

Supporting Documentation

Running churn model:

- The churn model is serialised and stored as a pickle file with name “churn_model.pkl”.
- Model is ready to predict in “model_deploy.ipynb” file.
- “model_deploy.ipynb” has tools to create a test set using SQL or pandas DataFrame.
- One way of testing is using SQL if the data is in the database. The “build_test_set” function can build a test set from raw transactional data that is ‘receipts’ and ‘receipt_lines’ table.
- Another way is by pandas dataframe. The dataframe must have 6 input features (purchase_since_2m, customer_age, total_visits, recent_purchased_day, month_of_year, ref_day) and 1 output feature (churn).
- There are 6 scoring options available. Accuracy, Precision and Recall are computed by default. Other metrics available are F1 score, Area under ROC curve and Balanced Accuracy.

No additional cleaning was taken place for the customer transaction tables.

Python methods for plotting:

1. Plotly:
For Exploratory Analysis and customer pen portrait. Various types of charts were used such as Bar chart, Line chart and Scatter Plot.
Documentation: <https://plotly.com/python-api-reference/>
2. PDP and SHAP:
Used for model understanding and exploration.
Documentation: https://pdpbox.readthedocs.io/en/latest/pdp_isolate.html
<https://shap.readthedocs.io/en/latest/api.html>

File Description:

Filename	Description
coursework_complete.ipynb	Python notebook with complete workflow of building a churn prediction model
model_deploy.ipynb	Python notebook for testing the model
churn_model.pkl	Serialized and trained churn prediction model
database.ini	Configurations of PostgreSQL database (must be setup first)
SQLConfig.py	A parser to parse the PostgreSQL configuration to Python