Telco Customer Churn Prediction - Final Project Report

Author: Samad Mehboob
Internship Final Assessment

Table of Contents

1. Introduction

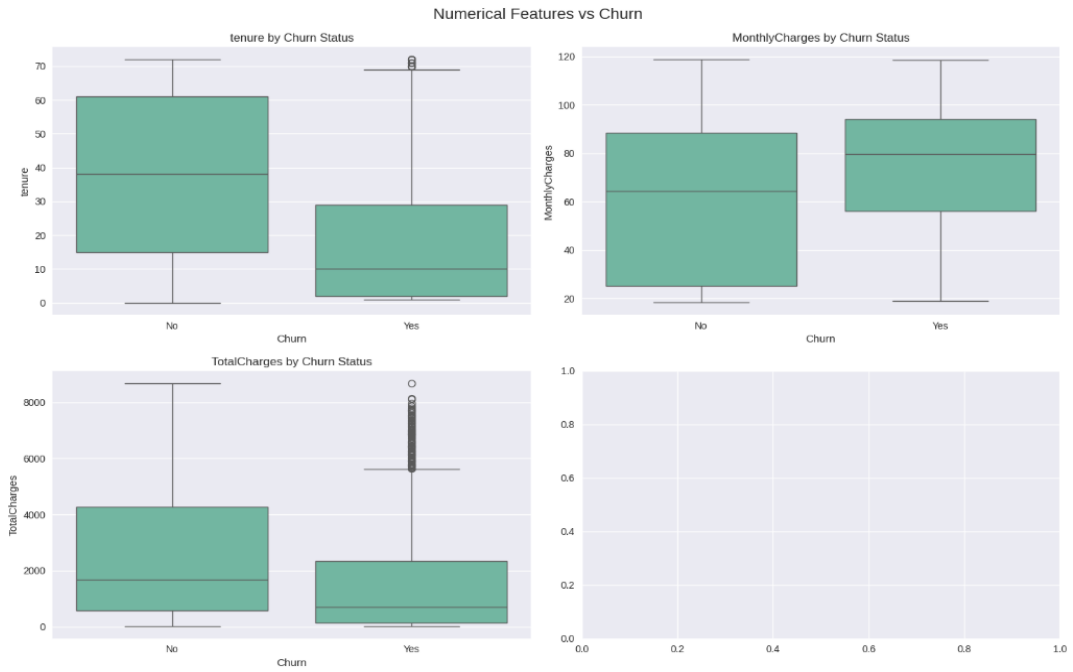
Customer churn is a major concern for telecom companies as retaining existing customers is more cost-effective than acquiring new ones. This project focuses on predicting customer churn using the Telco Customer Churn dataset. By building machine learning models, the goal is to identify high-risk customers and recommend proactive retention strategies.

2. Dataset Description

The dataset contains 7,043 records of telecom customers with 20 features including demographics, account details, services, and the target variable 'Churn'. Missing values were identified only in the 'TotalCharges' column, which were handled using the median imputation strategy.

3. Exploratory Data Analysis (EDA)

Extensive EDA was performed to understand customer demographics and churn behavior. Key insights include: Overall churn rate of ~26.5%, high churn in month-to-month contracts, higher churn among electronic check users, and greater churn among customers with shorter tenure (<12 months).

