**Customer Churn Prediction Documentation**

**Overview**

This notebook demonstrates the process of predicting customer churn in a telecommunications dataset using machine learning. The goal is to identify customers who are likely to leave the service.

**Sections**

1. **Introduction**
2. **Data Loading and Preprocessing**
3. **Exploratory Data Analysis (EDA)**
4. **Feature Engineering**
5. **Model Building**
6. **Model Evaluation**
7. **Conclusion**

**1. Introduction**

This section outlines the objective of the notebook: to predict customer churn using various machine learning models and to identify the model that provides the best performance.

**2. Data Loading and Preprocessing**

* **Loading Data**: The dataset is loaded into a Pandas DataFrame.
* **Handling Missing Values**: Missing values are identified and appropriately handled.
* **Encoding Categorical Variables**: Categorical variables are encoded using techniques like One-Hot Encoding or Label Encoding.

python

Copy code

import pandas as pd

# Load dataset

df = pd.read\_csv('customer\_churn\_data.csv')

# Display first few rows

df.head()

**3. Exploratory Data Analysis (EDA)**

This section involves visualizing and summarizing the main characteristics of the dataset to understand its structure and distribution.

* **Statistical Summary**: Generate summary statistics of the dataset.
* **Visualizations**: Plot distributions of various features, correlation matrices, and other relevant visualizations.

python

Copy code

import seaborn as sns

import matplotlib.pyplot as plt

# Summary statistics

df.describe()

# Correlation matrix

plt.figure(figsize=(12, 8))

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

plt.show()

**4. Feature Engineering**

Creating new features or modifying existing ones to improve model performance.

* **Feature Scaling**: Apply scaling techniques to normalize numerical features.
* **Feature Selection**: Select relevant features based on correlation and other criteria.

python

Copy code

from sklearn.preprocessing import StandardScaler

# Scaling numerical features

scaler = StandardScaler()

df\_scaled = scaler.fit\_transform(df)

**5. Model Building**

Building and training machine learning models to predict churn.

* **Model Selection**: Different models are chosen and evaluated.
* **Model Training**: Models are trained on the training dataset.

python

Copy code

from sklearn.model\_selection import train\_test\_split

from xgboost import XGBClassifier

# Splitting data into train and test sets

X = df.drop('Churn', axis=1)

y = df['Churn']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Training XGBoost Classifier

model = XGBClassifier()

model.fit(X\_train, y\_train)

**6. Model Evaluation**

Evaluating the performance of the trained models using various metrics.

* **Accuracy Score**: Calculate the accuracy of the model.
* **Classification Report**: Generate a report showing precision, recall, and F1-score.
* **Confusion Matrix**: Plot the confusion matrix to visualize performance.

python

Copy code

from sklearn.metrics import classification\_report, accuracy\_score

# Predicting on test data

y\_pred = model.predict(X\_test)

# Evaluation metrics

print(classification\_report(y\_test, y\_pred))

print('Accuracy:', accuracy\_score(y\_test, y\_pred))

**7. Conclusion**

Summarizes the findings and concludes which model performed best for the customer churn prediction task.

markdown

Copy code

In conclusion, the Telecom Customer Churn Prediction was best achieved with the XGBoost Classifier, yielding an accuracy score of 80.86%.

**Appendix**

**Dependencies**

List of libraries and dependencies used in the notebook:

* pandas
* numpy
* seaborn
* matplotlib
* scikit-learn
* xgboost

**Code for Reproducibility**

Ensure that the notebook includes all necessary code to reproduce the results.

python

Copy code

# Ensure reproducibility

import numpy as np

import random

seed = 42

random.seed(seed)

np.random.seed(seed)

This documentation provides a clear and structured overview of the Customer\_Churn\_Prediction.ipynb notebook, explaining its purpose, methodology, and key components. ​​

4o