Bitcoin Price Prediction

Swara Patel
B20AI044

I. INTRODUCTION

Cryptocurrency is a type of digital or virtual currency that is used in financial systems. It is protected by cryptography, which prevents counterfeiting or double-spending. Cryptocurrencies are not issued by a central authority or central banks and can be converted through cryptographic procedures. There are hundreds of cryptocurrencies in digital marketplaces, but Bitcoin is the most popular and is affected and interacted with by external forces such as news, social media, and minor cryptocurrencies with a limited market share, which is frequently overlooked by investors and traders. Because of price volatility and dynamism, predicting cryptocurrency values is challenging. This report aims to showcase the performance of different machine learning algorithms i.e. Linear regression, Support Vector Regression, and Long Short Term Memory RNN in predicting future bitcoin prices.

Dataset
The file bitcoin.csv is used as the training dataset.
Number of samples: 1556
Number of features: 7
Bitcoin Prices are from April 2013 to July 2017

II. METHODOLOGY

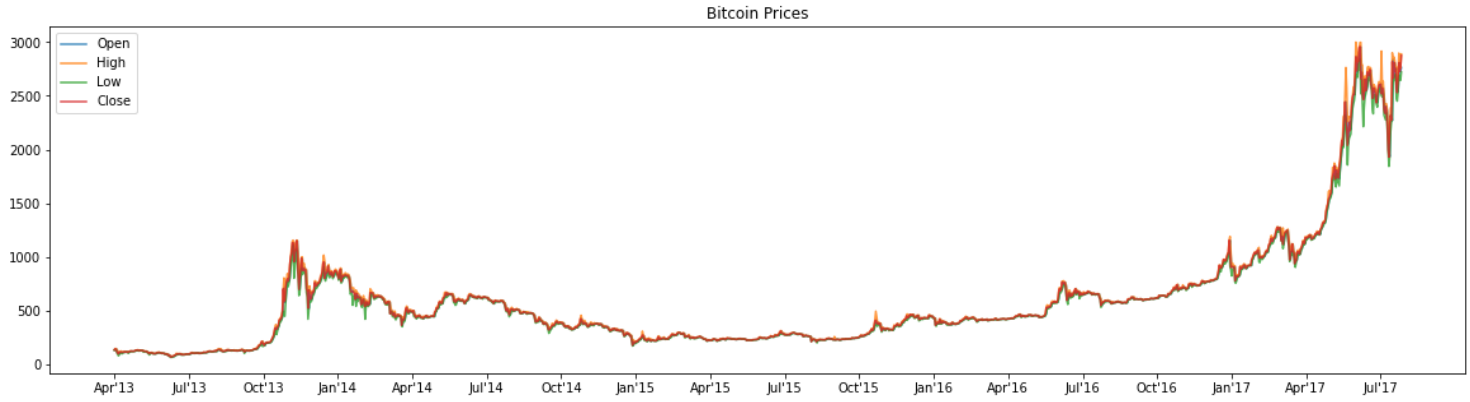
Overview

There are various classification algorithms present out of which we shall implement the following

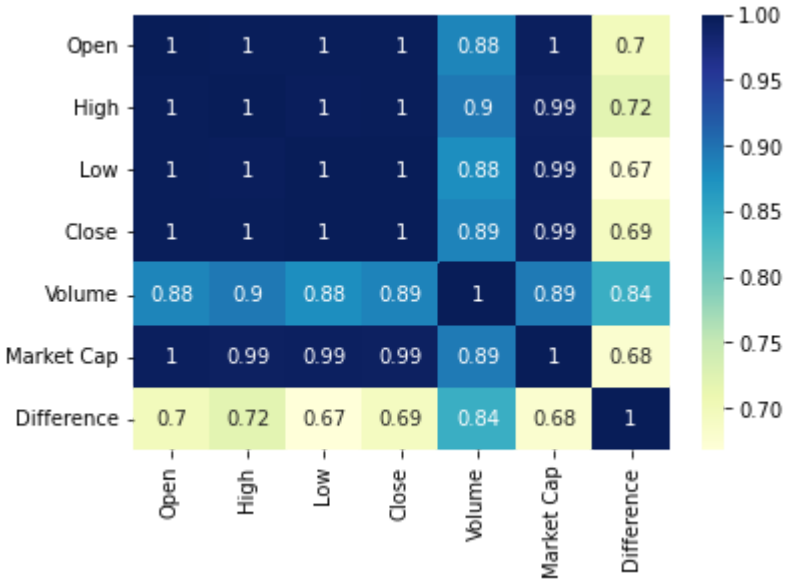
- *Linear Regression*
- *KNN*
- *Multilayer Perceptron*
- *Support Vector Machine*
- *Random Forest*
- *LightGBM*
- *LSTM*

Exploring the dataset and pre-processing

On counting the number of NULL values in the train dataset, it was found that there are 243 NULL values present which is around 15% of the samples, dropping the rows having null values would reault in loss of information, hence null values were interpolated using the ‘forward’ method. Next, converted the column containing Date string to Index column. Also created a column Difference, that contains difference between High and Low value of the sample. Then visualized the Open, Close, High and Low prices.

