

Final Project Report:

Telecom Churn Prediction Introduction Customer churn is a critical issue in the telecom industry, where retaining customers is often more cost-effective than acquiring new ones. In this project, I developed machine learning models to predict customer churn based on various features of the customers' profiles and behaviors. The models built included Logistic Regression, Random Forest, and Support Vector Classifier (SVC).

Problem Description: The objective of this project was to develop a predictive model that could identify customers who are likely to churn. The data provided included customer demographics, account information, and usage patterns. The problem is binary classification, where the target variable is whether a customer churned (1) or not (0).

Approach: The approach involved the following steps:

1. **Data Preprocessing:** The data was cleaned and preprocessed to handle missing values, categorical variables were encoded, and the dataset was normalized where necessary.
2. **Model Development:** I developed three different models to predict churn: - Logistic Regression: A baseline linear model which achieved an accuracy of 79%. - Random Forest: An ensemble model that combines the predictions of multiple decision trees, achieving the highest accuracy of 80%. - Support Vector Classifier (SVC): A non-linear model using a kernel trick, which also achieved an accuracy of 79%.
3. **Model Evaluation:** The models were evaluated based on accuracy. Random Forest slightly outperformed the other models with a balance of accuracy and interpretability.

Findings: The Random Forest model emerged as the best performing model with an accuracy of 80%. This model provides insights into the most important features contributing to churn, such as customer tenure, monthly charges, and contract type. The logistic regression and SVC models also performed well, both achieving an accuracy of 79%, suggesting that the problem is well-suited to both linear and non-linear models.

Recommendations: Based on our findings, we recommend the following:

1. **Targeted Retention Strategies:** Use the Random Forest model to identify customers at high risk of churning and offer them tailored retention strategies, such as personalized discounts or contract adjustments.
2. **Feature Analysis:** Focus on improving customer experience in areas highlighted as important by the model, particularly for customers with high monthly charges and those with short tenures.
3. Additionally, gathering more data on customer interactions and behaviors could enhance model performance.

Further Research: Future work, will be to exploring additional models or techniques, such as deep learning or advanced ensemble methods, to further improve accuracy

Conclusion: This project successfully demonstrated the application of machine learning models in predicting telecom customer churn. The Random Forest model proved to be the most effective, and the insights gained can be directly applied to reduce churn rates, thereby increasing customer retention and profitability.