11/7/2024

Telecom Customer Churn Prediction

yodahe teshome
HTTPS://GITHUB.COM/JODAHE1/ZINDI-KAGGLE-COMPETITIONS.GIT

This report analyzes the telecom-churn-prediction.ipynb" which I select from Kaggle ,and I try to make the focusing of the report on the algorithms and models used for predicting customer churn in the telecom industry.

1. Data Exploration and Preprocessing

The code begins by loading the training and testing datasets, displaying their shape and information.

It identifies the categorical and numerical features present in the data.

The code explores the distribution of categorical features using count plots, highlighting key observations like the high percentage of customers without international plans or voicemail plans.

Numerical features are analyzed using density plots to understand their distribution.

2. Algorithm Selection and Model Building

The code utilizes various machine learning algorithms for churn prediction. The algorithms used are:

Random Forest Classifier: A powerful ensemble method that combines multiple decision trees for improved accuracy.

Support Vector Machine (SVM): A robust algorithm that finds an optimal hyperplane to separate data points into different classes.

XGBoost Classifier: A gradient boosting algorithm known for its high performance and ability to handle complex datasets.

The code demonstrates the preprocessing steps required for each algorithm. These steps include:

Encoding Categorical Features: Using techniques like One-Hot Encoding to convert categorical variables into numerical representations.

Feature Scaling: Normalizing numerical features using MinMaxScaler to ensure all features have a similar scale.

Handling Class Imbalance: Applying SMOTE (Synthetic Minority Oversampling Technique) to address the imbalanced class distribution in the churn dataset.

3. Model Evaluation and Selection

The code evaluates the performance of each model using various metrics:

Accuracy: Measures the overall percentage of correctly classified instances.

Confusion Matrix: Provides a detailed breakdown of true positives, true negatives, false positives, and false negatives.

Classification Report: Summarizes precision, recall, F1-score, and support for each class.

Cohen's Kappa: Measures the agreement between the model's predictions and the actual churn labels.

The code compares the performance of different models based on the evaluation metrics, ultimately selecting the model with the best performance for churn prediction.

4. Insights and Recommendations

The analysis of the chosen model provides insights into the factors influencing customer churn.

The code uses these insights to recommend actions that telecom operators can take to reduce churn, such as:

Targeted Promotions: Offering special deals to customers identified as high-risk for churn.

Improved Customer Service: Addressing customer complaints and providing better support to enhance satisfaction.

Personalized Communication: Tailoring communication to individual customer needs and preferences.

Overall, the code provides a comprehensive approach to customer churn prediction in the telecom industry. It demonstrates the use of various machine learning algorithms, preprocessing techniques, and evaluation metrics to build an effective churn prediction model.