



Petmoraaa /
PREDICTING-CHURN-IN-TELECOMMUNICATION-INDUSTRY



<> Code Issues Pull requests Actions Projects Wiki Security Insights Setting



★ 0 stars 🍴 0 forks 👁 1 watching 🌿 Branches 📈 Activity
🏷 Tags

🌐 Public repository

 1 Branch 🏷 0 Tags  



Go to file + Add file ▾ Code ...



Petmoraaa Add files via upload

5d893e6 · 6 minutes ago 🔄



Phase_3_project.ipynb

Created using Colab

16 minutes ago



Presentation powerpoint for ...

Add files via upload

6 minutes ago



README.md

Update README.md

2 days ago

📖 README



PREDICTING-CHURN-IN-TELECOMMUNICATION-INDUSTRY

Customer Churn Prediction

Project Overview Companies lose millions due to customer churn. Understanding why customers leave can help businesses take action to retain them, such as offering promotions or improving service quality. This project aims to build a classification model to predict customer churn using historical data.

Objectives

1. Identify factors that contribute to customer churn.
2. Predict churn using multiple machine learning models.
3. Compare model performance and strategies to reduce churn.

Research Questions 1.What factors influence customer churn the most? 2.How accurately can we predict churn using available data? 3.Which classification model performs best for this problem?

Dataset 1.Source: [Kaggle - Churn in Telecoms Dataset](#) 2.File used: bigml_59c28831336c6604c800002a.csv 3.Key features:

- account_length : Number of days the customer has been with the company.
- total_day_minutes : Total minutes used during the day.
- customer_service_calls : Number of calls made to customer service.
- international_plan : Whether the customer has an international plan (Yes/No).

- `voice_mail_plan` : Whether the customer has a voicemail plan (Yes/No).
- Target Variable: `churn` (1 = churned, 0 = stayed)

Methodology

1. Data Preprocessing

- No missing values.
- Encoded categorical variables.
- Standardized numerical features.
- Removed highly correlated features using VIF.

2. Exploratory Data Analysis (EDA)

- Visualized feature distributions.
- Analyzed correlations between features and churn.
- Boxplots and histograms to understand churn patterns.

3. Modeling & Evaluation

- **Models Used:**
 - Logistic Regression
 - Decision Tree
 - Random Forest
 - K-Nearest Neighbors (KNN)
 - Naïve Bayes
 - Support Vector Machine (SVM)
- **Evaluation Metrics:**
 - Accuracy
 - Precision, Recall, F1-score
 - Confusion Matrix

4. Model Comparison

- Compared performance across all models using accuracy scores.
- Identified the best-performing model.

Results & Insights

1. Customers with higher total day minutes are more likely to churn.
2. More customer service calls correlate with higher churn.
3. The Random Forest model achieved the best accuracy.
4. Churn prediction can be improved with feature engineering and balanced datasets.

Conclusion This project successfully built and compared machine learning models to predict customer churn. The findings provide actionable insights for businesses to reduce churn through better customer engagement strategies.

Tools Used: Python, Pandas, Matplotlib, Seaborn, Scikit-learn, Google Colab

[Create a new release](#)

[Publish your first package](#)

● Jupyter Notebook 100.0%