kenjuma / **Phase-3-Project**   Public

☆ **0** stars   ⑂ **0** forks

| ☆ Star | ⊙ Watch |

<> **Code**   ⊙ Issues   ⑂ Pull requests   ▶ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈 Insights   ⚙ Se

⑂ **main** ▾   ···

🟪 **kenjuma**   ...   now   🕑

View code

README.md   ✎

Phase-3-Project

Syriatel Customer Churn Analysis

Introduction:

This project focuses on SyriaTel, a telecommunications company and one of the major players in the Syrian market. Like many companies in the Telco industry, SyriaTel faces the challenge of identifying potentially dissatisfied customers before they switch to another provider. This phenomenon is commonly known as customer churn, and it can significantly impact a company's profitability. To improve their bottom line and retain customers, SyriaTel aims to leverage machine learning techniques to predict customer churn. By predicting churn, SyriaTel can develop targeted strategies and action plans for different customer segments. Objective: The primary objective of this project is to utilize machine learning techniques to accurately predict customer churn for SyriaTel. By understanding which customers are more likely to churn, the company can proactively implement measures to mitigate customer attrition. This predictive capability will enable SyriaTel to take appropriate actions, such as targeted marketing campaigns, personalized offers, and improved customer service, to retain customers and enhance their overall satisfaction. Data Understanding In this study, we are conducting an analysis on behalf of Syriatel, a telecommunications company in Syria. However, the dataset provided for this analysis consists of US telco data. The dataset contains information on 3333 users, collected over a period of 256 days. It is important to note that geolocation and zip code data will be excluded when selecting predictors for the machine learning algorithms. The underlying assumption in this analysis is that all other data signals in the dataset can be safely utilized to draw meaningful inferences for the Syrian telco market.

Business Understanding

Customer churn analysis is crucial for telecommunication companies like SyriaTel due to its potential impact on profitability. By accurately identifying customers who are likely to churn, SyriaTel can allocate resources effectively to prevent customer loss and maximize customer retention. This proactive approach not only helps in reducing revenue loss but also allows the company to focus on providing better services, addressing customer concerns, and enhancing overall customer experience. By leveraging machine learning algorithms, SyriaTel can gain valuable insights into customer behavior and factors influencing churn, enabling them to make data-driven decisions and develop effective retention strategies. By undertaking this analysis, SyriaTel aims to achieve the following benefits: • Improved Customer Retention: By accurately predicting customer churn, SyriaTel can take timely actions to prevent attrition and retain valuable customers. • Increased Profitability: Retaining customers is more cost-effective than acquiring new ones. By reducing churn, SyriaTel can enhance profitability and maximize the lifetime value of each customer. • Enhanced Customer Experience: Understanding the factors that contribute to churn allows SyriaTel to address pain points, improve service quality, and provide personalized offerings to meet customer needs. Through this project, SyriaTel aims to leverage the power of data and machine learning to gain insights into customer churn behavior, develop effective retention strategies, and ultimately enhance customer satisfaction and profitability. Data Cleaning: Fortunately, the dataset for this analysis does not contain any missing values, and there are no significant outliers that require handling. However, it is important to note that there is a clear class imbalance in the target variable, as only 14.49% of the users have churned. Modelling Modeling In the modeling stage, we addressed a binary classification problem with a significant class imbalance. To evaluate the performance of different models, we primarily used the roc_score metric, which is well-suited for imbalanced datasets. After selecting the winning model, we further assessed its performance using additional metrics such as accuracy and f1-score in the final confusion matrix. However, considering the specific context of this project, we will explain in the next section why recall is likely the most important metric to consider after evaluating potential costs. Given the relatively small size of the dataset (3333 samples), we performed a single split into training and testing sets. Throughout the modeling process, we employed stratified k-fold cross-validation with a fold size of 10 to ensure robustness and minimize overfitting. We started with a baseline model using Naive Bayes, followed by implementing a Decision Tree, Random Forest, and finally a logistic regression. For all models, hyperparameter tuning was conducted using GridSearch CV to optimize their performance. Although the Random Forest model achieved the highest roc_auc_score among all models (0.90%), we ultimately selected the Decision Tree as the winning model. The decision was based on its higher interpretability of the features while still maintaining a slightly lower performance (roc_auc_score of 0.88%). Below is a snapshot of the Decision Tree winning model for reference:

