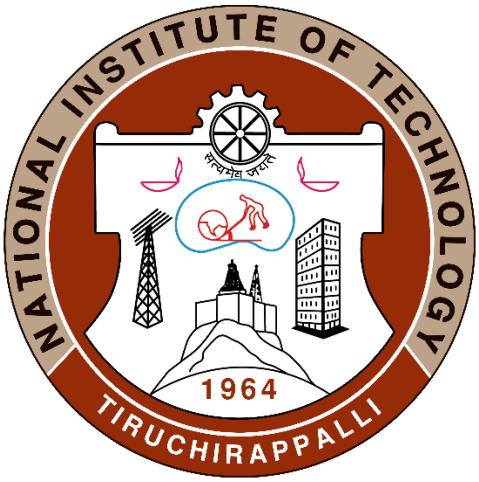
M.Tech. DATA ANALYTICS

Credit Based Flexible Curriculum
(Applicable from 2022-23)



Department of Computer Applications
National Institute of Technology
Tiruchirappalli - 620 015
Tamilnadu, India

About us:

The Department of Computer Applications is one of the pioneering departments of the institution that offers Information Technology courses such as MCA, M.Sc Computer Science and M.Tech (Data Analytics) and is one among the top five offering MCA courses in the country. It is committed to impart quality education in the sub-fields of IT, a field growing in leaps and bounds.

Vision:

Towards a school of Information Science and Technology conforming to international standards

Mission:

- To offer state-of-art education in Information Science and Technology
- To provide strong theoretical foundation complemented with extensive practical training
- To inculcate value-based, socially committed professionalism to the cause of overall development of students and society

M. Tech. Data Analytics

Objectives of the Programme:

- To provide students with a comprehensive foundation for applying statistical methods to resolve real-world problems in data analytics.
- To develop the ability to apply computational techniques to draw insight from big data in a variety of application domains.
- To prepare students for careers in data analytics with the ability to conduct in-depth data analysis, synthesis and evaluation.
- To manage real life complex data analytics projects.
- To prepare students for research in data analytics with the ability to provide solutions to recent research issues.

Qualification for Admission:

Bachelor's degree in Engineering / Technology in CSE / IT or MSc. (Computer Science / Information Technology / Software Engineering) or MCA from a recognized institution and a valid GATE score in Computer Science. The percentage and CGPA requirement for admission may be fixed as per NIT Trichy norms.

Head of the Department:

Dr. P. J. A. Alphonse

Members:

1. **Dr. S. R. Balasundaram, Professor**
2. **Dr. S. Nickolas, Professor**
3. **Dr. Michael Arock, Professor**
4. **Dr. P. J. A. Alphonse, Professor**
5. **Dr. S. Domnic, Associate Professor**
6. **Dr. G. R. Gangadharan, Associate Professor**
7. **Dr. B. Janet, Assistant Professor**
8. **Dr. S. Sangeetha, Assistant Professor**
9. **Dr. R. Eswari, Assistant Professor**
10. **Dr. U. Srinivasulu Reddy, Assistant Professor**
11. **Dr. B. Balaji, Assistant Professor**
12. **Dr. K. Selvakumar, Assistant Professor**
13. **Dr. Jitendra Kumar, Assistant Professor**
14. **Dr. Ghanshyam S. Bopche, Assistant Professor**

SYLLABUS

Semester I

Course Code	Course name	L	T	P	C
CA611	Mathematical Foundations for Data Analytics	3	0	0	**3**
CA613	Advanced Data Structures and Algorithms	3	0	2	**4**
CA615	Machine Learning Techniques	3	0	0	**3**
CA617	Big Data Analytics	3	0	0	**3**
CA6AX	Elective-1	3	0	0	**3**
CA6BX	Elective-2	3	0	0	**3**
CA601	Big Data Analytics Lab	0	0	4	**2**
CA603	Machine Learning Lab	0	0	4	**2**
Total					**23**

Semester II

Course Code	Course name	L	T	P	C
CA610	Distributed and Cloud Computing	3	0	0	**3**
CA612	Deep Learning and its Applications	3	0	0	**3**
CA6CX	Elective-3	3	0	0	**3**
CA6DX	Elective-4	3	0	0	**3**
CA6EX	Elective-5	3	0	0	**3**
CA6FX	Elective-6	3	0	0	**3**
CA602	Distributed and Cloud Computing Lab	0	0	4	**2**
CA604	Deep Learning Lab	0	0	4	**2**
Total					**22**

Semester III

Course Code	Course name	L	T	P	C
CA649	Project Work Phase I	0	0	0	**12**
Total					**12**

Semester IV

Course Code	Course name	L	T	P	C
CA650	Project Work Phase II	0	0	0	**12**
Total					**12**

LIST OF ELECTIVES

Course Code	Course name	L	T	P	C
	SEMESTER I				
CA6A1	Data Mining and Warehousing	3	0	0	**3**
CA6A2	Soft Computing Techniques	3	0	0	**3**
CA6A3	Next Generation Data Base Systems	3	0	0	**3**
CA6B1	DevOps	3	0	0	**3**
CA6B2	Analytics for Strategic Market Planning	3	0	0	**3**
CA6B3	Financial Risk Analytics and Management	3	0	0	**3**
	SEMESTER II				
CA6C1	Intelligent Systems and Process Automation	3	0	0	**3**
CA6C2	Cognitive Sciences	3	0	0	**3**
CA6C3	Blockchain Technology	3	0	0	**3**
CA6D1	Natural Language Processing	3	0	0	**3**
CA6D2	Social Network Analysis	3	0	0	**3**
CA6D3	Supply Chain Analytics	3	0	0	**3**
CA6E1	Web Analytics	3	0	0	**3**
CA6E2	Federated Learning	3	0	0	**3**
CA6E3	Human Resource Analytics	3	0	0	**3**
CA6F1	Data Visualization	3	0	0	**3**
CA6F2	Pattern Recognition	3	0	0	**3**
CA6F3	Image and Video Analytics	3	0	0	**3**

Semester – I

CA611 – MATHEMATICAL FOUNDATIONS OF DATA ANALYTICS

Objectives:

- To learn the mathematical foundations applicable to data analytics.

Systems of Linear Equations – Matrices – Solving Systems of Linear Equations – Vector Spaces – Linear Independence – Basis and Rank – Linear Mappings – Affine Spaces – Norms – Inner Products – Lengths and Distances – Angles and Orthogonality – Orthonormal Basis – Orthogonal Complement – Inner Product of Functions – Orthogonal Projections – Rotations

Determinant and Trace – Eigenvalues and Eigenvectors – Cholesky Decomposition – Eigen decomposition and Diagonalization – Singular Value Decomposition – Differentiation of Univariate Functions – Partial Differentiation and Gradients – Gradients of Vector - Valued Functions – Gradients of Matrices – Useful Identities for Computing Gradients

Discrete and Continuous Probabilities – Sum Rule, Product Rule – and Bayes' Theorem – Summary Statistics and Independence – Gaussian Distribution – Conjugacy and the Exponential Family – Change of Variables/Inverse Transform – Optimization Using Gradient Descent – Constrained Optimization and Lagrange Multipliers – Convex Optimization

Tabular data – Power and the computation of sample size – Advanced data handling – Multiple regression – Linear models – Logistic regression – Rates and Poisson regression Nonlinear curve fitting. Density Estimation – Recursive Partitioning – Smoothers and Generalised Additive Models – Survivals Analysis – Analysing Longitudinal Data – Simultaneous Inference and Multiple Comparisons – Meta-Analysis – Principal Component Analysis – Multidimensional Scaling Cluster Analysis

Sampling Distributions and Descriptive Statistics - The Central Limit Theorem – Sampling Distributions (Chi-Square, t, F, z) – Test of Hypothesis – Testing for Attributes – Mean of Normal Population – One-tailed and two-tailed tests – F-test and Chi-Square test – Analysis of variance ANOVA – One way and two-way classifications

REFERENCES:

1. Sheldon M. Ross, "Introduction to Probability and Statistics for Engineers and Scientists", Academic Press, 2009.
2. Marc Peter Deisenroth, A. Aldo Faisal and Cheng Soon Ong, "Mathematics for Machine Learning", Cambridge University Press, 2020.

