rajulayvonne-design / **Phase-3-project-customer-churn**

Code Issues Pull requests Agents Actions Projects Wiki Security Insights Settings

Watch 0

0 stars 0 forks 0 watching Branches Activity Tags

Public repository

m... 1 Branch 0 Tags Go to file t Go to file Add file Code

rajulayvonne-design first commit dce4d1e · 5 minutes ago

README.md	first commit	5 minutes ago
Reducing Customer Churn.pptx	first commit	5 minutes ago
Reducing customer churn noteboo...	first commit	5 minutes ago
Reducing customer churn pdf pres...	first commit	5 minutes ago
bigml_59c28831336c6604c800002...	first commit	5 minutes ago
phase3project.ipynb	first commit	5 minutes ago
~$Reducing Customer Churn.pptx	first commit	5 minutes ago

README

Customer Churn Prediction Project

Overview

The primary goal of this project is to build a predictive model that identifies customers at high risk of leaving a service provider (churning). By identifying these individuals accurately, the business can implement proactive retention strategies to save revenue and reduce the high costs associated with acquiring new customers.

Data Understanding

The analysis utilized a dataset containing 3,333 customer records with 21 initial features.

Key Features Analyzed: Usage Patterns: Day, evening, night, and international call minutes and charges.

Customer Interactions: Number of customer service calls made.

Account Information: Account length, area code, and specific service plans (International and Voice Mail).

Target Variable: churn (True/False).

Data Challenges:

Class Imbalance: The dataset is highly imbalanced, with only approximately 14.5% of customers having churned (483 churned vs. 2,850 stayed).

Feature Correlation: High correlations were found between minutes and their respective charges (e.g., total day minutes and total day charge), leading to the removal of redundant charge columns to improve model performance.

Methodology

The project followed a standard data science pipeline:

Data Preprocessing: Handled categorical variables by mapping them to numerical values and dropped non-predictive columns like phone number and state.

Addressing Imbalance: Utilized SMOTE (Synthetic Minority Over-sampling Technique) within an imbalanced-learn pipeline to ensure the models could effectively learn the characteristics of churning customers.

Modeling: Tested two primary algorithms:

Logistic Regression: To establish a baseline probability of churn.

Decision Tree Classifier: To capture non-linear relationships and provide logic-based decision rules.

Cross-Validation: Used StratifiedKFold to ensure consistent performance across different data subsets.

Results

Both the Logistic Regression and the Decision Tree models achieved an AUC (Area Under the Curve) of 0.81.

Discrimination Ability: An AUC of 0.81 indicates "good discrimination ability," meaning the model can correctly distinguish between a churned and a non-churned client 81% of the time.

Stability: The identical performance across different model types suggests the findings are stable and reliable for business use.

Conclusion & Recommendations

The models successfully identify at-risk clients, providing a data-backed foundation for a customer retention strategy.

Business Recommendations:

Monitor Service Calls: Customer service call frequency is a strong indicator of churn; high-frequency callers should be prioritized for outreach.

Usage-Based Incentives: Since high daytime usage correlates with churn, the company should consider offering loyalty rewards or plan upgrades for high-volume users.

Retention Pilot: Deploy the model to flag customers in real-time, allowing the marketing team to test targeted retention offers on identified "at-risk" segments.

- **Jupyter Notebook** 100.0%