

Abstract—In response to increasing competition and rising demands for digital services, telecommunications companies face heightened challenges in managing customer retention and reducing churn. Customer churn prediction has become crucial, as retaining existing customers is more cost-effective than acquiring new ones. This study investigates the effectiveness of machine learning algorithms, with a specific focus on the Random Forest model, in predicting customer churn within the telecommunications sector. Using the Telco Customer Churn dataset from Kaggle, the study performs extensive data preprocessing, including handling missing values, encoding categorical variables, scaling numeric features, and balancing data through the Synthetic Minority Oversampling Technique (SMOTE) to enhance model accuracy and interpretability. A comparative analysis of multiple algorithms, including Random Forest, Logistic Regression, K-Nearest Neighbors, Support Vector Classifier (SVC), and Gradient Boosting, is conducted based on ROC-AUC scores and additional classification metrics to evaluate predictive performance. Results indicate that Random Forest achieved the highest accuracy, with a ROC-AUC score of 0.85, underscoring its robustness and suitability for handling complex, high-dimensional datasets often encountered in churn prediction. Feature importance analysis highlights critical predictors, such as Total Charges, customer tenure, and Monthly Charges, providing actionable insights for developing targeted customer retention strategies. Moreover, incorporating cost-benefit analysis demonstrates the financial advantages of reducing churn rates, emphasizing the practical implications for telecommunication companies in designing effective retention strategies. This study highlights the efficacy of the Random Forest algorithm in telecommunications churn prediction, offering model-driven insights that support data-informed practices to enhance customer loyalty and optimize retention strategies.

Index Terms—Customer churn prediction, Random Forest, Telco Customer Churn, feature importance, SMOTE, machine learning, economic growth. .