BITCOIN PRICE PREDICTOR - BitPredict

A PROJECT REPORT

Submitted by

Rohan Rupela (18CP017)
Harsh Upadhyay (18CP027)
Ankur Patel (18CP066)

In partial fulfilment for the award of the degree of
B. TECH. (COMPUTER ENGINEERING)

4CP31: PROJECT-I

BIRLA VISHVAKARMA MAHAVIDYALAYA
(ENGINEERING COLLEGE) *(An Autonomous Institution)*
VALLABH VIDYANAGAR

Affiliated to

GUJARAT TECHNOLOGICAL UNIVERSITY, AHMEDABAD

Academic Year: 2021 – 2022

B. V. M. ENGINEERING COLLEGE, VALLABH VIDYANAGAR-388120

APPROVAL SHEET

The project work entitled "**Bitcoin Price Predictor - BitPredict**" carried out by "**Rohan Rupela (18CP017), Harsh Upadhyay (18CP027) and Ankur Patel (18CP066)**" is approved for the submission in the course 4CP31, Project-I for the partial fulfillment for the award of the degree of B. Tech. (Computer Engineering).

Date:

Place:

EXAMINER SIGNATURE

(Name & Designation)

CERTIFICATE

This is to certify that Project Work embodied in this project report titled "**Bitcoin Price Predictor - BitPredict**" was carried out by "**Rohan Rupela (18CP017), Harsh Upadhyay (18CP027) and Ankur Patel (18CP066)**" under the course 4**CP31, Project-I** for the partial fulfillment for the award of the degree of B. Tech. (Computer Engineering). Followings are the supervisors at the institute.

Date:

Place:

Prof. Prashant Swadas	**Prof. Dipa Soni**	**Dr. Darshak G Thakore**
Associate Professor	**Trainee Assistant Professor**	**Prof. (Dr.) & HOD**
Computer Engineering	**Computer Engineering**	**Computer Engineering**

DEPARTMENT OF COMPUTER ENGINEERING, B. V. M. ENGINEERING COLLEGE, VALLABH VIDYANAGAR-388120

DECLARATION OF ORIGINALITY

We hereby certify that we are the sole authors of this report under the course 4CP31: Project-I and that neither any part of this report nor the whole of the report has been submitted for a degree to any other University or Institution.

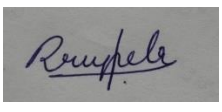
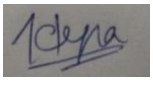
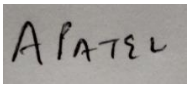
We certify that, to the best of our knowledge, the current report does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations or any other material from the work of other people included in our report, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that we have included copyrighted material that surpasses the boundary of fair dealing within the meaning of the Indian Copyright (Amendment) Act 2012, we certify that we have obtained a written permission from the copyright owner(s) to include such material(s) in the current report and have included copies of such copyright clearances to our appendix.

We declare that this is a true copy of report, including any final revisions, as approved by report review committee.

We have checked write up of the present report using anti-plagiarism database and it is in allowable limit. Even though later on in case of any complaint pertaining of plagiarism, we are sole responsible for the same and we understand that as per UGC norms, University can even revoke the degree conferred to the student submitting this report.

Date:

Institute code: 007

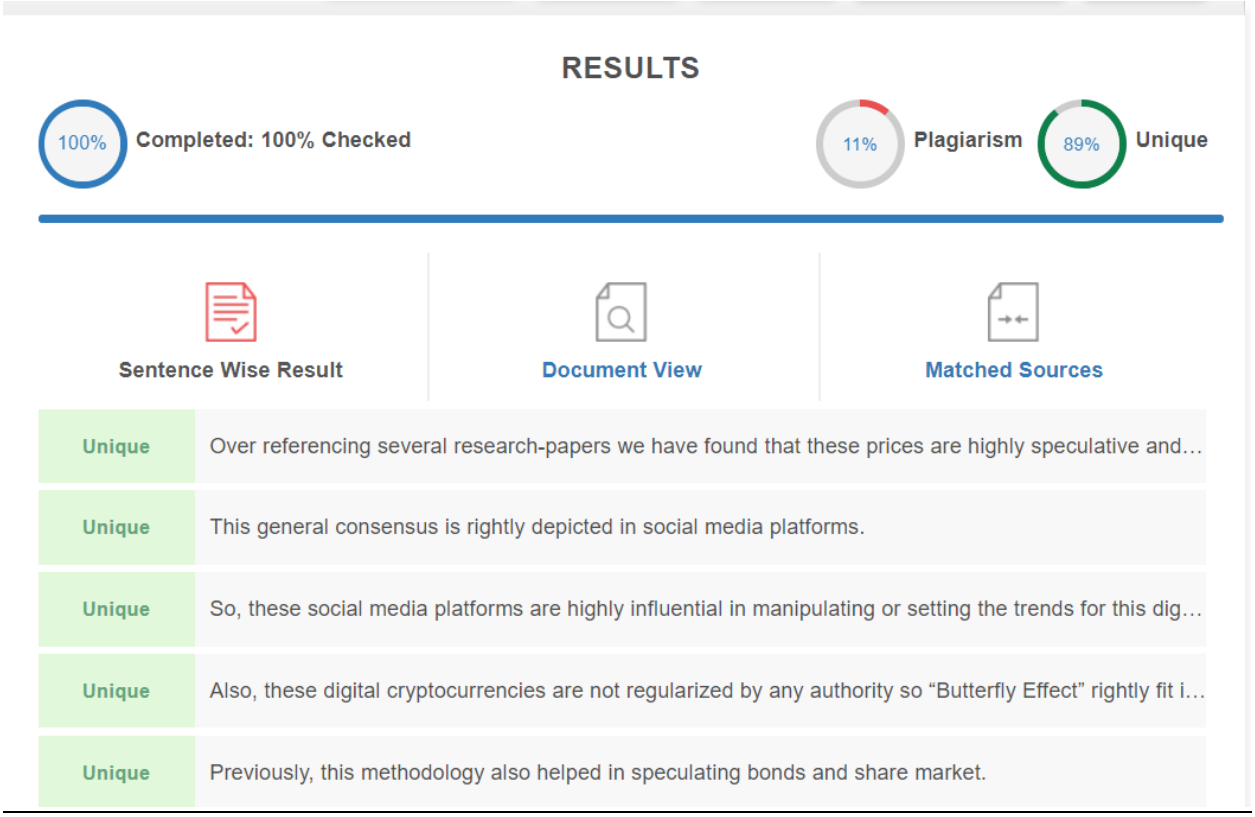
Rohan Rupela	**Harsh Upadhyay**	**Ankur Patel**
18CP017	**18CP027**	**18CP066**
		

ACKNOWLEDGEMENT

We would like to take this opportunity to thank all the persons who have genuinely helped us in any and every way to complete this project. Making use of technology, we have prepared Bitcoin Price Predictor. It is our sincere thanks to all enlisted people as well as not enlisted people, whose help and continuous inspiration leads to get these in a pleasant way, however it will be tough to thank them enough. I express my sincere thanks to the Head of Department of Computer Department Dr. Darshak G Thakore to allow us to do half semester internal project which really help us to learn and develop our skills and also thanks to my Internal Guides Prof. Prashant Swadas and Prof. Dipa Soni for helping me to do project successfully by not breaking any rules of our college and every help to solve any problem related to my project.

