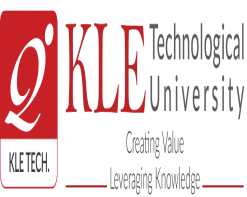
KLE Society's

KLE Technological University



**Course Project**

**On**

**Cryptocurrency Closing Price Prediction**

**Bachelor of Engineering**

**In**

**Computer Science and Engineering**

**Submitted By**

**Team No : 7**

|  |  |
| --- | --- |
| Name | srn |
| yash raj | 01fe19bcs229 |
| arpit arora | 01fe19bcs232 |
| sivanand mahapatra | 01fe19bcs230 |
| nitish kumar | 01fe19bcs227 |
| Nihad J | 01fe19bcs260 |

Faculty In charge

Prof. Soumya

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

HUBLI–580 031 (India).

Academic year 2021-22

**Introduction:**

Description

After the boom and bust of cryptocurrencies’ prices in recent years, cryptocurrencies have been increasingly regarded as an investment asset. Because of their highly volatile nature, there is a need for good predictions on which to base investment decisions. Different existing studies have leveraged machine learning for more accurate cryptocurrency price prediction. We are interested in applying different modeling techniques to samples with different data structures (qualitative and quantitative data) and dimensional features to achieve an optimization in price prediction.

Using the trading time series of a cryptocurrency’s price, in addition to a set of qualitative features (news, social impact, Twitter, Reddit, social media sentiment analysis), we would like to build a model that forecasts a cryptocurrency’s price. In this challenge, we are focusing on the trading time series and how we can optimize currency forecasting. We will predict future cryptocurrency prices.

There are many factors and constraints that can be taken into consideration when increasing or decreasing cryptocurrency prices by the different stakeholders. These factors can be directly seen in newspapers, related websites or social media, for that including these features in the model can add value and predict more accurate cryptocurrency prices.

The target value is the actual price. We have data extracted in an interval of 1h for a period of one year (from 1st of March 2020 to 1st of March 2021). We are interested to predict the values of cryptocurrency prices in specific timestamps that we have in the validation file.

There can be different ways to solve this problem. One can think about using the prices from the different trading platforms as the initial data (the provided dataset) and build forecasting models and/or Neural Networks ones.

The participant that will build a model with the most accurate results will be the winning one.

The challenge does not stop there, as our main goal is to reach and exceed a given threshold (a specified RMSE score) in the final developed model.

The goal is to have predictions that are accurate in a way that it’s mostly similar to the original validation file, to bypass the given threshold evaluation result, and come up with something that is more accurate.

**PROBLEM STATEMENT**

After the boom and bust of cryptocurrencies’ prices in recent years, cryptocurrencies have been increasingly regarded as an investment asset. Because of their highly volatile nature, there is a need for good predictions on which to base investment decisions.

We are interested in applying different modeling techniques to achieve optimization in price prediction.

**Data Set**

The challenge data comprises of the following datasets:

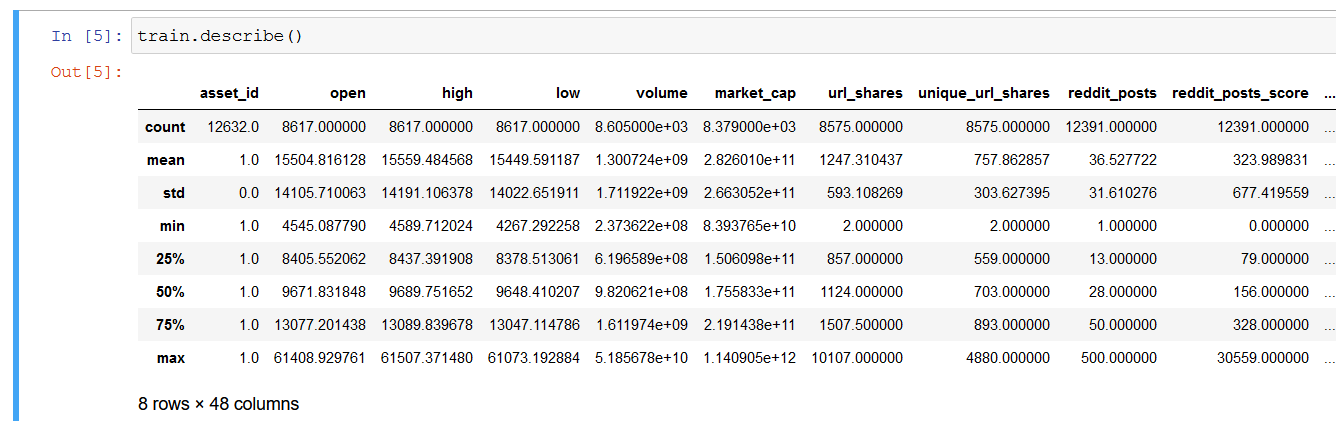
Train.csv (3.3MB) - contains the target. This is the dataset that we will use to train our model. (12632 rows X 49 columns)

