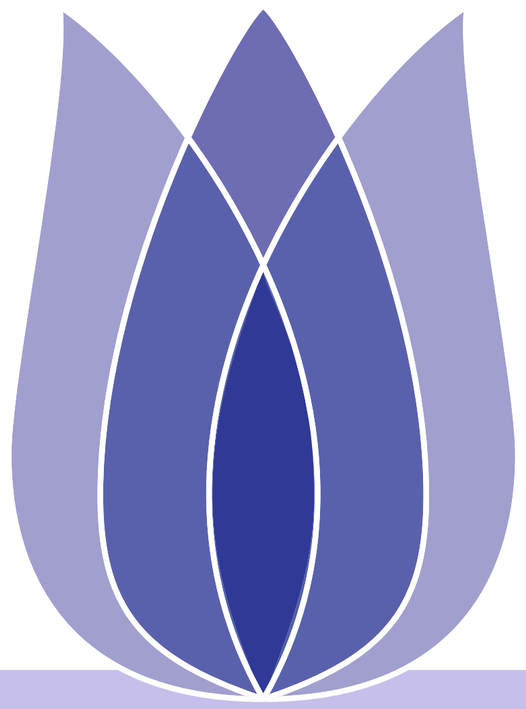
Identifying Customers

Xichen Tang
QUT

January 16, 2020



Directory

Introduce

Directory

Subject Introduce

Naive Bayes

Forecasts

Thanks and Question

Subject Introduce

Introduce

- The kaggle subject:Santander Customer Transaction Prediction

In this challenge, we need to identify which customers will make a specific transaction in the future, irrespective of the amount of money transacted.

Data

- train_data

ID_code	target	var_0	var_1	...	var_198	var_199
train_0	0	8.9255	-6.7863	...	12.7803	-1.0914
train_1	0	11.5006	-4.1473	...	18.3560	1.9518

- test.csv

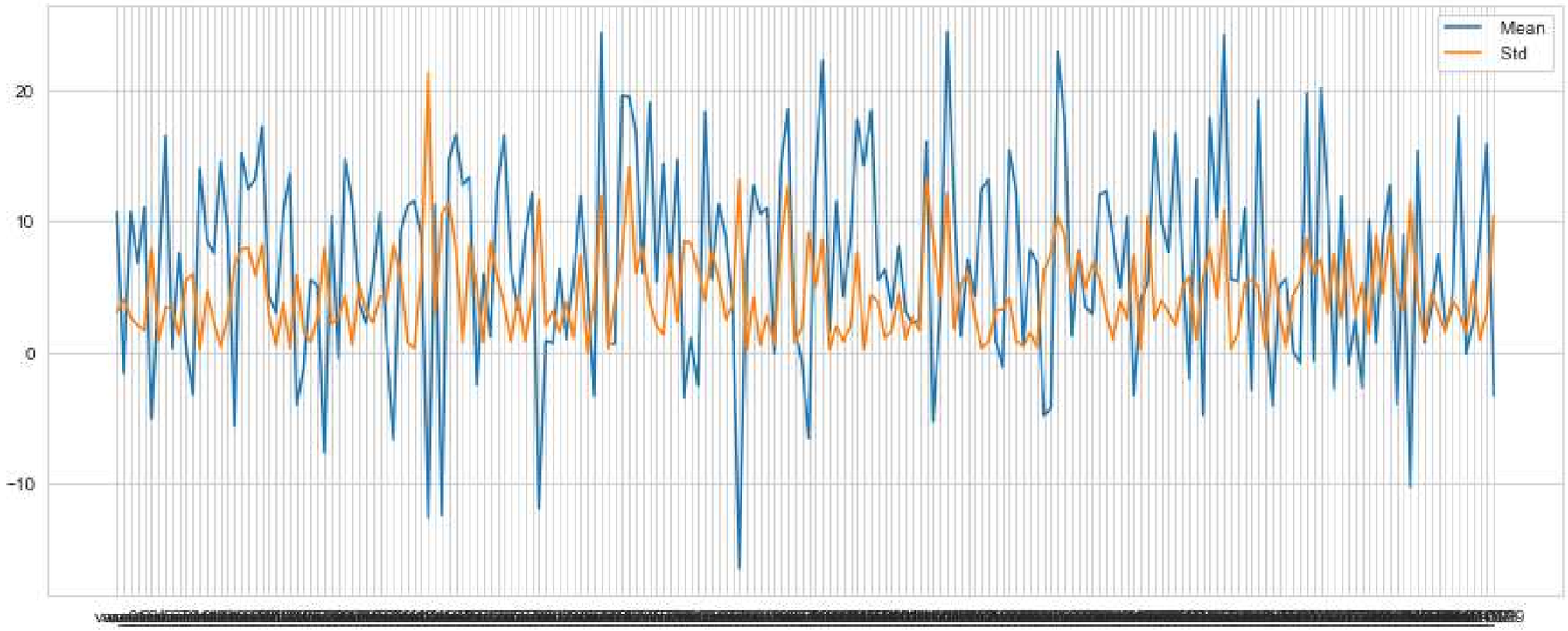
ID_code	var_0	var_1	...	var_198	var_199
test_0	8.9255	-6.7863	...	12.7803	-1.0914
test_1	11.5006	-4.1473	...	18.3560	1.9518