We hereby would like to take this opportunity to thank all persons who has involved generously in helping me and assisting me. We would like to convey our deepest gratitude towards our supervisor, Prof. Prashant Swadas and Prof. Dipa Soni for his support and supervision, and for the valuable knowledge that he shared with us.

Plagiarism Report



Abstract

Bitcoin Price Predictor - BitPredict

Over referencing several research-papers we have found that these prices are highly speculative and built by general consensus in the mind of people. This general consensus is rightly depicted in social media platforms. So, these social media platforms are highly influential in manipulating or setting the trends for this digital currency. Also, these digital cryptocurrencies are not regularized by any authority so "**Butterfly Effect**" rightly fit in the context of the cryptocurrencies. Previously, this methodology also helped in speculating bonds and share market.

Keeping all this in mind we have tried to implement machine learning technique called **Long Short-Term Memory – LSTM.** Our backend pulls up reddit posts, various cryptocurrency price like Bitcoin, Litecoin, Ethereum, news related to cryptocurrency form trending news platform and after calculating this sentiment of these data updates the dataset. Our model takes recent 15X24 entries from dataset, generates output and puts on **Google Firebase Firestore** database. All these processes are carried out in **Linux Ubuntu** instance in **Microsoft Azure**. We show the results in **Android Application** that we have developed specifically for the purpose. Our android application not only shows the speculative price but also current price of the currency as well as provides other utilities to manage portfolio, Articles related to cryptocurrencies and provides notification if the prices are going highly bullish.

Table of Contents

List of Figures

List of Symbols, Abbreviations and Nomenclature

UI - User Interface
DFD - Data Flow Diagram
API - Application Programming Interface
FSEP- Full Semester External Project
GUI - Graphical User Interface
CRUD - Create, Read, Update, and Delete
LSTM - Long Short-Term Memory
BTC - Bitcoin
ETH - Ethereum
LTC – Litecoin
RMSE – Root Mean Square Error

1. Introduction

Bitcoin has sparked a gigantic interest in cryptocurrency and blockchain technology. Since the inception of Bitcoin, cryptocurrency has gained the trust of the general population. Bitcoin has achieved the highest market capitalization among all of the cryptocurrencies. As of this writing Bitcoin market capitalization is more than 1.13 trillion US dollars.

As we know the opinion of masses matter in determining the impact of a commodity in the society. By conducting Sentiment Analysis, we can predict the future situation of the given commodity. Nowadays Social media marketing has a huge impact on the opinion of younger generations. As a result of butterfly effect opinion of a single person like Elon Musk can impact the overall situation of cryptocurrency.

Now for our app: It conducts sentiment analysis by fetching various posts on social media and along with the past monetary trend of bitcoin price it processes the data and gives us the future price prediction. Along with sentiments, we have included historical price and volume of Litecoin and Ethereum. We have trained machine learning model to learn about the correlation between all these features and predict the future price. We have also provided a section where the user can get the latest crypto news for reading. We have included a portfolio management section for storing your current bitcoin portfolio. We have provided the function of getting notifications of hourly predicted price and also give advice if it is advisable to but bitcoin at current price or not.

1.1 Domain Problem

Cryptocurrency is a monetary unstable commodity, due various variables its price keeps on changing periodically. People require some means to consult-to before buying it.

1.2 Project Objectives

Objectives of our Project is to build an android app where it conducts sentiment analysis by fetching various posts on social media and along with the past monetary trend of Bitcoin price, it processes the data and gives us the future Price Predictions.

2. Literature Survey

Purchase and Trade of Cryptocurrency has become a norm these days. As such general population has been trying to learn as much about cryptocurrencies.

We referred to various research-papers published in highly recognized journals and books and some mobile applications for developing UI/UX purpose. The list goes as:

1. "Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis" by Prasanga Neupane, Dibakar raj Pant presented in 2018 IEEE 3rd ICCCS.

2. "Bitcoin Price Prediction and Analysis Using Deep Learning Models" by Minakhi Rout, Suresh Satapathy from the book Communication Software and Networks, Proceedings of INDIA 2019.

3. "LSTM Based Sentiment Analysis for Cryptocurrency Prediction" by Xin Huang, Wenbin Zhang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liuand Ji Zhang.

4. "Predictive analysis of Bitcoin price considering social sentiments" by Pratikkumar Prajapati.

5. Android/iOS application "Binance" available on Google Play Store and iOS App Store.

Initial analysis was carried out over the large dataset of Reddit comments containing 'bitcoin' from 2009 to 2019 (having 4193519 rows and 9 columns) and historical price dataset containing hourly prices from Binance for bitcoin. Both datasets are available on Kaggle as open source.

	Unnamed: 0	datetime	date	author	subreddit	created_utc	score	controversiality	body
0	0	2014-06-26 16:49:41	2014-06-26	HonorConnor	AskReddit	1.403801e+09	162.0	0.0	How do you feel about Bitcoin? I don't really ...
1	1	2014-05-17 23:30:36	2014-05-17	Talvoren	AskReddit	1.400369e+09	75.0	0.0	[The guy who blew a huge portion of his and hi...
2	2	2014-12-18 04:54:48	2014-12-18	ninjoe87	Outdoors	1.418878e+09	-4.0	0.0	Because you are a damn liar.\n\nFrom my other ...
3	3	2014-04-30 00:24:16	2014-04-30	fathergrigori54	tf2	1.398817e+09	13.0	0.0	This was the bitcoin hat guy that ACTUALLY del...
4	4	2014-07-07 16:06:21	2014-07-07	Draber-Bien	starcraft	1.404749e+09	19.0	0.0	I'm sure this is *good* for bitcoin
...	...	...	...	...	...	...	...	...	...
4193514	4193355	2019-03-16 16:23:45	2019-03-16	Huntred	teslamotors	1.552753e+09	-1.0	0.0	Put 10% of those savings in Bitcoin.
4193515	4193356	2019-03-20 22:14:09	2019-03-20	KalEll66	Bitcoin	1.553120e+09	-1.0	1.0	Litecoin is another shitcoin used for testing ...
4193516	4193357	2019-03-23 21:08:12	2019-03-23	BonelessRedditor	technology	1.553375e+09	-1.0	0.0	Hahahahaha Bitcoin dudes are such dumbasses
4193517	4193358	2019-03-24 19:05:11	2019-03-24	WetPuppykisses	btc	1.553454e+09	-1.0	0.0	luke sees bitcoin as a collaborative effort be...
4193518	4193359	2019-03-30 02:25:53	2019-03-30	_0_1	Bitcoin	1.553913e+09	-1.0	1.0	1/1,000,000th of a bitcoin