Test.csv (1.6MB) - resembles Train.csv but without the target column. This is the dataset on which we will apply your model. (6222 rows X 48 columns)

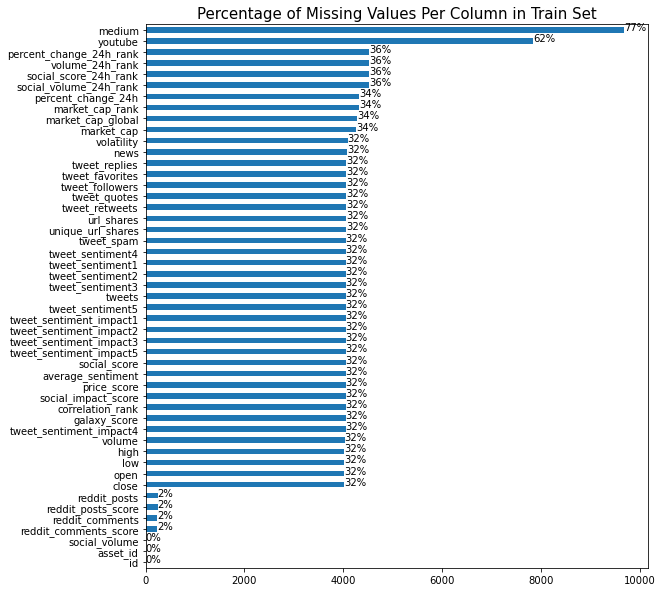
**Results and Discussion:**

**Exploratory Data Analysis:**

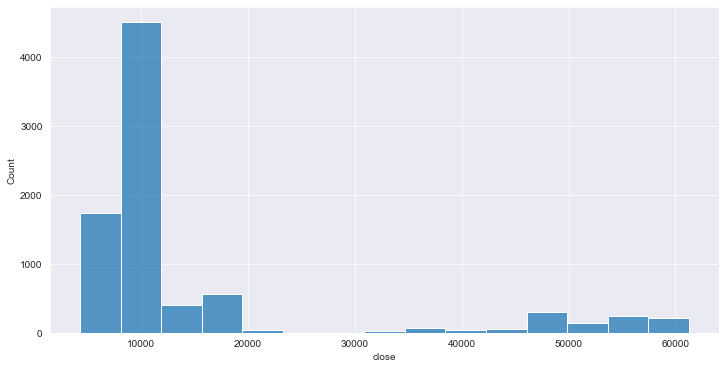
**UNDERSTANDING OF DATA**

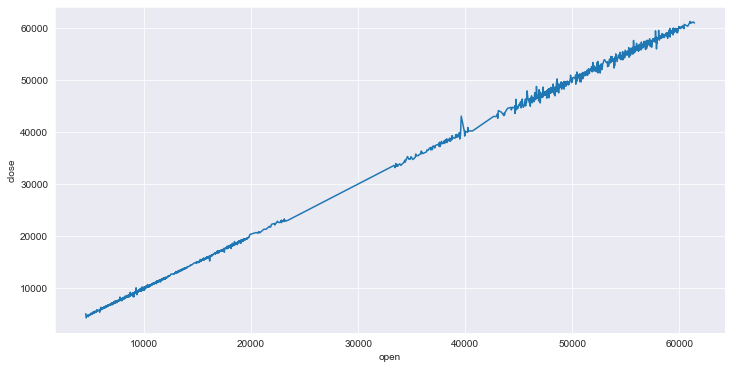


