

PREDICT FUTURE SALES

Xichen Tang
QUT, China

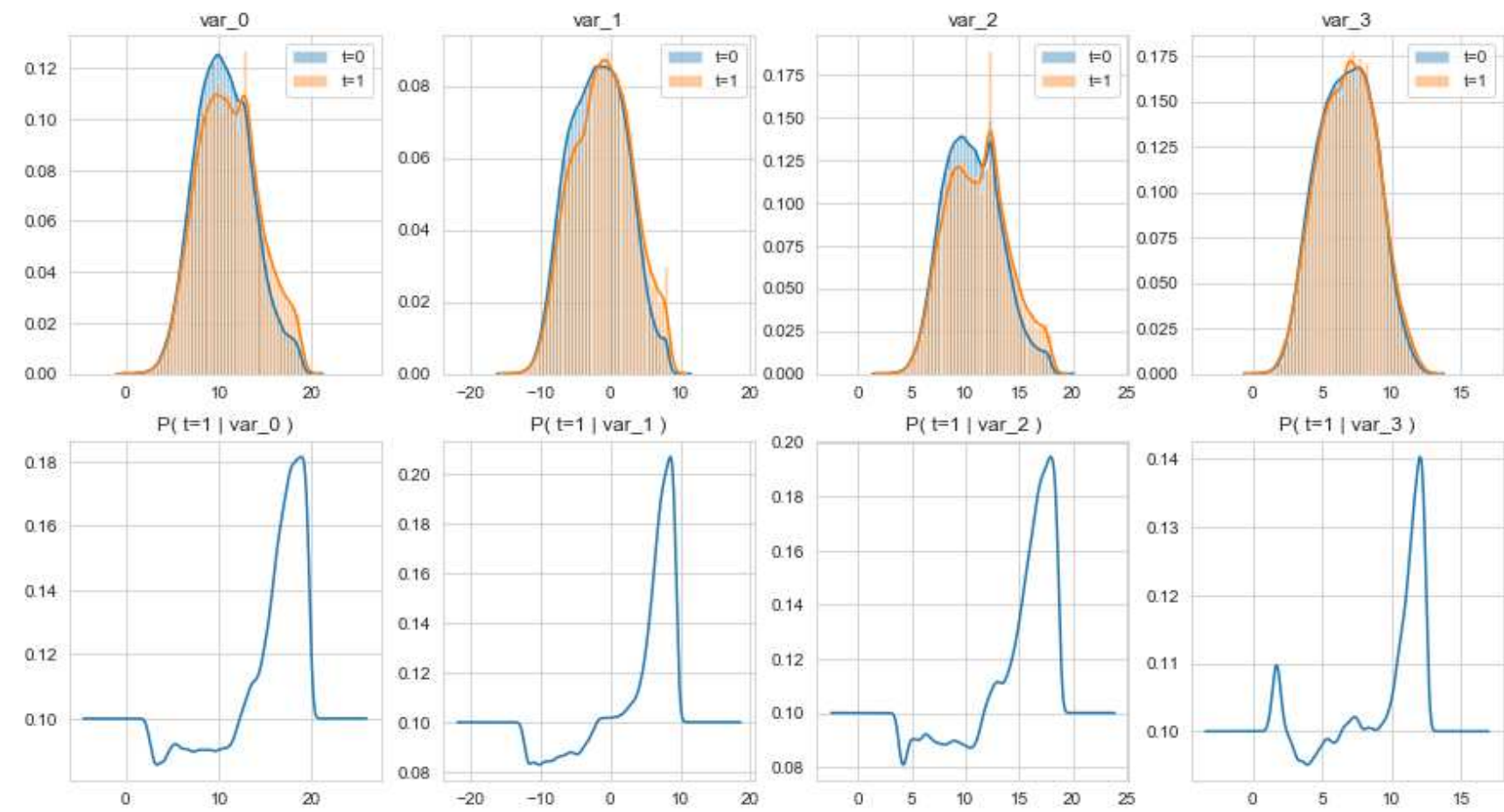
Introduction

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

ID_code
customer ID
target
Whether it will trade in the future, 0 means no, 1 means yes
var_0
Relevant information provided by the bank
...
Relevant information provided by the bank
var_199
Relevant information provided by the bank
Missing Values? False