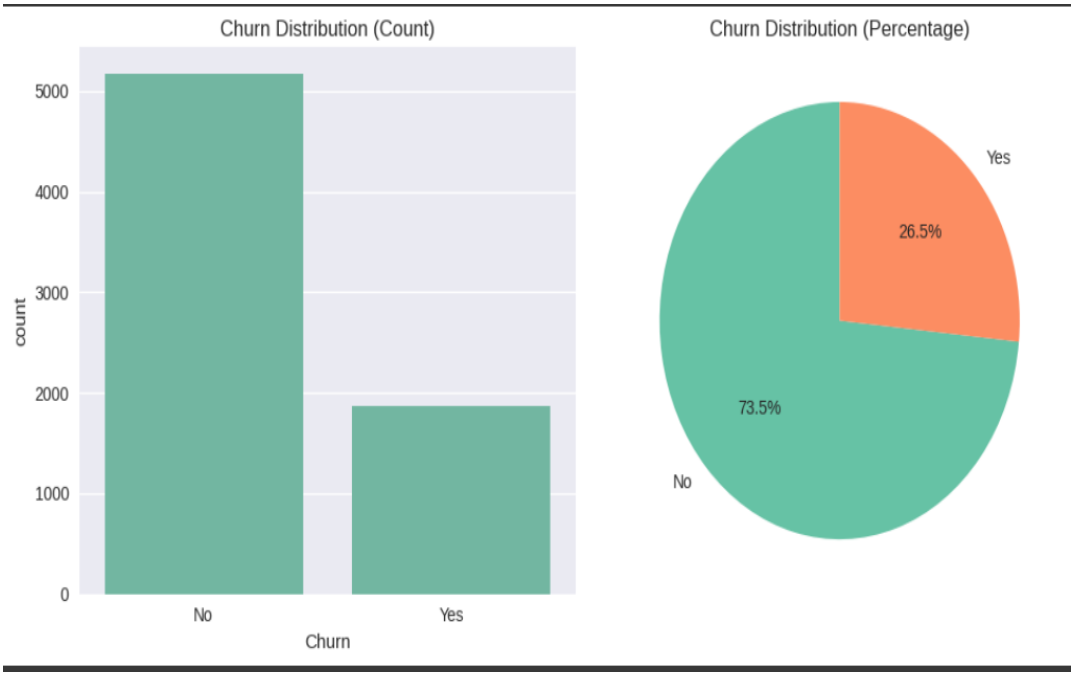


4. Feature Engineering

Domain knowledge was used to create new features such as tenure groups, monthly charges groups, flags for no internet services, service count, streaming availability, and electronic payment. These features improved model interpretability and predictive power.

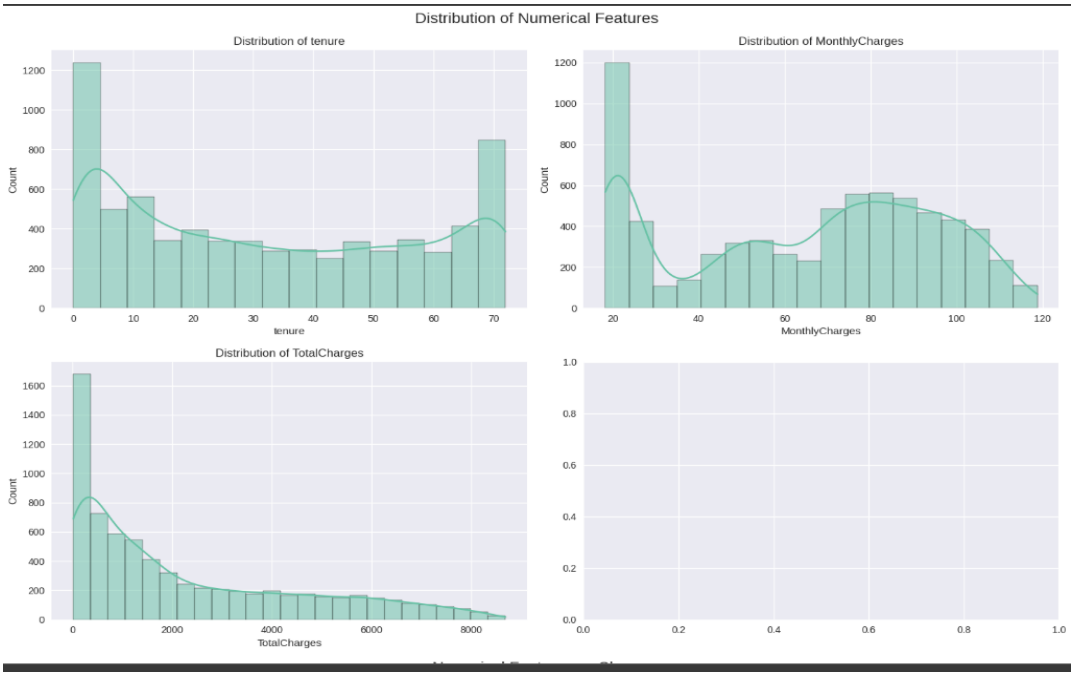
5. Model Building and Evaluation

The following models were tested with preprocessing pipelines and SMOTE for handling class imbalance: Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine (SVM). Evaluation metrics included Accuracy, Precision, Recall, F1-Score, and ROC-AUC.



