SYRIATEL CUSTOMER CHURN PREDITCTOR

BEGINNINING OVERVIEW

Churn is a measure of how many customers stop using a product or service due to some certain reasons making them not to use them. It is calculated by dividing the number of customers who have stopped using the product or service by the total number of customers who have ever used the product or service. The churn rate is usually expressed as a percentage. A high churn rate can indicate that there is a problem with the product or service. It can also be a sign that the company is not doing a good job of retaining customers.

In this project, we will try to determine the churn rate for future customers and find a way of minimizing it. This project basically, focuses on predicting customer churn for SyriaTel, a telecommunications company. The main goal is to develop a classifier that can identify customers who are likely to stop doing business with SyriaTel in the near future. By anticipating churn, SyriaTel can take proactive measures to retain customers, reduce financial losses, and improve customer satisfaction.

BUSINESS AND DATA UNDERSTANDING

- The primary audience for this project is SyriaTel's telecom business itself. The stakeholders include executives, marketing teams, and customer relationship management teams who are interested in reducing churn and optimizing customer retention strategies. The insights gained from the analysis will help the stakeholders understand the patterns and factors influencing churn and enable them to make data-driven decisions to mitigate customer churn.
- The dataset used for this project contains historical customer data from SyriaTel, including various features such as customer demographics, usage patterns, account information, and customer churn status. The dataset provides valuable insights into customer behavior and characteristics that can be leveraged to predict churn. By analyzing this dataset, we aim to identify predictable patterns and develop a robust churn prediction model. The dataset of our project has **3,333** number of rows and **21** number of columns.

MIDDLE MODELLING

- Classification modeling is a technique that categorizes customers as either churners or non-churners based on their characteristics and behavior. Therefore, in the modeling phase, we employed various classification algorithms to predict customer churn.
- By applying the classification models, we can analyze customer data and models that can learn from the patterns in order to predict future churn of which it will enable SyriaTel to identify customers who are most likely to churn, allowing them to take proactive measures to retain those customers.

We explored different classification models to predict customer churn. These models include:

- Logistic Regression: This model calculates the probability of a customer churning based on various customer attributes.
- Decision Tree Classifier: A decision tree model that splits the data based on specific attributes to classify customers as churners or non-churners.
- K-Nearest Neighbors (KNN): This model classifies customers based on similarity to their neighboring customers in terms of their characteristics.
- Random Forest Classifier: A powerful ensemble model that combines multiple decision trees to improve the accuracy of predictions.
- By leveraging these models, we can gain insights into the factors influencing churn and identify key features that contribute to customer attrition.

EVALUATION

In the evaluation stage, we assess the performance of our churn prediction models to determine their effectiveness in accurately identifying customers who are likely to churn.

To evaluate the models, we utilized various metrics that provide insights into their performance. These metrics include:

- Accuracy: Measures the overall correctness of the model's predictions.
- Precision: Indicates the proportion of correctly predicted churners among all customers predicted as churners.
- Recall: Represents the proportion of correctly predicted churners among all actual churners.
- F1-score: Harmonic mean of precision and recall, providing a balanced measure of the model's accuracy.

- Based on the models, we came up with the following results:

1. **Baseline Model (Logistic Regression):**

- The logistic regression model achieved an accuracy score of 73.3% on the test data.
- The cross-validation scores ranged from 71.3% to 75.1%, indicating some variance in model performance across different folds.
- The classification report shows that the model has higher precision and recall for predicting churn class 0 compared to class 1.
- The F1-score for predicting churn class 1 is relatively low, indicating that the model struggles to correctly identify churn instances.

2. **Decision Tree Classifier**

- The decision tree classifier achieved a higher accuracy score of 93.0% on the test data. - The model's performance on the training data was slightly lower with an accuracy score of 83.1%.
- The cross-validation scores ranged from 77.7% to 82.0%, showing relatively consistent performance across different folds.
- The classification report reveals that the model has good precision, recall, and F1-score for both churn classes, indicating balanced performance in predicting churn instances.

3. K-Nearest Neighbors (KNN) Model:

- The KNN model achieved an accuracy score of 79.2% on the test data.
- The model's performance on the training data was higher with an accuracy score of 92.4%. - The classification report shows that the model has higher precision and recall for predicting churn class 0 compared to class 1.
- The F1-score for predicting churn class 1 is relatively low, similar to the logistic regression model.

4. Random Forest Classifier:

- The random forest classifier achieved a relatively low accuracy score of 39% on the test data. The cross-validation scores ranged from 92.2% to 95.3%, suggesting consistent performance across different folds.
- The classification report shows that the model has higher precision and recall for predicting churn class 1 compared to class 0.
- The F1-score for predicting churn class 0 is relatively low, indicating poor performance in identifying non-churn instances.

END RECOMMENDATIONS

- Based on the evaluation, we recommend using the random forest model as it consistently delivered the highest accuracy and precision for churn prediction. Its classification report indicates excellent performance across various metrics. By implementing the random forest model, SyriaTel can effectively identify potential churners and take proactive measures to retain customers.
- However, it's important to note that model selection should also consider other factors such as interpretability, scalability, and implementation feasibility. Further analysis and testing may be necessary to ensure the selected model aligns with SyriaTel's specific business requirements and constraints.
- Overall, the random forest model presents a strong choice for predicting customer churn and reducing financial losses for SyriaTel.

NEXT STEPS

- Incorporate Additional Data: To enhance the predictive power of our models, we can consider integrating external data sources. This could include customer feedback, social media sentiment analysis, or economic indicators. By incorporating a broader range of data, we can capture a more comprehensive view of customer behavior and improve the accuracy of our predictions.
- Deploy Real-time Monitoring: Implementing a real-time monitoring system will enable us to continuously track customer churn indicators. By leveraging advanced analytics and machine learning, we can identify churn risks in real-time and take proactive measures to retain valuable customers.
- Collaborate with Stakeholders: To ensure the successful implementation of our churn prediction models, it is crucial to collaborate with various stakeholders within SyriaTel. Engaging executives, marketing teams, customer service representatives, and data scientists will foster a data-driven culture and ensure alignment with business objectives.

THANK YOU

- Thank you for your time and attention. We are to explore this project with the aim of helping SyriaTel to reduce and optimize customer churn and maintaining its long-term success over a long period of time