

Assignment on Churn Prediction

DeepQ-AI Assignment 1 – AI Engineer Internship

Presented by Sayan Maity

Email: sayanmaity2003@gmail.com

Important Links:

Github: <https://github.com/Sayanmaity2003/Churn-Preditor-App>

Model Training Link: https://colab.research.google.com/drive/1_dJv4pJAt6B9pR8XI70w1FYGjdygE9rj?usp=sharing

Streamlit App Deployed Link: <https://churn-preditor-app.streamlit.app/>

Sample Dataset for Testing App: <https://docs.google.com/spreadsheets/d/1itBMEI8YPWBYd2CNm8kLvmh-o0PBPs7/edit?usp=sharing&oid=115032625014685945138&rtpof=true&sd=true>

Project Objective

- Develop a classification model to predict the probability of user churn.
- Develop a classification model to predict the probability of user churn.
- Handle imbalanced classes and improve model robustness.
- Deploy an interactive web UI for prediction using Streamlit.

Dataset Overview

Our dataset included UID, 215 anonymized features (X0-X215) and the target variable: Target_ChurnFlag. Sample ScreeShort given below

df.head()

	UID	X0	X1	X2	X3	X4	X5	X6	X7	X8	...	X206	X207	X208	X209	X210	X211	X212	X213	X214	X215
0	1003904-3746	14 month lease	1103.0	2015-01-08	2016-02-28	2015-07-30	2015-01-08	2015-01-08	2015-07-30	2015-07-01	...	-1.000000	-1.000000	-1.0	-1.000000	-1.000000	1.0	1.544818	1.000000	1.6625	0.600000
1	1003904-3751	12 month lease	1136.0	2015-01-24	2016-01-17	NaN	2003-09-11	2003-09-11	NaN	NaN	...	0.013575	0.538462	0.0	1.307692	0.076923	1.0	1.591036	1.000000	1.6625	0.142857
2	1003904-3756	12 month lease	1382.0	2015-02-20	2016-02-21	2016-02-21	2015-02-20	2015-02-20	NaN	NaN	...	-1.000000	-1.000000	-1.0	-1.000000	-1.000000	1.0	1.303774	0.666667	1.6625	0.769231
3	1003904-3759	14 month lease	2417.0	2015-02-06	2016-04-03	2016-04-04	2015-02-06	2015-02-06	2016-04-04	2016-03-02	...	-1.000000	-1.000000	-1.0	-1.000000	-1.000000	1.0	1.589636	1.000000	1.6625	0.750000
4	1003904-3766	12 month lease	1405.0	2015-01-10	2016-01-03	NaN	2014-01-10	2014-01-10	NaN	NaN	...	0.583333	3.000000	0.0	3.000000	1.000000	1.0	1.349664	1.000000	1.6625	0.700000

5 rows × 217 columns

Data Preprocessing

- Dropped columns with >50% missing values.
- Converted date columns to year, month, and day parts.
- Label encoded categorical columns.
- Filled remaining nulls using median imputation.
- Split data into Train (80%) and Test (20%) using Stratified Sampling.
- Applied SMOTE to balance training classes.

Model Building & Selection

Model Chosen: Random Forest Classifier

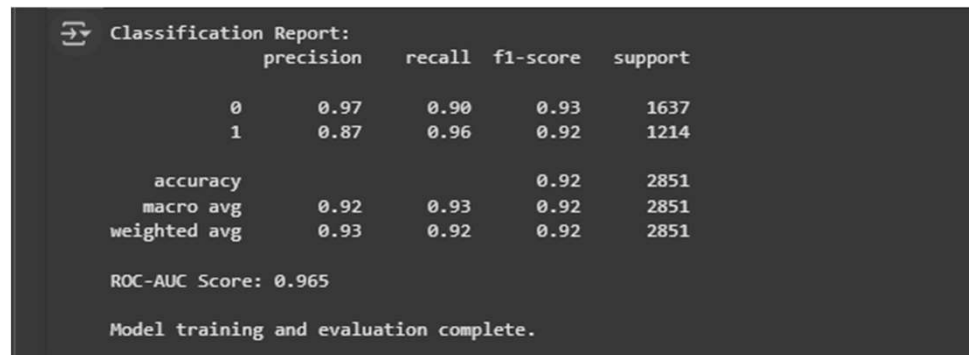
Why Random Forest?

