

Credit Score Prediction

C R E D I T S C O R E



Presented by
Artem Ramus

Introduction

About the data from a publisher:

We provide you with a data set in CSV format. The data set contains 8,000 train instances and 2000 test instance There are 304 input features, labeled x001 to x304.

The target variable is labeled y.

Task Create a model to predict the target variable y.

A report - A Power point presentation Any custom code you used Instructions for me to run your model on a separate data set.

Link to the dataset: <https://www.kaggle.com/prasy46/credit-score-prediction>

Methodology

- The data set is checked for duplicates, null values and homogeneous features.
- Constant and quasi-constant features were deleted.
- Correlated features were deleted.

Performance of the Model

Performance of the following models was estimated based on Pearson coefficient and the loss:

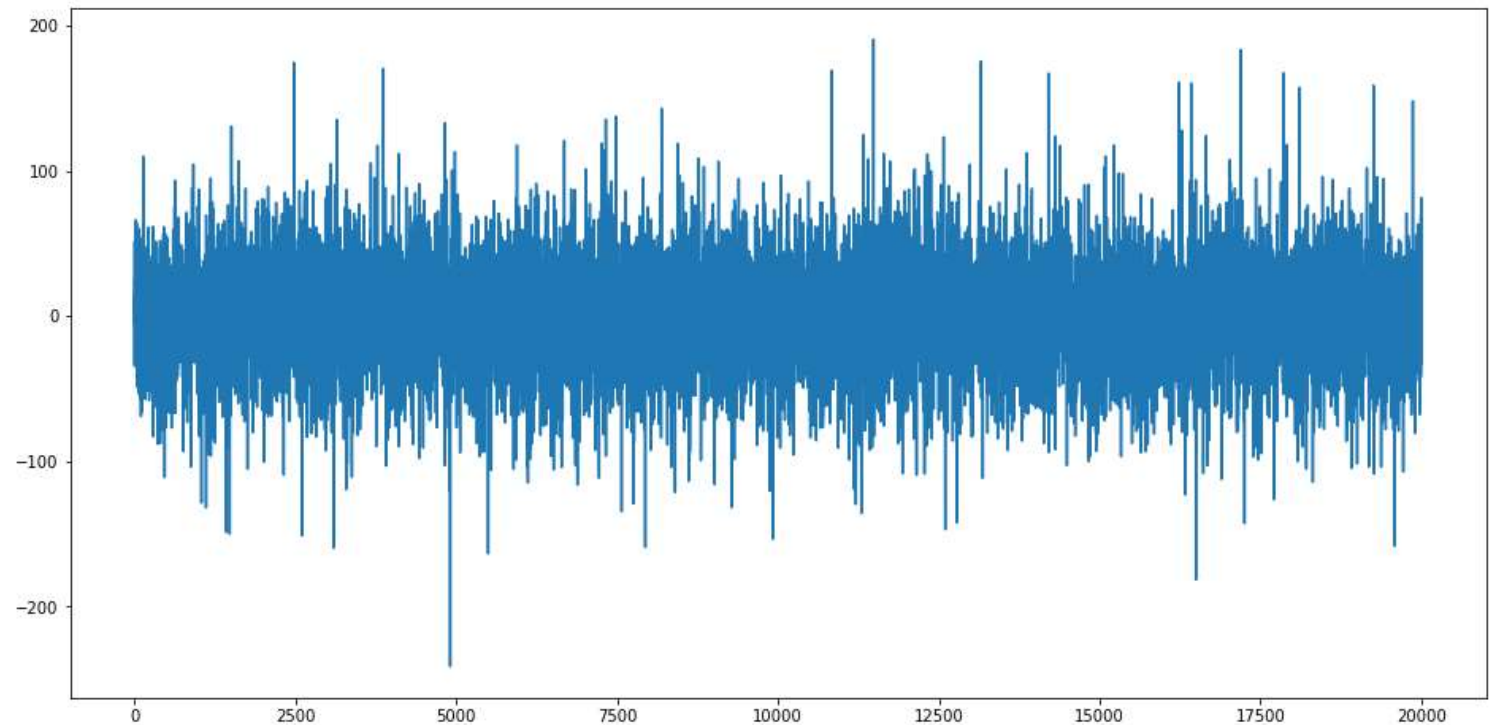
- Random forest
- XGBoost

Summary and Conclusions

XGBoost regression R2 score is 0.94.

Given limit criteria 25, percent of true prediction is 69%.

Figure: loss for
all the cases



The end

Thank you for your attention!