

JoanGatharia / Phase-3-project

<> 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

t

Go to file

+

Add file

Code

JoanGatharia Add files via upload5e0458c · 19 minutes ago🕒

Churn_tel_data.csv	Phase3Projectmaterials	37 minutes ago
Customer Churn Presentatio...	Phase3Projectmaterials	37 minutes ago
Phase_3_Project (4).ipynb	Add files via upload	19 minutes ago
Phase_3_Project.Notebook p...	Phase3Projectmaterials	37 minutes ago
README.md	README.md	31 minutes ago
phase_3_project.py	Phase3Projectmaterials	37 minutes ago

README

# Phase-3-project

## Customer Churn Prediction Project

### Project Overview

This project focuses on predicting customer churn in the telecom industry. The goal is to **identify customers at risk of leaving** and implement proactive strategies to improve retention.

The dataset contains customer-related attributes such as **account length**, **call usage**, **service plans**, and **customer service interactions**. Various machine learning models were tested, with **Random Forest** and **Decision Tree** classifiers performing best after hyperparameter tuning.

# Business Understanding

---

## Business Problem

Customer churn negatively impacts revenue and increases acquisition costs. By predicting churn, the company can implement **data-driven retention strategies**.

## Project Objectives

1. **Predict churn using classification models** to identify high-risk customers.
2. **Determine key churn drivers** like service plans and customer service interactions.
3. **Implement data-driven retention strategies** based on model insights.
4. **Improve customer service** by analyzing the impact of service interactions on churn.
5. **Enhance loyalty and marketing strategies** through targeted offers.

## Research Questions

1. What are the most significant factors influencing customer churn?
2. Does frequent customer service interaction indicate a higher risk of churn?
3. Do customers with an international plan have a higher churn rate?
4. Can a machine learning model accurately predict churn using available features?

## Data Overview

---

- **Dataset Name:** Churn\_tel\_data.csv
- **Records:** 3,333 rows
- **Features:** 21 (including account length , total call minutes , international plan , etc.)
- **Target Variable:** Churn (Binary: 0 = No, 1 = Yes)

## Data Preprocessing

---

### 1. Data Cleaning

- Converted churn from boolean to integer.
- Encoded categorical variables ( international plan , voice mail plan ).
- Dropped **irrelevant columns** (e.g., phone number , area code ).

### 2. Outlier Detection & Handling

- Used **Z-score** to detect extreme values.
- Applied **log transformation** to normalize skewed features.

### 3. Feature Engineering & Selection

- Created new variables (e.g., total charge derived from minutes ).
- Used correlation analysis & Chi-square tests to remove redundant features.

4. Handling Missing Values

- No missing values detected in the dataset.

## Exploratory Data Analysis (EDA)

### Key Insights

- High churn rates among customers with frequent customer service calls.
- Customers with international plans tend to churn more.
- Night call duration is a significant churn predictor.
- Churn rate varies by state but not by area code.

## Machine Learning Models

Model	Accuracy	Recall	Precision	F1-Score
Logistic Regression	84.7%	5.1%	55.6%	9.3%
Decision Tree (Tuned)	92%	69%	79%	74%
Random Forest (Tuned)	93%	59%	91%	72%
K-Nearest Neighbors	82.9%	2%	14.2%	3.5%

### Best Model: Decision Tree (Tuned)

- Accuracy: 92%
- Recall: 69%
- Precision: 79%
- F1-Score: 74%

### Hyperparameter Tuning

- Used GridSearchCV to optimize:
  - max\_depth
  - min\_samples\_split
  - min\_samples\_leaf
- Applied Class Weighting to handle class imbalance.

## Tableau Visualizations

---

Created an interactive dashboard to visualize: **Churn Rate by State & Area Code** **Customer Service Call Analysis** **Churn Trends Based on Call Minutes & Charges** **Demographic Insights on Churn Behavior**

## Conclusion & Recommendations

---

### Key Findings

Customers with high service calls have increased churn risk. International plan users show higher churn rates. Night call duration is a significant churn predictor.

### Business Recommendations

**Improve Customer Support:** Reduce complaints with better service. **Targeted Retention Offers:** Provide discounts to high-risk customers. **Optimize Service Plans:** Create flexible plans for international callers. **Monitor Key Metrics:** Regular churn analysis for proactive decision-making.

## References

---

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*.
- Towards Data Science. (2023). *Customer Churn Prediction: Best Practices in Machine Learning*.

## Future Improvements

---

Test **deep learning models** like Neural Networks. Deploy as a **real-time churn prediction API**. Improve **class balance techniques** for better recall.

### Releases

No releases published

[Create a new release](#)

### Packages


No packages published  
[Publish your first package](#)


Languages


Jupyter Notebook 98.9% Python 1.1%

Suggested workflows

Based on your tech stack

 **Python package** Configure  
Create and test a Python package on multiple Python versions.

 **Python application** Configure  
Create and test a Python application.

 **Publish Python Package** Configure  
Publish a Python Package to PyPI on release.

[More workflows](#)

[Dismiss suggestions](#)