CS-301T   # Data Mining Techniques   **Credits 4(3-1-0)**

UNIT 1: Introduction: Types of data, learning of Computers, Self-learning & Methodology of science, Understanding of Concept learning using A kangaroo in mist example. KDD: Definition, Golden rules to setup KDD environment, KDD-processes; Introduction to Data mining, Data Mining System Architecture, Different phases of Datamining process, Types of Datamining techniques and Application areas; Data Preprocessing (Cleaning: Handling of missing values and Noisy data; Data Integration and Transformation, Data Reduction), **Data Warehouse:** problem definition, characteristics, OLAP, Multidimensional Data Analysis and data warehouse schemas, 3-level Architecture of data warehouse, Bigdata, DW vs Data Lake VS Lake house.

UNIT 2:Data Science: Problem definition, data, data types and data exploration, data science and data science process; **Artificial Intelligence:** Overview, AI For Data Mining, AI evolution, AI canonical Architecture, Machine learning deep Dives, Machine learning for datamining, Ensemble learning such as Random Forests, Bagging and AdaBoost; **Deep Learning:** Problem definition, McCulloch–Pitts model, Activation functions, single layer and multilayer Perceptions and examples,Anatomy of NN and design choices, major types of Neural networks.

UNIT 3: Data Mining Tasks-1: Problem definition, **Regression:** Introduction, Statistical modeling (KNN), model fitness and comparison; Linear Regression Model, estimate of the regression coefficient, model selection,Bias vs Variance;Logistic Regression Model, Estimation in Logistic regression, classification using logistic model, multiple logistic regression model; **Classification:** Introduction, regression vs classification, Types of classification Techniques, ensembled classifier, types of ensembled classifier; Decision Tree:Introduction, Tree construction principle, Basic Algorithm for Decision tree Induction, Understanding of Entropy and Gini Index using ID3, C-4.5 and CART algorithms;

UNIT 4: Data Mining Tasks-2: Problem definition, **Clustering:** Problem definition, clustering paradigms, Measurethe quality of clustering, Major Clustering Approaches: Partitioning clustering method [K-means and k-medoids (PAM, CLARA, & CLARANS) algorithms], Hierarchical clustering method [(BIRCH, CURE)], Density Based Method (DBSCAN), Model based clustering method, and Constraint based clustering method, outlier analysis; **Association Rules:** Problem definition, Frequent item set generation, Rule generation, Challenges, Apriori algorithm, Partition algorithms, Pincer search algorithms, Incremental algorithms, Border algorithm.

UNIT 5: Other Data Mining Tasks: Web Mining: Problem definition, Types of Web Mining (Web content mining, Web structure mining, Web users mining); **Text Mining:** Types and Challenges, Intelligent Information Retrieval for Text mining, Text classification; **Search engine for web mining**: Basics and its component, Web search product and services, Search Strategies for search Engine [word Frequency, Popularity, and expert database), Approaches for Ranking the Web pages (Page Rank, Hubs and Authorities).

References:
      1) Data Mining techniques by Arun Pujari, Universities Press (India) P Ltd.
      2) Data Mining by Pieter Adriaans, Dolf Zantinge, Addison-Wesley.
      4) Datawarehousing, Data mining, OLAP, by Alex Berson & Stephen J. Smith, TMH Edition
      5) Berson, "Data Warehousing, Data-Mining & OLAP", TMH
      6) Margaret H. Dunham, "Data-Mining. Introductory & Advanced Topics", Pearson Education
      7) Deep learning, I. Good fellow and Y. Bengio and A. Courville, MIT press Book Press