS: from Data to Models: Introduction, Data Structure, Data Filtering, Knowledge Models: Descriptive models, Associative models, Discriminant Models, Predictive models, Uncertainty Models, Probabilistic models, Fuzzy models, Post-Processing and Model Validation: Post-processing techniques, Validation, Statistical Methods for Hypotheses Verification

Tools and Applications, Software Tools for IDSS Development, Application of IDSS to real-world problems, Future Trends in IDSS and Conclusions

References:

Textbooks:
1. Intelligent decision support methods: the science of knowledge work - DHAR, Vasant; STEIN, Roger, Prentice Hall, 1997. ISBN: 978-0135199350

2. Decision Support Systems in the Twenty-first Century. - MARAKAS, G.M., Upper Saddle River, NJ: Prentice-Hall, 2003. ISBN: 978-0130922069

Reference books:

3. Decision Support Systems and Intelligent Systems - TURBAN, E.; ARONSON, J.E.; LIANG T-P, Pearson/Prentice Hall, 2005.

4. Decision Support Systems: concepts and resources for managers - POWER, Daniel J., Greenwood Publishing Group, 2002.

Course Title	Course Code	Structure (I-P-C)		
Cloud Computing	CS520T	**3**	**0**	**3**

Prerequisite: Nil
Course outcomes:

CO1	Articulate the main concepts, key technologies, strengths, and limitations of cloud computing and the possible applications for state-of-the-art cloud computing
CO2	Identify the architecture and infrastructure of cloud computing, including SaaS, PaaS, IaaS, public cloud, private cloud, hybrid cloud, etc.
CO3	Choose a suitable technique to address security, interoperability issues
CO4	Provide the appropriate cloud computing solutions and recommendations according to the applications used

Syllabus:

Introduction to Cloud Computing, Gartner's Hype Cycle for Emerging Technologies, Comparisons: Cluster, Grid and Cloud, Cloud Computing at a Glance, Vision, A Close Look, The NIST Model, Cloud Cube Model, Types: Deployment and Service Models, Public, Private, Hybrid and Community Cloud, IaaS, PaaS, SaaS, Characteristics, Applications, Benefits, Disadvantages, Web 2.0, The Laws of Cloudonomics, Obstacles, Cloud Adoption, Measuring the Costs, Service-Level Agreement, Cloud Architecture, Virtual Appliances, Connecting to the Cloud, IaaS Workloads, Open SaaS and SOA, On Demand vs. On Premises IT, Bird's-Eye View of Cloud Computing Vendors, Virtualization, Characteristics of Virtualized Environments, Taxonomy of Virtualized Techniques, Full Virtualization, Paravirtualization, Partial Virtualization, Pros and Cons of Virtualization, Hypervisor

Cloud issues and challenges - Properties - Characteristics - Service models, Deployment models Virtualization – Virtual Machines, Resource Allocation, Leases: Advance Reservation, Best Effort, Immediate, Deadline Sensitive and Negotiated, Swapping and Backfilling, Resource Allocation Measures, Task Scheduling, Task: Dependent and Independent, Job, Application, Workflow: Montage, Epigenomics, SIPHT, LIGO, CyberShake, Machine: Homogeneous and Heterogeneous, Mode: Immediate, Intermediate and Batch, Expected Time to Compute Matrix, Manager Server, Data Center, Virtual Machine, Server, Makespan, Resource Utilization, Average Execution Time, Uncertainty

Introduction to Energy Efficient Task Consolidation, Energy-Conscious Task Consolidation, MaxUtil, Energy-Aware Task Consolidation, Virtual Cluster, CPU Utilization Threshold, Sleep or Power Saving Mode, High-Throughput Computing: Task Computing and Task-based Application Models, Market-Based Management of Clouds, Green Cloud Computing Architecture, Federated Clouds, Pricing Mechanism, SLA Violation.

Introduction to Cloud Security, Case Studies: Manjrasoft Aneka, Amazon Web Services, Google Cloud Platform, Microsoft Azure, Programming support of Google App Engine, Virtual Machine and its Provisioning, Time and Space-shared Provisioning.

References:

Textbooks:

1. R. Buyya, C. Vecchiola and S. T. Selvi, Mastering Cloud Computing Foundations and Applications Programming, Morgan Kaufmann, Elsevier, 2013.

2. B. Sosinsky, Cloud Computing Bible, Wiley, 2011.

Reference books:

3. D. N. Chorafas, Cloud Computing Strategies, CRC Press, Taylor and Francis Group, 2011.

4. Kai Hwang, Geoffrey C. Fox and Jack J. Dongarra, “Distributed and cloud computing from Parallel Processing to the Internet of Things”, Morgan Kaufmann, Elsevier , 2012.