COURSE OUTCOMES:

The students will be able to:

- Deduce complex tasks by various Mathematical logic.
- Solve the problems using the concepts of Linear Equations, Discrete and Continuous Probabilities.

CA613 – ADVANCED DATA STRUCTURES AND ALGORITHMS

Objectives:

- To introduce different advanced data structures and algorithms.

Advanced Data Structures (Part I) – Top-Down Splay Trees – Red-Black Trees – Bottom-Up Insertion – Top-Down Deletion – Treaps – Suffix Arrays and Suffix Trees – Linear-Time Construction of Suffix Arrays and Suffix Trees – Trie structures – The *k*-d Trees.

Advanced Data Structures (Part II) – B-trees: Definition – operations – Fibonacci heaps: Definition – operations – van Emde Boas Trees: Preliminary approaches – A recursive structure – The van Emde Boas tree – Data Structures for Disjoint Sets: Disjoint-set operations – Linked-list representation of disjoint sets – Disjoint-set forests.

Greedy Algorithms – Elements – An activity-selection problem – Huffman codes and Minimum Spanning tree algorithms – Graph Algorithms – Single source shortest paths problem – All-pairs shortest paths problem – Flow Networks.

Backtracking and Branch-and-Bound strategies with applications – Randomized algorithms – Examples.

NP concepts – Introduction – NP-hard and NP-complete problems – Definitions and Properties – Satisfiability problem – Reducibility – Cook's Theorem (without proof) – Approximation algorithms.

REFERENCES:

1. M.A.Weiss, "Data Structures and Problem Solving using Java", 4th Edition, Addison Wesley, 2009.
2. T.H. Cormen, C.E. Leiserson, R.L. Rivest and C.Stein, "Introduction to Algorithms", 3rd Edition, MIT Press, 2009.
3. Robert Sedgewick and Kevin Wayne, "Algorithms", 4th Edition, Addison Wesely, 2011.
4. Steve S. Skiena, "The Algorithm Design Manual", 2nd Edition, Springer, 2008.
5. George T. Heineman, Gary Pollice and Stanley Selkow, "Algorithms in a Nutshell", 2nd Edition, O'Reilley, 2016.
6. Kleinberg and Tardos, "Algorithm Design", Pearson, 2006.
7. Udi Manber, "Introduction to Algorithms: A Creative Approach", Addison Wesley, 1989.
8. Anany Levitin, "The Design and Analysis of Algorithms", 3rd Edition, Pearson, 2012.

COURSE OUTCOMES:

The students will be able to:

- Write structured pseudo code for the given problem.
- Use advanced data structures to solve real-time problems.

CA615 - MACHINE LEARNING TECHNIQUES

Objectives:

- To introduce the basic concepts and techniques of Machine Learning.
- To develop the skills in using recent machine learning software for solving practical problems
- To be familiar with a set of well-known supervised, semi-supervised and unsupervised learning algorithms

INTRODUCTION – Learning – Designing a Learning System – Traditional Learning vs Machine Learning – Various types of Machine Learning: Supervised Learning – Unsupervised Learning – Reinforcement Learning – Machine Learning workflow – Machine Learning issues and challenges – Machine Learning Applications in real world problems.

SUPERVISED LEARNING – Predictive Models: Regression – Multivariate Regression – Types of Regression Models – Estimation of Regression coefficients – issues and challenges – applications. – Classification Models: Introduction – Different types of classifiers: Perceptron – Naive Bayes – Decision Tree – Logistic Regression – K-Nearest Neighbor – Artificial Neural Networks – Support Vector Machine – Evolution metrics for supervised learning – Issues and challenges – applications.

UNSUPERVISED LEARNING –Clustering Models: Partitioning based clustering – Hierarchical based clustering – Density based clustering – Grid based clustering – Mixture Models and EM Algorithm – Fuzzy k-Means Algorithm – Evolution metrics for clustering models - Dimensional Reduction Techniques: Need – Various types: PCA – ICA – FA – t-SNE - Case studies.

ENSEMBLE LEARNING – Boosting – AdaBoost Algorithm – Bagging – Random Forest – No-Free-Lunch Theorem – XGBoost Algorithm – Stacking – Voting – Ensemble Diversity – Error Decomposition – Diversity Measures – Evaluating Ensembles of Classifiers – Case studies.

REINFORCEMENT LEARNING – Introduction to Reinforcement Learning – Learning Task – Example of Reinforcement Learning in Practice – Learning Models for Reinforcement – (Markov Decision process – Q Learning – Q Learning function – Q Learning Algorithm – SARSA algorithm) – Nondeterministic Rewards and Actions – Applications of Reinforcement Learning.

REFERENCES:

1. Ethem Alpaydin, Introduction to Machine Learning, 4th edition, MIT Press 2020.
2. Bishop, Christopher M., Pattern Recognition and Machine Learning. Springer-Verlag, 2006.
3. Zhi-Hua Zhou, Ensemble Methods: Foundations and Algorithms, CRC Press, 2012.
4. Lior Rokach, Ensemble Learning: Pattern Classification using Ensemble Methods, 2nd ed., World Scientific, 2019.
5. Tom M. Mitchell, Machine Learning, McGraw-Hill Education (India) Private Limited, 2013.
6. Stephen Marsland, Machine Learning: An Algorithmic Perspective, CRC Press, 2014.

7. Andrew Glassner, Deep Learning from basics to practice. Volume1 & 2, Kindle Edition, 2018.

COURSE OUTCOMES:

The students will be able to:

- Select real-world applications that needs machine learning based solutions.
- Implement and apply machine learning algorithms.
- Select appropriate algorithms for solving a particular group of real-world problems.
- Recognize the characteristics of machine learning techniques that are useful to solve real-world problems.

CA617 - BIG DATA ANALYTICS

Objectives:

- To understand the Big Data Platform and its Use cases.
- To provide an overview of Apache Hadoop Ecosystem.
- To provide HDFS Concepts and Interfacing with HDFS.
- To understand Map Reduce concepts.

Introduction to Big Data and Hadoop: Types of Digital Data – Introduction to Big Data – Big Data Analytics – History of Hadoop – Analysing Data with Hadoop – Hadoop Streaming – Hadoop Eco System – Applications of Big Data: marketing – fraud detection – risk assessment – credit risk management – healthcare – medicine – advertising.

HDFS (Hadoop Distributed File System): Design of HDFS – HDFS Concepts – Command Line Interface – Hadoop file system interfaces – Data flow – Data Ingest with Flume and Sqoop – Hadoop I/O: Compression – Serialization – Avro and File-Based Data structures.

Map Reduce: Anatomy of a Map Reduce - Job Run – Failures – Job Scheduling – Shuffle and Sort – Task Execution – Map Reduce Types and Formats – Map Reduce Features - Composing map reduce calculations.

Hadoop Ecosystem: Introduction to PIG – Execution Modes of Pig – Comparison of Pig with Databases – Grunt – Pig Latin – User Defined Functions – Data Processing operators - Hbase – data model and implementations – Hbase clients – Hbase examples – Hive – data types and file formats – HiveQL data definition – HiveQL data manipulation – HiveQL queries.

Mining Data Streams: Stream Data Model – Sampling Data in the Stream – Filtering Streams – Counting Distance Elements in a Stream – Estimating Moments – Counting Ones in Window – Decaying Windows.

