Project Title: Customer Churn Analysis in the Telecommunications Industry

---

Module Code and Title:
7COM1039-0206-2024 - Advanced Computer Science Masters Project

Your Name and Student ID:
Kinjal Atulkumar Raval - 23037922

Aim of the Project

This project aims to analyze customer churn in the telecommunications industry by identifying key factors that influence customer retention and attrition. By leveraging **machine learning techniques and statistical analysis**, this study will provide actionable insights to help businesses improve customer engagement, reduce churn, and enhance service offerings.

Research Question/Hypothesis

- What are the key **demographic, contractual, and service-related** factors that influence customer churn in a telecom company?
- Can **machine learning models** predict customer churn with high accuracy?
- Can **data-driven insights** help in designing effective customer retention strategies?

Objectives

1. **Data Collection & Preprocessing**: Clean and preprocess the Kaggle **Telco Customer Churn Dataset** by handling missing values and encoding categorical variables.
2. **Exploratory Data Analysis (EDA)**: Identify patterns in customer behavior and key factors contributing to churn.
3. **Feature Engineering**: Create meaningful variables that enhance predictive performance.
4. **Model Development**:
    - Compare different machine learning models such as **Logistic Regression, Decision Trees, Random Forest, and XGBoost**.
    - Evaluate performance using **accuracy, precision, recall, and F1-score**.
5. **Interpretation & Business Insights**:

- Use **SHAP (SHapley Additive exPlanations)** values to understand feature importance.
- Develop strategies to **minimize churn** and improve customer retention.

Short Description of Project Idea

Customer churn is a major concern in the telecommunications industry, leading to revenue loss and increased acquisition costs. This project aims to analyze customer behavior patterns using **statistical and machine learning models** to predict and reduce churn. By leveraging insights from data, telecom companies can develop **targeted marketing strategies, improve customer satisfaction, and reduce churn rates**.

Research Methodology

1. **Dataset Selection**:
   - **Telco Customer Churn Dataset** from Kaggle (Dataset Link).
   - Contains customer details, service subscriptions, contract types, payment methods, and churn labels.
2. **Data Preprocessing**:
   - Handle missing values and inconsistencies.
   - Encode categorical variables (e.g., contract type, payment method).
   - Normalize numerical variables (e.g., tenure, monthly charges).
3. **Exploratory Data Analysis (EDA)**:
   - Identify **churn trends** across different customer segments.
   - Visualize relationships using **histograms, boxplots, and correlation heatmaps**.
4. **Feature Engineering & Selection**:
   - Select features with high predictive value using **correlation analysis and feature importance techniques**.
5. **Model Development**:
   - Train and compare **multiple machine learning models**.
   - Tune hyperparameters for optimal performance.
   - Use **cross-validation** to ensure robustness.
6. **Evaluation & Interpretation**:
   - Assess models using **confusion matrix, ROC curve, and precision-recall analysis**.
   - Interpret results using **SHAP values** to understand decision-making.
7. **Business Recommendations**:
   - Provide **actionable insights** for customer retention strategies.
   - Suggest personalized offers, contract changes, and proactive customer support.

Citations

1. Fader, P. S., & Hardie, B. G. (2013). **How to Project Customer Retention**. Journal of Marketing Research, 50(2), 263-280.
2. Verbeke, W., Martens, D., & Baesens, B. (2012). **Social Network Analysis for Churn Prediction**. IEEE Transactions on Knowledge and Data Engineering, 25(3), 431-444.
3. Kaggle. (2024). **Telco Customer Churn Dataset**. Retrieved from Kaggle.

Considerations & Feasibility

Data Availability: The dataset is publicly available on Kaggle, ensuring easy access.
Realistic Timeline: The project scope is achievable within the academic timeframe.
Business Impact: The insights can help telecom companies **enhance customer retention strategies**.
Scalability: The methodology can be extended to other industries facing customer churn challenges.

Conclusion

This project will provide a structured approach to **analyzing and predicting customer churn** in the telecom industry using **data science and machine learning techniques**. The findings will be valuable in **improving customer retention, reducing churn rates, and enhancing service offerings**. By integrating predictive analytics with business strategy, telecom providers can make **data-driven decisions to optimize customer experience and maximize profitability**.