Fig. 2.1 - Reddit comments containing 'bitcoin' 2009 to 2019 dataset

For sentiment analysis, vaderSentiment python module was used. After cleaning up the dataset, sentiment analysis of each comment was carried out. The sentiment was summed up and the price was taken as average for each month, plotted onto a graph to visualize the correlation between the sentiment and price. The result is given below:

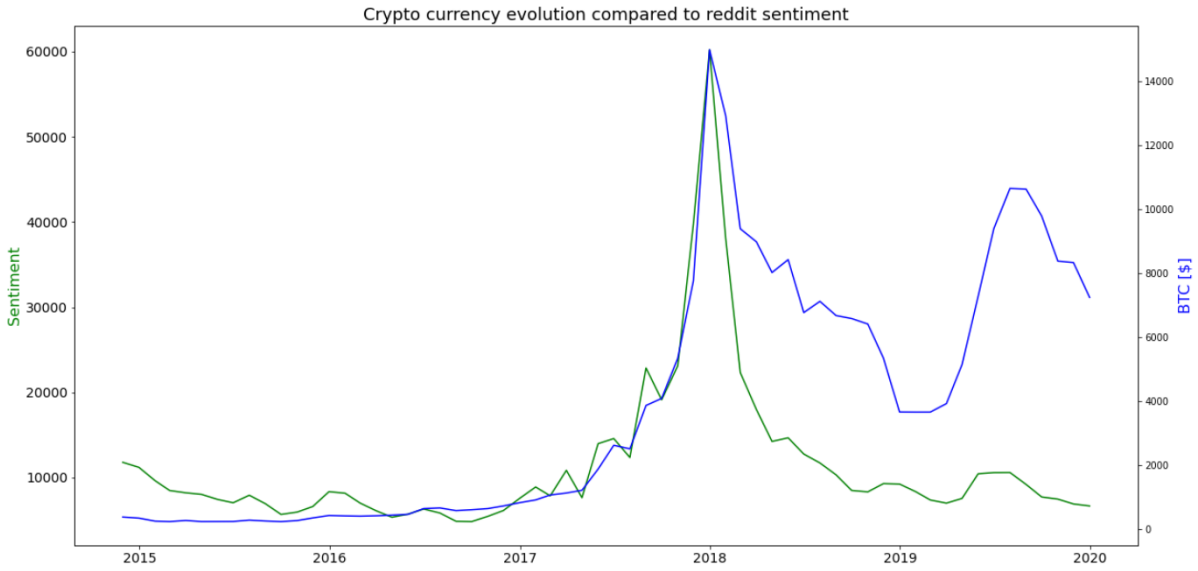


Fig. 2.2 – Relation between Sentiment and Bitcoin Price

Sentiment score is on left axis and plotted throughout using the green line. Bitcoin price is on right axis and plotted throughout using the blue line.

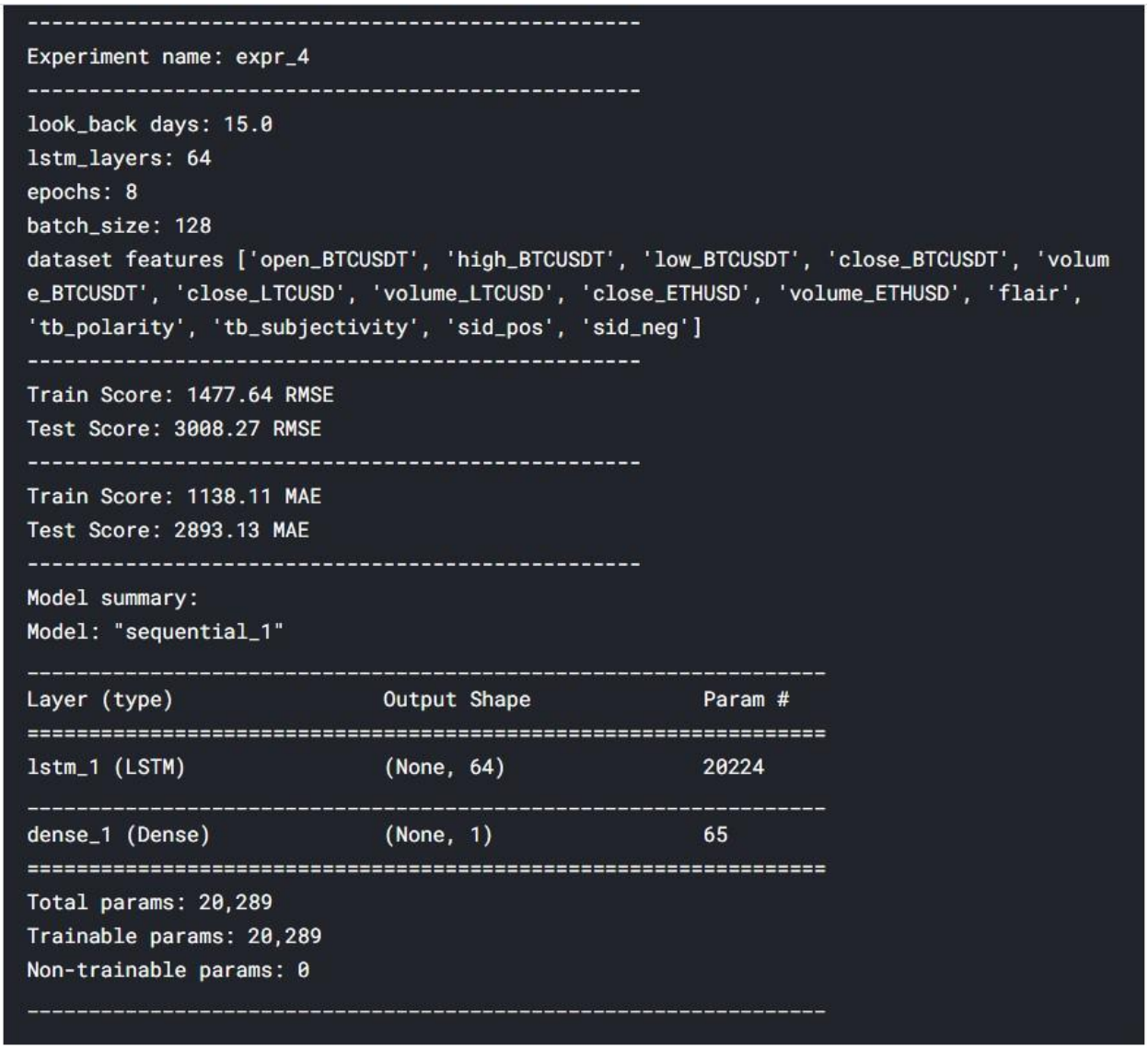
For each year, some amount of correlation can be observed from this analysis, which draws light on the possibility of predicting future trends based on sentiment and historical pricing.