REFERENCES:

1. Tom White “Hadoop: The Definitive Guide” Third Edition, O’reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data and Analytics" Wiley Publication, 2015.
3. Raj Kamal, Preeti Saxena, “Big Data Analytics: Introduction to Hadoop, Spark, and Machine-Learning”, McGraw Hill, 2018.
4. Jay Liebowitz, “Big Data and Business Analytics” CRC press, 2013.
5. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
6. Jure Leskovec, Anand Rajaraman, Jeffrey Ullman. “Mining of Massive Datasets.” Cambridge University Press. 2014.

COURSE OUTCOMES:

The students will be able to:

- Describe big data and use cases from selected business domains
- List the components of Hadoop and Hadoop Eco-System
- Access and Process Data on Distributed File System
- Manage Job Execution in Hadoop Environment
- Develop Big Data Solutions using Hadoop Eco System

CA601 - BIG DATA ANALYTICS LAB

Objectives:

- To optimize business decisions and create competitive advantage with Big Data analytics.
- To Impart the architectural concepts of Hadoop and introduce map reduce paradigm.
- To introduce PIG & HIVE in Hadoop ecosystem.
- To Develop Big Data applications for streaming data using Apache Spark.

Exercises:

Problems to

- Perform setting up and installing Hadoop
- Implement file management tasks
- Run different MapReduce programs
- Install and run Pig and write Pig Latin scripts
- Install and run Hive and use Hive
- Install and deploy Apache Spark cluster and run apache spark applications
- Develop Big data applications for streaming data using Apache Spark

COURSE OUTCOMES:

The students will be able to:

- Preparing for data summarization, query, and modeling
- Applying data modeling techniques to large data sets
- Creating applications for Big Data analytics
- Building a complete business data analytic solution

CA603 - MACHINE LEARNING LAB

Objectives:

- To introduce basic machine learning techniques.
- To develop the skills in using recent machine learning software for solving practical problems in high-performance computing environment.
- To develop the skills in applying appropriate supervised, semi-supervised or unsupervised learning algorithms for solving practical problems

- Problem Selection
- Literature Review and Data Collection
- Data Preprocessing
- Model Selection (Machine Learning Models)
- Model Evaluation
- Model Deployment
- Results Analysis

COURSE OUTCOMES:
The students will be able to:

- Implement and apply machine learning algorithms to solve problems.
- Select appropriate algorithms for solving a of real-world problems.
- Use machine learning techniques in high-performance computing environment to solve real-world problems

Semester – II

CA610 – DISTRIBUTED AND CLOUD COMPUTING

Objectives:

- To explore basic concepts and practices of distributed computing.
- To understand Cloud Computing concepts, technologies, architecture and applications.
- To understand different cloud programming platforms and tools to develop and deploy applications on cloud.

Fundamentals of Distributed Computing – Inter Process Communications – Distributed Computing Paradigms - Distributed Objects

Fundamentals of Parallel Computing – Classification of Parallel Computers – Parallel Computer Architectures – Performance Analysis of Parallel Computing – Parallel Computational Model – Introduction to Parallel Algorithms – OpenMP and CUDA

Introduction to virtualization – Different Approaches to Virtualization – Server – Storage – Network Virtualization – VM Migration – Hypervisors – Case Studies: VMware – KVM – Xen – Containerization

Cloud Computing Properties and Characteristics – Business Drivers for Adopting Cloud Computing – Cloud Computing Architecture – Cloud Computing Service Delivery Models: Infrastructure as a Service (IaaS) – Platform as a Service (PaaS) – Software as a Service(SaaS) Deployment Models: Public cloud – Private cloud – Hybrid cloud – Data Center Design and Management – Case Studies: Amazon AWS – Microsoft Azure – Amazon EC2 – Google Cloud

Service Level Agreements (SLAs) – Pricing Models of Cloud – Migrating to Cloud – Cloud Simulators – Cloud Security Risks – Emerging Trends in Cloud Computing

REFERENCES:

1. Kai Hwang, Jack Dongarra, Geoffrey C. Fox, "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things", Morgan Kaufmann, 2013
2. Rajkumar Buyya, Christian Vecchiola, and Thamarai Selvi, "Mastering Cloud Computing", McGraw Hill, 2017.
3. Ajit Singh, "Parallel and Distributed Computing", Kindle Version, e-Book, 2021.
4. Bertil S, Jorge G-D, Christian H and Mortiz S, "Parallel Programming: Concepts and Practice", Morgan Kaufmann, 2018.
5. Sunita Mahajan and Seema Shah, "Distributed Computing", 2nd Edition, Oxford Press, 2013.

COURSE OUTCOMES:

The students will be able to:

- Acquire knowledge of the concepts and technologies of distributed and cloud Computing.
- Demonstrate where to apply parallel and distributed techniques.
- Use various performance criteria to evaluate the quality of the cloud architecture.

CA612 – DEEP LEARNING AND ITS APPLICATIONS

Objectives:

- To introduce the techniques of deep learning.

INTRODUCTION – Learning – Various types Learning – Machine Learning: issues and challenges – CPU vs GPU – massive parallelism – Introduction to Deep Learning – Deep Learning Models: CNN – RNN – AE – GAN – real world applications of Deep Learning – Packages used for Deep Learning.

DEEP LEARNING – Introduction – shallow neural networks – Deep neural networks – Architecture Design – Convolutional Neural Networks – Introduction – Convolution (1D and 2D) – Pooling – Training of network – Hyper parameter tuning – pre-trained models: AlexNet – GoogleNet – Resnet – VGG-16 – VGG-19 – ImageNet – Case study of CNN (Healthcare – Agriculture – Stock Market – Weather Forecasting.

SEQUENCE MODELING – Recurrent Neural Network (RNN) Model – Types of RNNs – Vanishing Gradients with RNN – Gated Recurrent Unit – Long Short Term Memory (LSTM) – Deep Recurrent Neural Networks – RNN for Time Series – Transformer Network Models - Case Studies on Recent Real World Problems.

AUTOENCODERS AND GENERATIVE ADVERSARIAL NETWORKS - Autoencoders – Architecture of Autoencoders – Types of Autoencoders – Applications of Autoencoders – Generative Models – Generative Adversarial Networks – Applications – Autoencoders vs Generative Adversarial Networks – Use Cases.

DEEP REINFORCEMENT LEARNING- Foundations of Reinforcement Learning – Value-based models – Policy-based models – Multi-Agent Reinforcement Learning – Deep Q-Learning – SARSA Learning – Real World Applications.

REFERENCES:

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press 2016.
2. Andrew Glassner, Deep Learning from basics to practice. Volume 1 & 2, Kindle Edition, 2018.
3. François Chollet, Deep Learning with Python, Manning Publications, 2018.

COURSE OUTCOMES:

The students will be able to:

- Explore the essentials of Deep Learning and Deep Network architectures
- Define, train and use Deep Neural Networks for solving real world problems that require artificial intelligence based solutions

CA602 – DISTRIBUTED AND CLOUD COMPUTING LAB

Objectives:

- To introduce basic distributed and cloud computing techniques.
- To develop the skills in applying appropriate distributed and cloud computing techniques for solving practical problems.

- MPI-code parallel program to multiply two large scale N X N matrices
- Problem on Load Sharing Facility (LSF) for cluster Computing.
- Problem related to analysis of the performance of Xen live migration for I/O read-intensive applications.
- Problem related to analysis of the performance of Xen live migration for I/O write-intensive applications.
- Problem on fat-tree concept. (scalability related)
- Program that uses publish-subscribe to communicate between entities developed in different languages.
- Program to run the GAE platform for backup storage of a large amount of different type of data (text, images, video, audio etc.).
- MPI code on any GT$-enable grid platform (set up one with a research team). Write a report to explain the grid experimental setting using the Globus tools.
- Problem related to issues in P2P computing.
- Problem related to WSN.

