

Vee2002 / phase_3_project

Q

Type to search

>_

+

▼

○

∞

✉

⋈

<> Code

○ Issues

∞ Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Settings

⋈

phase_3_project

Public

🔑 master

🔑 1 Branch

🏷 0 Tags

🔑

📁

Q

Go to file

t

Go to file

+

🔑 Add file

Code

⋮

⚙

⋈

Vee2002

Update README.md

4d36a17 · now

🕒 9 Commits

⋮

📁	churn_dataset	Project	1 hour ago
📄	Customer_Churn_A...	Project	1 hour ago
📄	README.md	Update README.md	now
📄	notebook.pdf	Notebook pdf	2 minutes ago
📄	presentation.pdf	Presentation pdf	1 hour ago
📄	presentation.pptx	Project	1 hour ago

No description, website, or topics provided.

- 📖

Readme
- 📈

Activity
- ★

0 stars
- 👁

1 watching
- 🔑

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 2

- ⋈

VivianMosomi432
- ⋈

Vee2002

Languages

- Jupyter Notebook 100.0%

📖 README

✎

OVERVIEW

Customer churn analysis stands as a critical aspect in various industries, enabling businesses to identify and predict customer attrition. Syria Tel is facing customer attrition and has decided to delve into churn prediction. By leveraging advanced machine learning models, the company aims to forecast potential churners and facilitate proactive strategies to retain customers, thereby strengthening customer loyalty and business sustainability.

DATA UNDERSTANDING

The data used has 3333 rows and 21 columns which are total day,evening and night charge, voice mail and itnenational plan, customer service calls, account number and others. Our target variable is churn.

MODELLING AND EVALUATION

After data preparation techniques such as data cleaning, encoding, scaling and solving class imbalance, I went ahead to carry out prediction using the following models: LogIstic Regression, Decision Trees, Random Forest and XGBoost Classifier. Focusing on recall and f1-score, the above models had the following results: Recall and F1-score Logistic Regresssion 78% 52% Decision Trees 71% 69% Random Forest 75% 78% XGBoost Classifier 78% 84%

XGBoost had the highest Recall and F1-score

ROC Curve Analysis also showed that XGBoost was the best classifier as it had an AUC of 0.88 followed by Random Forest with an AUC of 0.86

HYPERPARAMETER TUNING

On performing hyperparameter tuning on Decision Trees, Random Forest and XGBoost so as to boost performance, the following were the results: Recall and F1-score Decision Trees 73% 70% Random Forest 77% 80% XGBoost Classifier 78% 85%

XGBoost and Random Forest were the highest classifier with and AUC score of 0.93

CONCLUSION

Using feature importance results from our best classifiers which are Random Forest and XGBoost Classifier, the following were shown to be the most important; Total day charge, total day minutes, number of voice mail messages and area code And from that, it can be seen that customer service related issues are causing high churn rate and Syria Tel should work towards that.

RECOMMENDATIONS

- 1. Improving customer service quality so as to reduce the number of customer service calls syria tel receives that may be the cause of high churn rate
- 2. Reviewing Pricing Strategies to prevent customer attrition
- 3. Looking into the states with high churn rate such as New Jersey, California and Texas and offer better network coverage or offer promotions to customers to decrease customer churn
- 4. Marketing in area codes with high churn rate.