3. Technologies and Tools

Following are the Tools and Technologies which we were learnt by us and implemented in the Bitcoin Price Predictor project.

3.1 LSTM

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs.



```
--------------------------------------------------
Experiment name: expr_4
--------------------------------------------------
look_back days: 15.0
lstm_layers: 64
epochs: 8
batch_size: 128
dataset features ['open_BTCUSDT', 'high_BTCUSDT', 'low_BTCUSDT', 'close_BTCUSDT', 'volum
e_BTCUSDT', 'close_LTCUSD', 'volume_LTCUSD', 'close_ETHUSD', 'volume_ETHUSD', 'flair',
'tb_polarity', 'tb_subjectivity', 'sid_pos', 'sid_neg']
--------------------------------------------------
Train Score: 1477.64 RMSE
Test Score: 3008.27 RMSE
--------------------------------------------------
Train Score: 1138.11 MAE
Test Score: 2893.13 MAE
--------------------------------------------------
Model summary:
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 64)                20224
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
Total params: 20,289
Trainable params: 20,289
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3.1 LSTM Model Training and Testing

Our Model takes 14 attributes as an input. LSTM needs recent data of several days our model takes 15 days as lookback which means recent 15X24 entries. For training and testing of model we have considered values from 1st Jan 2018 to 29 Sept 2021 of the dataset. Of which 80% of the considered dataset is used for the training purpose and remaining 20% of the dataset is used for the testing purpose.

This figure shows details such as parameters, keras model, training and testing score and model summary.

3.2 Microsoft Azure