5. D. Janakiram, Grid Computing, Tata McGraw-Hill, 2005.

Course Title	Course Code	Structure (I-P-C)		
Data Mining	CS511T	**3**	**0**	**3**

Prerequisite: Basic courses in Programming, Data Structures, Algorithms

Course outcomes:

CO1	Analyze Algorithms for Frequent item sets and frequent patterns
CO2	Design algorithms for various types of sequential patterns
CO3	Extract patterns from time series data
CO4	Develop algorithms for Temporal Patterns
CO5	Extend the Graph mining algorithms to Web Mining
CO6	Comprehend Trajectory Data Mining patterns and techniques

Syllabus:

Introduction to Data Mining, KDD process, Data Mining functionalities, Pre-processing of Data

Association Rules: Market Basket problem, Frequent Itemsets, Interesting measures, Apriori and FP growth algorithm, Algorithms for Closed and Maximal frequent itemsets, Quantitative Association Rules.

Sequential Pattern Mining concepts, primitives, GSP algorithm, scalable methods-Prefix Span, SPADE; Closed Sequential Patterns- BIDE algorithm. Transactional Patterns and other temporal based frequent patterns.

Mining Time series Data, Periodicity Analysis for time related sequence data, Trend analysis, Similarity search in Time-series analysis

Graph Mining, Mining frequent sub-graphs, gpan algorithm, finding clusters, hub and outliers in large graphs, Web Mining

Trajectory Pattern Mining: Moving together patterns, Sequential Pattern mining from trajectories, Trajectory Clustering.

References:

Textbooks:

1. Jiawei Han, Michelin Kamber and Jian Pei - Data Mining : Concepts and Techniques, Third Edition, Elsevier Publication, 2011.

2. Pang-Ning Tan, Michael Steinbach, Vipin Kumar - Introduction to Data Mining, Pearson 2016.

Course Title	Course Code	Structure (I-P-C)		
Predictive Analytics Practice	CS507P	**0**	**3**	**2**

Prerequisite: Nil

Course outcomes:

CO1	Structure a business scenario, devise a framework and implement using Data Science methods and Python
CO2	Practical nuances to manage data, execute various data treatment strategies, implement feature engineering, and model development and evaluation steps
CO3	Hands on experience to develop predictive models for number of real life scenarios

Syllabus:

Data Ingestions in Python, Feature Engineering and Data Management Scenarios, Regression Models using Python, Credit Card Spend Estimation – Modeling in Python

Building a Decision Tree Model for Customer Attrition, Tree based Models for a business scenario and compare performance, Unstructured Data and Modeling in Python

Forecasting Model development in Python, Hands on model development using Exponential smoothing techniques, (ARIMA) Model development for a business scenario

Hands on Image Classification using Python, Hands on Emotion Detection from Images, Text Data Manipulations in Python, Practical Model Development or Emotion Detection from Text, Model Development to Deployment in Python

References:

1. The Elements of Statistical Learning Artificial Intelligence for Marketing: Practical Applications
2. S. Dey. Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data. Packt Publishing Ltd, 2018.

Course Title	Course Code	Structure (I-P-C)		
Machine Learning Practice	CS509P	**0**	**3**	**2**

Prerequisite: Nil

Course outcomes:

CO1	Use python programming in solving machine learning (ML) tasks
CO2	Use cloud platforms like Google colab for implementing the ML algorithms over real-world (large-scale) datasets
CO3	Implementation of shallow and deep learning methods on spatio-temporal datasets
CO4	Introduction to some open areas of research and finding their possible solutions through a mini-project

Syllabus:

1. Developing codes for well-known machine learning algorithms in Python: K-nearest neighbour, K-means
2. Evaluating the confusion matrix programmatically to find out performance measures like accuracy, recall, precision and F1 score
3. Evaluating the dimensionality reduction techniques in Python and k-fold cross-validation
4. Implementation of basic artificial neural networks like SLP, MLP and SOM using tensor flow
5. Implementation of deep learning techniques like deep feed-forward networks, CNN on standard computer vision datasets
6. Application of generative models in computer vision tasks
7. A mini-project on real-world deep learning problems
8. Implementation of Association Rule Mining Algorithms
9. Implementation of Sequential Pattern Mining Algorithms

References:

Textbooks:

1. A. Géron. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 2019.
2. M. Fenner. Machine learning with Python for everyone. Addison-Wesley Professional, 2019.

Reference books:

3. A. C. Müller and S. Guido. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.", 2016.