Kinyanjui287 / **Phase-_3_Project**

[<>] **Code** [⊙] Issues [⑂] Pull requests [▷] Actions [▦] Projects [📖] Wiki [⊘] Security [📈] Insights [⚙] Settings

[H] **Phase-_3_Project** Public

[⑂] **1 Branch**   [⬚] **0 Tags**   [Go to file   t]   Go to file   [+]   Add file ▾ About   [Code]   [...]   [⚙]

[H] Kinyanjui287 Changes Made	3e9e5d5 · 1 minute ago	[⏱] **2 Commits**
📁 Data	Changes Made	1 minute ago
📄 .gitignore	Initial commit	2 days ago
📄 LICENSE	Initial commit	2 days ago
📄 Notebook.pdf	Changes Made	1 minute ago
📄 Presentation.pdf	Changes Made	1 minute ago
📄 README.md	Changes Made	1 minute ago
📄 index.ipynb	Changes Made	1 minute ago

[📖] **README**   [⚖] **MIT license**   [✎] [☰]

Customer Churn Prediction in Telecommunication

[📖] Readme

[⚖] MIT license

[⎘] Activity

[☆] 0 stars

[👁] 1 watching

[⑂] 0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Phase-_3_Project

Customer Churn Prediction in Telecommunication

Project Criteria

This project will follow the CRISP_DM Criteria

1. Business understanding
2. Data Understanding
3. Data preparation
4. Modeling
5. Evaluation 6 Deployment

Business Understanding

Project Overview

- Churn occurs when customers are leaving a company's services in pursuit of better services from other network providers.
- This is caused by dissatisfaction of the company's services or competitors offering better prices.
- Churn causes loss of the revenue to the company and it makes it hard to retain customers.
- Identifying potential churners will help to retain customers and improve customer satisfaction.

Business Problem

The business objective is to identify customers with a high likelihood of churning and develop effective strategies to retain them. This involves analyzing key factors contributing to customer dissatisfaction and churn, such as network quality, customer service issues, or pricing concerns. Additionally, the goal is to segment customers based on their behavior and churn likelihood, enabling tailored marketing and retention strategies that address each group's unique needs and preferences.

Objectives

1. Churn Prediction: Develop machine learning models to predict customers likely to churn by analyzing customer data and features.
2. Model Performance Assessment: Evaluate and compare machine learning models to identify the most accurate prediction model.
3. Increase Revenue: Retaining more customers will lead to higher revenue and an increase in market share.
4. Feature Insights: Analyze individual features to uncover key factors driving customer churn in the telecommunications company.

Data Source

> My project utilizes data obtained from [Kaggle](), it is about customer churn in a telecommunication company.

Data Description

1. State: The U.S. state in which the customer resides, represented by a two-letter abbreviation.

2. Account Length: The duration (in days) that the customer has been with the service provider.

3. Area Code: The telephone area code associated with the customer's phone number.

4. International Plan: Indicates whether the customer has subscribed to an international calling plan ('yes' or 'no').

5. Voice Mail Plan: Indicates whether the customer has a voice mail feature ('yes' or 'no').

6. Number Vmail Messages: The count of voice mail messages the customer has.

7. Total Day Minutes: The total number of minutes the customer has used during the day.

8. Total Day Calls: The total number of calls made by the customer during the day.

9. Total Day Charge: The total charges incurred by the customer for daytime calls.

10. Total Eve Minutes: The total number of minutes the customer has used during the evening.

11. Total Eve Calls: The total number of calls made by the customer during the evening.

12. Total Eve Charge: The total charges incurred by the customer for evening calls.

13. Total Night Minutes: The total number of minutes the customer has used during the night.

14. Total Night Calls: The total number of calls made by the customer during the night.

15. Total Night Charge: The total charges incurred by the customer for nighttime calls.

16. Total Intl Minutes: The total number of minutes the customer has used for international calls.

17. Total Intl Calls: The total number of international calls made by the customer.

18. Total Intl Charge: The total charges incurred by the customer for international calls.

19. Customer Service Calls: The number of calls the customer has made to customer service.

20. Churn: Indicates whether the customer has discontinued the service ('yes' for churned, 'no' for active).

Evaluation

The evaluation metrics that I will focus on are:

- Accuracy
- Precision
- Recall
- F1 Score

Accuracy

Logistic Regression = 86.21%

Decision Tree = 91.75%

Random Forest = 93.7%

The Random Forest model performs the best with 93.7% accuracy, followed by the Decision Tree (91.75%) and Logistic Regression (86.21%).

Precision

Logistic Regression = 41.3%

Decision Tree = 53.38%

Random Forest = 82%

The Random Forest model has the highest precision, indicating it made the most accurate positive predictions. The Decision Tree performed better than Logistic Regression in this aspect.

Recall

Logistic Regression = 75.25%

Decision Tree = 74.26%

Random Forest = 81.19%

The Random Forest model again performs the best in recall, identifying 81.19% of positive cases. The Logistic Regression has the highest recall compared to the Decision Tree.

F1 Score

Logistic Regression = 53.3%

Decision Tree = 60.7%

Random Forest = 81.6%

The Random Forest model again outperforms the other models, showing the best balance between precision and recall. The Decision Tree follows, and Logistic Regression has the lowest F1 score.

summary of the models

Random Forest model consistently outperforms the other two models across all metrics, making it the best choice among the three

Conclusion

Model Performance: I tested 3 models with Random Forest Classifier emerging as the top performer, achieving a remarkable 95% accuracy and well-balanced precision and recall.

Key Features: The analysis showed some influential features: "customer_service_calls", "total_day_minutes","total day charge", "total intl calls" and "total eve charge" highlighting their importance in predicting churn.

In summary, the analysis recommends Random Forest Classifier for predicting customer churn.

Recommedation

Improve Customer Service:

- High customer service calls correlate with churn.
- Reduce customer service calls and improve quality of customer service by offering comprehensive training to customer service representatives.
  Pricing Structure Evaluation:
- Evaluate the pricing structure for day, evening, night, and international charges.
- Adjusting pricing plans or introducing discounted packages would address the concerns related to higher charges, which contribute to customer churning