**Exploring the dataset**

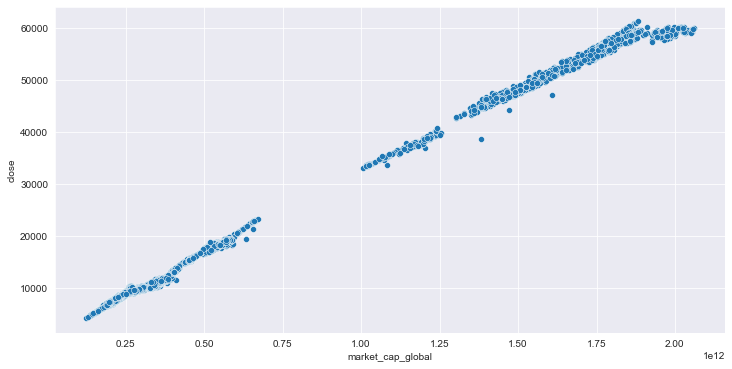


**Data Visualization**



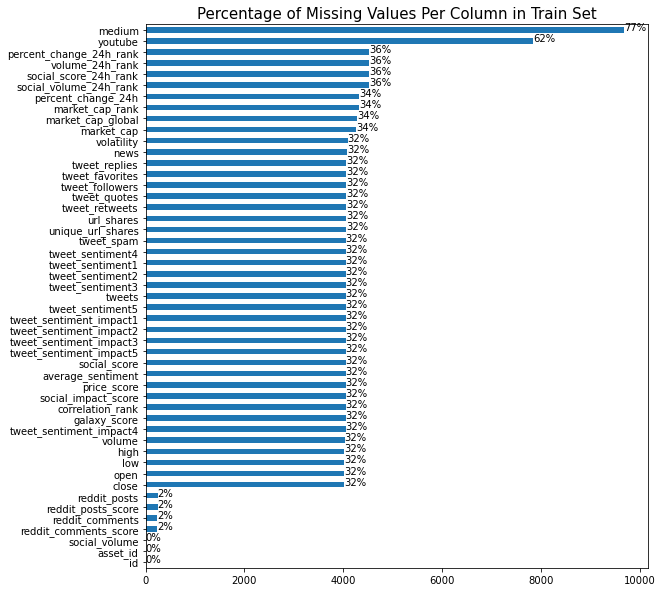


….…………

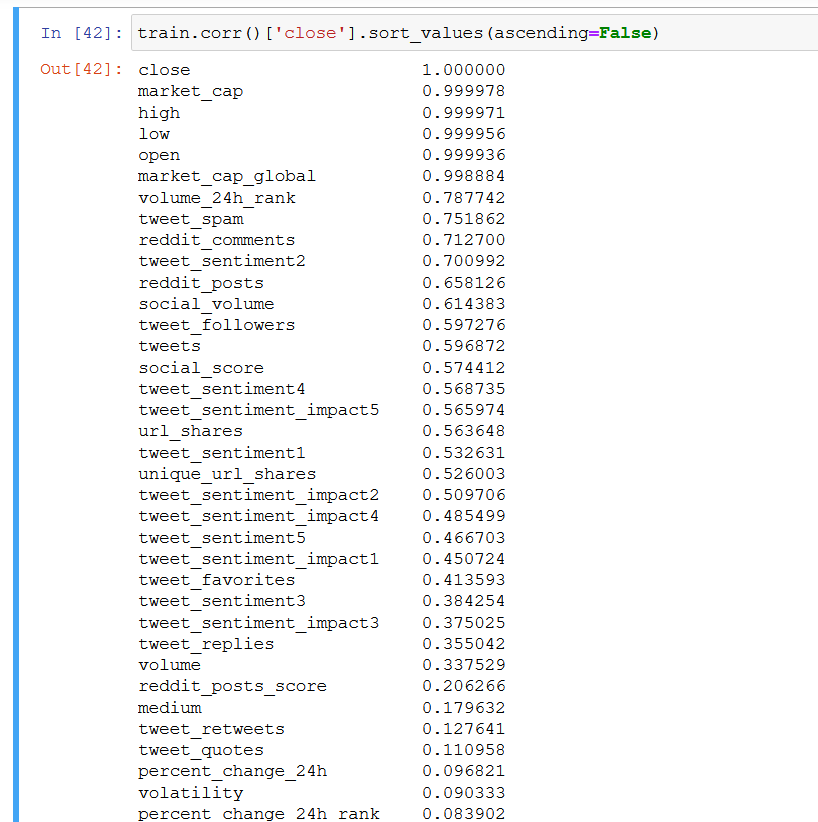




**Data Preprocessing**



**Correlation Analysis**

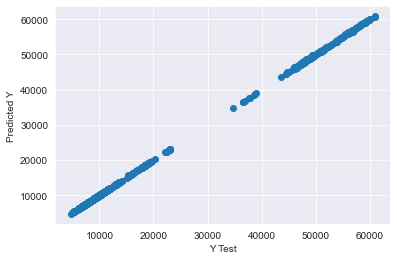


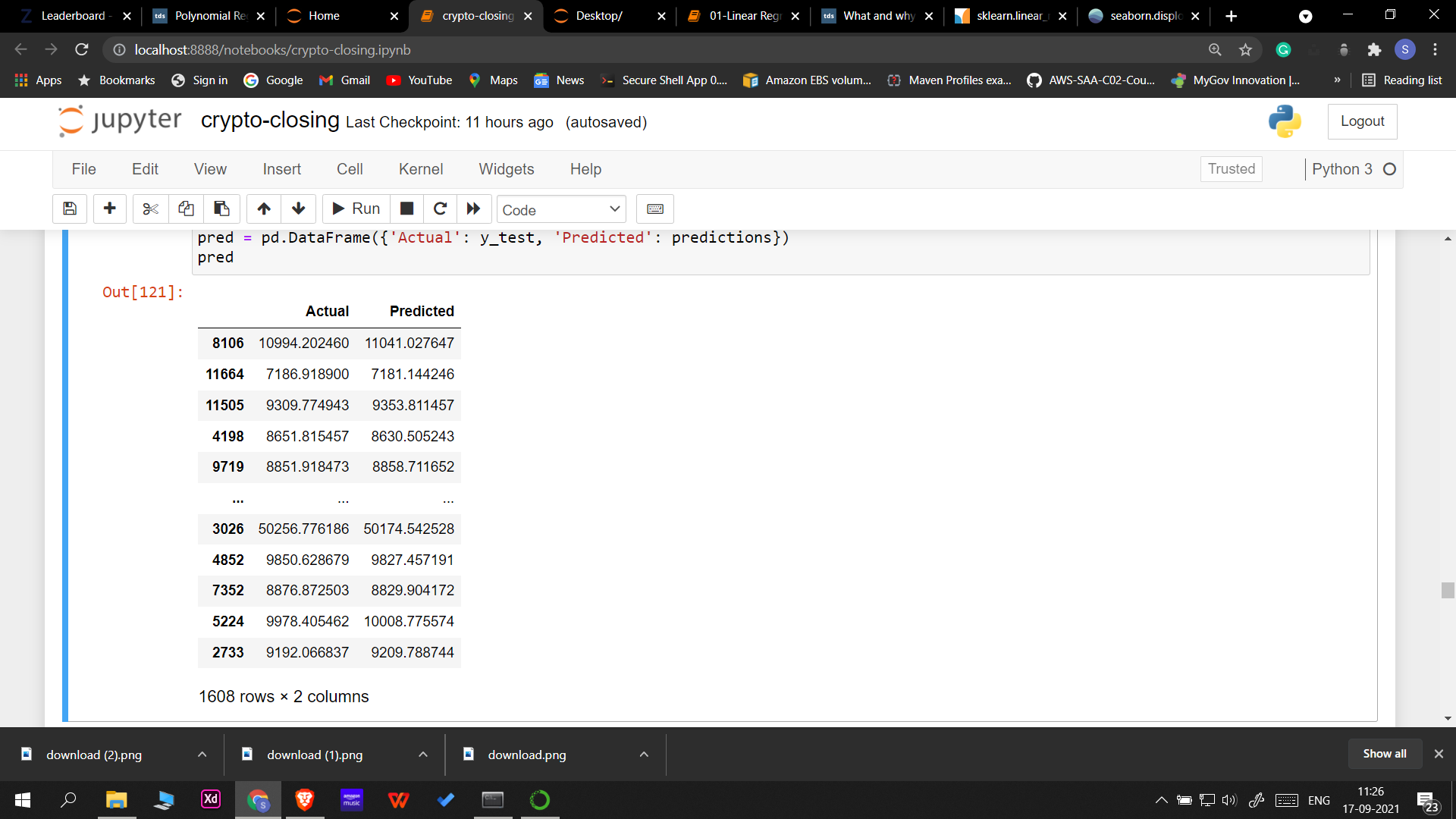
**Data Cleaning**

1. Columns which have percentage of missing values higher than 40% has been removed.
2. Dropped rows where entries of most important colums were missing.
3. Discarded the irrelevant social media attributes.
4. Replaced all the null values with ‘0’ as we had to predict closing price of the cryptocurrency for rows which had null values also in the submission file.

