

**COURSE NAME: INTRODUCTION TO MACHINE LEARNING**

**COURSE ID: CS4640**

**INSTRUCTOR: DR. ASIM WAGAN**

**ID: FA21-BSCS-0030**

**NAME: SYED MUSTAFA HASSAN**

**ASSIGNMENT-02**

**TITLE:**

Customer Churn Analysis

**PROBLEM UNDERSTANDING AND DATA EXPLORATION:**

**Customer churn is a significant issue for businesses, especially those relying on a subscription-based model. Churn refers to customers stopping their use of a product or service, leading to revenue loss. Predicting churn enables businesses to take proactive measures to retain customers, thereby reducing turnover and increasing customer lifetime value.**

**Objective:** The goal is to predict whether a customer is likely to churn based on their usage patterns and demographic information.

**Dataset:** The dataset used for this analysis is the Telco Customer Churn dataset from Kaggle, which contains customer data including demographic information, account details, and service usage.

**DATA PREPROCESSING:**

**Handling Missing Values:**

The dataset is checked for missing values, and appropriate strategies are applied if any are found.

**Dropping Irrelevant Features:**

The ***customerID***column is dropped as it is not useful for prediction.

**Encoding Categorical Variables:**

Convert categorical variables to numerical using ***LabelEncoder***.

**Feature Scaling:**

Numerical features are standardized to ensure they have a mean of 0 and a standard deviation of 1.

**Splitting Data:**

The data is split into training and test sets with balanced representation of churn and non-churn cases.

**FEATURE ENGINEERING:**

**Creating New Features:** Additional features that might be predictive of churn, such as interaction frequency and subscription length, can be created.

**Feature Selection:** Use SelectKBest to identify the most important features for prediction.

**MODEL DEVELOPMENT:**

**Model Selection:** Various machine learning algorithms are used, such as Logistic Regression, Random Forest, and Gradient Boosting.

**Hyperparameter Tuning:** Hyperparameters are optimized using techniques like GridSearchCV or RandomizedSearchCV.

**MODEL EVALUATION:**

**Performance Metrics:** Evaluate models using accuracy, precision, recall, F1-score, and ROC-AUC.

**MODEL INTERPRETATION AND DEPLOYMENT:**

**Feature Importance:** Use feature importance scores to understand which features are most influential.

**Deployment Considerations:** Discuss how the model can be integrated into existing systems and monitored over time.

**BUSINESS INSIGHTS AND RECOMMENDATIONS:**

Based on model predictions, provide actionable insights for reducing churn, such as improving customer service, offering incentives, and identifying at-risk customers early.