CS-524
Creating a BitCoin Price Prediction Model Using AWS SageMaker
Final Report - Team 5

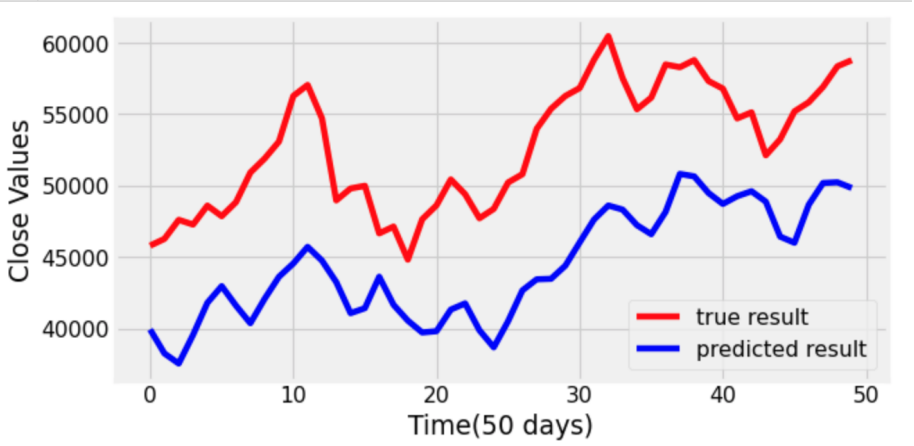
David DeLaus, Abhishek Desai, Vishal Kuchadi
Jayant Kumar, Jiahan Liu, Hardi Patel, Malika Thakur

Objectives and Results

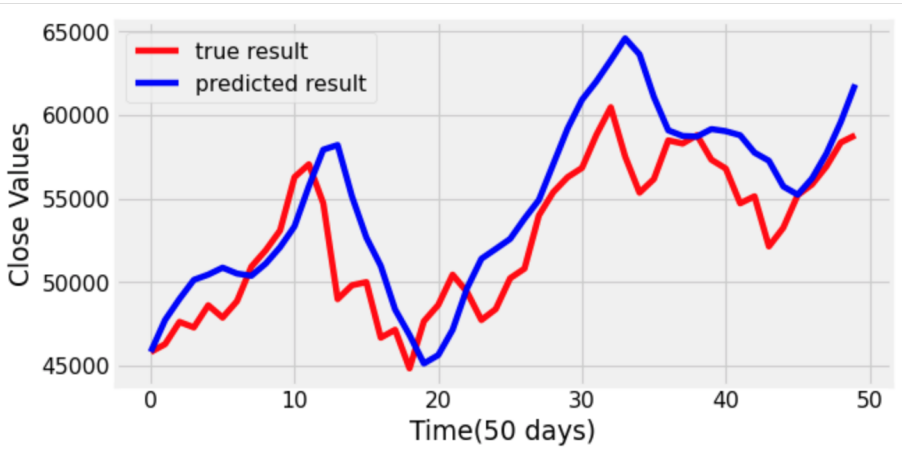
The objective of the project is to build a machine learning model that can accurately predict the price of Bitcoin over a period of time using AWS services. AWS Sagemaker has been an incredible arrangement for our predictive analysis algorithm to achieve a really start to finish ML arrangement. It dealt with abstracting a huge load of programming improvement abilities important to achieve the errand while being exceptionally powerful and adaptable and practical. Above all, it helped us focus on the core Deep Learning experiments and supplements the remaining necessary skills with easy abstracted tools similar to our existing workflow.

We have used AWS S3 bucket for storage of data, AWS IAM for encryption, AWS sagemaker for creating the project and using the Deep learning libraries, AWS sagemaker Jupyter notebook instance for creating Python files and training our model with variousDeep learning algorithms. LSTM and a simple RNN. We found that the LSTM model was better suited for this task and gave us more accurate results as opposed to the RNN which had issues giving us accurate price predictions. LSTM is more suited for this type of modeling due to it having a larger control capacity and keeping vital data to the prediction modeling in memory. We have created a viable working machine learning model that can calculate the close time price of bitcoin. The graphs below show the model results.

Simple RNN:



LSTM:



Methods and Tools

For this project we used three main pieces of technology, AWS S3 Bucket, AWS SageMaker, and JupyterLab.

AWS S3 Bucket:
The S3 Bucket was used to allow us to access the data to train our algorithms from the SageMaker instance. It was vital to have a fast and easily accessible way to store the training and testing data in order to make proper use of it and the S3 Bucket perfectly satisfied this role.

AWS SageMaker:
SageMaker is Amazon's premier machine learning service. It allows users to easily implement and train machine learning algorithms using their simple ui and powerful computers. We decided that SageMaker was a perfect fit due to the computing power and its ease of use.

JupyterLab:
JupyterLab is a service that allows you to compile and run python code line by line, it is very commonly used in the field of machine learning. It allowed us to easily implement and test our machine learning algorithms as we could just upload the file to SageMaker as well as easily analyze our data through visualization.

The project was broken down into 3 main stages, AWS setup, Data Analysis and Cleaning, and Algorithm implementation.

AWS setup:

1. First we needed to setup our S3 bucket with default settings
    1.1. Create a bucket in the US east as to have less latency
2. Upload our training and testing data to the bucket in csv format
3. Configure our bucket's access list so SageMaker has permission to access it
4. Create our SageMaker instance
    4.1. Create a notebook instance so we can utilize JupyterLab notebooks
    4.2. Give the "notebook instance name", "notebook instance type" , enable "IAM role" and keep the root access "enable"

Data Analysis and Cleaning:

1. We opened the SageMaker Jupyter notebook
2. Import the data from the S3 bucket as .csv to a pandas dataframe
3. We then examined the data and either deleted the blank entries or filled them in depending on what was needed
4. We then created different visualizations for all the different data feature correlations to examine what we should use in our training data.
5. We determined that all the data features could be used to train and test the model

Algorithm Implementation:

1. We implemented LSTM and simple RNN modules to create working models with these two algorithms
2. We conclude LSTM predicts better than simple RNN

Individual Contributions

Documentation:

David DeLaus:
As the appointed team leader, I helped to organize meeting times and worked on creating and formatting a formal project report. I worked on creating the structure of a short report and will be working on structure of presentation preparation for the final demo.

S3 bucket and Data Extraction:

Jiahan Liu: S3 bucket understanding, setting permissions, ACL list permissions, extracting and storing data.

Data Cleaning and Visualization:

Abhishek Desai: Worked on fetching data from the cloud and data cleaning on a python jupyter notebook by removing the redundant, NAN data values.

Hardi Patel: Worked on data visualization, plotted various curves for each feature w.r.t price prediction for better understanding of data values and added the same in the report.

AWS sagemaker and Predictive algorithms:

Vishal Kuchadi: Understanding of Bitcoin concepts and sharing knowledge with the team,Studied the Bitcoin Kaggle historical dataset and decided the target column for prediction. Also worked on implementing the deep learning algorithms and writing the report.

Jayant Kumar: worked on AWS sagemaker instance and permissions and learning and understanding **RNN** deep learning library, hypertuning and implementing the same on sagemaker Jupyter notebook and adding screens to the report for the Predictive models.

Malika Thakur: pitching sagemaker idea, worked on AWS sagemaker instances and IAM roles and learning and understanding **LSTM** deep learning library, hypertuning and implemented the same on sagemaker jupyter notebook by creating a python file for training our model and adding results to the report.