



**Ahmedabad  
University**

**CSE523: Machine Learning**

**Weekly Report-4**

**Submitted to: Prof Mehul Raval**

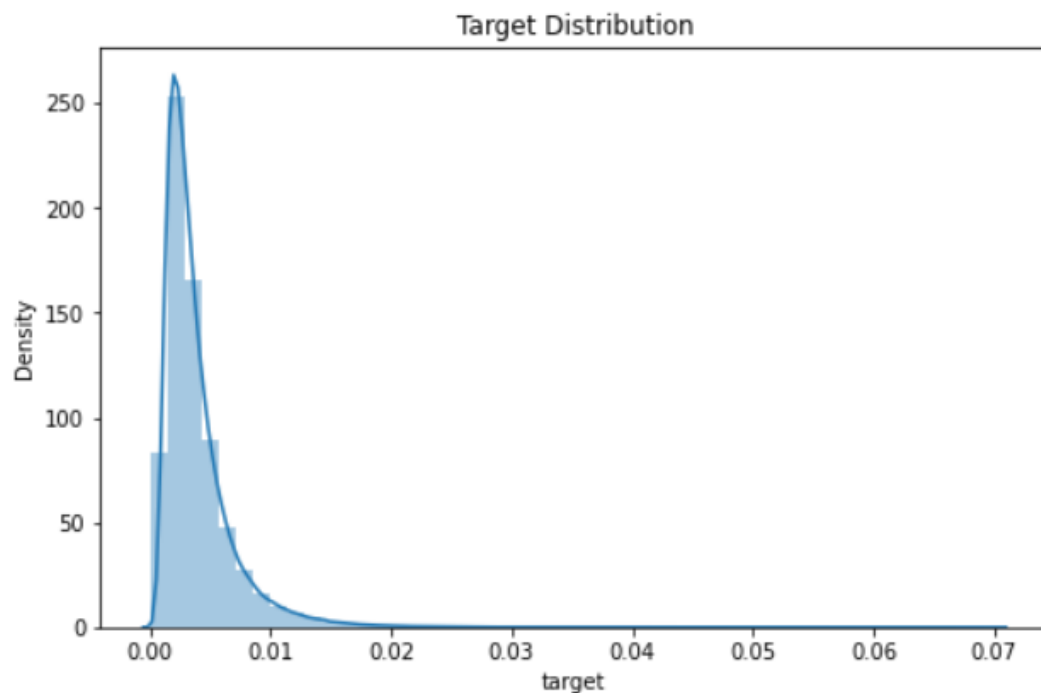
**Date of submission: 01/03/2022**

**Group No.4**

**Optimizers**

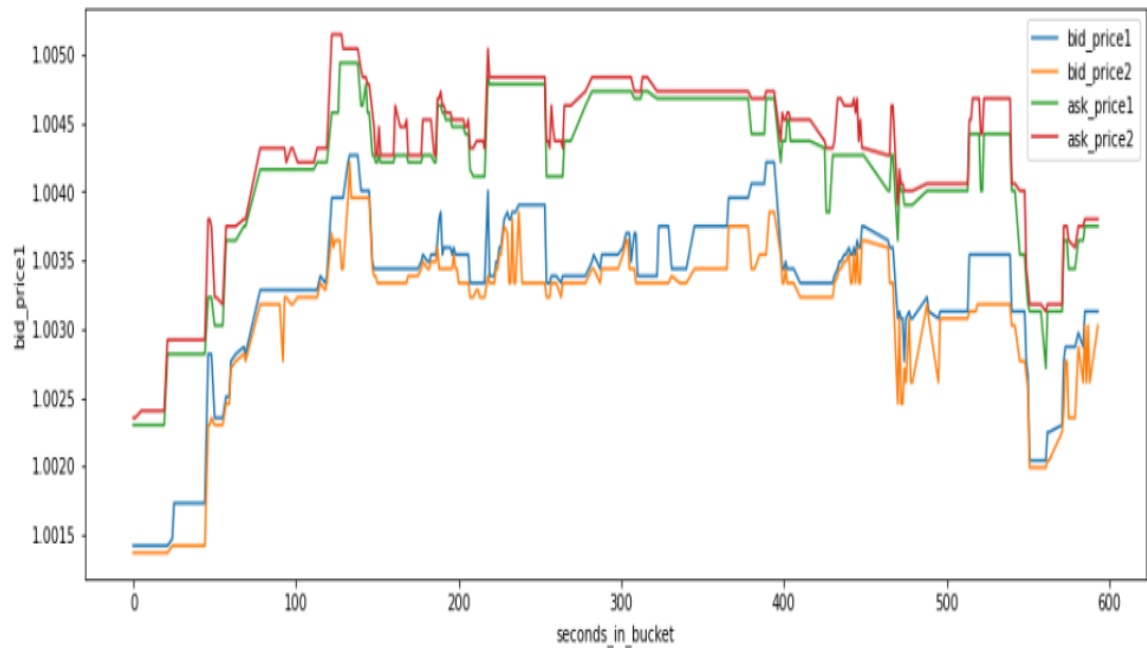
Name	Enrolment Number
Digvijaysinh Gohil	AU1940199
Satya Shah	AU1940288
Smit Shah	AU1940291
Tirth Kanani	AU1920144

We performed Exploratory data analysis further to get more inferences from the data.



We plotted density histogram for our target data. As we know target is the realized volatility which we have to predict, hence having clear idea of its distribution is very important. As seen in the graph above the target is highly skewed to the right and there are some outliers in data as it should be because some of the stocks can be more volatile than the other. Further we also have intuition that the same stock can be more volatile than the other in a particular time bucket, however we were not able to represent this visually. We will try again next week.

There are approximately 112 stock ids, 3830-time ids and 414278 target which are having unique values in train data.



- Time\_id – ID code for the time bucket. Time ID s are not necessarily sequential but are consistent across all stocks.
- Seconds\_in\_bucket- Number of seconds from the start of the bucket. It starts from 0
- Bid\_price- This is the highest price a buyer will pay to buy number of shares of stocks.
- Ask\_price- This is the lowest price a seller will sell the stock
- *Always the bid price is lower than the ask price.*

The above graph is about price vs seconds\_in\_bucket for 600 seconds. We took the data for time\_id=5 of the train data. As we can see that the bid prices increased for the first 100 seconds then it remained constant and again decreased at 550 seconds. The same goes with the ask price.

For next week:

We will be trying model building next week but from our initial observations it looks like linear model's won't work very accurately. Nonetheless we will try them out next week and note down the observations. Further we have been looking at the ARIMA model which seems to be the ideal model to go around with the time series problems.

Reference:

<https://ieeexplore.ieee.org/document/8626097>

<https://towardsdatascience.com/time-series-forecasting-predicting-stock-prices-using-an-arima-model-2e3b3080bd70>