----------------------------------------------------------------------------------------------------------------------------

Code:

- sql

In total, 18 sql files, which include:

1) put csv data into DB,

2) join two or three tables,

3) add/delete/change columns,

4) compute features

- python

1) FILLmissing.py

We fill missing values in user demographic information, e.g. age\_range, gender.

2) TRENDfeature.py

We compute the slope features for user, merchant and user-merchant interaction, in total 12 features generated.

3) blend\_model.py

We train our models(LR, RF, SVM, GBM, Ada, NN) and blend all these models together. Calculate the AUC score for each model based on each feature group.

4) GBM.py, SVM.py, LR.py, RF.py, Ada.py, NN.py, Xgb.py

We use grid search method to find best parameters for each model based on each feature group.

5)featureRanking\_xgboost.py

Use xgboost to analyze importance of feature group and rank features according to gain scores in each feature group. Train LR and RF models with new features.

6)PCA\_ feature.py

Calculate PCA features for merchants

----------------------------------------------------------------------------------------------------------------------------

Key Results:

1) Data Management.png

We managed all the data in PostgreSQL, and stored in total 14 tables

2) purchase\_gender.png

It shows the relationship between #purchases and gender.

3) purchase\_age.png

It shows the relationship between #purchases and age.

4) slope\_feature\_example.png

It shows the example of the slope feature we used for classification.

5) ui.png, up.png, mp.png, um.png, pca.png, slope.png, double11.png, age.png, xgboost.png

Show the results of feature importance and rankings for each feature group

6) LR.png, RF.png, GBM.png, SVM.png, NN.png, Ada.png, BL.png

The evaluation results for each model based on different feature groups

* Screenshots:

1) results\_with\_uifeatures.PNG:

The training result of each model based on user demographic features.

2) results\_with\_upfeatures.PNG:

The training result of each model based on user purchase behavior features.

3) results\_with\_umfeatures.PNG:

The training result of each model based on user merchant interaction features.

4) results\_with\_mpfeatures.PNG:

The training result of each model based on merchant-related features.

5) results\_with\_doublefeatures.PNG:

The training result of each model based on double 11-related features.

6) results\_with\_complexfeatures.PNG:

The training result of each model based on complex features.

7) results\_with\_allfeatures.PNG:

The training result of each model based on all features.

8) RF\_gridsearch\_undermpfeatures.PNG, svm\_gridsearch\_undermpfeatures.PNG, GBM\_gridsearch\_undermpfeatures.PNG, xgb\_gridsearch\_undermpfeatures.PNG, Adaboosting\_gridsearch\_undermpfeatures.PNG, NN\_gridsearch\_undermpfeatures.PNG

The grid search result for each model based on merchant-related features

9)screenshot\_xgboost\_gridsearch.png

Grid search result for XGBoost model when evaluating the feature importance.