marykamithi Rename github.pdf to notebook.pdf c0e7890 · now

README.md	Revise README for SyriaTel Custo...	14 minutes ago
bigml_59c28831336c6604c...	the dataset used	last week
notebook.pdf	Rename github.pdf to notebook.pdf	now
presentation.pdf	Rename SyriaTel_Customer_Churn_...	7 minutes ago

README

📊 SyriaTel Customer Churn Prediction

🧠 Project Overview

SyriaTel, a major telecom provider, faces an annual churn rate of ~15%, impacting revenue and customer acquisition costs. This project aims to build predictive models that identify high-risk customers and uncover key drivers of churn using historical usage and service data.

🎯 Objectives

- Predict which customers are likely to churn.
- Identify the most influential features contributing to churn.
- Provide actionable insights to reduce churn and improve retention strategies.

📁 Dataset

- **Source**: SyriaTel internal customer usage records.
- **Size**: 3,333 entries, 21 features.
- **Target Variable**: `churn` (binary: 1 = churned, 0 = retained).

Key Features:

- Account length, service plans (international, voicemail), call durations, charges, support interactions.
- Engineered features: total minutes, total charges, average call duration, service interaction rate.

🧹 Data Preparation

- Cleaned column names and removed duplicates.
- Converted categorical variables to numeric (e.g., yes/no → 1/0).
- Handled class imbalance using SMOTE.

- Scaled numerical features and one-hot encoded categorical ones.

📊 Exploratory Data Analysis

- Univariate and bivariate visualizations (boxplots, histograms).
- Correlation matrix to assess feature relationships.
- Identified strong churn indicators: customer service calls, total charge, international plan.

🛠️ Feature Engineering

- Created composite metrics:
  - `total_minutes`, `total_calls`, `total_charge`
  - `avg_minutes_per_call`, `service_calls_per_length`
  - `tenure_estimate`, `engagement_score`, `service_interaction_rate`

🤖 Modeling & Evaluation

Models Used:

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost (baseline and tuned)

Performance Metrics:

- Precision, Recall, F1-Score, ROC AUC
- Threshold tuning to optimize recall vs precision trade-offs

Highlights:

- **XGBoost** achieved 94% accuracy and strong ROC AUC.
- **Decision Tree with SMOTE** improved recall for churners.
- Feature importance consistently ranked `total_charge`, `customer_service_calls`, and `international_plan` as top predictors.

🔍 Insights

- High service interaction correlates with churn—potential dissatisfaction.
- Customers with international plans and high charges are more likely to leave.
- Imbalanced data requires careful handling to avoid biased predictions.

📈 Next Steps

- Deploy model for real-time churn prediction.
- Integrate with SyriaTel's CRM for proactive retention campaigns.
- Explore ensemble methods and time-series behavior for deeper insights.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package