COURSE OUTCOMES:

The students will be able to:

- Implement and apply distributed and cloud computing techniques to solve problems.
- Select appropriate algorithms for solving real-world problems.

CA604 – DEEP LEARNING LAB

Objectives:

- To introduce basic deep learning techniques.
- To develop the skills in using recent deep learning software for solving practical problems in high-performance computing environment.
- To develop the skills in applying appropriate deep learning algorithms for solving practical problems.

- Problem Selection
- Literature Review and Data Collection
- Data Preprocessing
- Model Selection (Deep Learning Models)
- Model Evaluation
- Model Deployment
- Results Analysis

COURSE OUTCOMES:

The students will be able to:

- Implement and apply deep learning algorithms to solve problems.
- Use deep learning techniques in a high-performance computing environment to solve real-world problems

Semester – III

CA649 PROJECT WORK – PHASE I

- Internal project work of 6 Months duration to be extended in phase II.

Semester – IV

CA650 PROJECT WORK – PHASE II

- Internal project work of 6 Months duration with submission of thesis and viva-voce examination

Semester – I: ELECTIVES

CA6A1: DATA MINING AND WAREHOUSING

Objectives:

- To develop and understand the necessity of Data Mining, the strengths and limitations of popular data mining techniques and how they can be applied in business applications.
- To understand the overall architecture of a data warehouse and methods for data gathering and data pre-processing.
- To explore data mining and data warehousing applications in various real-world datasets.

Fundamentals of data mining and Data Preprocessing: Motivation – Importance – Definition of Data Mining - Data Mining Functionalities - Classification of Data Mining systems - Data Mining Task Primitives - Integration of a Data Mining System with a Database or a Data Warehouse System - Major issues in Data Mining. Types of Data Sets and Attribute Values - Basic Statistical Descriptions of Data - Data Visualization - Measuring Data Similarity Data – Pre-processing: Need for Pre-processing the Data - Data Cleaning - Data Integration and Transformation - Data Reduction - Discretization and Concept Hierarchy Generation

Data Warehouse and OLAP Technology for Data Mining: Data Warehouse - Multidimensional Data Model - Data Warehouse Architecture - Data Warehouse Implementation - Further Development of Data Cube Technology - From Data Warehousing to Data Mining Data Cube Computation and Data Generalization: Efficient Methods for Data Cube Computation – Further Development of Data Cube and OLAP Technology

Mining Frequent Patterns – Associations and Correlations: Basic Concepts on Frequent Itemsets - Efficient and Scalable Frequent Item set Mining Methods - Mining various kinds of Association Rules – Apriori Algorithm – FP-tree algorithm - From Association Mining to Correlation Analysis – Constraint-Based Association Mining

Classification and Prediction: Issues Regarding Classification and Prediction - Classification by Decision Tree Induction - Bayesian Classification - Rule-Based Classification - Classification by Back propagation - Support Vector Machines – Prediction - Accuracy and Error measures - Evaluating the accuracy of a Classifier or a Predictor - Ensemble Methods

Clustering Methods: Cluster Analysis Introduction - Types of Data in Cluster Analysis - A Categorization of Major Clustering Methods- Partitioning Methods – Hierarchical Methods – Density-Based Methods – GridBased Methods – Model-Based Clustering Methods - Clustering

High-Dimensional Data – Constraint Based Cluster Analysis; Outlier Analysis; Mining Stream – Time-Series and Sequence Data – Graph Mining – Mining the Web Data

REFERENCES:

1. Jiawei Han & Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Elsevier, 2012.
2. Margaret H Dunham, Data Mining Introductory and Advanced Topics, 2nd edition, Pearson Education, 2006.
3. Arun K Pujari, Data Mining Techniques, Universities Press, 2010
4. Amitesh Sinha, Data Warehousing, Thomson Learning, 2007.
5. Xingdong Wu, Vipin Kumar, The Top Ten Algorithms in Data Mining, CRC Press, 2009.

COURSE OUTCOMES:

The students will be able to:

- Learn the task of data mining as an important phase of the knowledge discovery process
- Understand the Data warehouse and its implementation
- Understand the mining methodologies
- Study and Implement machine learning algorithms
- Apply Data Mining for Knowledge extraction from real-world databases.

CA6A2 – SOFT COMPUTING TECHNIQUES

Objectives:

- To learn the techniques of soft computing.
- To know artificial neural network concepts.
- To learn the applications of evolutionary and genetic algorithms.
- To know the design of fuzzy controller and rough sets.

Soft Computing: Introduction – Soft Computing versus Hard Computing – Need – Tools and Techniques – Applications

Artificial Neural Network (ANN): Biological Foundation of Neural Network – Neural Model and Network Architectures – Perceptron Learning – Supervised Hebbian Learning – Back-propagation – Associative Learning – Competitive Networks – Hopfield Network – Computing with Neural Nets – and applications of Artificial Neural Network

Evolutionary and Genetic Algorithms: Concepts – Operators – Function Optimization – Dominance – Swarm Intelligence – Colony systems – Modeling Collective Behavior in Social Insects – Division of Labor and Task Allocation – Applications

Fuzzy Logic: Fuzzy Sets – Operations on Fuzzy sets – Fuzzy Relations – Fuzzy Systems – Fuzzy controller – Fuzzy Decision Making – Fuzzy Clustering – Applications. Rough Sets: Imprecise Categories Approximations and Rough Sets – Reduction of Knowledge, Decision Tables – Applications

Neuro Fuzzy and Soft Computing: Adaptive Neuro-Fuzzy Inference System Architecture – Hybrid Learning Algorithm – Learning Methods that Cross-fertilize ANFIS and RBFN – Coactive Neuro Fuzzy Modeling – Framework – Neuron Functions for Adaptive Networks – Neuro Fuzzy Spectrum - Hybridization

REFERENCES:

1. J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing, PHI, 2015
2. G. J. Klir, and B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, 2011
3. S. Rajasekaran and G. A. V. Pai, Neural Networks, Fuzzy Logic and Genetic Algorithms, PHI, 2017
4. D. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Pearson Education, 1989

5. E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, Oxford Press, 1999
6. K. L. Du, M. N. S. Swamy, Neural Networks in a Soft computing Framework, Springer, 2008
7. R. A. Aliev, R. R. Aliev, Soft Computing and Its Applications, World Scientific, 2001
8. D. Ruan, Intelligent Hybrid Systems, Springer, 2011.

COURSE OUTCOMES:

The students will be able to:

- Explain the basics of soft computing and their suitable industry related applications
- Apply neural network principles and algorithms for given problems
- Apply the principles of fuzzy and hybrid algorithms for real time applications
- Solve problems using evolutionary algorithms

CA6A3 – NEXT GENERATION DATABASE SYSTEMS

Objectives:

- To explore the fundamental concepts of next generation database systems.
- To design and develop NoSQL database management systems.

Database Revolutions – Relational Databases – Advanced SQL queries - Transaction Management – Data warehouse and Data Mining – Information Retrieval.

Big Data Revolution – CAP Theorem – Birth of NoSQL – NSQL Databases Concepts – Types and Examples – Performance Analysis.

Document Databases – MongoDB – Couch DB – Key value Databases – DynamoDB – Riak.

Graph Databases – Graph Models – Properties – Neo4J – IBMGraph – Column Oriented Databases – Cassandra – HBase – In-memory Databases – Redis – SQLite.

Distributed Database Patterns – Consistency Models – Data Models and Storage – Languages and Programming Interfaces – Emerging Trends in Databases.

REFERENCES:

1. Dan Sullivan. NoSQL for Mere Mortals. Addison-Wesley Professional. 2015.
2. Guy Harrison. Next-Generation Databases. Apress. 2016.

COURSE OUTCOMES:

The students will be able to:

- Understand how NoSQL databases differ from relational databases from a theoretical perspective.
- Design NoSQL database management systems.
- Select a particular NoSQL database for specific use cases.