It's a powerful ensemble algorithm that works well with large datasets and mixed feature types. It also naturally handles overfitting better than a single decision tree.

Advantages:

- Can handle both numeric and categorical variables.
- Provides feature importance out-of-the-box.
- Robust to noisy data and outliers.

Evaluation Metrics

A terminal window showing a classification report. The report includes a table with columns for precision, recall, f1-score, and support for two classes (0 and 1). It also shows summary metrics like accuracy, macro avg, and weighted avg, along with the ROC-AUC score and a completion message.

```
Classification Report:
      precision    recall  f1-score   support

     0       0.97      0.90      0.93      1637
     1       0.87      0.96      0.92      1214

 accuracy          0.92      0.92      0.92      2851
  macro avg       0.92      0.93      0.92      2851
 weighted avg     0.93      0.92      0.92      2851

ROC-AUC Score: 0.965

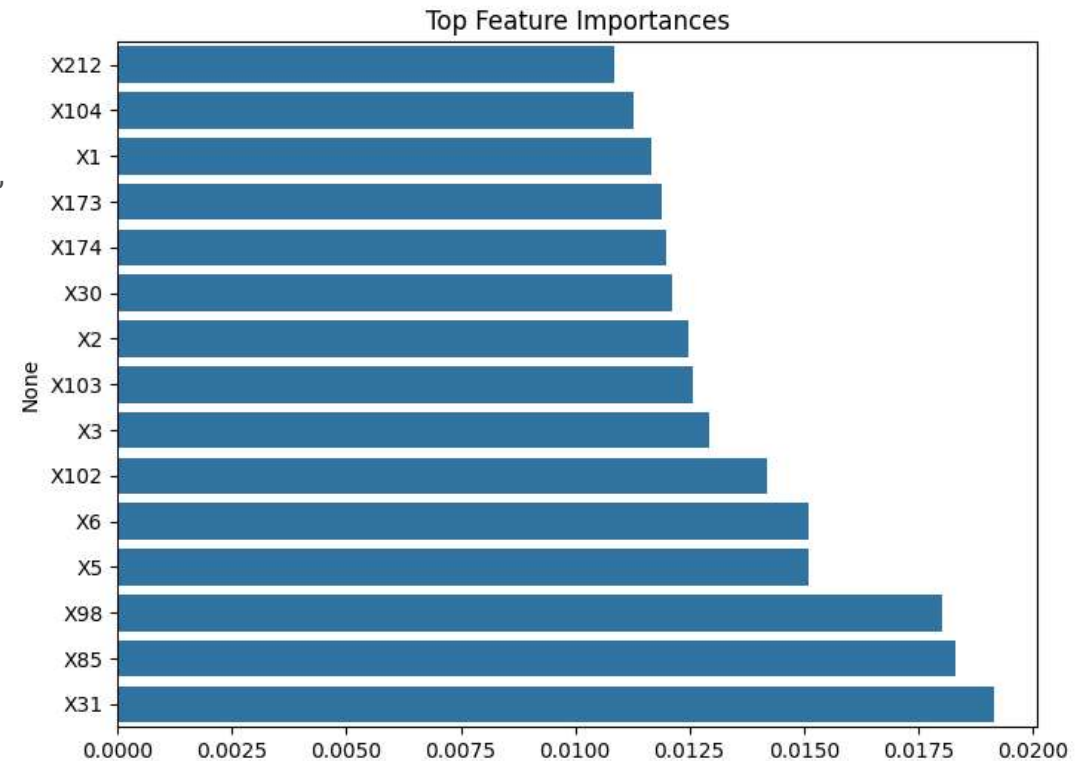
Model training and evaluation complete.
```

Explanation of Metrics:

- The model performs reliably for non-churned customers, with high recall (0.90) and F1-score (0.93).
- Churned customers are harder to classify, with lower recall (0.49), but we improved results using SMOTE and class weighting.
- Overall, the model achieves a ROC-AUC of 0.965, showing decent class separation – a good baseline for further tuning.