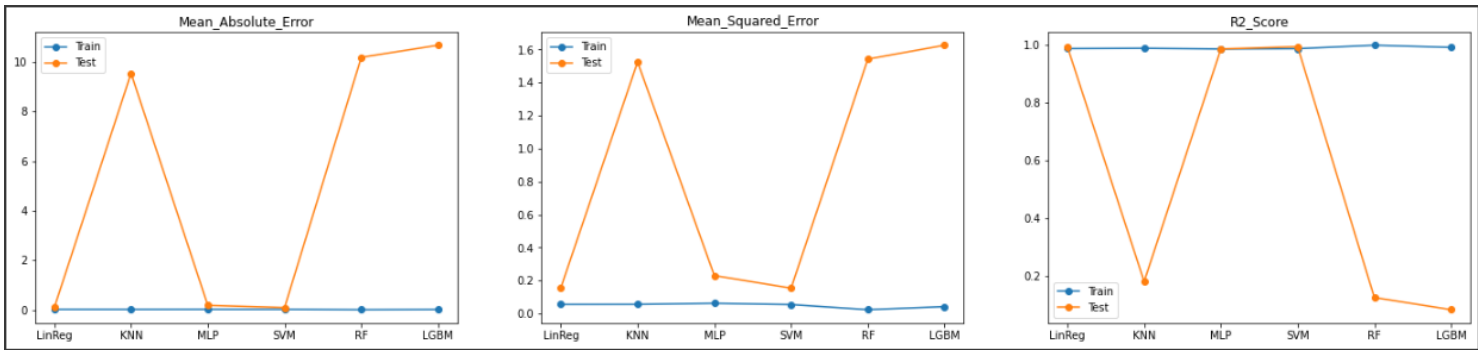


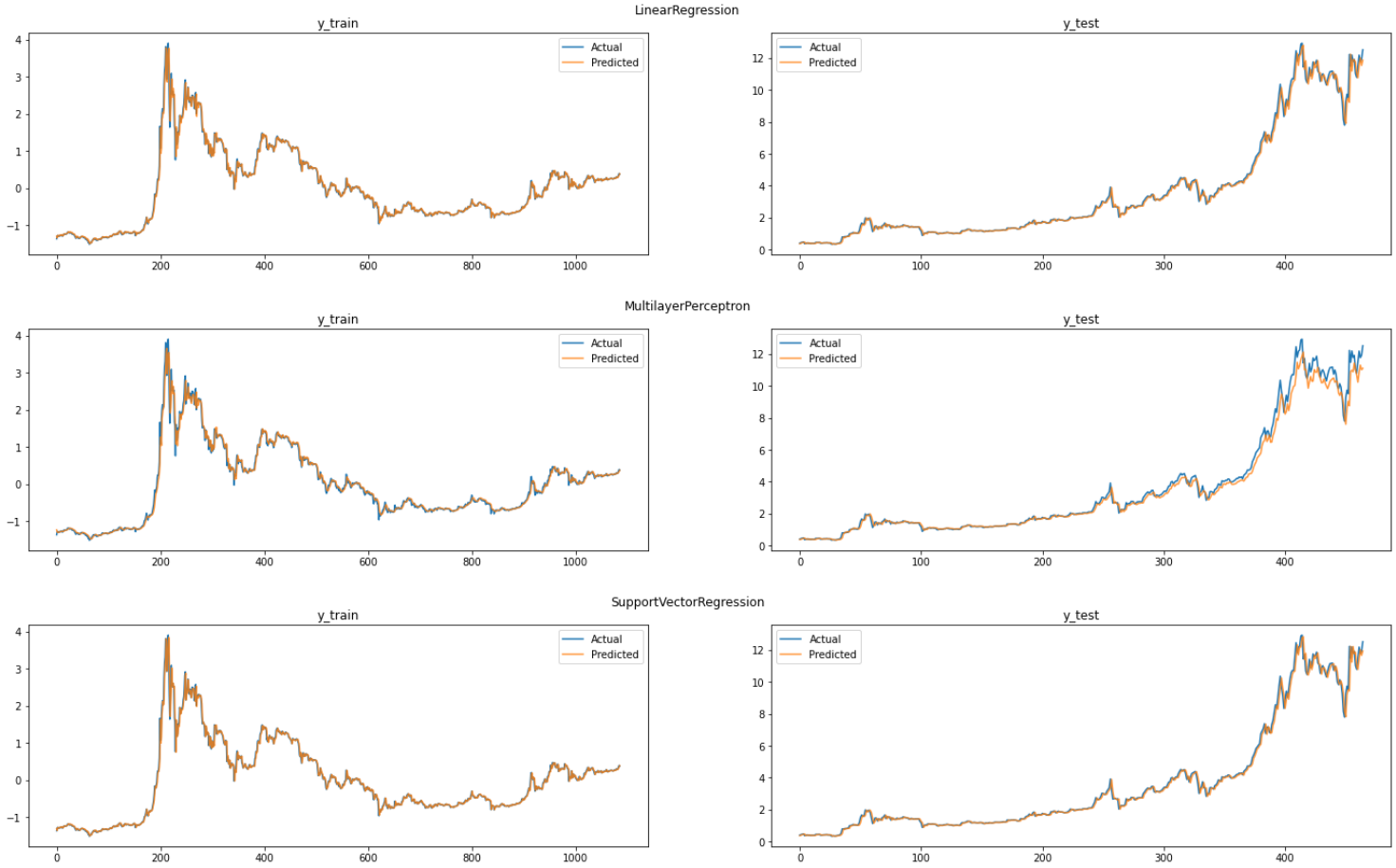
Next, plotted the covariance matrix heatmap to check correlation among all the feature columns