CA6B1 – DEVOPS

Objectives:

- To learn the rise of agile methodologies
- To define and design purpose and applications of DevOps

Rise of Agile Methodologies: The Advent of Software Engineering - Waterfall method - Developers vs IT Operations conflict - Agile Vs Waterfall Method - Iterative Agile Software Development - Individual and team interactions over processes and tools – Working software over -comprehensive documentation - Customer collaboration over contract negotiation - Responding to change over following a plan

Basics of DevOps: Introduction to DevOps - DevOps and Agile - Minimum Viable Product - Application Deployment - Continuous Integration - Continuous Delivery

CAMS (Culture, Automation, Measurement, And Sharing): CAMS – Culture - CAMS – Automation - CAMS – Measurement - CAMS – Sharing - Test-Driven Development - Configuration Management - Infrastructure Automation - Root Cause Analysis – Blamelessness - Organizational Learning

Continuous Delivery and Continuous Integration: Rebase frequently from the mainline - Check-in frequently - Frequent build – Automate the testing as much as possible - Automate the deployment - Development operations - Use a version control system - Use repository tools - Use a Continuous Integration tool - Automate the packaging - Automating the deployments - Automating the testing

Applied Devops: DevOps Tools and Platforms - Real World Applications of DevOps - DevOps Practical Examples - Case Studies

REFERENCES:

1. Gene Kim, Jez Humble, Patrick Debois, John Willis and Nicole Forsgren, The DevOps Handbook, IT Revolution Press. 2021
2. Paul Swartout, Continuous Delivery and DevOps A Quickstart Guide, Packt Publishing Limited 2018
3. Len Bass, Ingo Weber, Liming Zhu, DevOps: A Software Architect's Perspective Addison-Wesley. 2015

COURSE OUTCOMES:

The students will be able to:

- Understand the basic concepts DevOps.
- Understand the real world applications of DevOps.

CA6B2 – ANALYTICS FOR STRATEGIC MARKET PLANNING

Objectives:

- To build marketing response models for strategic marketing decisions.

Marketing Analytics: Basics of marketing analytics - marketing decisions models - characteristics - types and benefits of marketing decisions models - Response models - types - calibration - objectives - interactions effects - dynamic effects - competitive effects - models in individual levels - shared experience and qualitative models.

Segmentation and Targeting: The segmentation process and defining the market with models - Segmentation research - methods using factors analysis and cluster analysis - behaviour-based segmentation: cross classification - regression and choice-based segmentation - customer heterogeneity- issues and challenges.

Positioning: Differentiation and positioning - perceptual maps: developing perceptual map – multi dimensional scaling - techniques – attribute based and similarity based - joint space mapping.

Strategic Market Analysis: Strategic marketing decisions - market demand and trend analysis - product life cycle - cost dynamics: scale and experience effects.

Models for Strategic Decision Making: Market entry and exit decisions - PIMS: shared experience models - product portfolio models: BCG - GE etc. - financial models - analytical hierarchy process.

REFERENCES:

1. Lilien, Gary L. and Arvind Rangaswamy, *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*. Revised Second Edition, Trafford Publishing, 2004

COURSE OUTCOMES:

The students will be able to:

- Take data empowered strategic marketing decisions by using analytical techniques.
- Sharpen their analytical skills by getting exposure to computer-based marketing models and tools for decision making.

CA6B3 – FINANCIAL RISK ANALYTICS AND MANAGEMENT

Objectives:

- To identify the different risks involved in Finance arena.
- To understand and solve the different risks pertaining to the stock market and its instruments.
- To analyze the legal issues affecting the business.

Introduction to Risk: Understanding Risk - Nature of Risk - Source of Risk - Need for risk management - Benefits of Risk Management - Risk Management approaches - Risk Classification - Credit risk - Market risk - Operational Risk and Other Risk

Risk Measurements: Measurement of Risk – Credit Risk Measurement - Market Risk Measurement - Interest Rate Risk Measurement - Asset Liability Management - Measurement of Operational Risk

Risk Management: Managing Credit Risk - Managing Operational Risk - Managing Market Risk.

Risk in Instruments: Tools for Risk Management – Derivatives - Combinations of Derivative Instruments - Neutral and Volatile Strategies - Credit Derivatives - Credit Ratings - Swaps.

Regulation and Other Issues: Other Issues in Risk Management - Regulatory Framework - Basel Committee - Legal Issues - Accounting Issues - Tax Issues - MIS and Reporting - Integrated Risk Management.

REFERENCES:

1. Dun, Bradstreet, "Financial Risk Management", TMH, 2006.
2. John C Hull, "Risk Management and Financial Institutions", Pearson, 2015.
3. Aswath Damodharan, "Strategic Risk Taking", Pearson, 2008.

COURSE OUTCOMES:

The students will be able to:

- Identify and categorize the various risks faced by an organization.
- Explore the tools and practices needed to assess and evaluate financial risks.
- Explore risk management practices in an industry.
- Identify and solve legal issues that impact financial and other risk affecting business.

Semester – II: ELECTIVES

CA6C1 - INTELLIGENT SYSTEMS AND PROCESS AUTOMATION

Objectives:

- To know about AI and problem solving techniques
- To understand various search techniques and knowledge representation schemes
- To learn the concepts and working principles of RPA
- To gain knowledge in RPA related tools, technologies and services

Introduction to AI – Problem Solving by Search - Tree Based – Graph based – Breadth First – Depth First – Backtracking – Depth Bounded – Iterative Deepening – Branch and Bound – Bidirectional Search – Heuristic Search – Iterative Search – MIN-MAX – Alpha Beta Pruning

Knowledge Representation and Reasoning – Propositional Logic – FOPL – Rule Based Systems – Chaining Systems – Expert Systems – Managing Uncertainty in Rule Based Systems – Fuzzy Expert Systems

History of Automation - RPA vs Automation - Processes & Flowcharts - Programming Constructs in RPA - What Processes can be Automated - Types of Bots - Workloads which can be automated - RPA Advanced Concepts - Standardization of processes - RPA Development methodologies - Difference from SDLC - Robotic control flow architecture - RPA business case - RPA Team - Process Design Document/Solution Design Document - Industries best suited for RPA - Risks & Challenges with RPA - RPA and Emerging Ecosystem

Introduction to RPA Tool - The User Interface - Variables - Control Flow - - The Assign Activity - The Delay Activity - The Do While Activity - Data Manipulation - Gathering and Assembling Data

Recording Introduction - Web Recording - Input/Output Methods - Screen Scraping -Data Scraping - Scraping advanced techniques - Selectors - Image, Text & Advanced Citrix Automation - Excel Data Tables & PDF - Data Tables in RPA – Extracting Data from PDF - Anchors - Handling User Events & Assistant Bots - Exception Handling – Triggers – Debugging.- Deploying And Maintaining The Bot -Publishing - Creation of Server - Using Server to control the bots - - Connecting a Robot to Server - Deploy the Robot to Server - Managing packages.

REFERENCES:

1. Crina Grosan and Ajith Abraham, “Intelligent Systems – A Modern Approach”, Springer, 2011.
2. Tom Taulli, “The Robotic Process Automation Handbook”, Apress, 2020.

3. Alok Mani Tripathi, “Learning Robotic Process Automation”,Packt, 2018.
4. Husan Mahey, “Robotic Process Automation with Automation Anywhere”, Packt, 2020.

COURSE OUTCOMES:

The students will be able to:

- Understand the Significance of AI and various techniques used in intelligent applications.
- Apply basic principles of Expert systems and their fundamentals, and learning.
- Understand the RPA and various platforms available.
- To show the importance of RPA in solving real world problems.

CA6C2 - COGNITIVE SCIENCE

Objectives:

- To understand the structure, features and functions of brain
- To know about brain's ability to understand symbols, visuals, language and communication
- To understand the algorithms used in information processing at various levels involving thinking process.
- To model artificial intelligent systems involving cognitive aspects.