- train_data.info

RangeIndex:	200000 entries	0 to 199999
Columns:	202 entries	ID_code to var_199

Avg and Std

- by describe()

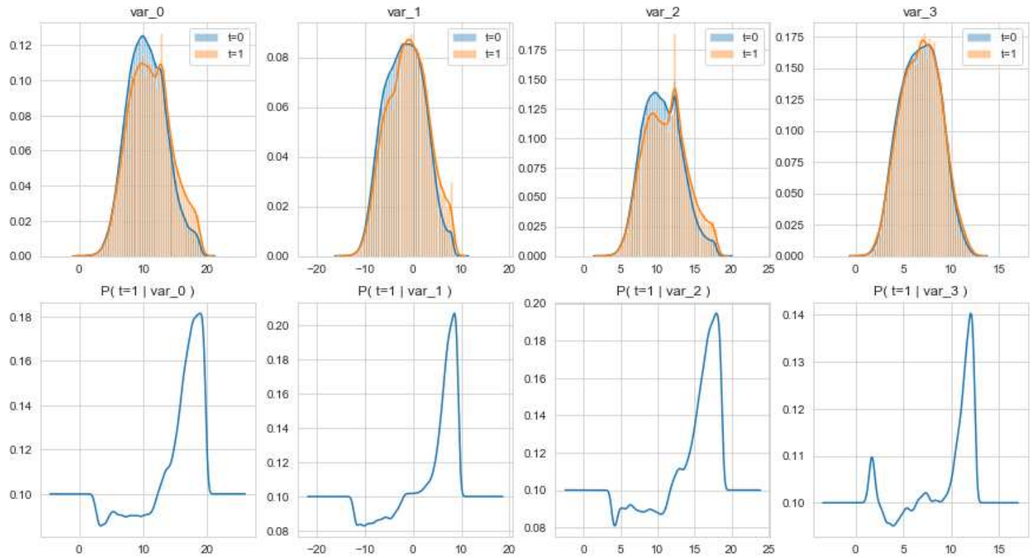


Naive Bayes

Statistical Functions

- Calculate Prob