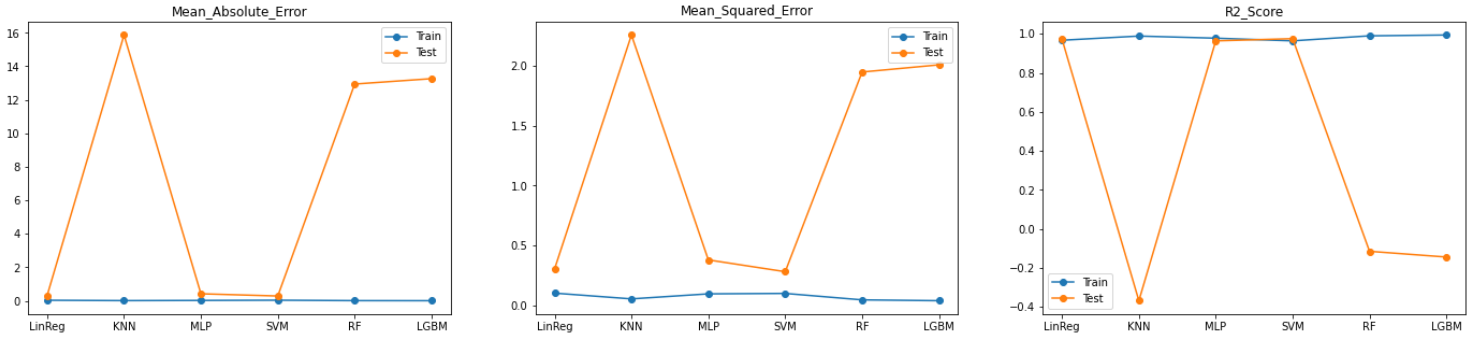
Transforming the dataset and Implementation of different algorithms