Behaviourism in Psychology – Cognitive Maps – Plans and Complex Behaviors – Algorithms and Turing Machines – Linguistics and the Formal Analysis of Language – Information-Processing Models in Psychology – Flow of Information – Role of Auditory Localization in Attention and Memory Span – Perception and Communication – Language and Micro-worlds – Mental Imagery – Cognitive Systems as Functional Systems – Anatomy of the Brain and the Primary Visual Pathway – Extending Computational Modeling to the Brain – Mapping the Stages of Lexical Processing – Tomographic Studies of the Cortical Anatomy of Single Word Processing – Studying Memory for Visual Events – Functional Neuroimaging with fMRI.

Symbols and Symbol Systems – Transforming Symbol Structures – Intelligent Action and the Physical Symbol System – Neurally Inspired Models of Information Processing -Neurons and Network Units – Single-Layer Networks and Boolean Functions – Learning in Single-Layer Networks: The Perceptron Convergence Rule - Linear Separability and the Limits of Perceptron Convergence - Multilayer Networks – Backpropagation Algorithm – Information Processing in Neural Networks.

Dynamical Systems – Hypothesis – Motor Control – A-Not-B Error – Bayesianism – Perception as a Bayesian Problem – Architectures for Artificial Agents – Hybrid Architectures: Example of ACT-R – Structure and Function in the Brain – Exploring Anatomical Connectivity – Studying Cognitive Functioning Techniques – Mapping the Brain's Electrical Activity – Locus of Selection Problem – Networks for Attention – Visuospatial Attention – From Data to Maps.

Models of Language Learning – Language and Rules – Understanding a Language and Learning a Language – Language Learning in Neural Networks – Bayesian Language Learning Probabilities in Word and Phrase Segmentation - Understanding Pronouns - Learning Linguistic Categories – Object Permanence and Physical Reasoning in Infancy – Modeling Object Permanence – Modeling the Balance Beam Problem.

Expert Systems and Machine Learning – Robotics: From GOFAI to Situated Cognition and Behavior-Based Robotics - Subsumption Architecture – Multiagent Programming – Neural Prosthetics – Cognitive Science and the Law – Autonomous Vehicles – Combining Deep Learning and Intuitive Knowledge.

REFERENCES:

1. Jose Luis Bermudez, “Cognitive Science: An Introduction to the Science of the Mind”, Cambridge University Press, 2014.
2. Richard Passingham, “Cognitive Neuroscience : A Very Short Introduction”, Oxford University Press, 2016.
3. Alfonzo Doyle, “Advances in Cognitive Science”, Excelic Press, 2019.

COURSE OUTCOMES:

The students will be able to:

- Understand the principles and techniques behind cognitive science.
- Design applications involving cognitive science.

CA6C3 - BLOCKCHAIN TECHNOLOGY

Objectives:

- To understand blockchain systems (Bitcoin and Ethereum)
- To design, build, and deploy smart contracts and distributed applications
- To integrate ideas from blockchain technology into projects.

Distributed Database – Byzantine General problem and Fault Tolerance – Hadoop Distributed File System – Distributed Hash Table – ASIC resistance – Turing Complete – Cryptography: – Hash function – Digital Signature - ECDSA – Memory Hard Algorithm – Zero Knowledge Proof.

Blockchain Introduction – Advantage over conventional distributed database – Blockchain Network – Mining Mechanism – Distributed Consensus – Merkle Patricia Tree – Gas Limit – Transactions and Fee – Anonymity – Reward – Chain Policy – Life of Blockchain application – Soft & Hard Fork – Private and Public blockchain.

Distributed Consensus – Nakamoto consensus – Proof of Work – Proof of Stake – Proof of Burn – Difficulty Level – Sybil Attack – Energy utilization and alternate.

Cryptocurrency – Distributed Ledger – Bitcoin protocols – Mining strategy and rewards – Ethereum – Construction – DAO – Smart Contract – GHOST – Vulnerability – Attacks – Sidechain – Namecoin.

Cryptocurrency Regulation – Stakeholders – Roots of Bit coin – Legal Aspects-Crypto currency Exchange – Black Market and Global Economy – Applications: – Internet of Things – Medical Record Management System – Domain Name Service and future of Blockchain

REFERENCES:

1. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder, Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction, Princeton University Press (2016).
2. Bashir Imran, Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks, Packt Publishing; 1st edition, 2017.

COURSE OUTCOMES:

The students will be able to:

- Explain design principles of Bitcoin and Ethereum.
- Explain the Simplified Payment Verification protocol.
- Interact with a blockchain system by sending and reading transactions.
- Design, build, and deploy a distributed application.

- Evaluate security, privacy, and efficiency of a given blockchain system.

CA6D1 - NATURAL LANGUAGE PROCESSING

Objectives:

- To introduce language processing techniques to enable Text data processing
- To impart knowledge on text data processing using Statistical and Machine learning models
- To describe the various Embeddings and Deep learning models for NLP
- To introduce real world applications of Language processing

Natural Language Processing – Applications of NLP – Linguistic Background – NLP tasks – Ambiguities in NLP tasks – Finite state automata – Regular Expressions – Corpus – Text Normalization – Edit Distance – Boundary Determination – Tokenization – Stemming - Lemmatization.

Morphological Analysis – Part of speech tagging – Shallow parsing – Dependency parsing – WordNet- Sematic similarity measures – Semantic representation – Coreference Resolution – Tools – Natural Language Toolkit – Stanford CoreNLP.

Language model – n-gram language models – Hidden Markov Model – Conditional random Fields – Topic models – Graph Models – Machine Learning for NLP – Language Features – Maximum Entropy classifier – Phrase Based clustering.

Deep Learning for NLP - Neural Networks – Vector Representations – Word Embeddings – Attention model – Encoder Decoder – Transformer based models – Tools: PyTorch.

Applications and Case Studies – Question Answering – Machine Translation – Information retrieval – Information Extraction

REFERENCES:

1. Anders Søgaard, Ivan Vulić, Sebastian Ruder, Manaal Faruqui , Cross-Lingual Word Embeddings (Synthesis Lectures on Human Language Technologies), Morgan & Claypool Publishers, 2019
2. Delip Rao, Brian McMahan, Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning, O'Reilly, 2019
3. Daniel Jurfsky, James H. Martin Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Second Edition, Pearson 2013
4. Christopher D. Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999.

COURSE OUTCOMES:

The students will be able to:

- Identify the patterns in text and pre-process the large text corpus
- Describe and work with basic NLP tasks
- Use statistical and machine learning models with feature selection for NLP
- Adopt Deep learning models for NLP along with embeddings
- Apply the concepts for solving NLP Applications.

CA6D2 - SOCIAL NETWORK ANALYTICS

Objectives:

- To enable students to understand and visualize the social network structure.
- To impart knowledge on mining the user link structure in the social network.
- To enable students to analyse the social network content

Introduction – Properties of Social Networks – Network Analysis – Network Measures – Discussion Networks – Blogs and Online Communities – Semantic Web

Modeling and Visualization – Visualizing Social Networks – Graph Representation – Node-Edge Diagrams – Modelling and Aggregating Social Network Data – Random Walks – Tools: NetworkX – SocioViz – NetworkKit – Infinite Graph

Mining Communities – Aggregating and Reasoning with Social Network Data – Detecting Communities in Social Networks – Evaluating Communities – Community Welfare – Collaboration networks – Co-Citation Networks - Link Prediction – Rank Aggregation and Voting Theory – Tools: Ghephi – Neo4J – Mathematica

Semantic Web – Resource Description Framework – Linked Open Data – Ontological Representation of Social Individuals and Relations – Entity Linking – Entity Resolution – Graph based Event Detection

Social Media Content Analytics – Text Mining in Social Networks – Opinion Extraction – Sentiment Classification and Clustering – Irony Detection – Review Classification – Recommendation Systems – Hashtags – Information Diffusion

REFERENCES:

1. Matthew A. Russell. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More, 3nd Edition, O'Reilly Media, 2019.
2. Jennifer Golbeck, Analyzing the social web, Morgan Kaufmann, 2013.
3. GuandongXu, Yanchun Zhang and Lin Li, Web Mining and Social Networking – Techniques and applications, 1st edition, Springer, 2011.
4. BorkoFurht, Handbook of Social Network Technologies and Applications, 1st edition, Springer, 2010.
5. John G. Breslin, Alexander Passant and Stefan Decker, The Social Semantic Web, Springer, 2009.