$$P(A \mid B)=\frac{P(AB)}{P(B)}$$

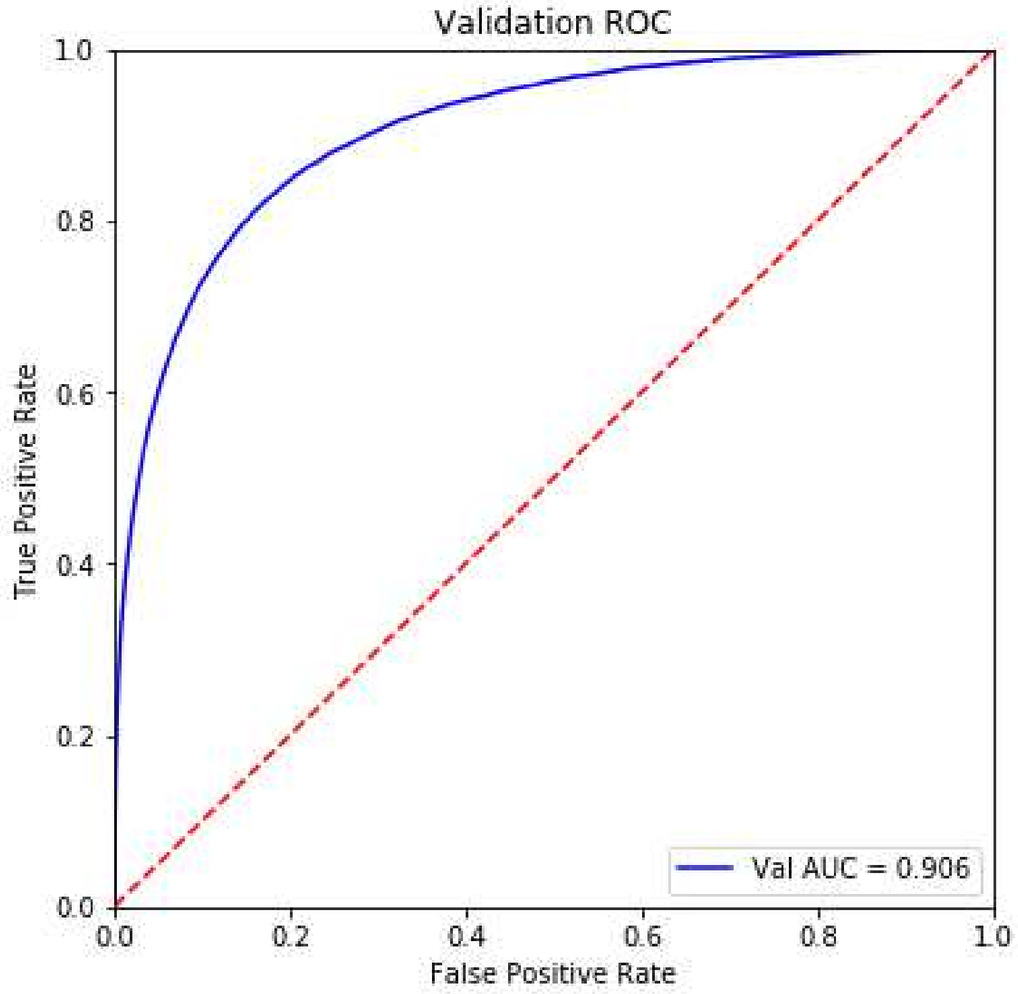


- Smoothing

  If the probability value to be estimated is 0, the calculation result of posterior probability will be affected. The solution to this problem is to use smoothing