Microsoft Azure, often referred to as Azure is a cloud computing service operated by Microsoft for application management via Microsoft-managed data centers. It provides software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) and supports many different programming languages, tools, and frameworks, including both Microsoft-specific and third-party software and systems.

3.3 Google Firebase

Firebase is a development platform known originally for its real-time database that's still at its core a multi-node, key-value database optimized for synchronizing data, often between user machines or smartphones and centralized storage in the cloud. The Firebase Realtime Database lets you build rich, collaborative applications by allowing secure access to the database directly from client-side code. Data is persisted locally, and even while offline, real-time events continue to fire, giving the end user a responsive experience.

3.4 BeautifulSoup

Beautiful Soup is a Python library that is used for web scraping purposes to pull the data out of HTML and XML files. It creates a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner.

3.5 VADER

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only talks about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is.

3.6 TextBlob

TextBlob is an open-source python library for processing textual data. It performs different operations on textual data such as noun phrase extraction, sentiment analysis, classification, translation, etc.

3.7 NLTK

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.

3.8 Flair

Flair is a simple Natural Language Processing (NLP) library developed and open-sourced by Zalando Research. It is used to build machine learning models for text classification and speech recognition. Flair can be used to build models used in language translation applications and speech recognition.

4. Diagrams and Modules

4.1 Diagrams

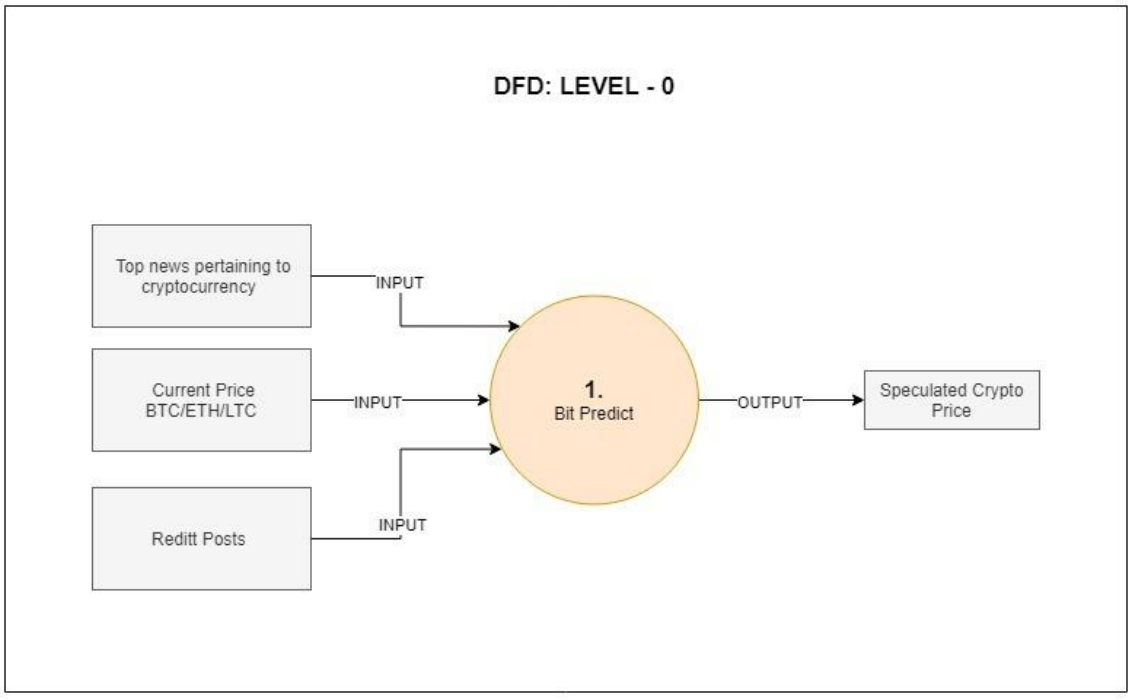
4.1.1 DFD

DFD (Data Flow Diagram)