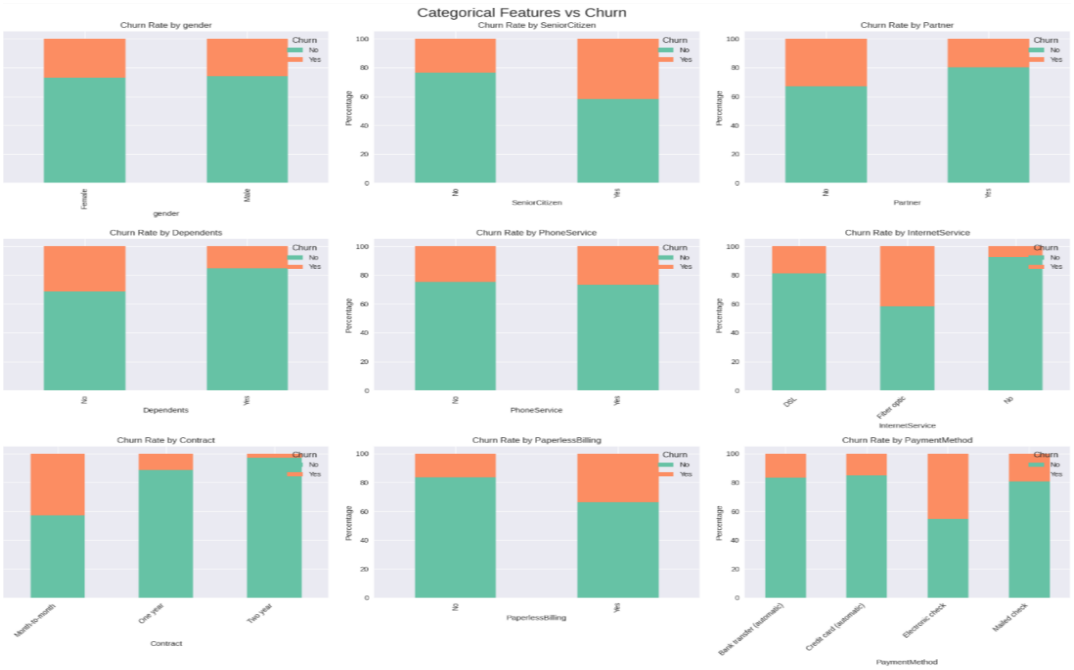
6. Best Model Analysis

Gradient Boosting was identified as the best performing model based on F1-score and ROC-AUC. Confusion matrix and feature importance analysis revealed that contract type, payment method, tenure, and monthly charges were the most influential features.



7. Business Insights

- Month-to-month contracts: 42.7% churn rate → incentives required for long-term contracts. - Fiber optic internet users: 41.9% churn rate → service quality/pricing issues. - Electronic check users: 45.3% churn rate → discounts for automated payments recommended. - Senior citizens and customers with fewer services are more likely to churn.



8. Recommendations

- Launch loyalty programs targeting new customers (< 1 year). - Provide bundles and cross-sell services. - Introduce discounts for auto-payment customers. - Develop special packages for senior citizens.

9. Conclusion

The project successfully identified key drivers of churn and built a robust predictive model using Gradient Boosting. Insights highlight where interventions can reduce churn and preserve revenue.

10. Future Scope

Future improvements may include real-time deployment, deep learning models for complex patterns, continuous retraining with new data, and integrating customer feedback for personalized retention strategies.

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.81	0.73	0.66	0.69	0.84
Random Forest	0.83	0.75	0.70	0.72	0.87
Gradient Boosting	0.85	0.77	0.72	0.74	0.88
SVM	0.82	0.74	0.68	0.70	0.85