



Customer Churn Prediction - Detailed Project Report



Project Overview

The objective of this project is to **build a classification model that predicts customer churn** for a telecom company. By predicting which customers are likely to leave, the company can implement strategies to improve customer retention.



Dataset Information

- **Dataset Name:** Telecom Customer Churn Dataset
- **Description:** Contains customer demographic information, account details, services opted, and whether the customer churned.
- **Target Variable:** Churn (Yes / No)
- **Features:** Includes:
 - **Customer Info:** Gender, SeniorCitizen, Partner, Dependents
 - **Account Info:** Tenure, MonthlyCharges, TotalCharges
 - **Services:** PhoneService, InternetService, Contract type, PaymentMethod, etc.



Tools & Technologies

- **Language:** Python
- **Libraries:**
 - Pandas for data manipulation
 - NumPy for numerical operations
 - Matplotlib & Seaborn for data visualization
 - Scikit-learn for machine learning models and evaluation
 - XGBoost for advanced boosting-based classification



Data Preprocessing

1. **Loading the Data**
 - Loaded CSV data using Pandas.
2. **Data Cleaning**
 - Handled missing values in TotalCharges.
 - Converted data types where required.
3. **Feature Encoding**
 - Applied **Label Encoding** for binary categories.
 - Applied **One-Hot Encoding** for multiclass categorical features.
4. **Feature Scaling**
 - Used **StandardScaler** to normalize numeric features for Logistic Regression.
5. **Data Splitting**
 - Split data into **Training (80%)** and **Testing (20%)** using `train_test_split`.

Modeling

◆ Logistic Regression

- **Scaler:** StandardScaler
- **Hyperparameter:** Increased `max_iter=2000` to ensure convergence.
- **Accuracy:** ~80%
- **Evaluation:**
 - Confusion Matrix
 - Classification Report (Precision, Recall, F1-Score)

◆ XGBoost Classifier

- Model trained without label encoding (as it's deprecated in newer XGBoost versions).
- **Hyperparameters:** Default (with `eval_metric='logloss'`)
- **Accuracy:** Higher than Logistic Regression (Typically ~82-85%)
- **Feature Importance:** Extracted to understand the most impactful features.

Results

Model	Accuracy
Logistic Regression	~80%
XGBoost	~82-85%

- **Key Metrics:** Precision, Recall, F1-Score evaluated on test data.
- **Feature Insights:** Contract type, tenure, and monthly charges are significant indicators of churn.

Visualizations

- **Correlation Heatmap** for feature relationships.
- **Churn Distribution Plot**
- **Bar plot for Feature Importance from XGBoost**

Conclusions

- **XGBoost outperforms Logistic Regression** in churn prediction.
- The model can aid telecom companies to proactively target customers likely to churn.
- Business strategies can focus on improving contract offerings and reducing monthly charges.

Tools & Environment

- Python 3.11
- Jupyter Notebook & Python Scripts
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost

Future Enhancements

- Hyperparameter tuning (e.g., Grid Search, Randomized Search) for XGBoost.
- Explore more complex models like Random Forest or Neural Networks.
- Deploy model via a web application for real-time predictions.

