Ngaire44 / **PROJECT**

‹› **Code**   ⊙ Issues   ⅄ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈 Insights   ⚙

⭐ 0 stars   ⅄ 0 forks   ⊙ 0 watching   ⅄ Branches   ⋀ Activity
                                                        🏷 Tags

🌐 Public repository

⅄   ⅄ **1 Branch** 🏷 **0 Tags**   ⅄   🏷 | 🔍 Go to file   t | Go to file   Add file ⊹   **Code**   ⋯

🔲 Ngaire44 final	1da1910 · 1 minute ago 🕐
📄 README.md	final	1 minute ago
📄 bigml.csv	final project	33 minutes ago
📄 clean_customer_churn.csv	final project	33 minutes ago
📄 notebook.ipynb	final project	33 minutes ago
📄 presentation.pdf	presentation	13 minutes ago

📖 **README**   ✏

Customer Churn Prediction Project Overview

This project focuses on predicting customer churn for SyriaTel, a telecommunications company. Customer churn is when clients stop using the company's services, leading to revenue loss. The goal is not only to predict churn but also to understand the key factors driving churn, so that SyriaTel can take proactive steps to retain customers.

Objectives

Predict which customers are most likely to churn.

Identify behavioral and service-related factors influencing churn.

Build machine learning models and select the best-performing one for deployment.

Address class imbalance and prevent overfitting.

Provide actionable recommendations for business stakeholders.

Dataset

The dataset was cleaned, preprocessed, and split into training and testing sets.

Class imbalance was handled using SMOTE.

Features such as call durations, charges, and customer service calls were analyzed.

Models Used

Logistic Regression

Decision Tree

Random Forest

XGBoost (Best Performing Model)

Best Model

XGBoost was selected as the final model because:

It achieved the highest recall score (93.26%), aligning with the business need to correctly identify as many churners as possible.

Balanced precision and recall, ensuring fewer false negatives.

Showed high stability with minimal overfitting (Train Accuracy: 100% vs Test Accuracy: 95%).

Feature Importance

Key drivers of churn include:

voicemail plan, total charge, area code, customer service calls, International plan, Number of voicemails, total international calls

Observations

Class imbalance was successfully handled using SMOTE, improving recall performance.

Overfitting was reduced by hyperparameter tuning and model regularization.

XGBoost outperformed other models across almost all metrics (Accuracy, Recall, AUC).

Recommendations

Focus retention strategies on customers with high service call frequency.

Reassess the International Plan since it strongly influences churn likelihood.

Offer targeted promotions for customers with high day-time usage.

Monitor high-risk customers predicted by the model and intervene early.

Project Artifacts

Jupyter Notebook (notebook.ipynb) - Contains full EDA, preprocessing, modeling, and evaluation.

Final Report (Churn_Prediction_Report.pdf) Summary of results, conclusions, and recommendations.

Author

Prepared by: Gathoni

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● **Jupyter Notebook** 100.0%