1. Created a dataframe where Close price of 5 days is given as input and Close price of 6th day is to be predicted.

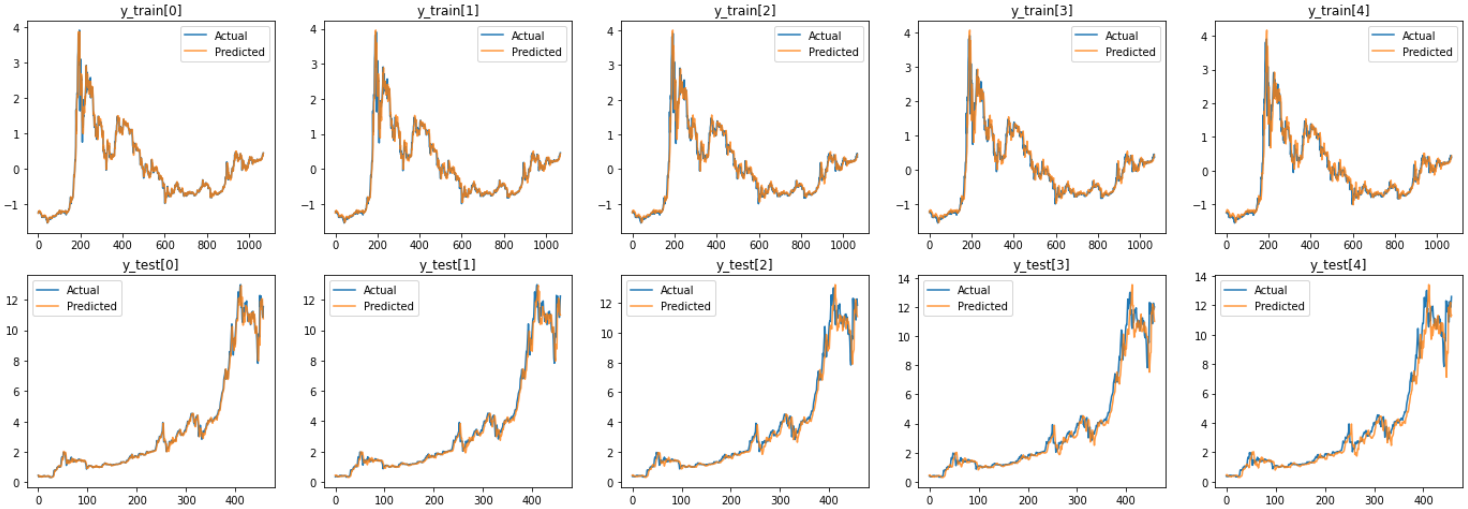




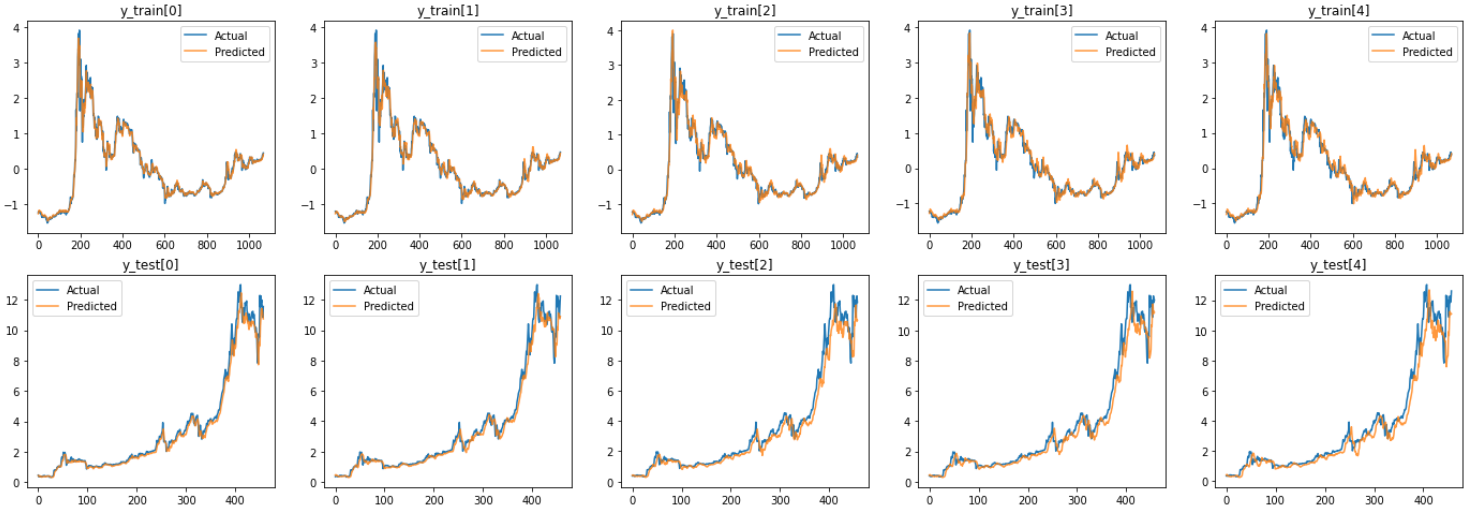
2. Created a dataframe Close Price of 25 previous days is given as input and Close price next 5 days is to be predicted.