Author

Meet Limbachiya
July 2025 | OutriX Internship Project

TERMINAL OUTPUTS:

```
EXPLORER CUSTOMER_CHURN_PREDICTION.ipynb CUSTOMER_CHURN_PREDICTION.py x
OPEN EDITORS CUSTOMER_C... CUSTOMER_C...
OUTLINE
TIMELINE
CUSTOMER CHURN PRE...
CUSTOMER_CH...
Figure_1.png
Figure_2.png
telecom_custom...

CUSTOMER_CHURN_PREDICTION.py > ...
1 # CUSTOMER_CHURN_PREDICTION.py
2
3 import pandas as pd
4

(churn-prediction) meetlimbachiya@192 CUSTOMER CHURN PREDICTION % /Users/meetlimbachiya/miniconda3/bin/python "/Users/meetlimbachiya/Desktop/CODE/OutriX Internship/CUSTO
MER CHURN PREDICTION/CUSTOMER_CHURN_PREDICTION.py"
Data Shape: (7843, 21)
First 5 rows:
customerID gender SeniorCitizen Partner Dependents tenure ... Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
0 7590-WIVGG Female 0 Yes No 1 ... Month-to-month Yes Electronic check 29.85 29.85 No
1 5575-QNVDE Male 0 No 34 ... One year No Mailed check 56.95 1889.50 No
2 3668-QPYBK Male 0 No 2 ... Month-to-month Yes Mailed check 53.85 108.15 Yes
3 7795-CFOCM Male 0 No 45 ... One year No Bank transfer (automatic) 42.30 1848.75 No
4 9237-HQITU Female 0 No 2 ... Month-to-month Yes Electronic check 70.70 151.65 Yes

[5 rows x 21 columns]
/Users/meetlimbachiya/Desktop/CODE/OutriX Internship/CUSTOMER_CHURN_PREDICTION/CUSTOMER_CHURN_PREDICTION.py:29: FutureWarning: A value is trying to be set on a copy of a
DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform t
he operation inplace on the original object.

df['TotalCharges'].fillna(df['TotalCharges'].median(), inplace=True)

After Encoding:
gender SeniorCitizen Partner Dependents tenure PhoneService ... Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
0 0 0 1 0 1 0 ... 0 1 2 29.85 29.85 0
1 1 0 0 0 34 1 ... 1 0 3 56.95 1889.50 0
2 1 0 0 0 2 1 ... 0 1 3 53.85 108.15 1
3 1 0 0 0 45 0 ... 1 0 0 42.30 1848.75 0
4 0 0 0 0 2 1 ... 0 1 2 70.70 151.65 1

[5 rows x 20 columns]

--- Logistic Regression ---
Accuracy: 0.815471965933286
Classification Report:
precision recall f1-score support
0 0.86 0.90 0.88 1036
1 0.68 0.58 0.62 373

accuracy 0.77 0.74 0.82 1409
macro avg 0.77 0.74 0.75 1409
weighted avg 0.81 0.82 0.81 1409

Confusion Matrix:
[[933 103]
 [157 216]]
/Users/meetlimbachiya/miniconda3/lib/python3.13/site-packages/xgboost/core.py:158: UserWarning: [21:59:14] WARNING: /var/folders/c/_qfmhj66j0tn016nkx_th4hxm0000gp/T/abs_
Ln 13, Col 1 Spaces: 4 UTF-8 LF () Python 3.13.2 (base)
```

```
EXPLORER CUSTOMER_CHURN_PREDICTION.ipynb CUSTOMER_CHURN_PREDICTION.py x
OPEN EDITORS CUSTOMER_C... CUSTOMER_C...
OUTLINE
TIMELINE
CUSTOMER CHURN PRE...
CUSTOMER_CH...
Figure_1.png
Figure_2.png
telecom_custom...

CUSTOMER_CHURN_PREDICTION.py > ...
1 # CUSTOMER_CHURN_PREDICTION.py
2
3 import pandas as pd
4

(churn-prediction) meetlimbachiya@192 CUSTOMER CHURN PREDICTION % [
After Encoding:
gender SeniorCitizen Partner Dependents tenure PhoneService ... Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
0 0 0 1 0 1 0 ... 0 1 2 29.85 29.85 0
1 1 0 0 0 34 1 ... 1 0 3 56.95 1889.50 0
2 1 0 0 0 2 1 ... 0 1 3 53.85 108.15 1
3 1 0 0 0 45 0 ... 1 0 0 42.30 1848.75 0
4 0 0 0 0 2 1 ... 0 1 2 70.70 151.65 1

[5 rows x 20 columns]

--- Logistic Regression ---
Accuracy: 0.815471965933286
Classification Report:
precision recall f1-score support
0 0.86 0.90 0.88 1036
1 0.68 0.58 0.62 373

accuracy 0.77 0.74 0.82 1409
macro avg 0.77 0.74 0.75 1409
weighted avg 0.81 0.82 0.81 1409

Confusion Matrix:
[[933 103]
 [157 216]]
/Users/meetlimbachiya/miniconda3/lib/python3.13/site-packages/xgboost/core.py:158: UserWarning: [21:59:14] WARNING: /var/folders/c/_qfmhj66j0tn016nkx_th4hxm0000gp/T/abs_
bgqkllz_uq/croot/xgboost-split.1724873748391/work/src/learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

warnings.warn(msg, UserWarning)

--- XGBoost Classifier ---
Accuracy: 0.7835344215755855
Classification Report:
precision recall f1-score support
0 0.83 0.88 0.86 1036
1 0.61 0.51 0.56 373

accuracy 0.72 0.70 0.78 1409
macro avg 0.72 0.70 0.78 1409
weighted avg 0.77 0.78 0.78 1409

Confusion Matrix:
[[913 123]
 [182 191]]
2025-07-19 21:59:26.317 python[26145:6081280] The class 'NSSavePanel' overrides the method identifier. This method is implemented by class 'NSWindow'
(churn-prediction) meetlimbachiya@192 CUSTOMER CHURN PREDICTION % [
Ln 13, Col 1 Spaces: 4 UTF-8 LF () Python 3.13.2 (base)
```

GRAPHS:

