Predicting Customer Churn

By: Veom Nemade

Predicting customer churn enables businesses to implement proactive measures to retain customers and enhance customer satisfaction.

Objective

- Develop and evaluate predictive models to accurately identify customers at risk of churn.

- Create a comparative analysis to understand each algorithm's strengths, limitations, and suitability.

- Provide valuable insights for businesses seeking to optimize customer retention strategies and improve overall performance.

Summary

This project uses the Telco Customer Churn dataset to predict churn and helps identify the most suitable algorithm through comparative analysis.

- Dataset includes customer demographics, services, and churn status
- Analysis includes data exploration, preprocessing, model training, and evaluation

Machine Learning Algorithms

- Logistic Regression
- Decision Trees
- Gradient Boosting
- Random Forests
- Neural Networks

Metrics

- Accuracy: This metric represents the ratio of correctly predicted instances (both true positives and true negatives) to the total number of instances.

- Precision: It measures how many predicted positive instances are positive.

- Recall: It measures how many of the actual positive instances are correctly predicted by the model.

- F1-score: It balances precision and recall, considering false positives and false negatives.

- ROC-AUC: It measures the area under the ROC curve, which plots the true positive rate against the false positive rate at various threshold settings.

Insights

- Logistic Regression was the most reliable choice, balancing accuracy and interpretability. It was the best at predicting customer churn, offering high accuracy, precision, and F1-score. It is straightforward to understand, making it useful for businesses.

- Decision Trees, Random Forests, and Neural Networks did well in specific aspects like recall and ROC-AUC.

- Each algorithm has its strengths and weaknesses, highlighting the importance of considering trade-offs in model selection.

Insights into Customer Churn

The implemented algorithms effectively addressed the task of predicting customer churn. The rigorous evaluation of multiple algorithms provided a comprehensive understanding of their performance and implications for businesses.

What Works

- The Telco Customer Churn dataset offered valuable insights into customer demographics, service subscriptions, and churn behavior.

- Using multiple machine learning algorithms enabled a thorough analysis of churn prediction performance.

- Comparing algorithms pinpointed the most effective churn prediction method.

- The project revealed factors impacting customer churn and the success of predictive modeling methods.

What Does Not Work

- Some models might overfit due to dataset complexity or insufficient regularization techniques

- The dataset might lack certain predictive features, limiting the models' effectiveness in churn prediction.

- Imbalanced churned and non-churned customer distributions could have biased predictions in some models.

Lessons Learned

- Effective preprocessing, like handling missing values, encoding categorical variables, and scaling features, is vital for model performance.
- Not all algorithms excel in every task. Experimenting with various ones and assessing their performance is crucial for selecting the best approach.
- Understanding various evaluation metrics is crucial for accurately interpreting model performance.
- Simple models like logistic regression may outperform complex ones like neural networks, especially when interpretability is crucial.
- The project highlighted the importance of continuous learning and staying updated with the latest developments in machine learning techniques and methodologies.

References

Here is the Demo link which has a well-rounded view of all the implementations which you can test yourself given you download the csv file associated with the dataset and then do a read.csv on that file:

Dataset Link: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Demo Link:
https://colab.research.google.com/drive/1Gydkt6TmZpgmebmoilieYNDGA1vv6ZUn?usp=sharing