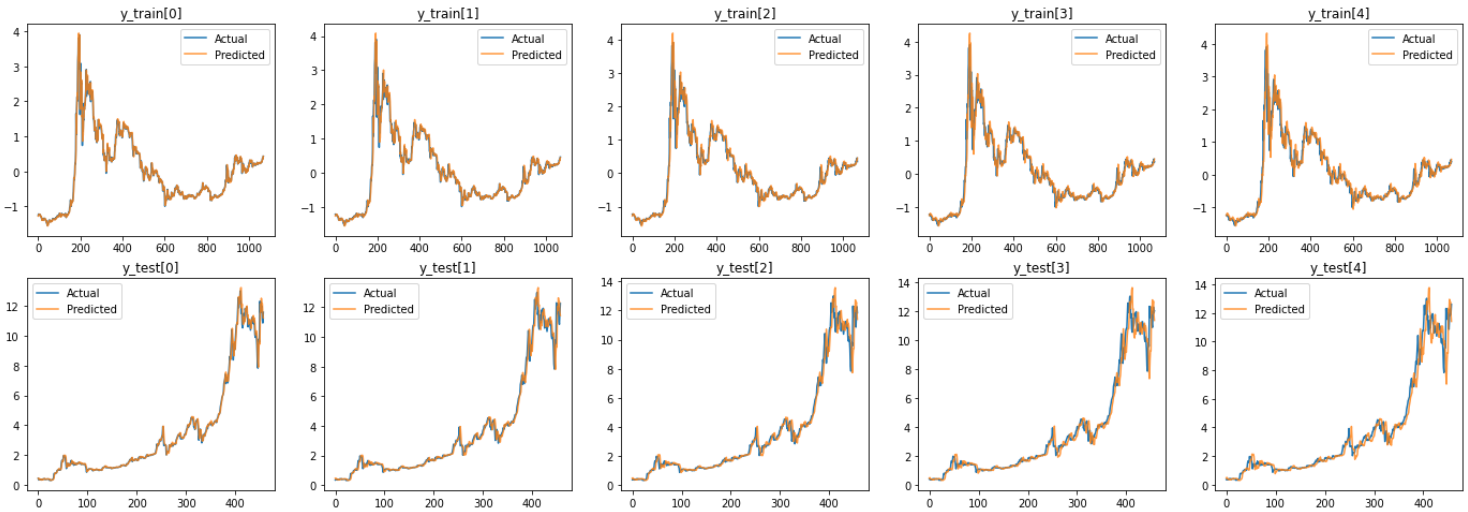
LinearRegression



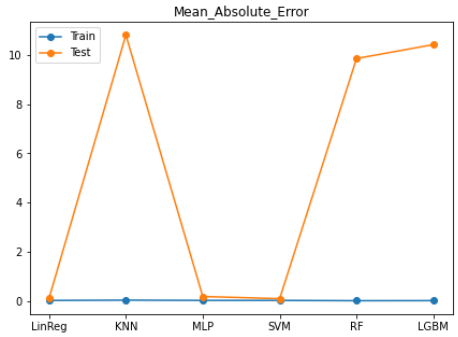
MultilayerPerceptron

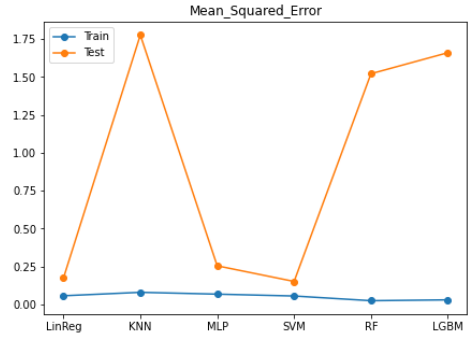


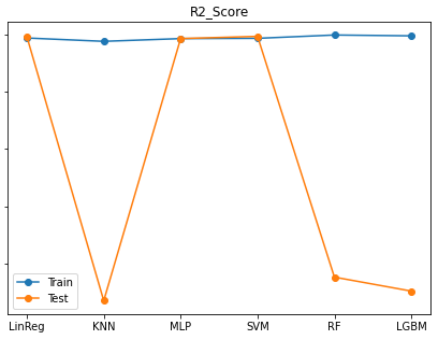
SupportVectorRegression

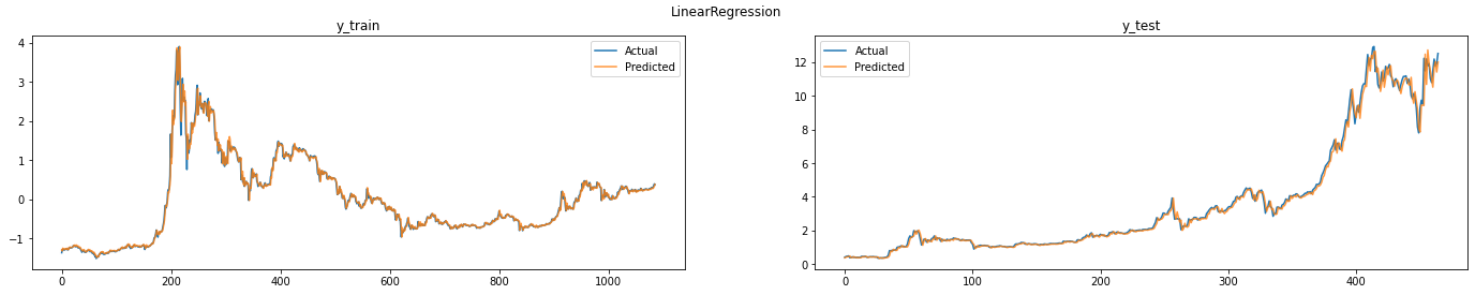


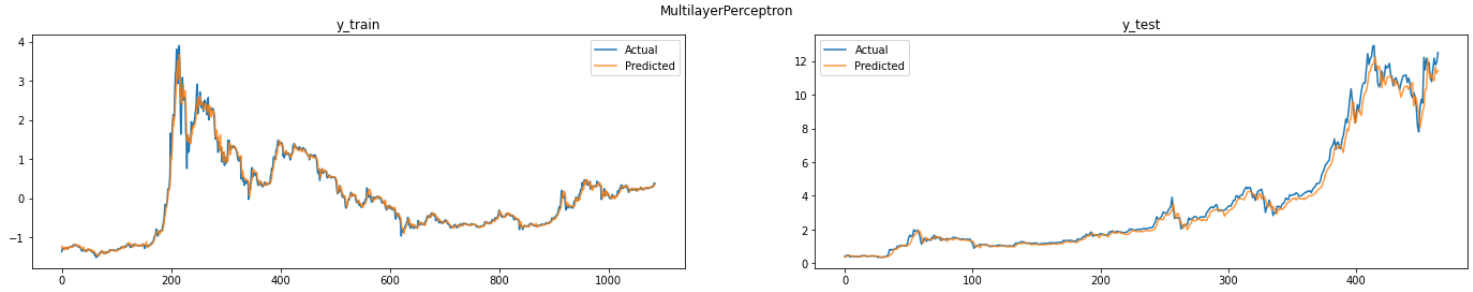
