

PRML Bonus Project Bitcoin Price Prediction

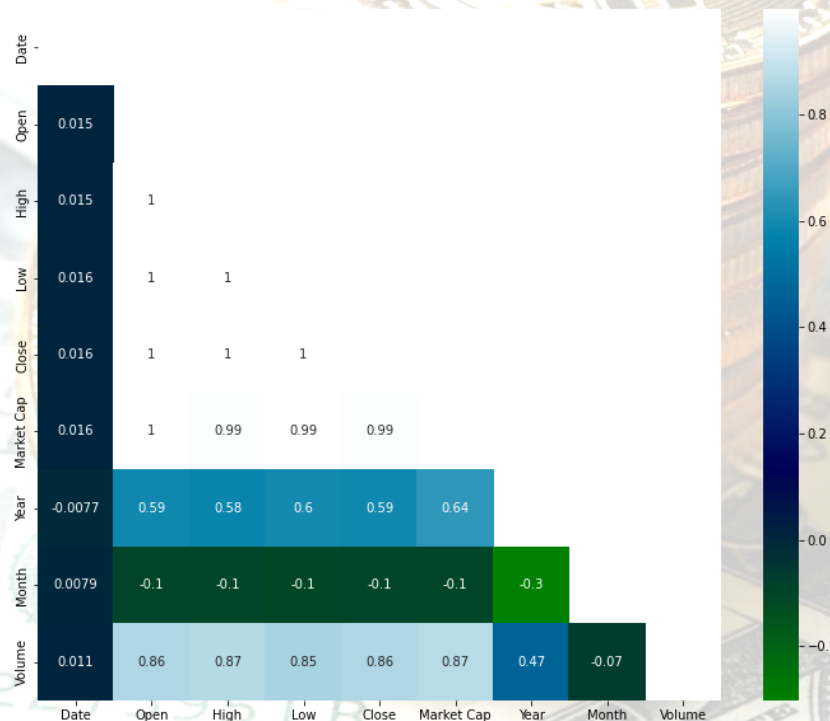
Pratul Singh (B20CS095)

Preprocessing:-

- Made new columns for date, month and Year.
- Removed commas from the string values in Volume and Market Cap and then converted them to int data type.
- Checked for missing values and filled the missing values in Volume column.
- Used the rows which had volume values and then trained Random forest classifier on them To predict the missing values.

