Customer Churn Prediction Report

Github : https://github.com/NovasusTV/Dorsret

Introduction

This report focuses on predicting customer churn, i.e., whether a customer will cancel their subscription to a service, based on usage patterns, demographics, and other features. We utilized the Telco Customer Churn dataset, which contains information on customers of a telecommunications company and their associated churn status. In this project, we applied the K-Nearest Neighbors (KNN) algorithm to predict customer churn.

K-Nearest Neighbors (KNN) Algorithm

The K-Nearest Neighbors (KNN) algorithm is a non-parametric classification algorithm that determines the class of an unseen example based on its proximity to the nearest neighbors in the training set. In our case, KNN identifies customers with similar characteristics to predict whether a customer is likely to churn or not.

Data Exploration and Preprocessing

We started by exploring the Telco Customer Churn dataset, examining its structure and gaining insights into the data. We performed data preprocessing steps such as dropping unnecessary columns ('CustomerID', 'Count', 'Lat Long') and converting categorical variables to numerical using one-hot encoding. We also handled missing values by dropping rows with missing data.

KNN Model Training and Evaluation

We split the preprocessed data into training and testing sets using the train_test_split function. Before training the KNN model, we standardized the feature variables using the StandardScaler to ensure they have a similar scale. We then instantiated a KNeighborsClassifier and fit it to the training data.

To evaluate the model's performance, we made predictions on the test set and calculated the accuracy score. The accuracy score represents the proportion of correctly classified instances. The higher the accuracy score, the better the model's performance.

Model Performance Improvement

To improve the model's performance, we conducted hyperparameter tuning using GridSearchCV. We defined a parameter grid containing different values for 'C' (regularization parameter), 'gamma' (kernel coefficient), and 'kernel' (kernel type). Grid search exhaustively searched through the parameter combinations and evaluated the model's performance using cross-validation. The best combination of hyperparameters was identified based on the highest accuracy score.

We created a final model using the best hyperparameters obtained from grid search and trained it on the entire training set. The accuracy of the final model was evaluated on the test set, providing a measure of its performance on unseen data.

Model Training and Evaluation Visualization

To visualize the training process, we can plot the accuracy of the KNN model against the number of neighbors (k) considered. This graph shows how the model's accuracy changes as we vary the number of neighbors. It helps in understanding the impact of the choice of k on the model's performance.

Additionally, we can create a confusion matrix to visualize the performance of the final model. The confusion matrix provides insights into the number of true positives, true negatives, false positives, and false negatives. This visualization helps in assessing the model's ability to correctly predict churn and non-churn instances.

Conclusion

In conclusion, the application of the K-Nearest Neighbors (KNN) algorithm to predict customer churn based on usage patterns, demographics, and other features has proven successful. By accurately identifying customers at risk of canceling their subscriptions, businesses can take proactive measures to improve customer retention and satisfaction.

Through data exploration, preprocessing, and model evaluation, we obtained valuable insights into customer churn behavior. The hyperparameter tuning process further enhanced the model's accuracy and performance. Visualizations of the training process and model evaluation provided intuitive representations of the model's capabilities.

Overall, this project demonstrates the effectiveness of using machine learning techniques, specifically KNN, to address customer churn prediction. The findings contribute to informed decision-making, enabling businesses to develop targeted strategies for customer retention and business growth in the telecommunications industry.