M.Sc. IT Data Management and Visual Insight(5 Years Integrated)

Centre for Professional Courses,
Gujarat University Navarangpura, Ahmedabad-380009, Gujarat

Assignment
Paper: - DMVIMSC41
Machine Learning Algorithms

1	**Implementing Boosting (Adaboost & Gradient Boosting)** 1. Load a classification dataset (e.g., UCI Credit Approval dataset). 2. Implement **AdaBoost** and **Gradient Boosting** classifiers. 3. Compare their performance with a base Decision Tree classifier. 4. Use ROC curves and confusion matrices to analyze the results.
2	**Regularization Techniques (L1 & L2 Regularization)** 1. Load a regression dataset (e.g., Boston Housing dataset). 2. Implement **Linear Regression**, **Lasso (L1 Regularization)**, and **Ridge (L2 Regularization)**. 3. Compare their performance in terms of mean squared error (MSE) and coefficient values. 4. Visualize the effect of regularization strength on model performance.
3	**Performance Measurement and Model Evaluation** 1. Train a classification model (Random Forest, Logistic Regression, or any of choice) on the UCI Adult Income dataset. 2. Evaluate it using different performance metrics: a. Confusion Matrix b. Precision, Recall, F1-score c. ROC-AUC curve d. Cross-validation scores 3. Interpret the results and discuss the model's performance.
4	**Hyperparameter Tuning & Model Optimization** 1. Perform **GridSearchCV** or **RandomizedSearchCV** for one of the above models using the dataset from UCI Adult Income dataset or UCI Breast Cancer dataset. 2. Analyze how hyperparameter tuning improves performance.
5	**Implementing Apriori Algorithm** 1. Load a transactional dataset (e.g., Online Retail Dataset from UCI). 2. Perform necessary data preprocessing (handling missing values, formatting transactions, etc.). 3. Implement the **Apriori algorithm** using the mlxtend library. 4. Extract frequent itemsets with different support thresholds. 5. Generate association rules based on confidence and lift.
6	**Market Basket Analysis on Real-world Dataset** 1. Load the Market Basket Dataset from Kaggle. 2. Perform necessary preprocessing (grouping transactions, filtering relevant data, etc.). 3. Implement Apriori to find associations between items. 4. Visualize the top rules using network graphs or heatmaps.

Assignment

Paper: - DMVIMSC41

Machine Learning Algorithms

7	Load the store_data.csv into the DataFrame find the association rule with minimum support =0.0095 and min_confidence= 0.2.