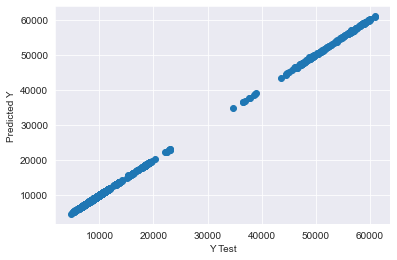
**Learning Models**

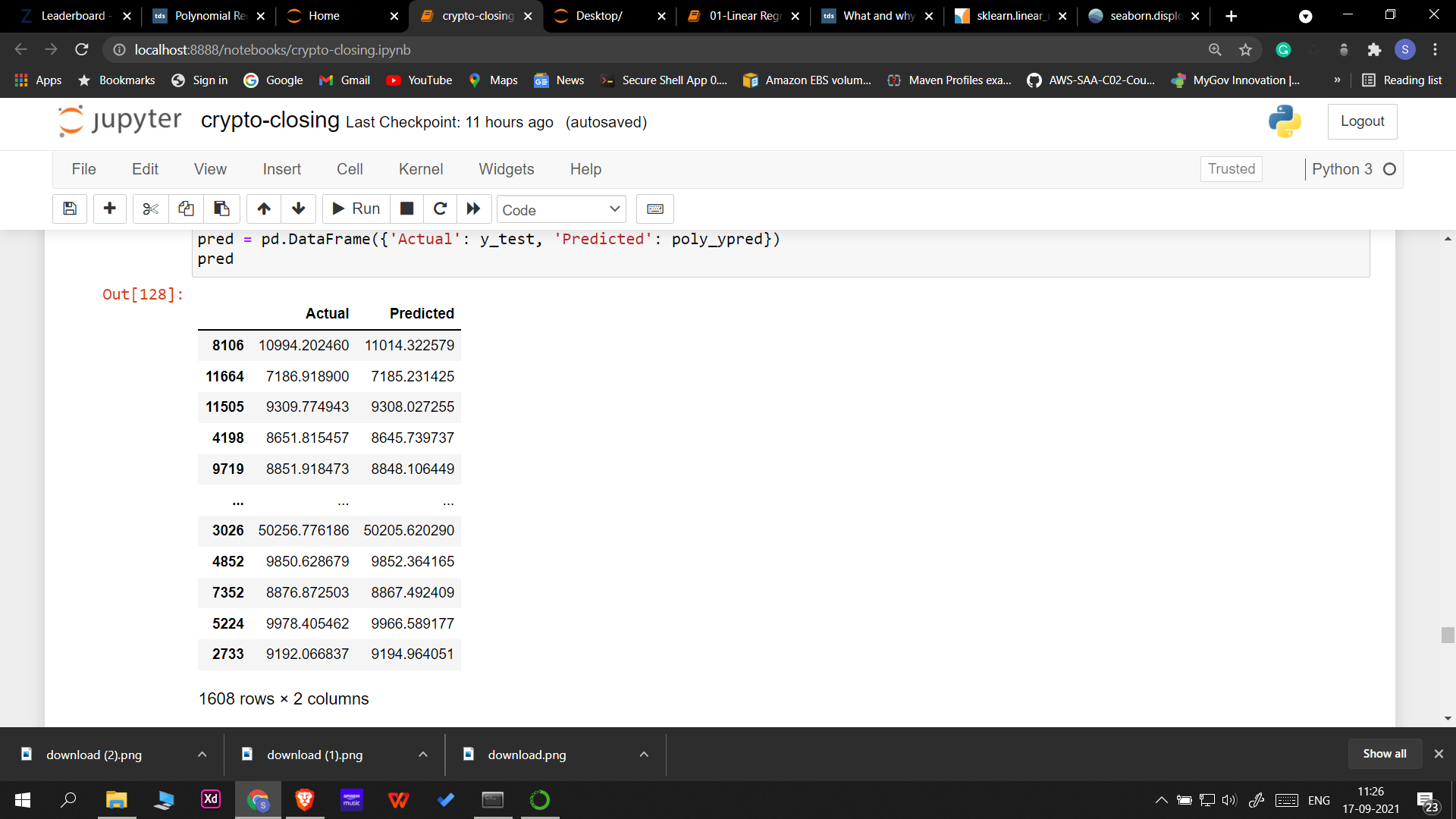
1. **Linear regression**



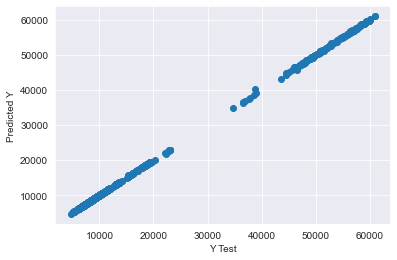


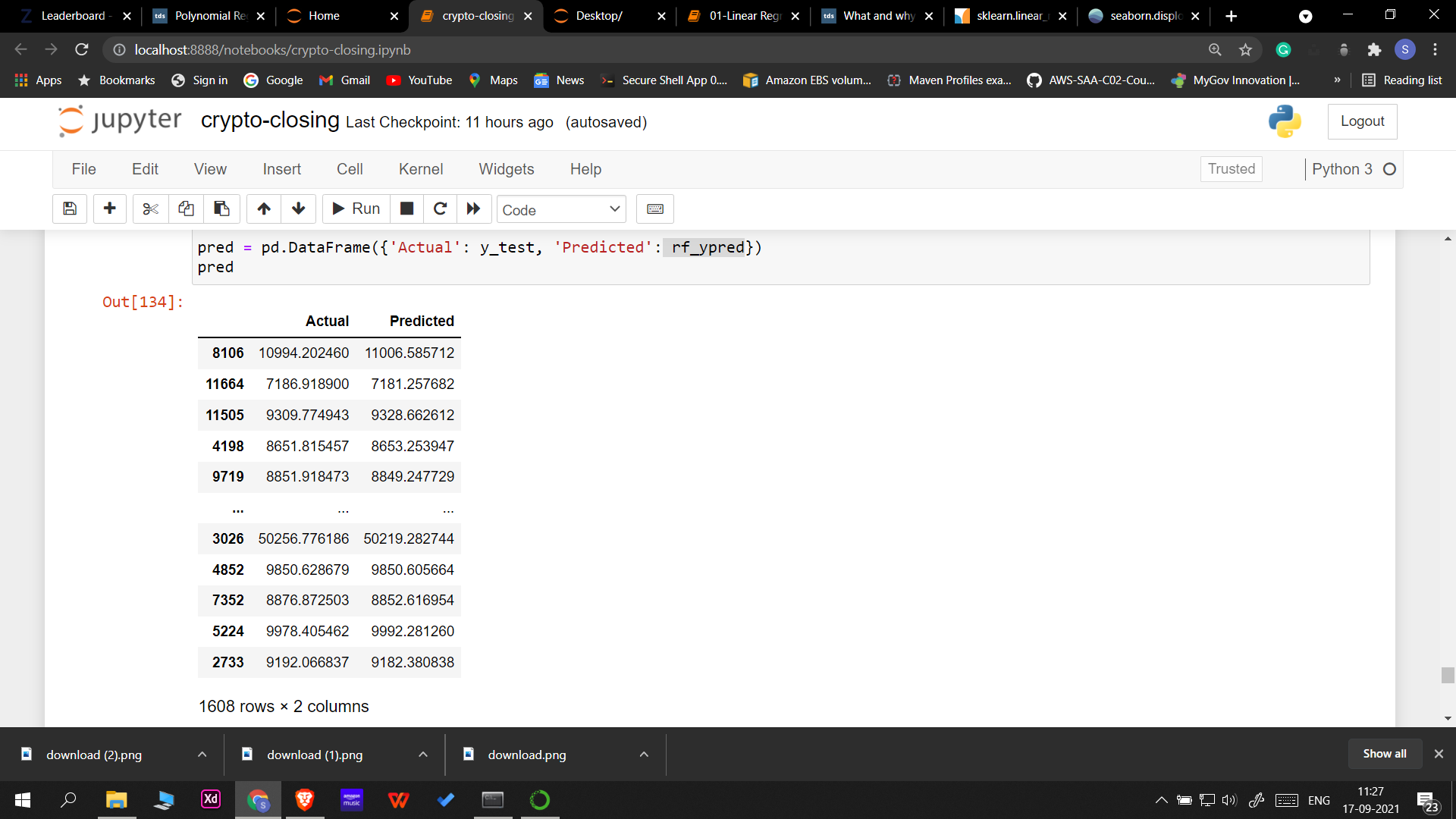
1. **Polynomial regression**



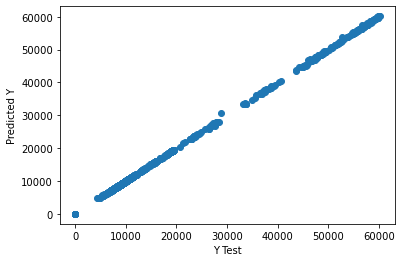