DFD (data flow diagram) can be drawn to represent the system of different levels of abstraction. Higher-level DFDs are partitioned into low levels-hacking more information and functional elements. Levels in DFD are numbered 0, 1, 2 or beyond. Here, we will see mainly 3 levels in the data flow diagram, which are: 0-level DFD, 1-level DFD, and 2-level DFD.

Level: 0

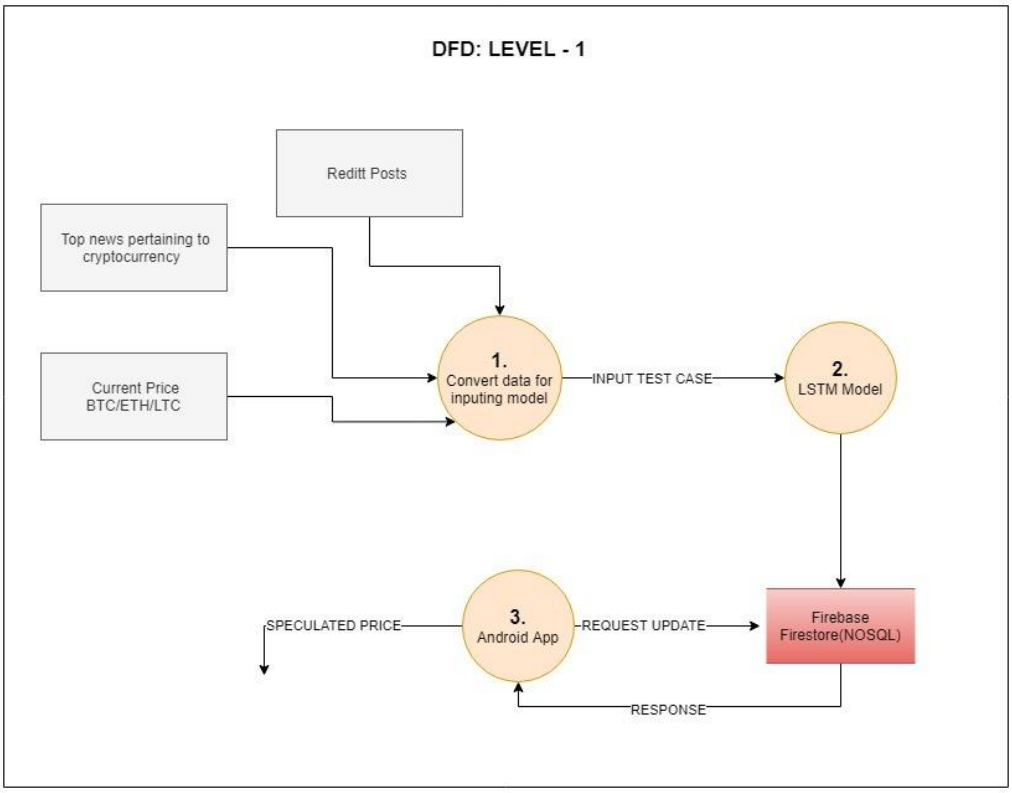
It is also known as a context diagram. It's designed to be an abstraction view, showing the system as a single process with its relationship to external entities. It represents the entire system as a single bubble with input and output data indicated by incoming/outgoing arrows.



4.1.2 DFD – 1

Level 1:

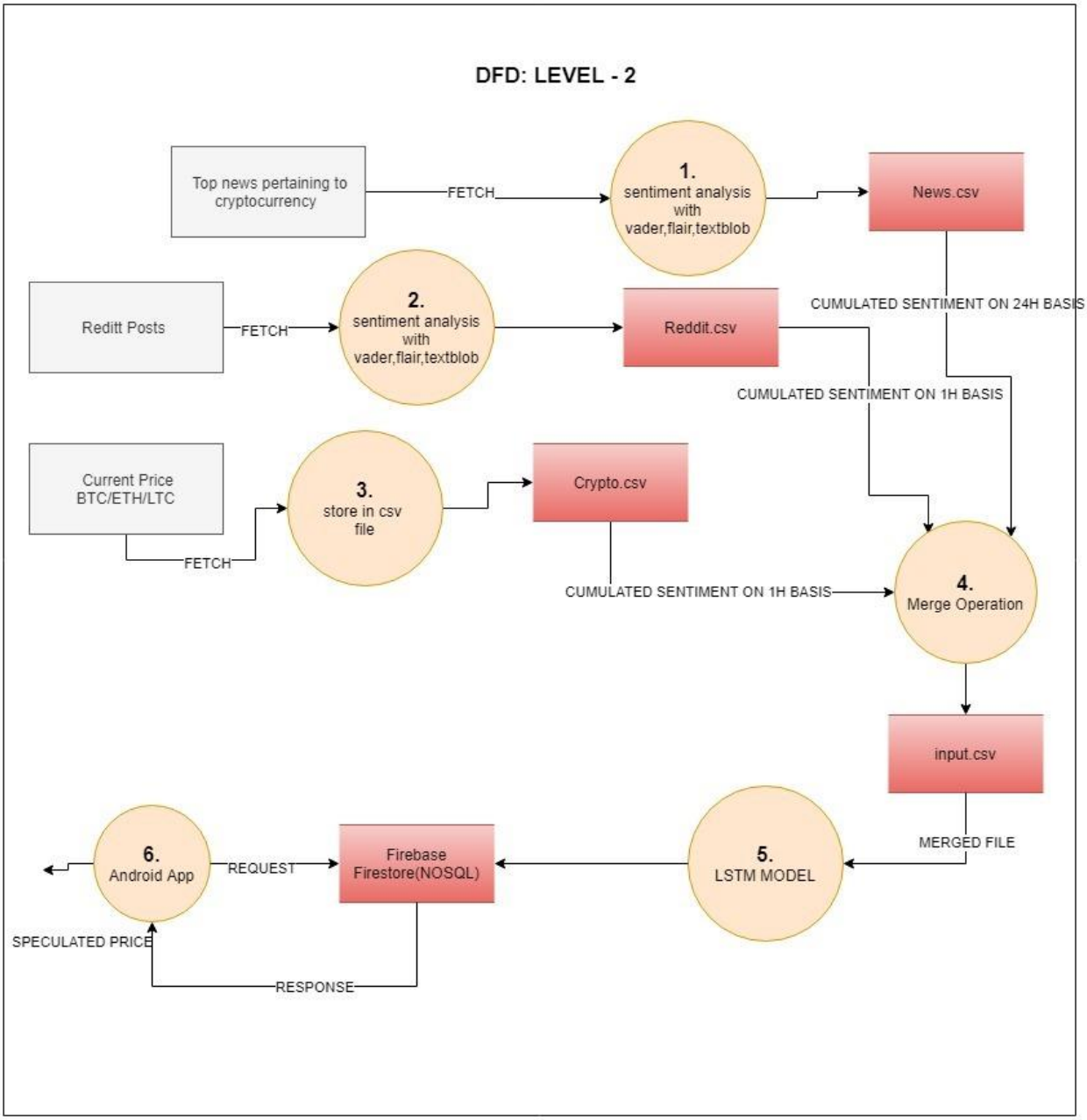
In 1-level DFD, the context diagram is decomposed into multiple bubbles/processes. In this level, we highlight the main functions of the system and breakdown the high-level process of 0-level DFD into subprocesses.



4.1.3 DFD – 2

Level 2:

2-level DFD goes one step deeper into parts of 1-level DFD. It can be used to plan or record the specific/necessary detail about the system's functioning.



4.2 Modules

4.2.1 Fetching, Cleaning and Storing Data

- **Reddit**
  We use Pushshift API for fetching the data of reddit posts related to bitcoin and store it in a dataset.

- **News**
  We use BeautifulSoup for web scrapping Google News, later it is cleaned and sorted.

- **Crypto Price Data**
  We download the past data of various cryptocurrency (like BTC, ETH, LTC) from Binance and clean and store them in new dataset.

4.2.2 Sentiment Analysis on stored data

- **Reddit**
  We carry out Sentiment Analysis on Reddit data using models like VADER, Flair and TextBlob.
- **News**
  We carry out Sentiment Analysis on Google news data using models like VADER, Flair and TextBlob.