Naive Bayes

We use different methods to predict, and finally through comparison, we get the best value



Validation AUC = 0.905571412599524

Gaussian naive Bayes

Calculation of prior probability,use Counter() maybe more convenient. Calculate likelihood.Using probability density function of Gaussian distribution to calculate likelihood and then multiply to get likelihood We can get Raw data, trend data, periodic data, random variables

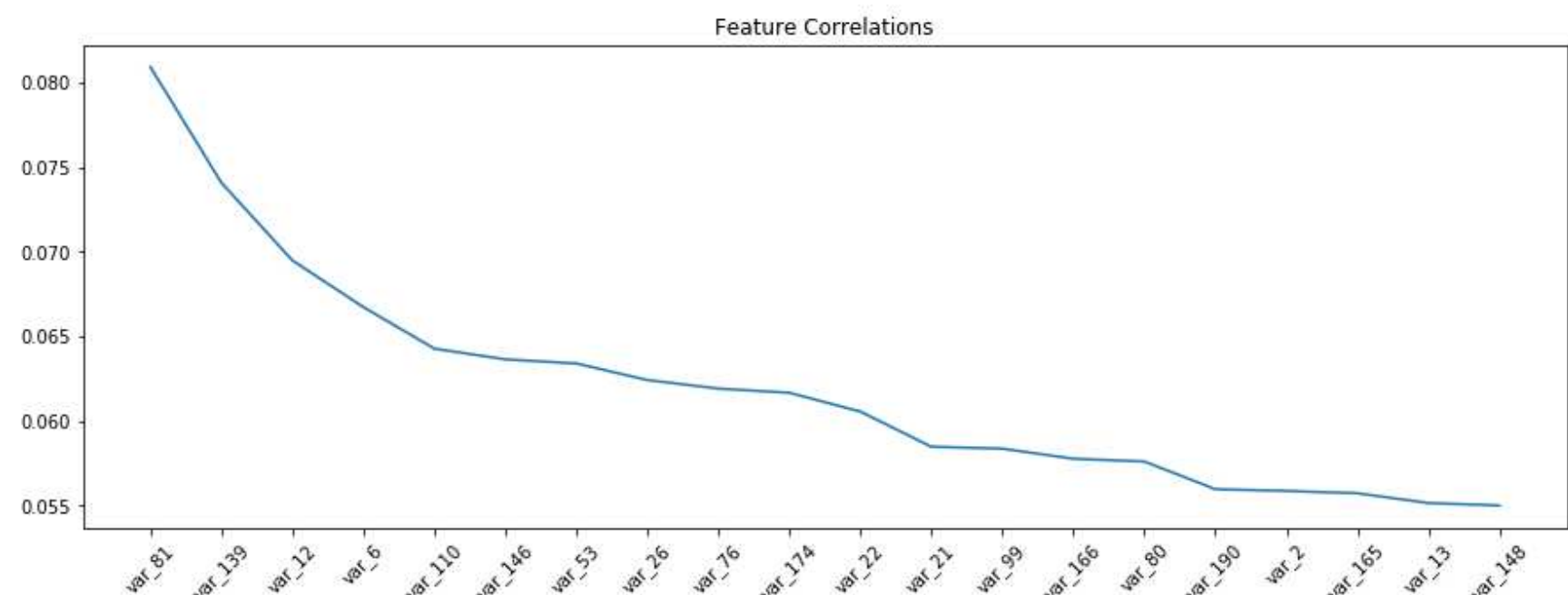
Validation AUC = 0.8051607443604657.