3. Created a dataframe where Close price of 5 days along with Difference (High - Low) is given as input and Close price of 6th day is to be predicted

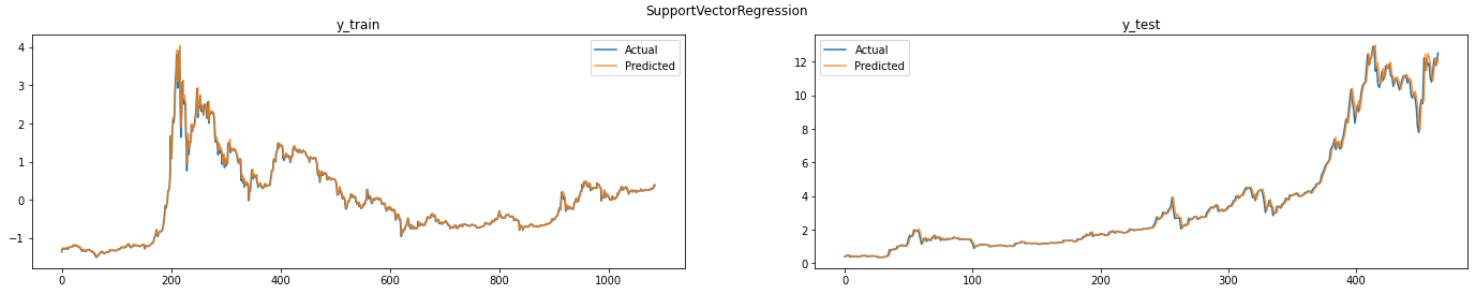












4. LSTM

   Trained an LSTM model on Transformation 1 with following hyperparameters:

   Epochs: 100

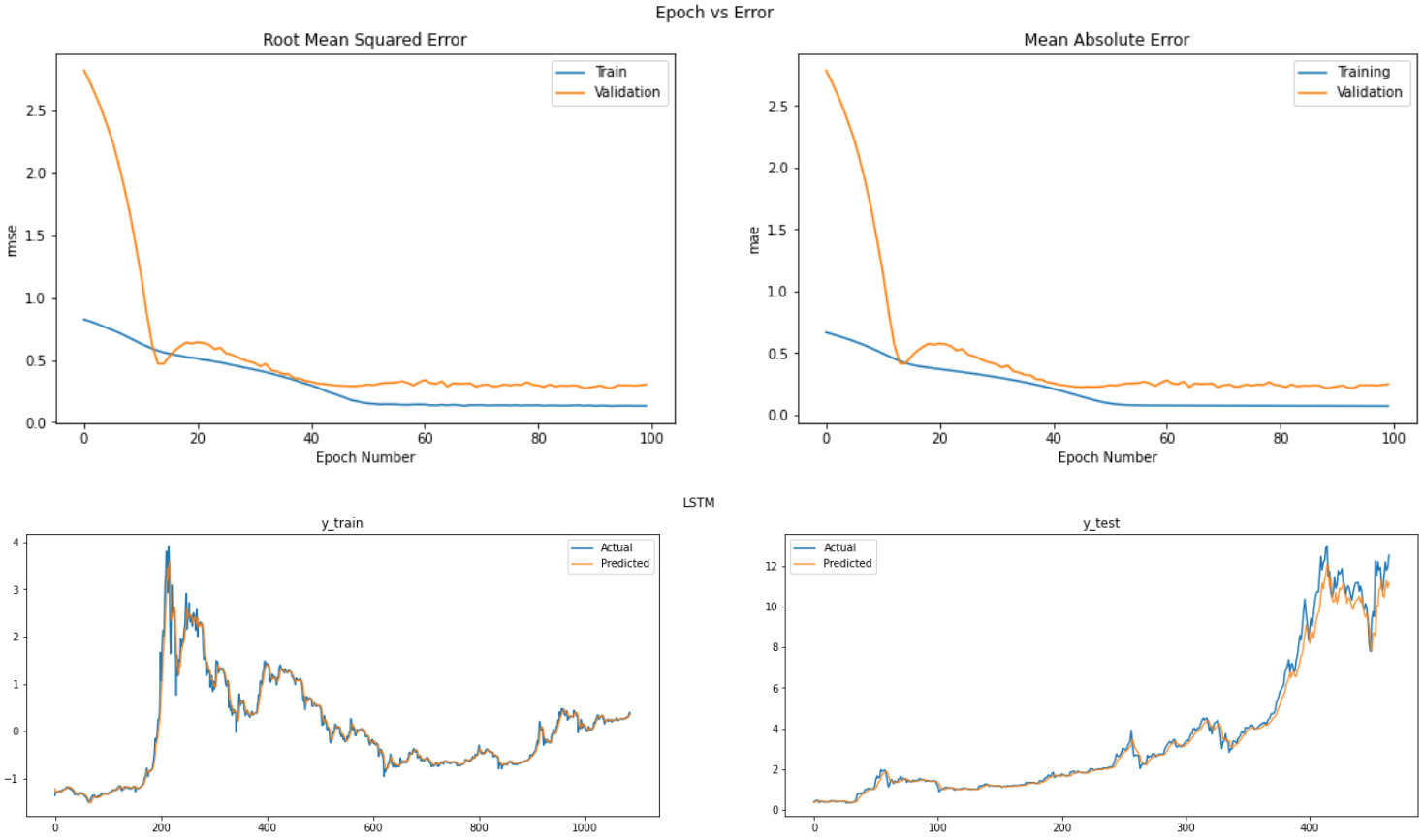
   Batch size: 32

   Optimizer: Adam

   Loss: Root Mean Squared Error

   Metrics: Mean Absolute Error

   Activation: ReLu



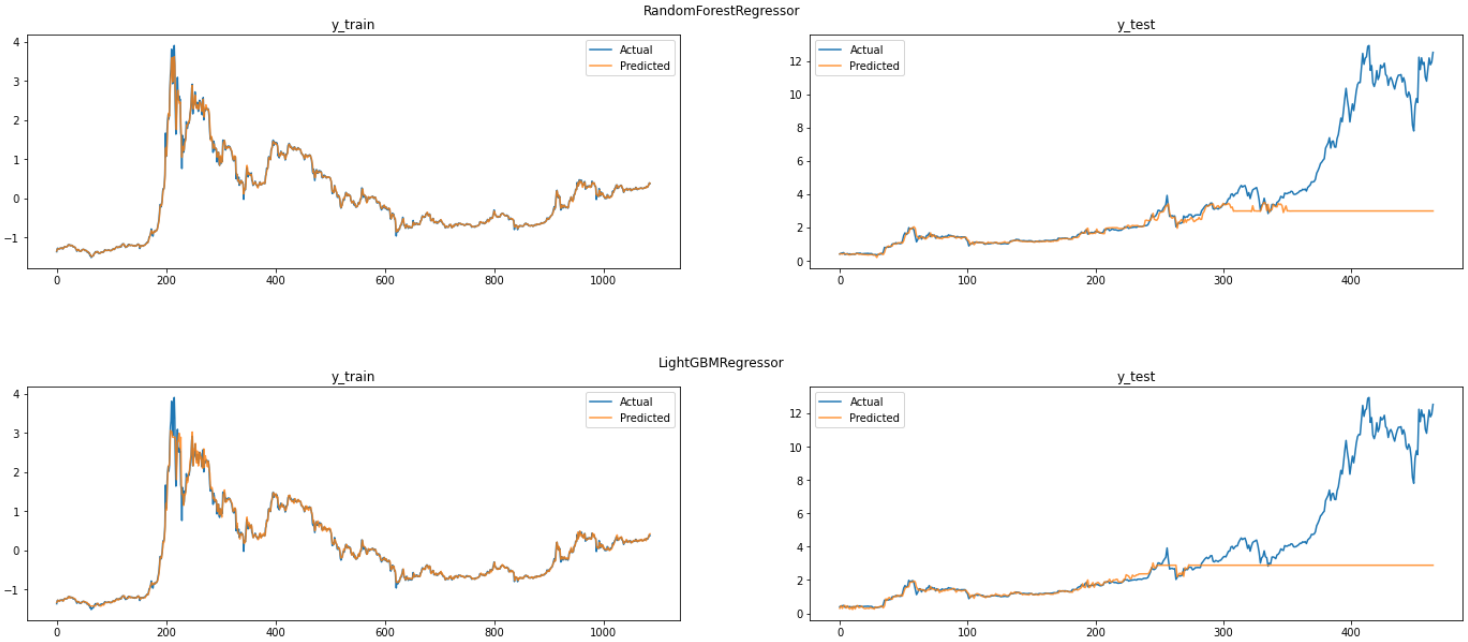
III. EVALUATION OF MODELS

Following table contains evaluation metrics of different models when applied on Transformation 1

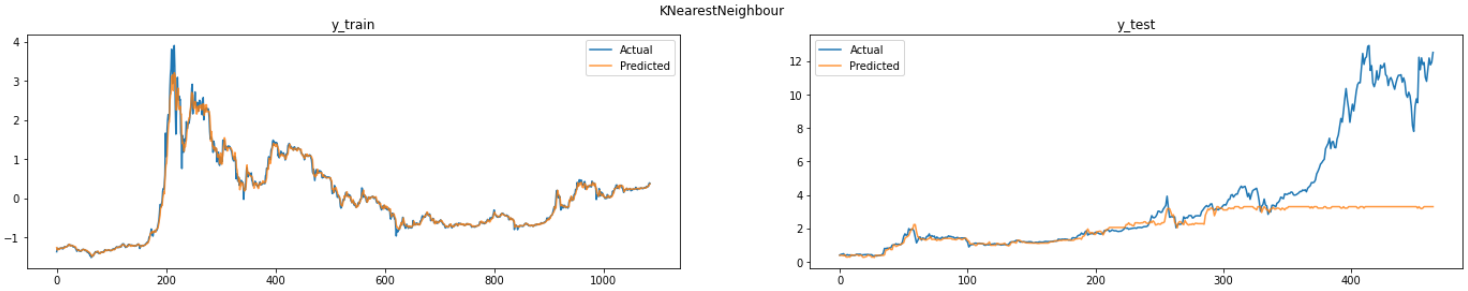
	Train			Test		
	MSE	MAE	R2	MSE	MAE	R2
Linear Regression	0.013	0.055	0.986	0.087	0.156	0.992
KNN	0.012	0.056	0.987	9.541	1.525	0.180
MLP	0.016	0.063	0.983	0.231	0.263	0.980
SVM	0.014	0.054	0.985	0.082	0.146	0.993
Random Forest	0.002	0.024	0.997	10.348	1.558	0.111
LightGBM	0.009	0.041	0.990	10.674	1.626	0.083
LSTM	0.020	0.069	0.979	0.209	0.246	0.982

Linear Regression, MLP, SVM & LSTM - are the models that work great for bitcoin price prediction.

Random Forest and LightGBM tend to overfit greatly since they are ensemble models.



And KNN, the algorithm of selecting nearest neighbors and predicting value according to it, won't work for bitcoin price prediction, since it tends to average out the high fluctuations in the graph



IV. RESULTS AND ANALYSIS

The table shows that Linear Regression, MLP, SVM and LSTM performed equally good. SVM and Linear Regression outperformed the other models with R2_Score of 0.993 and 0.992 respectively.