Evaluation Through comprehensive research, we conducted an analysis of the potential costs associated with the predictive power of our model. Our findings revealed that the highest potential cost lies in the "False Negatives" predictions, where the model incorrectly labels a customer as a non-churner when they actually would churn. This is significant because acquiring new customers is, on average, five times more expensive than retaining existing ones. Consequently, our model prioritizes this specific class and places greater emphasis on "Recall" or True Positive Rate (TPR), making it the most important metric in this context. Further details on this can be found in the notebook, particularly in the section on threshold selection. Upon evaluating the model's performance on the test set, we achieved favorable results, with an overall accuracy of 87% and a Recall rate of 83%. This means that our model can successfully identify potential churners eight times out of ten, thanks to its high Recall rate. To estimate the potential benefits of implementing this model, we conducted a brief analysis. Based on our calculations, by correctly identifying future churners and taking appropriate actions to retain them, Syriatel could potentially save $24,915 for every 1000 customers.

The graph provides a clear visualization of the features that have the greatest impact on customer churn. Among these features, customer service calls, international plan, and total day charge stand out, as they have approximately three times more weight than any other feature in the dataset. In the final section, we developed custom functions to delve deeper into the magnitude and direction of these three influential features. We discovered that when the number of customer service calls reaches four, the churn percentage significantly increases. This finding suggests that Syriatel should consider flagging this threshold as a potential indicator of provider dissatisfaction. It is recommended to closely monitor customers who exceed this threshold within a 256-day period, as they are more likely to churn.

Findings and Recommendations: Churn Rate Analysis: Syriatel aims to reduce their churn rate, which refers to the number of customers who unsubscribe or discontinue a specific service. The focus is on improving the phone service sector while keeping internet services separate for now. Model Development: The goal is to develop a model that minimizes customer churn without Syriatel being aware of their intention to churn. Four classification models (Naive Bayes, Decision Tree, Random Forest, and Logistic Regression) were evaluated to achieve this goal. Model Evaluation: The chosen metric for evaluating model performance was roc_auc_score, considering the heavy class imbalance in the binary classification problem. The Decision Tree model was selected as the winning model due to its higher interpretability and slightly lower performance compared to Random Forest. Cost Evaluation and Insights: False negatives, which refer to incorrectly labeling someone as a non-churner when they would churn, were identified as having the highest potential cost. Therefore, the model prioritizes recall (True Positive Rate) as the most important metric. This is because acquiring new customers is more expensive than retaining existing ones. Performance Evaluation: The Decision Tree model achieved an overall accuracy of 87% and a recall of 83% on the test set. This means that the model successfully identifies potential churners 8 times out of 10. This performance indicates the model's ability to effectively identify customers at risk of churn. Cost Savings: By implementing this model, Syriatel could potentially save $24,915 for every 1000 customers by correctly identifying future churners and taking proactive actions to retain them. This demonstrates the value of using the model to optimize customer retention strategies. Feature Importance: The Decision Tree model's feature importance analysis revealed that customer service calls, international plan, and total day charge have the highest impact on churn. These features should be closely monitored and used as signals for potential churn. Customer service calls, in particular, showed a significant increase in churn percentage after four calls, indicating a potential dissatisfaction with the provider's service.

Based on these findings, the following recommendations are provided: Syriatel should focus on improving customer service quality and addressing issues related to customer service calls. Promptly addressing customer concerns and providing satisfactory resolutions can help reduce churn. International plan offerings should be carefully evaluated and optimized based on customer needs and preferences. Providing competitive and attractive international plans can contribute to customer retention. Total day charge should be monitored and managed effectively. Offering cost-effective plans and providing transparent billing practices can help mitigate churn. The model's predictions should be integrated into Syriatel's customer retention strategies. By identifying potential churners early on, Syriatel can proactively engage with customers to address their concerns, offer incentives, or tailor personalized offers to convince them to stay. Regular monitoring and evaluation of the model's performance should be conducted to ensure its effectiveness in reducing churn. Continuous improvements and updates to the model can further enhance its predictive capabilities. Syriatel should consider expanding the analysis to include internet services and evaluate the factors specific to that segment. Understanding the unique drivers of churn in the internet services sector can help develop targeted strategies for customer retention. By implementing these recommendations and leveraging the insights gained from the winning model, Syriatel can optimize their customer retention efforts and ultimately reduce churn, leading to improved business performance and customer satisfaction.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package