LinearRegression

Merge test/train datasets,add more features
Normalize the data,Standardization of normal distribution,then Square the value, cubic the value,Cumulative normal percentile,Normalize the data,again.Do linear regression,Write submission file
Validation AUC = 0.8025517936065763

Catboost

Feature Correlations

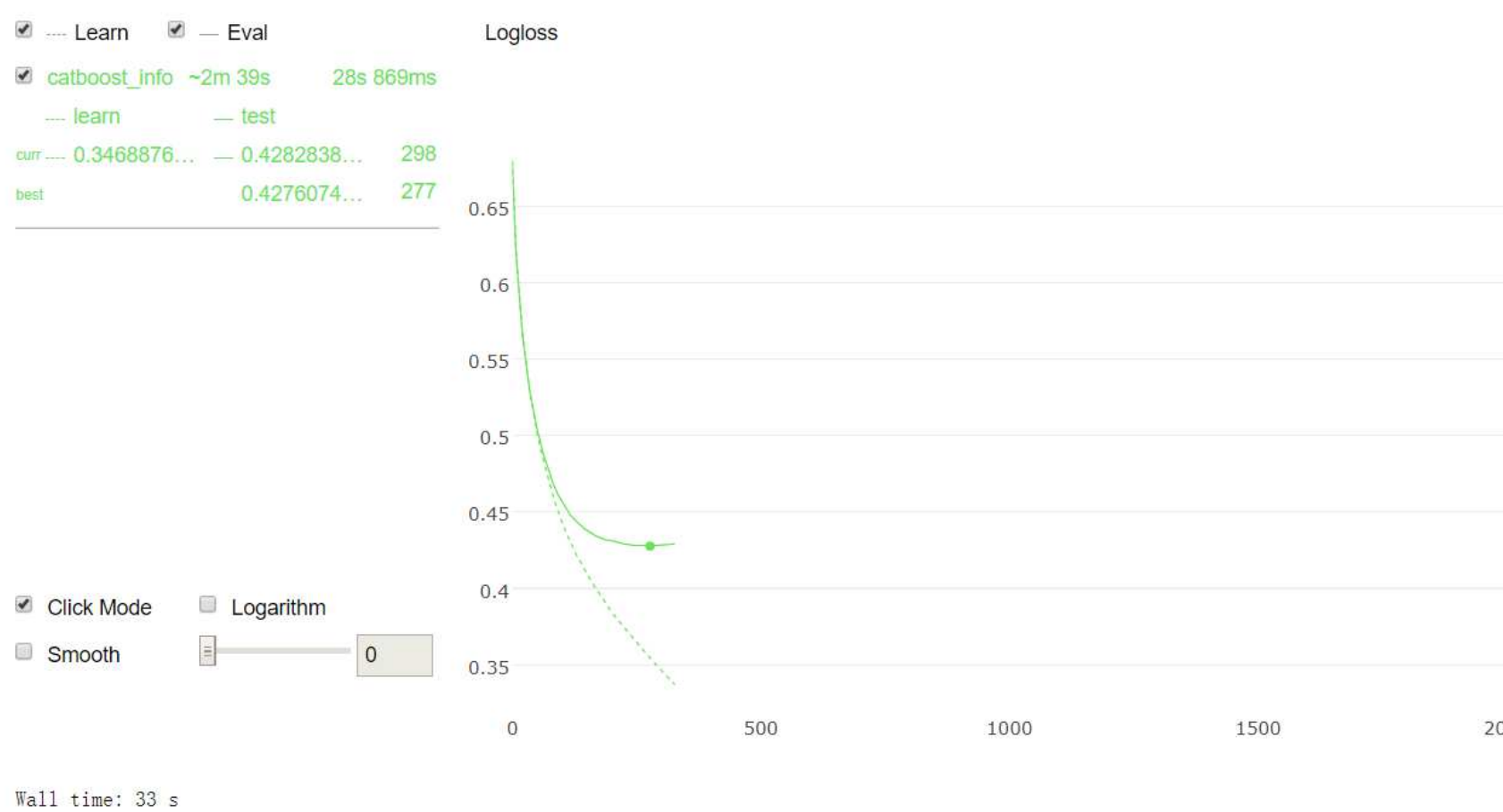


Get the features,Get the top 100 features,merge them and divide the training set and test set

process data

In catboost, you don't have to worry about this at all. You just need to tell the algorithm which features belong to category features, and it will help you deal with them automatically

Finally, we feed the data to the algorithm and train it



Conclusion

Naive Bayes
AUC: 0.9055714
Gaussian naive Bayes
AUC: 0.8051607
LinearRegression
AUC: 0.8025517
Catboost
AUC: 0.8039915

Acknowledgement
•International Cooperation Project (Y7Z0511101)
of IIE, Chinese Academy of Sciences