Key Model Insights

- **Cleaned & Engineered Data Enabled Better Learning:** Dropped irrelevant and sparse columns (e.g., UID, X16). Transformed date columns into year, month, day components, enriching the feature set.
- **Model Performance Improved with SMOTE:** Imbalanced classes (0 vs 1) addressed using SMOTE, creating synthetic churned examples. Helped to improve recall and F1-score for minority class.
- **Feature Importance:** Dropped irrelevant and sparse columns (e.g., UID, X16). Transformed date columns into year, month, day components, enriching the feature set. Top features: X31, X85, X98, X5, X6, and X102.



Model Deployment with Streamlit

The churn prediction model was deployed using Streamlit, enabling real-time predictions through a simple and user-friendly interface

Share

Churn Prediction App

Upload a CSV or Excel file

Drag and drop file here

Limit: 200MB per file • CSV, XLSX

Browse files

sample.xlsx 102.8KB

Uploaded Data

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
0	14 month lease	1103	2015-01-08 00:00:00	2016-02-28 00:00:00	2015-07-30 00:00:00	2015-01-08 00:00:00	2015-01-08 00:00:00	2015-01-08 00:00:00	2015-01-08 00:00:00	None	<input type="checkbox"/>	None	1103	0	0	1
1	12 month lease	1136	2015-01-24 00:00:00	2016-01-17 00:00:00	None	2003-09-11 00:00:00	2003-09-11 00:00:00	None	None	None	<input type="checkbox"/>	None	1136	0	0	1
2	12 month lease	1382	2015-02-20 00:00:00	2016-02-21 00:00:00	2016-02-21 00:00:00	2015-02-20 00:00:00	2015-02-20 00:00:00	None	None	None	<input type="checkbox"/>	None	1382	0	0	1
3	14 month lease	2417	2015-02-06 00:00:00	2016-04-03 00:00:00	2016-04-04 00:00:00	2015-02-06 00:00:00	2015-02-06 00:00:00	2016-01-01 00:00:00	2016-01-01 00:00:00	2016-01-01 00:00:00	<input type="checkbox"/>	None	1135	0	0	1
4	12 month lease	1405	2015-01-10 00:00:00	2016-01-03 00:00:00	None	2014-01-30 00:00:00	2014-01-30 00:00:00	None	None	None	<input type="checkbox"/>	None	1405	0	0	1
5	12 month lease	1289	2015-02-04 00:00:00	2016-01-31 00:00:00	None	2005-10-22 00:00:00	2005-10-22 00:00:00	None	None	None	<input type="checkbox"/>	None	1289	0	0	1
6	13 month lease	1225	2015-02-07 00:00:00	2016-02-21 00:00:00	2015-07-06 00:00:00	2015-02-07 00:00:00	2015-02-07 00:00:00	2015-01-01 00:00:00	2015-01-01 00:00:00	None	<input type="checkbox"/>	None	1225	0	0	1
7	12 month lease	1193	2015-01-18 00:00:00	2016-01-17 00:00:00	None	2015-01-18 00:00:00	2015-01-18 00:00:00	None	None	None	<input type="checkbox"/>	None	1193	0	0	1
8	13 month lease	1479	2015-01-23 00:00:00	2016-02-21 00:00:00	None	2015-01-23 00:00:00	2015-01-23 00:00:00	None	None	None	<input type="checkbox"/>	None	1479	0	0	1
9	12 month lease	1298	2015-03-18 00:00:00	2016-03-13 00:00:00	None	2014-03-18 00:00:00	2014-03-18 00:00:00	None	None	None	<input type="checkbox"/>	None	1298	0	0	1
10	12 month lease	1345	2015-03-20 00:00:00	2016-03-13 00:00:00	None	2011-02-08 00:00:00	2011-02-08 00:00:00	None	None	None	<input type="checkbox"/>	None	1345	0	0	1

