**CUSTOMER CHURN ANALYSIS FOR TELECOM INDUSTRY**

**ABSTRACT:**

This project addresses the challenge of customer churn in the telecommunications industry by developing a machine learning model to predict whether a customer is likely to leave the company. Utilizing a customer dataset, the analysis involves preprocessing, exploratory data analysis, model building using a Random Forest Classifier, and model interpretability techniques such as SHAP values. The final model achieved satisfactory accuracy and interpretability, offering valuable insights for retention strategies.

**INTRODUCTION:**

Customer churn is a critical concern in highly competitive industries like telecommunications. Understanding the factors that lead customers to discontinue service enables companies to implement effective retention strategies. This project aims to develop a predictive model for customer churn using historical data, with the goal of identifying at-risk customers before they leave.

**WORKING:**

The Telecom Customer Churn Prediction project operates by leveraging customer data to identify individuals who are likely to stop using the service. It begins with data preprocessing, where missing or inconsistent values in fields like TotalCharges are handled and categorical data is encoded numerically for model compatibility. Following this, exploratory data analysis (EDA) is conducted to visualize patterns and relationships in the data—such as how tenure, contract type, and monthly charges relate to customer churn. The cleaned and transformed data is then split into training and testing sets. A Random Forest Classifier is trained on the training data to learn the underlying patterns that distinguish churned customers from those who remain. This model is fine-tuned using GridSearchCV to find the best hyperparameters. The performance of the model is evaluated using metrics like accuracy, confusion matrix, ROC-AUC, and precision-recall curves. Finally, SHAP (SHapley Additive exPlanations) values are used to interpret the model, showing which features most influenced the predictions, thereby making the system transparent and actionable for business use.

These insights empower telecom companies to implement proactive measures aimed at customer retention. The model’s interpretability ensures that decision-makers can understand and trust the reasoning behind predictions. Overall, this end-to-end pipeline highlights how machine learning can be effectively applied to anticipate business challenges and support data-driven strategies.

**CONCLUSION:**

The Telecom Customer Churn Prediction project successfully demonstrates the application of machine learning techniques to address a real-world business challenge. By leveraging a combination of demographic, behavioral, and service-related features, the model accurately predicts which customers are likely to churn. Through comprehensive exploratory data analysis and a robust Random Forest classifier, the project identified key patterns—such as higher churn rates among customers with short tenure and high monthly charges. The use of SHAP values for interpretability was particularly valuable, offering insights into the influence of individual features and increasing transparency for business stakeholders.

The model’s strong performance across metrics like accuracy, ROC-AUC, and precision-recall highlights its potential for real-world deployment. Telecom companies can use such a tool to identify high-risk customers and implement targeted retention strategies such as discounts, personalized communication, or contract changes. Overall, this project provides a practical, data-driven solution to reduce churn and improve customer satisfaction.