COURSE OUTCOMES:

The students will be able to:

- Interpret the social network structure
- Model and visualize the social network.
- Mine the user relations in the social network.
- Analyse social network content

CA6D3 - SUPPLY CHAIN ANALYTICS

Objectives:

- To provide the knowledge and necessary skills for carrying out job roles in the domain of Supply chain analytics

Demand Planning: Demand Planning- Review of Forecasting and planning concepts - Defining KPIs- Forecasting Model Building - Discrete and continuous manufacturing

Supply planning: Supply planning - Procurement and Strategic Sourcing - Inventory Modelling – aggregate planning and resource allocation decisions - Procurement Analytics - Production modelling

Demand Fulfilment: Demand Fulfilment - DC location and network design - optimizing inventory levels in distribution network - Logistics & Network Modelling - Transportation modelling - delayed differentiation - mass customization

Integrated supply chain: Advanced and business supply chain related topics like CPFR - DDSN - Make/Buy Case Study - Total Supply Chain Cost - computation of transfer prices - revenue management-yield management - product changes/economies of scale

Project Development: Undertaking projects with industry inputs- validation of models – frameworks – Review of data analytics techniques - choice of tools and designing solution approach to specific applications

REFERENCES:

1. Raman, A & Fisher, M., *How Analytics Are Transforming the Supply Chain and Improving Performance,* HBS Press. 2010.
2. Tayur, S. Ganeshan, R. & Michael, M. (editors). *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers. 1999.

COURSE OUTCOMES:

The students will be able to:

- Understand the concepts of supply chain analytics.
- Apply various supply chain management concepts.
- Improve an existing supply chain and design an efficient supply chain in alignment with the strategic goals of the company.

CA6E1 - WEB ANALYTICS

Objectives:

- To know the methodologies for collecting, reporting and analysis of website data
- To understand various analysis associated with website and web usage data
- To understand the features of Web Analytics to improve the business in the web, mobile and cloud domains**.**

Web Analytics Approach – Know about web site – Gather Data – Transform – Balancing Time and the Need for Certainty – Context; How Web Analytics Works – Log File Analysis – Page Tagging – Cookies – Accuracy – Accounts and Profiles – Click Analytics – Metrics and Dimensions – Interacting With Data In Google Analytics – Goals and Conversions – Conversion Rate – Goal Reports – Performance Indicators.

Learning about Users – Visitor Analysis – Demographics – Behavior – Technology – Mobile – Custom based – Traffic Analysis – Source and Medium (Dimensions) – Organic Search – Search Query Analysis – Referral and Direct Traffic – Content Usage Analysis – Website Content Reports – Pageviews – Low and High Time – Pageviews Ratio Bounce rate – Page Value – Click – Path Analysis – Relationships between Pages – Navigation Summary – Visitors Flow – Examples.

Segmentation – Google Analytics' Advanced Segments – AND – OR – and Sequence of Filters – Metrics – Dimensions – Based on Page – User Traits – Information need – User's Goal Completion – Page Related Landing and Viewing – Pairing Analytics Data with UX Methods – Personas – Usability Testing and Inspection – Measuring Effects of Changes – Reframe as a Rate – Types of Changes – Conversion Rate – Redirect Traffic – Page Time – Other Continuous Metrics – Reporting.

Measuring Behavior within pages – Google In – Page Analytics – Measurable Clicks – Analytics Tools – Event Data Analysis – Virtual Page Views – A/B Testing – Analytics Profiles – Filters – Reporting Culture – Making Case for Usability Activities – Design Changes – User Research – Mobile Application Analytics – Cross – Device Measurement – On Page Behavior – Connecting to other data sources – Google Analytics – E – Commerce Tracking – Online Campaign Tracking – Event Tracking – Customizing GATC – Hacks – Methodologies.

Google Tag Managers – Building blocks – Enhancing websites with Tag Manager – Using Google Analytics with BigQuery – Tag Manager and Analytics APIs.

REFERENCES:

1. Michael Beasley, "Practical Web Analytics for User Experience", Elsevier, 2013

2. Beatriz Plaza “Advanced Web Metrics with Google Analytics”, Koros Press, 2017
3. Todd Kelsey, “Introduction to Google Analytics: A Guide for Absolute Beginners”, Apress, 2017.
4. Jonathan Weber et al., “Practical Google Analytics and Google Tag Manager for Developers”, Apress, 2015.

COURSE OUTCOMES:

The students will be able to:

- Understand the factors influencing the business through the web site data
- Understand various types of analysis involved in web site data (users, business data, etc.)
- Practice applying Google Analytics APIs to improve the business process.

CA6E2 - FEDERATED LEARNING

Objectives:

- To know the basics of federated learning.
- To learn the applications of federated learning
- To know the privacy preserving deep learning.

Introduction to Federated Learning – Privacy Preserving – Distributed Machine Learning – Threats to Federated Learning – Data Valuation.

Horizontal Federated Learning – Vertical Federated Learning and Federated Transfer Learning.

Federated Optimization for Heterogeneous Networks – Deep Networks from Decentralized Data – Federated Multi-task Learning – Personalized Federated Learning.

Federated Learning Applications: Recommendation in Health Care and Finance - Mobile Keyboard Prediction – Learning of Out-of-vocabulary Words.

Adaptive Personalized Federated Learning – Privacy Preserving Deep Learning – Advances and Open Problems.

REFERENCES:

1. Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, Han Yu, Federated Learning, Morgan & Claypool Publishers, 2019.
2. Kim-Kwang Raymond Choo, Ali Dehghantanha, Handbook of Big Data Privacy, Springer Nature Switzerland, 2020.

COURSE OUTCOMES:

The students will be able to:

- Knowledge of the basic concepts, architecture and applications of Federated Learning.
- Understanding of new research and application trends in Federated Learning.
- Ability to deploy real-world Federated Learning projects.
- Hands-on experience in applying Federated Learning tools to solve privacy-preserving AI challenges.

CA6E3 - HUMAN RESOURCE ANALYTICS

Objectives:

- To provide the knowledge on various frameworks and hands-on analytical approaches for identifying the business contributions of the HR function in the organizations.

Introduction to HR Analytics: Overview of HR Process - HR Analytics – An overview - Role of analytics in HR – HRM in changing context - Transition from transaction orientation to analytics orientation - HR Analytics Framework – People Capability Maturity Model – LAMP framework – HCM 21 framework – Talent analytics maturity model.

Organization-Wide Alignment Analytics: Alignment Analytics – link HR process and organization process - identify alignment opportunities; Human Capital Strategy - HR Alignment Inventory - workforce planning - Measurement Map – Lead and Lag indicators.

HR Metrics and Audits: Formulation of key performance indicators and key result areas; HR Metrics – Recruitment metrics – Training and development metrics - Talent retention metrics –HR cost benefit metrics – Career Progression Metrics - Performance metrics – Diversity and Inclusion Metrics - Human capital ROI - Designing and Implementing HR Scorecard - Conducting HR Practice Audits

Descriptive Analytics: Descriptive Analytics – Exploring the people data – slice and dice of data – HR Dashboards - Segmentation - Business Insights - KPI Catalogue Creation.

