**CUSTOMER CHURN PREDICTION USING MACHINE LEARNING**

**Abstract:**

Customer churn prediction is crucial for businesses to retain customers and improve profitability. This project aims to develop a machine learning model to predict customer churn based on various behavioral and demographic factors. By using data preprocessing, feature engineering, and classification techniques, we aim to build an accurate predictive model.

**Introduction**

Customer churn refers to the process where customers stop using a company's services. Predicting churn helps businesses take proactive measures to retain customers. This project leverages machine learning models to analyze customer data and predict potential churners. Various classification algorithms such as Logistic Regression, Decision Trees, Random Forest, and XGBoost are used for prediction.

## ****System Analysis****

### **Existing System:**

* Many businesses rely on traditional methods such as surveys and customer feedback to analyze churn.
* These methods are time-consuming and may not provide real-time insights

### **Proposed System**

* A machine learning-based system that automates churn prediction using historical customer data.
* Uses various classification models and data balancing techniques to improve accuracy.

## ****System Specification****

### **Software Requirements:**

* Programming Language: Python
* Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Imbalanced-learn (SMOTE), XGBoost, Pickle.
* IDE: VS Code

## ****Methodology****

### **1. Data Collection**

* Customer transaction data, service usage, and demographic information.

### 2. **Data Preprocessing**

* Handling missing values.
* Encoding categorical features using Label Encoding.
* Balancing the dataset using SMOTE (Synthetic Minority Over-sampling Technique).

### **3. Feature Selection & Engineering**

* Splitting data into training and testing sets using *train\_test\_split*.
* Training multiple classification models:
  + 1. Logistic Regression
    2. Decision Tree Classifier
    3. Random Forest Classifier
    4. XGBoost Classifier

### **5. Model Evaluation**

* Measuring performance using:

1. Accuracy Score
2. Confusion Matrix
3. Classification Report (Precision, Recall, F1-score)
4. Cross-validation

### **6. Model Deployment**

* Saving the best-performing model using pickle for future use.

## ****Module Description****

### **1. Data Preprocessing Module**

* Handles missing values and encodes categorical features.
* Uses SMOTE to balance the dataset.

### **Feature Engineering Module**

* Selects the most important features for the model.

### **Model Training & Evaluation Module**

* Trains multiple classifiers and evaluates performance.

### **Prediction & Deployment Module**

* Saves the model using *pickle* and allows real-time predictions.

## ****Conclusion****

This project successfully demonstrates how machine learning can be used to predict customer churn. The implementation of multiple models allows for performance comparison, and the best model can be used to make accurate predictions.

## ****Future Enhancement****

* Integration with real-time customer data streams.
* Deployment as a web application for easy business use.
* Implementation of deep learning techniques for improved accuracy.
* Use of additional customer behavioral metrics for better insights.