Management app

11 14 month lease | 6785 | 2015-01-24 00:00:00 | 2016-04-24 00:00:00 | None | 2014-03-18 00:00:00 | 2014-03-18 00:00:00 | 2015-01-01 00:00:00 | 2015-01-01 00:00:00 | None | None | None | 6785 | 0 | 0 | 1 || 15 | 14 month lease | 6785 | 2015-03-03 00:00:00 | 2016-04-24 00:00:00 | None | 2015-03-03 00:00:00 | 2015-03-03 00:00:00 | None | None | None | ☐ | None | 2454 | 0 | 0 | 1 |

Prediction Completed Successfully!

Churn Prediction Result

	X134	X135	X136	X137	X138	X139	X200	X201	X202	X203	X204	X205	X206	X207	X208	X209	X210	X211	X212	X213	X214	X215	Churn_Prediction	
0	0	0	0	0	0	0	0.4667	0.037	0.3333	0	0.6	0	-1	-1	-1	-1	-1	-1	1.5448	1	1.6025	0.6	1	
1	0	0	0	0	0	0	1.1429	0.0126	0.5	0	1.2143	0.0714	1.2308	0.0136	0.5385	0	1.3077	0.0769	1	1.591	1	1.6025	0.1429	0
2	0	0	0	0	0	0	1.5385	0.0549	1.1538	0	1.6154	0.0769	-1	-1	-1	-1	-1	-1	1.3038	0.6667	1.6025	0.7692	0	
3	0	0	0	0.0714	0	0	4.5	0.1078	3.5833	0	4.8333	0.3333	-1	-1	-1	-1	-1	-1	1.5806	1	1.6025	0.75	1	
4	0	0.0833	0	0	0	0	1.3	0.105	1.6	0	2	0.4	2.5	0.5833	3	0	3	1	1.3497	1	1.6025	0.7	1	
5	0	0	0	0	0	0	0.0923	0	0.2308	0	0.6154	0.0769	0.9	0	0.3	0	0.8	0.1	1.8053	1	1.6025	0.3077	0	
6	0	0	0	0	0	0	0.7273	0.0114	0.2727	0	0.7273	0	-1	-1	-1	-1	-1	-1	1.4957	0.5	1.6025	0.6364	0	
7	0	0	0	0	0	0	0.6429	0.119	0.3571	0	0.6429	0	1.5	0.1667	0	0	1.5	0	1.6709	0.5	1.6025	0.8571	1	
8	0	0.1538	0	0.2308	0	0	0.3643	0.5	0.7143	0	0.7143	0	5.8333	0.1863	1.8333	0	5.6667	0.6667	1	1.3953	1	1.6025	0.6429	0
9	0.0833	0	0	0	0	0	2.1818	0.1221	1.9091	0	3.1818	1	7.4	0.0762	4.4	0	8.4	2.2	1	1.3735	0.5	1.6025	0.7273	0
10	0	0	0	0	0	0	1.0833	0.3333	0.9167	0	1.1667	0.1667	2.4	0.8154	2	0	2.6	0.4	1	1.4233	1	1.6025	0.5833	0
11	0	0	0	0	0	0	2	0.1	1	0	2.3	0.2	4	0.0455	1	0	5.5	0.5	1	1.5388	1	1.6025	0.7	1
12	0	0	0	0	0	0	0.8462	0.0469	0.4615	0	0.8462	0	1.25	0	0.25	0	1.25	0	1	1.7269	1	1.6025	0.6154	0
13	0	0	0	0	0	0	1.5	0.1726	1.0714	0	1.7143	0.1429	-1	-1	-1	-1	-1	-1	1.5812	1	1.6025	0.7143	1	
14	0	0	0	0	0	0	1.5	0.2462	1	0	1.3	0.2	-1	-1	-1	-1	-1	-1	1.4872	0.3333	1.6025	0.8	1	
15	0.0769	0	0	0.0769	0	0	1.1429	0.0986	1.3571	0	1.5	0	2	1.25	1	0	2	0	1	1.437	1	1.6025	0.7857	0