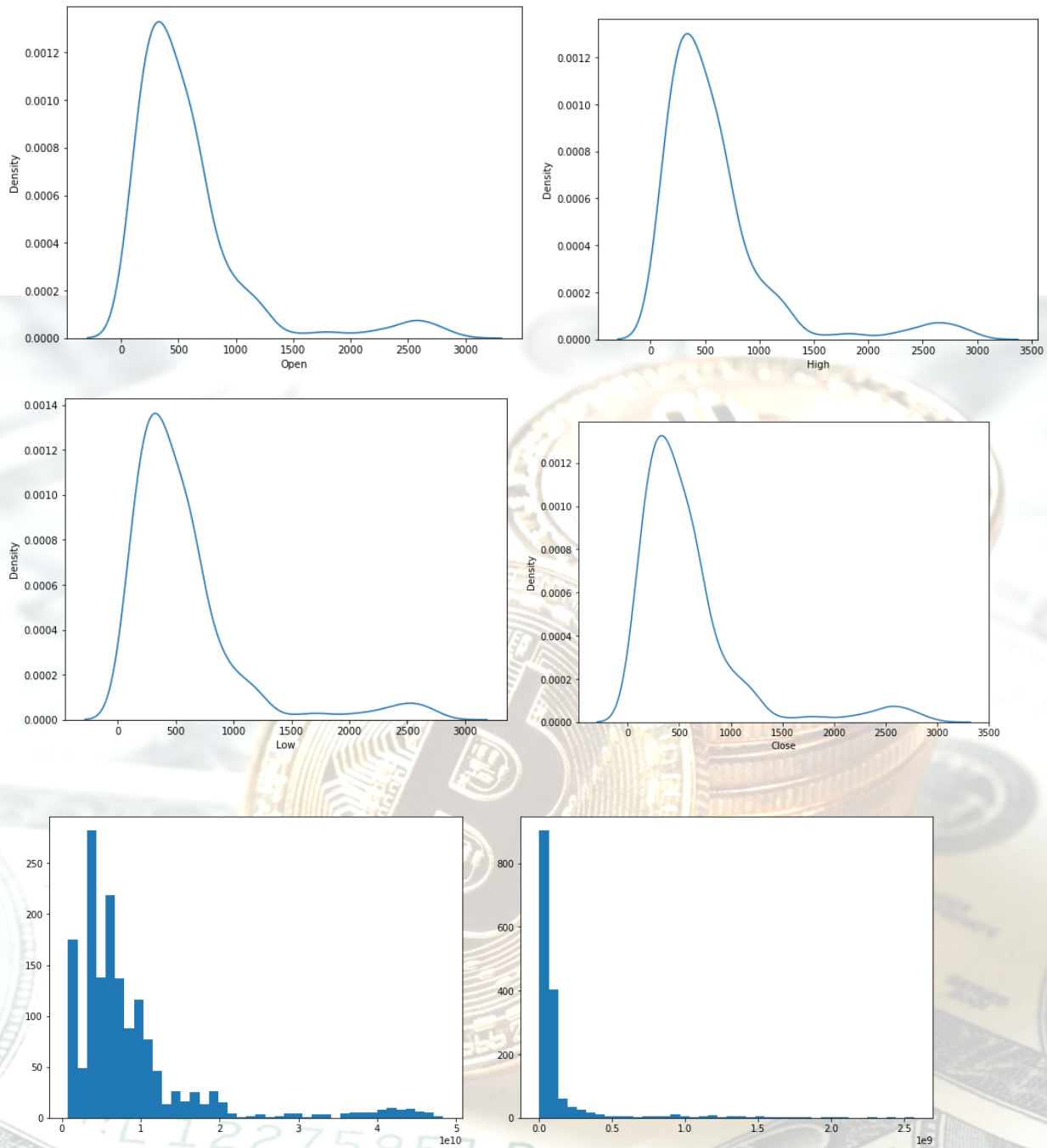
Exploratory Data Analysis:-

- Plotted heatmap to find correlation between the columns.



We can see that the columns, except the date columns, are highly correlated. Since the data is continuous and columns are highly correlated, we are going to use regression models.

- Then I plotted the distribution graphs for the columns.



Modeling:

- ☐ Split the data into train and test and then train the train data on the following models:
 - Decision tree Regressor
 - AdaBoostRegressor
 - Support Vector Regressor
 - Linear Regression
 - LGBM Regressor
 - KNN Regressor

- ☐ Without hyperparameter tuning, following r2 scores came:

Model Name	R2 score
Decision Tree Regressor	0.994245502649019
AdaboostRegressor	0.9894969181440969
SVR	0.21988519252738392
Linear Regression	0.9981253907791198
LGBM Regressor	0.9961591936587169
KNeighborsRegressor	0.9859725192424257

Hyper Parameterization:

The metrics of the base models show that they require hyper tweaking to achieve better performance (especially SVR). The only exception is Sklearn's Linear Regression model, which already has a very high R2 score in comparison to the other models. Linear Regression cannot be hypertuned because it lacks any configurable parameters for the relevant problems.

- ☐ Tuned the hyperparameters of all the models.
- ☐ The following r2 scores came after tuning the models.

Model Name	R2 score
Decision Tree Regressor	0.9944293156003446
AdaboostRegressor	0.9901297888496786
SVR	0.9961591936587169
Linear Regression	0.9981253907791198
LGBM Regressor	0.9971205465469176
KNeighborsRegressor	0.9881876588054498

Best Model:

- The best model, as we can see from the above table, is Linear Regression. It was also the best expected model as we can see the columns were linearly correlated.

- Graph for the prediction and the actual y_{test} when using Linear Regression.