4.2.3 Model Training and Testing

Finally, after carrying out sentiment analysis on reddit post and google news, all those datasets are merged in a final dataset which is then given to model for training.

The final dataset is given to the LSTM model for training the model and the new data which is updated in the final dataset is used for predicting the future predictions.

4.2.4 Uploading and fetching Predicted data in Firebase

After getting the Predicted price from the model, it is upload on Firebase Cloud Database along with Current price and Timestamp.

We have published the python code for the above actions on Microsoft Azure which keeps on running the code periodically at hourly basis and thus the model gets updated every hour and gives us the latest predictions.

4.2.5 Fetching and Displaying News related to Cryptocurrency

We have used NewsAPI.org to fetch the latest Crypto News from various sources and displayed in a separate section in our app.

4.2.6 Android Application

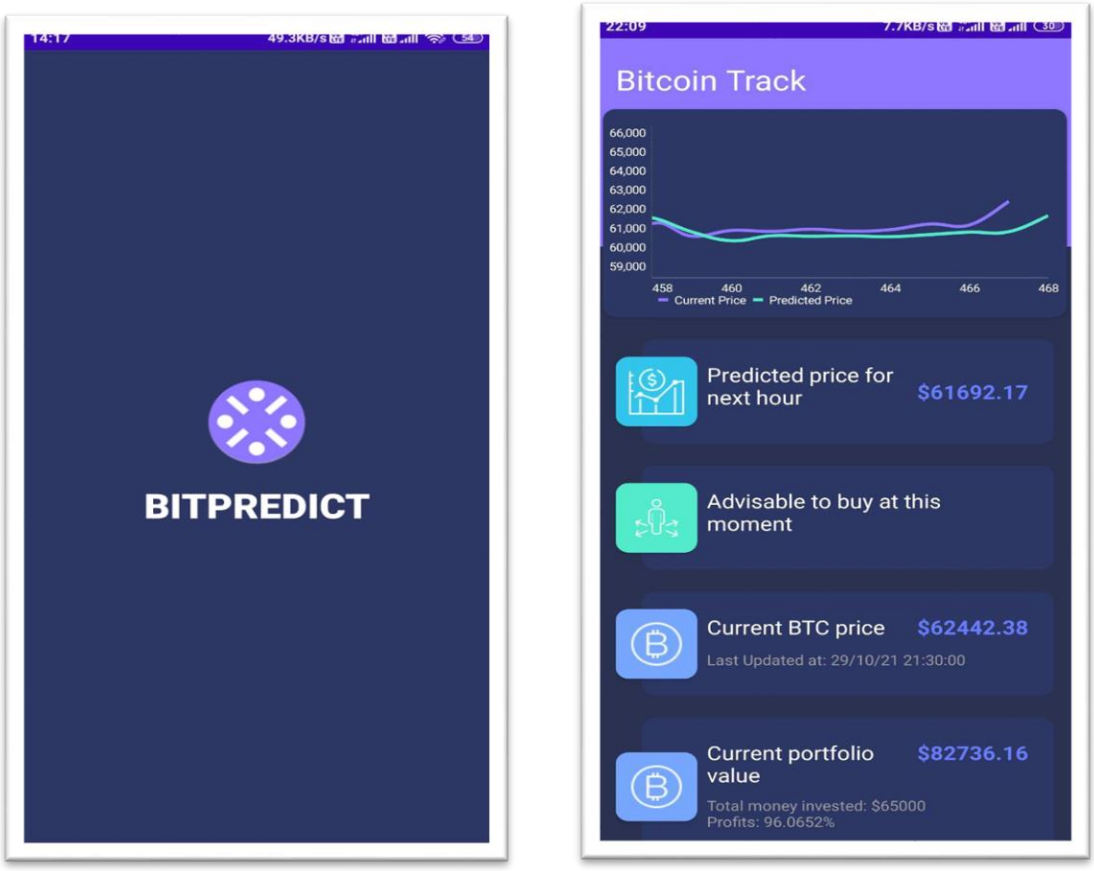
We fetch the various past Predicted and Current price from firebase and plot them into Line Chart for the reference of the user. There are various section on the Application Home Page for displaying Predicted price, Current price, Advice on buying and your current Portfolio. There is a separate section for displaying Latest News and Setting your portfolio.

5. Implementation

5.1 Front end

5.1.1 Home Page

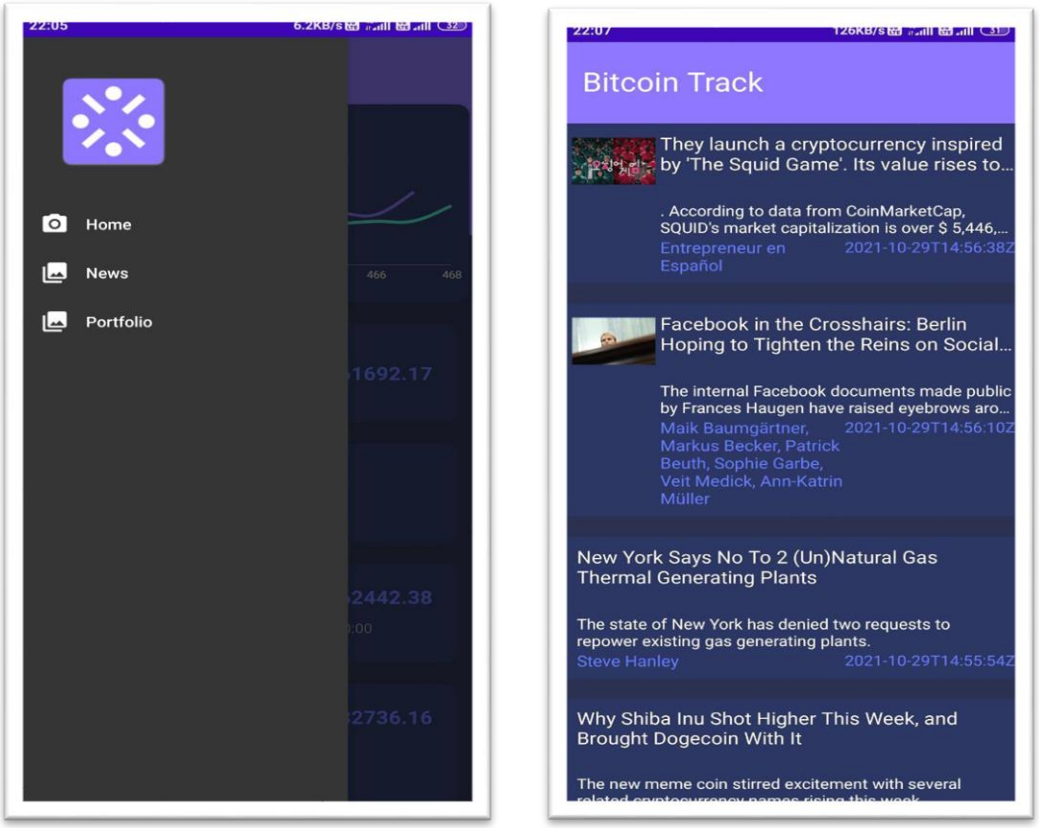
Here as you can see the home page contains a Line Chart depicting the graph of Current price vs Predicted price. It also includes the Predicted price, Current price, Advice and your Current Portfolio value.



5.1.1 – Home Page

5.1.2 News Page

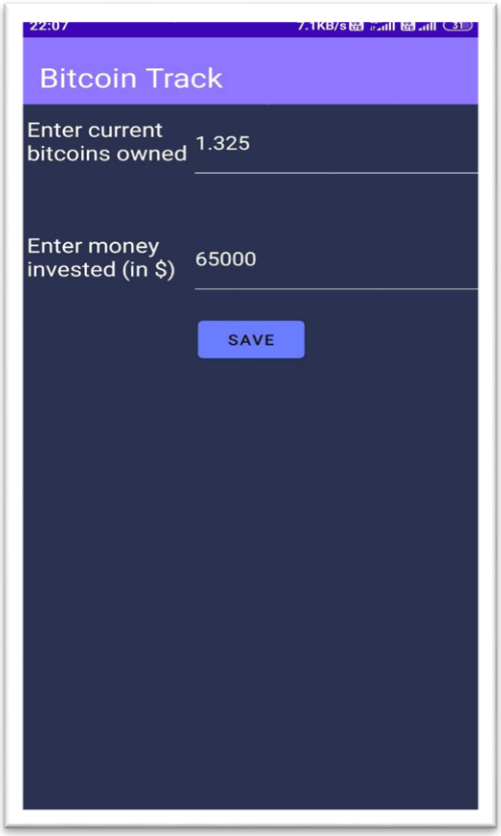
It contains Latest Crypto News from various sources.



5.1.2 – News Page