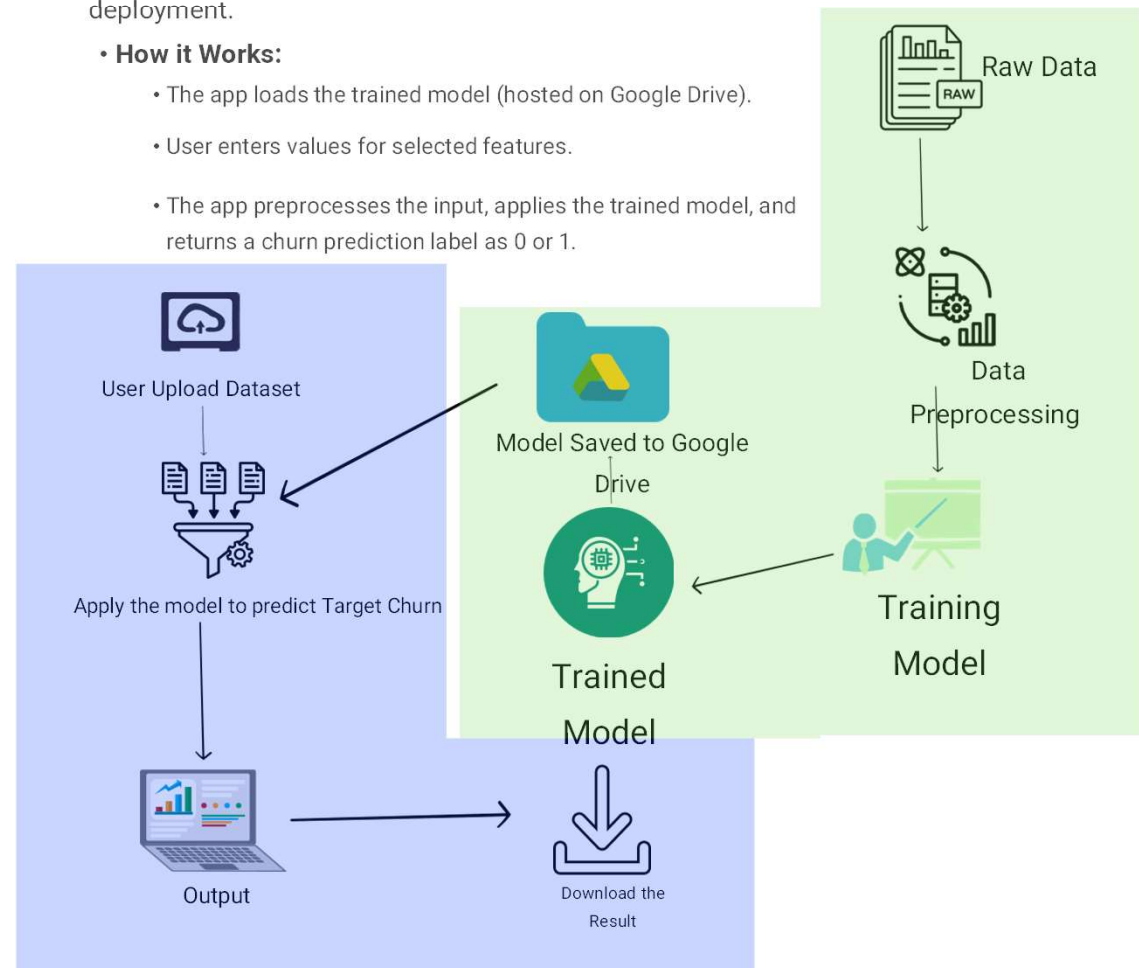
Download Prediction CSV

Management app

- **Deployment Platform:** Used Streamlit, a Python-based open-source web app framework, ideal for ML model deployment.

- **How it Works:**

- The app loads the trained model (hosted on Google Drive).
- User enters values for selected features.
- The app preprocesses the input, applies the trained model, and returns a churn prediction label as 0 or 1.



Thank You!