AUC

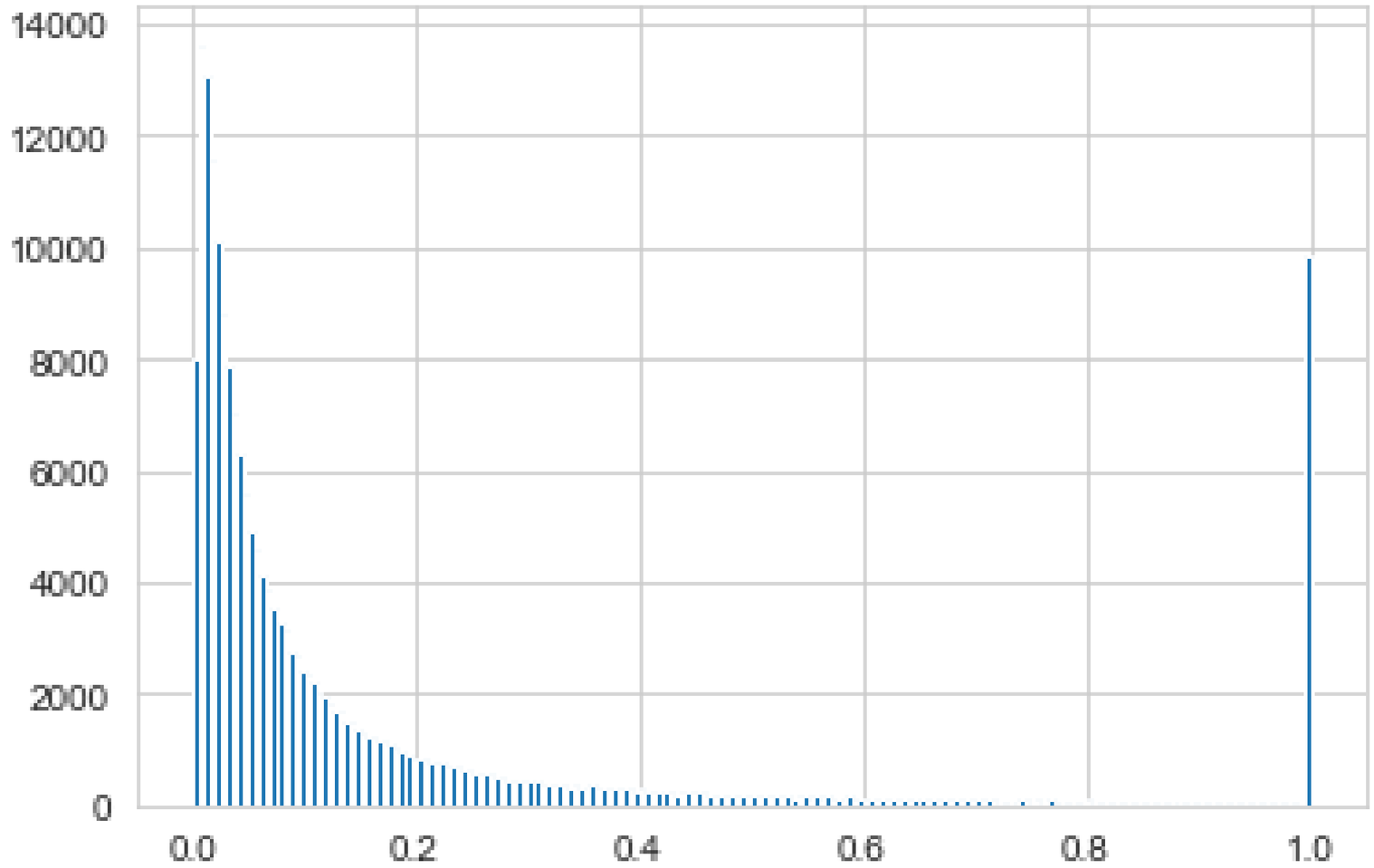
Validation AUC



Validation AUC = 0.9055714125995 24

Result

1. Probability



Decomposition of the graphics

- Result
  We can get Raw data, trend data, periodic data, random variables

Test for stationarity

- The Measure
  seasonal_decompose
- The Result
  There is a 15% probability that the sequence is non-stationary

Remove Seasonalization

- remove seasonalization

Remove Seasonalization

- Result
  The p value is very small, and the sequence after the difference is considered stable
  Now after the transformations, our p-value for the DF test is well within 5 %. Hence we can assume Stationarity of the series

Forecasts

Modle

- modle
- result
  sm.tsa.arma_order_select_ic
  get the best p and q values (time-consuming) by passing in the qualified maximum,It takes too long
- get result
  Input the start time and end time for data prediction then Restore the predicted value

Thanks and Question