**1. Introduction**

The purpose of this analysis is to understand customer churn in a telecom company by exploring different machine learning techniques. The dataset includes various customer attributes such as call usage, charges, customer service interactions, and additional service plans. The goal is to identify patterns that contribute to customer churn and determine the most effective analytical approach.

**2. Dataset Overview**

The dataset contains multiple features related to customer behavior, including:

* **Call Usage:** Day, evening, night, and international minutes.
* **Charges:** Corresponding charges for the usage.
* **Customer Service Calls:** Number of times a customer contacted support.
* **Additional Services:** Subscription to voicemail and international plans.
* **Churn Status:** Whether a customer left the service (target variable).

**3. Methods Used**

During the semester, various machine learning techniques were applied, including:

**3.1 Supervised Learning (Classification)**

* **Logistic Regression:** Used to predict churn based on probability but may not handle complex relationships effectively.
* **Decision Trees:** Effective for capturing non-linear patterns but prone to overfitting.
* **Random Forest:** An ensemble method reducing overfitting by averaging multiple decision trees, making it more effective.

**3.2 Unsupervised Learning (Clustering)**

* **K-Means Clustering:** Used to segment customers based on call usage and service interactions, helping identify groups of similar customers.
* **Evaluation using Elbow Method & Silhouette Score:** Determined the optimal number of clusters.

**3.3 Dimensionality Reduction (PCA)**

* Applied to visualize high-dimensional data and improve clustering performance.

**4. Results & Interpretation**

* **Clustering Outcome:** K-Means provided insights into customer segmentation but did not directly predict churn.
* **Classification Performance:** Random Forest outperformed other models by effectively predicting churn and identifying key influencing factors.
* **Key Findings:** Customers with higher service call interactions and international plan subscriptions showed higher churn tendencies.

**5. Conclusion & Recommendation**

* **Most Effective Model:** **Random Forest**, as it provided the highest prediction accuracy and feature importance insights.
* **Business Insights:** Targeted retention strategies should focus on customers with frequent service interactions and additional service plans.
* **Future Work:** Further improvements can be made using advanced ensemble methods and deep learning approaches.

This report highlights the importance of supervised learning in churn prediction while acknowledging the role of clustering for segmentation. The findings can assist telecom companies in designing better customer retention strategies.