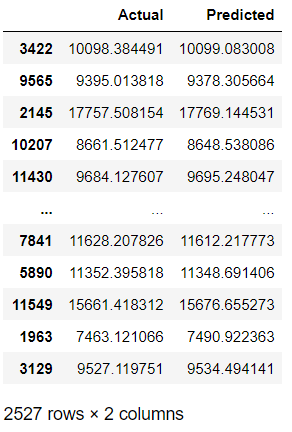
1. **Random Forest regressor**



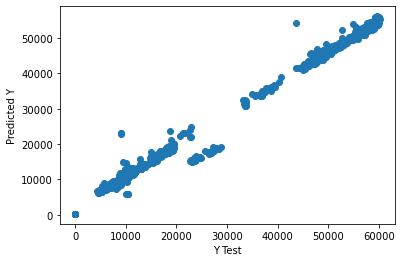


1. **XG boost**





1. **Elastic Net Regressor**





**Conclusion**

From the RMSE score we got for all the models we submitted, it is evident that linear regression model was the best suited model for this problem.

**Leaderboard Ranking**



**References**

1. https://zindi.africa/competitions/cryptocurrency-closing-price-prediction
2. https://pandas.pydata.org/pandas-docs/stable/index.html
3. https://seaborn.pydata.org/index.html
4. https://scikit-learn.org