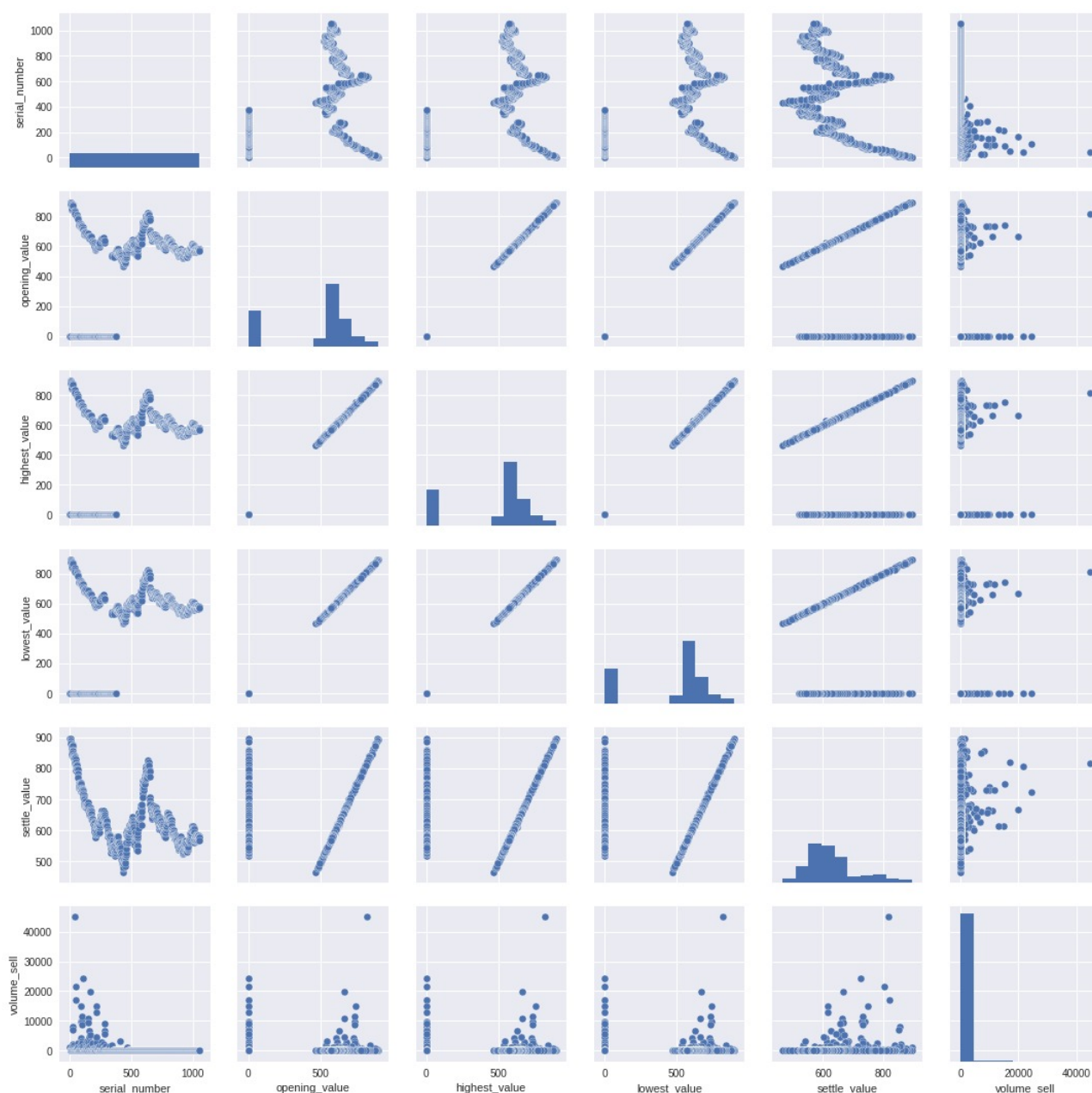


After observing the data, we tried different models i.e. Time series, Xgboost, LSTM (neural network). We found LSTM (Long Short Term Memory) model to fit the best. These are steps which we followed to implement LSTM model:

- The training data and testing data are imported using pandas library.
- Pair plots are plotted for the given training data to find outliers. We observed the variation of an entity with the other. It is found that Row no. 42 is the outlier.



- Then feature engineering is performed. In this, the data is converted into 1 column data by taking average of opening\_value, highest\_value, lowest\_value and settle\_value. This is the approach used in prediction of stock prices.
- All values have been normalised between 0 and 1.
- Training data is splitted in 3:1 ratio. The model is trained using 75% of data and is tested on remaining 25% of data.
- The input of the model is average of OHLC and the output is volume\_sell.
- Two sequential LSTM layers have been stacked together and one dense layer is used to build the model using keras deep learning library. Since this is a regression task, 'linear' activation has been used in final layer.
- Then validation is done on validation data i.e. 25% of training data and RMSE is calculated.
- The output is predicted for the given testing data.

### Why LSTM?

Since our data is time series data we need to remember the trends in previous data, that is exactly what an LSTM is used for. This helps retain important insights from the previous time steps and predictions is more accurate than using normal statistics.

