MACHINE LEARNING-BASED TELECOM CUSTOMER CHURN PREDICTION AND SEGMENTATION

A PREDICTIVE APPROACH TO ENHANCE CUSTOMER LOYALTY AND REDUCE CHURN

INTRODUCTION:

- **What is Customer Churn?**

  Customer churn refers to when customers stop using a company's services. In the telecom industry, churn means losing subscribers who switch to competitors.

- **Why is it important for telecom companies?**

  High churn rates lead to revenue loss and increased marketing costs to acquire new customers. Retaining customers is more cost-effective and critical for sustained growth.

- **Project objective:**

  Build a machine learning model to predict which customers are likely to churn, enabling proactive retention strategies.

DATASET OVERVIEW:

- **Dataset description:**

Source: Telecom customer churn dataset

Records: Approximately [1.] customers

Features: Demographics, account info, services, payment methods, tenure, charges, etc.

- **Target variable:**

‘Customer Status’ column with values such as 'Active', 'Churned', 'Exited', and 'Inactive'. Mapped to binary labels: 0 for active, 1 for churned.

DATA PREPROCESSING:

- **Mapping churn labels:** Converted categorical statuses into binary format for classification.
- **Handling missing values:** Converted 'Total Charges' to numeric and filled missing values with median.
- **Encoding categorical variables:** Used Label Encoding for binary categories and One-Hot Encoding for multi-class categorical features.
- Removed 'Customer ID' as it does not contribute to prediction. **Dropping irrelevant columns:**

EXPLORATORY DATA ANALYSIS (EDA) HIGHLIGHTS:

- Visualized distribution of churned vs active customers.
- Identified key factors influencing churn such as contract type, payment
- method, tenure, and monthly charges.
- Used seaborn plots and heatmaps to understand feature correlations.

MODEL SELECTION AND TRAINING:

- **Model used:** Random Forest Classifier
- **Train-test split:** 80% training, 20% testing with stratification to maintain
- class distribution
- **Why Random Forest?**

  Robust to overfitting, handles non-linear relationships well, and performs
- well with mixed data types.

MODEL EVALUATION:

- **Accuracy:** Achieved [1.] on test data
- **Classification report:** Precision, recall, and F1-score showed balanced performance
- **Confusion matrix:** Heatmap visualization of true vs predicted labels confirmed model effectiveness

CUSTOMER SEGMENTATION:

- Created three segments based on predicted churn and tenure:
- **At Risk:** Predicted to churn
- **Loyal:** Active customers with tenure > 30 months
- **Dormant:** Active customers with shorter tenure
- Pie chart showed distribution of these segments, helping target retention efforts.

CHALLENGES AND LIMITATIONS:

- **Class imbalance:** Majority of customers were active, causing model bias toward the dominant class.
- **Missing values:** Required cleaning, especially in numeric columns.
- **Limited hyperparameter tuning:** Used default Random Forest parameters; further tuning may improve performance.

FUTURE WORK:

- Implement oversampling techniques like SMOTE or class weights to address imbalance.
- Experiment with other ML models like XGBoost, Logistic Regression, or Neural Networks.
- Use model explainability tools like SHAP or ELI5 to interpret predictions.
- Develop a real-time prediction API for deployment.
- Improve feature engineering with SQL queries to derive new insights.

CONCLUSION:

- Successfully built a predictive model to identify customers likely to churn.
- Created meaningful customer segments to guide retention strategies.
- Business impact: Enable telecom companies to proactively reduce churn and increase customer loyalty.
- Open to questions.