Predictive and Prescriptive Analytics: Predicting future performance - Techniques to capture the fallouts of HR Practices – Data driven decision making - Organization change and improvement.

REFERENCES:

1. Bassi, Laurie Jo, Rob Carpenter, and Dan McMurrer. *HR analytics handbook*. Reed Business. 2010.
2. Fitz-Enz, Jac. *The New HR Analytic. Predicting the Economic Value of Your Company's Human Capital Investments*. American Management Association. 2010.

COURSE OUTCOMES:

The students will be able to:

- Identify necessary skills to carry out the personnel roles in the domain of HR.
- Identify and develop metrics to improve employer-employee relationship and improve employee retention.
- Identify skilled personnel and job tasks to achieve mission-critical goals.
- Align organization's mission and goals with key metrics and benchmarks.
- Apply HR analytics to improve organizational performance by providing better insights on human resources data.

CA6F1 - DATA VISUALIZATION

Objectives:

- To introduce students with fundamental concepts, approaches, and applications of data visualization
- To familiarize students with different visualization and analytic tools and techniques
- To develop skills and competencies required by the professionals

Data – Perception and Data Visualization – Graphical Excellence – Data Maps – Time-Series – Graphics of Space and Time – Graphical Integrity – Distortion in a Data Graphic – Design and Data Variation – Visual Area and Numerical Measure

Data-Ink and Graphical Redesign – Data-Ink – Maximizing the Share of Data-Ink – Two Erasing Principles – Application of the Principles in Editing and Redesign - Data-Ink Maximization and Graphical Design – Redesign of Box Plot – Histograms and Scatterplot – Data Density

Visualizing Amounts – Bar Plots – Grouped and Stacked Bars – Dot Plots and Heatmaps - Visualizing Distributions – Histograms and Density Plots – Q-Q Plots – Skewed Distributions - Visualizing Proportions – Pie Charts - Visualizing Time Series – Individual Time Series – Multiple Time Series – Time Series with Multiple Response Variables

Mapping – Connections and correlations – Multivariate Analysis – Scatterplot Maps – Trees – Hierarchies and Recursion – Networks and Graphs

Visual Analytics – Visual Analytics vs Data Visualization – Visual Analytics Process – Visual Analytics Applications – Case Studies - Python Data Visualization Libraries – Tableau

REFERENCES:

1. Edward R. Tufte, The Visual Display of Quantitative Information, 2nd edition, Graphics Press, 2007
2. Claus O. Wilke, Fundamentals of Data Visualization, O'Reilly, 2019
3. Ben Fry, Visualizing Data: Exploring and Explaining Data with the Processing Environment, O'reilly, 1st edition, 2008
4. A. Kerren, J. T. Stasko, J. Fekete, C. North, Information Visualization, Springer, 2008
5. Alex Campbell, Data Visualization: Ultimate Guide to Data Mining and Visualization, 2020
6. A. Loth, Visual Analytics with Tableau, Willey, 2019
7. Kieran Healy, Data Visualization: A Practical Introduction, Princeton University Press, 2018
8. Alexandru C. Telea, Data Visualization: Principles and Practice, Second Edition, CRC Press, 2015

COURSE OUTCOMES:

The students will be able to:

- Understand and apply the fundamental design principles of data visualization
- Present data by applying different data visualization techniques for better communication
- Analyse data and communicate information through visual representation

CA6F2 - PATTERN RECOGNITION

Objectives:

- To develop the mathematical tools required for the pattern recognition.
- To learn about supervised and unsupervised pattern classifiers.
- To familiarize about different feature extraction techniques

Basics of Probability – Random Processes and Linear Algebra – Probability– independence of events – conditional and joint probability – Bayes' theorem - Random Processes – Stationary and nonstationary processes – Expectation – Autocorrelation – Cross-Correlation – spectra; Linear Algebra– Inner product – outer product – inverses – eigen values – eigen vectors– Bayes Decision Theory.

Bayes Decision Theory – Minimum-error rate classification – Classifiers – Discriminant functions – Decision surfaces – Normal density and discriminant functions – discrete features Parameter Estimation Methods – Maximum-Likelihood estimation: Gaussian case – Maximum a Posteriori estimation– Bayesian estimation - Gaussian case.

Unsupervised learning and clustering – Criterion functions for clustering – Algorithms for clustering – K-Means – Hierarchical and other methods – Cluster validation – Gaussian mixture models – Expectation– Maximization method for parameter estimation – Maximum entropy estimation.

Sequential Pattern Recognition – Hidden Markov Models (HMMs) – Discrete HMMs - Continuous HMMs – Nonparametric techniques for density estimation – Parzen-window method – K-Nearest Neighbour method

Dimensionality reduction – Fisher discriminant analysis – Principal component analysis – Factor Analysis – Linear discriminant functions – Gradient descent procedures – Perceptron – Support vector machines.

REFERENCES:

1. S.Theodoridis and K.Koutroumbas, Pattern Recognition, 4th Ed., Academic Press, 2009
2. R.O.Duda, P.E.Hart and D.G.Stork, Pattern Classification, John Wiley, 2001
3. C.M.Bishop, Pattern Recognition and Machine Learning, Springer, 2006
4. P.A. Devijer & J. Kittler, Pattern Recognition - A Statistical Approach, Prentice-Hall.
5. Christopher. M. Bishop, 'Pattern recognition and machine learning, Springer, 2006.

COURSE OUTCOMES:
The students will be able to:

- Summarize the various techniques involved in pattern recognition
- Categorize the various pattern recognition techniques into supervised and unsupervised.
- Illustrate the artificial neural network-based pattern recognition
- Discuss the applications of pattern recognition in various applications

CA6F3 - IMAGE AND VIDEO ANALYTICS

Objectives:

- To teach the fundamentals of digital image processing, image and video analysis.
- To understand the real time, use of image and video analytics.
- To demonstrate real time image and video analytics applications and others.

Digital image representation – Visual Perception – Sampling and Quantization – Basic Relations between Pixels – Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matric Operations – Image Transforms (DFT – DCT – DWT – Hadamard).

Fundamentals of Spatial Filtering - Spatial Correlation and Convolution – Smoothing – Blurring – Sharpening – Edge Detection – Basics of Filtering in the Frequency Domain: smoothing – blurring – sharpening – Histograms and Basic Statistical Models of Image.

Image Segmentation – Region based Segmentation – Image and Video Segmentation – Image and Video Enhancement – Image and Video Compression.

Object Detection – CNN – LSTM – Transformers – Object Recognitions Models – Image and Video Classification Models.

Applications and Case Studies – Face Detection and Recognition – Automatic Image Captioning – Image Augmentation – Object Tracking and Motion Estimation – Human Action Recognition.

REFERENCES:

1. R.C. Gonzalez and R.E. Woods." Digital Image Processing". 3rd Edition. Addison Wesley, 2007.
2. W. Härdle, M. Müller, S. Sperlich, A. Werwatz, "Nonparametric and Semi parametric Models", Springer, 2004.
3. Rick Szelisk, "Computer Vision: Algorithms and Applications", Springer 2011.
4. Jean-Yves Dufour, "Intelligent Video Surveillance Systems", Wiley, 2013.
5. Caifeng Shan, Fatih Porikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business Intelligence", Springer, 2012.
6. AsierPerallos, Unai Hernandez-Jayo, Enrique Onieva, Ignacio Julio García Zuazola, "Intelligent Transport Systems: Technologies and Applications", Wiley, 2015.
7. Basudeb Bhatta, "Analysis of Urban Growth and Sprawl from Remote Sensing Data", Springer, 2010.
8. Ian Good Fellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT Press, 2017

COURSE OUTCOMES:

The students will be able to:

- Describe the fundamental principles of image and video analysis and have an idea of their application.
- Apply image and video analysis in real world problems.