



INFORMATION TECHNOLOGY  
ENGINEERING

# IT ENGINEERING SEM VI

# DATA MINING & BUSINESS

# INTELLIGENCE

Programming & development

Course Curriculum



# DATA MINING & BUSINESS INTELLIGENCE SEM VI

## **Module 1 : Data Warehouse (DWH) Fundamentals with Introduction to Data Mining**

- DWH characteristics
- Dimensional modelling
  - Star
  - Snowflakes
  - OLAP operation
  - OLTP vs OLAP
  - Data Mining as a step in KDD
- Kind of patterns to be mined
- Technologies used
- Data Mining applications

## **Module 2: Data Exploration and Data Preprocessing**

- Types of Attributes
- Statistical Description of Data
- Measuring Data Similarity and Dissimilarity
- Why Preprocessing?
- Data Cleaning
- Data Integration
- Data Reduction
  - Attribute Subset Selection
  - Histograms
  - Clustering
  - Sampling
  - Data Cube aggregation
- Data transformation and Data Discretization
  - Normalization
  - Binning
  - Histogram Analysis

## **Module 3: Classification**



- Basic Concepts
- Classification methods
  - Decision Tree Induction
    - Attribute Selection Measures
    - Tree pruning
  - Bayesian Classification
  - Naïve Bayes Classifier
  - Prediction
    - Structure of regression models
- Simple linear regression
- Accuracy and Error measures
- Precision
- Recall
- Holdout
- Random Sampling
- Cross Validation
- Bootstrap
- Introduction of Ensemble methods
- Bagging
- Boosting
- AdaBoost and Random forest



## **Module 4: Clustering and Outlier Detection**

- Cluster Analysis
  - Basic Concepts
- Partitioning Methods
  - K-Means
  - K Medoids
- Hierarchical Methods
  - Agglomerative
  - Divisive
  - BIRCH
- Density-Based Methods
  - DBSCAN
- What are outliers?
  - Types
  - Challenges
- Outlier Detection Methods



- Supervised
- Semi Supervised
- Unsupervised
- Proximity based
- Clustering Based

## **Module 5: Frequent Pattern Mining**

- Basic Concepts
- Market Basket Analysis
- Frequent Itemset
- Closed Itemset
- and Association Rules
- Frequent Itemset
- Mining Methods
  - The Apriori Algorithm
    - Finding Frequent Itemset Using Candidate Generation
    - Generating Association Rules from Frequent Itemset
    - Improving the Efficiency of Apriori
    - A pattern growth approach for mining Frequent Itemset
    - Mining Frequent Itemset using vertical data formats
- Introduction to Advance Pattern Mining
  - Mining Multilevel Association Rules and Multidimensional Association Rules

## **Module 6: Business Intelligence**

- What is BI?
- Business intelligence architectures
- Definition of decision support system
- Development of a business intelligence system using Data Mining for business Applications like Fraud Detection
- Recommendation System