PHASE 3 PROJECT

Olive-Mirriam Cleaned up embedded repo and added project files	c69db5c · 1 hour ago	
customer-churn-analysis	Cleaned up embedded repo and ad...	1 hour ago
.gitignore	Cleaned up embedded repo and ad...	1 hour ago
.gitignoreTAB	Cleaned up embedded repo and ad...	1 hour ago
README.md	Cleaned up embedded repo and ad...	1 hour ago
bigml_59c28831336c6604c...	Cleaned up embedded repo and ad...	1 hour ago
customerchurn.ipynb	Cleaned up embedded repo and ad...	1 hour ago

README

Customer Churn Analysis

1. Overview

This project focuses on understanding customer churn—why customers leave a business—and developing a data-driven solution to predict and reduce it. Using customer data, we aim to uncover key factors that contribute to churn and build a model that can help the business proactively retain customers.

Customer churn is a critical metric for subscription-based and service industries. High churn rates translate to lost revenue, increased marketing costs, and reduced lifetime value. Our goal is to empower the business with actionable insights and a predictive model to address churn effectively.

2. Business and Data Understanding

Business Understanding

The primary business question is: **“Which customers are at risk of leaving the service, and what can we do to retain them?”**
Retaining customers is often more cost-effective than acquiring new ones. Therefore, understanding churn behavior is not just analytical—it’s strategic.

Data Understanding

We used the **Telco Customer Churn** dataset, which contains over 7,000 entries of customer demographics, services used, billing information, and churn status. The dataset includes:

- **Customer Demographics**: gender, senior citizen status, partner/dependents
- **Service Details**: phone service, internet service, streaming options
- **Account Details**: contract type, tenure, paperless billing, payment method
- **Financial Metrics**: monthly and total charges
- **Target Variable**: whether the customer churned or not

This dataset is ideal for our goal because it mirrors typical customer behavior in a service-oriented business and provides the variables needed to identify churn predictors.

3. Modeling

We applied a **classification modeling approach** to predict whether a customer is likely to churn. The target variable was binary (Churn: Yes/No). The general steps included:

- **Data preprocessing**: Cleaning and encoding categorical features, handling missing values
- **Feature engineering**: Selecting and transforming relevant features
- **Model selection**: Multiple models were considered, including logistic regression and tree-based algorithms
- **Training and testing**: The dataset was split into training and testing subsets for unbiased performance evaluation

We selected the model that provided the best trade-off between accuracy and interpretability.

4. Evaluation

The final model was evaluated using several performance metrics including:

- **Accuracy**: How often the model correctly predicted churn
- **Precision & Recall**: Important to balance false positives and false negatives
- **Confusion Matrix**: Provided a clear view of misclassifications
- **ROC Curve and AUC**: Assessed model performance across thresholds

The model achieved an accuracy close to **80%**, which is a strong result for a churn prediction problem. More importantly, it provides **actionable churn scores** for each customer, enabling proactive retention strategies.

5. Conclusion

The analysis revealed clear patterns in customer behavior. For example, customers with month-to-month contracts, shorter tenures, and higher monthly charges were significantly more likely to churn. These insights, combined with the predictive model, equip the business to:

- Identify high-risk customers early
- Design personalized retention strategies
- Monitor churn trends over time

This project serves as a bridge between technical insights and business strategy, aligning data science outcomes with operational goals.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%