5.1.3 Portfolio Page

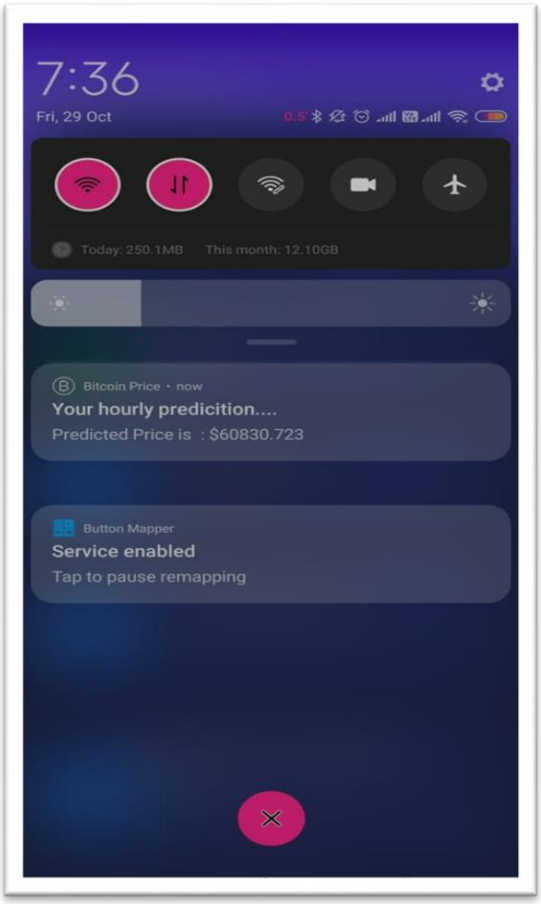
It stores the value of your current portfolio.



5.1.3 – Portfolio Page

5.1.4 Background Notification

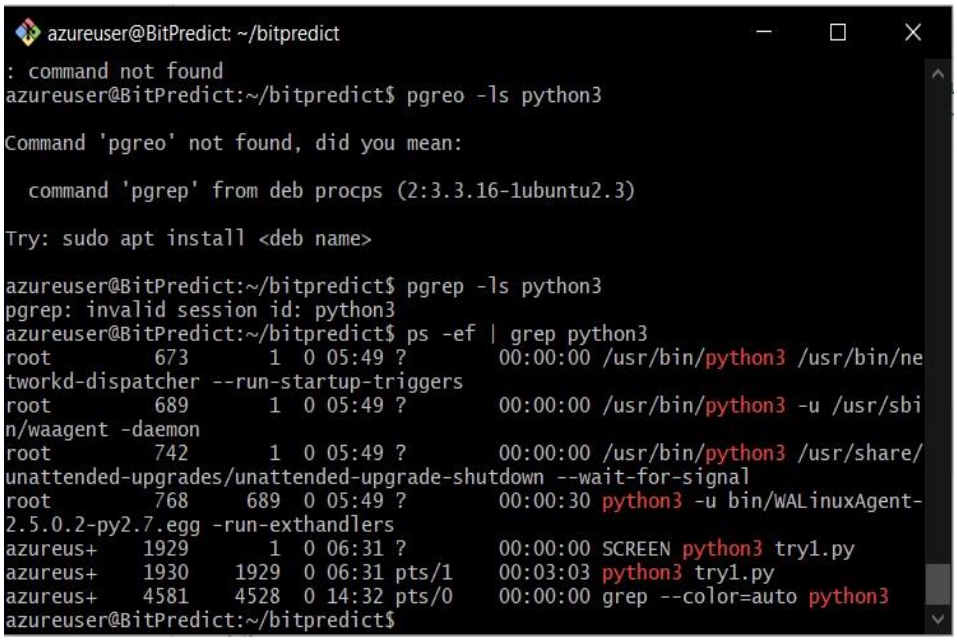
It gives you hourly Prediction notification.



5.1.4 – Hourly Notification

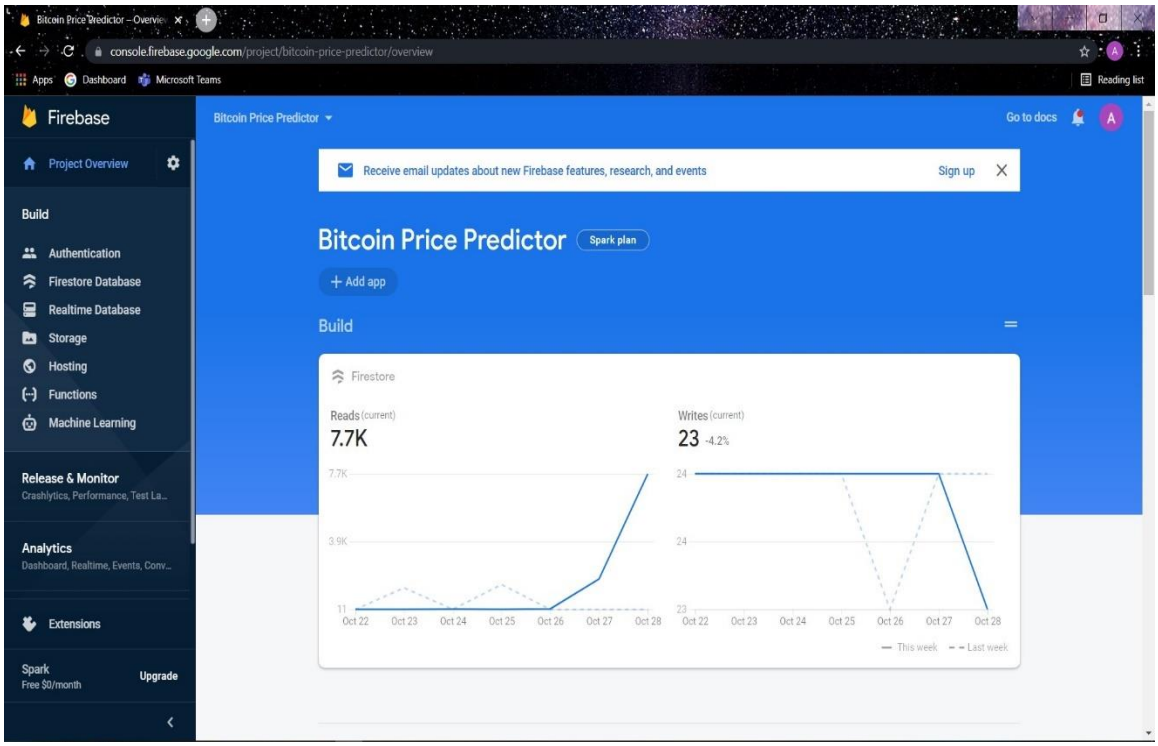
5.2 Back end

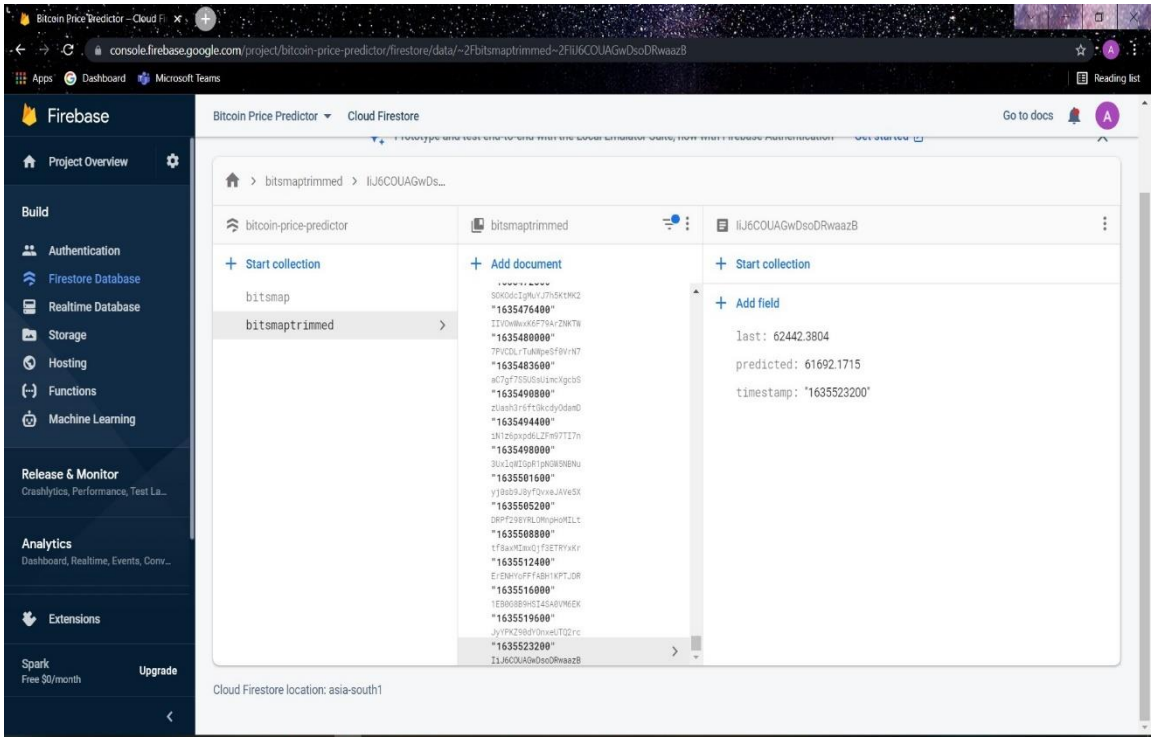
5.2.1 Microsoft Azure



5.2.1 – Running try1.py (file for prediction) periodically on Azure

5.2.2 Google Firebase Firestore Database





5.2.2 – Various instances of the Predicted value updated on the Firebase Cloud

6. Conclusion

6.1 Limitations and Future work

There are certain limitations to this project such as: Currently it only shows the prediction of Bitcoin, in future we can also provide predictions for various other cryptocurrencies like Dogecoin, Ethereum, Litecoin, etc. We can make the News section UI more user-friendly. We can add section for inputting multiple portfolios for other Cryptocurrencies.

6.2 Summary

BitPredict app is a useful app for knowing the current trend of Bitcoin in the market. It provides the latest predictions on the bitcoin prices and helps you manage your bitcoin portfolio. Along with that it also includes a separate section for Crypto News for the quick referral of the user. All in all, it satisfies the aim of an app useful to user for referring the current trend in Bitcoin.

References

1 – Current Bitcoin market capitalization & price - link

2 – “LSTM Based Sentiment Analysis for Cryptocurrency Prediction” - link

3 – “Bitcoin Price Prediction and Analysis Using Deep Learning Models” - link

4 – “Predictive analysis of Bitcoin price considering social sentiments” - Pratikkumar Prajapati link

5 – Crypto App - Widgets, Alerts, News, Bitcoin Prices - link

6 – Bitcoin Price Prediction App - link

7 – Crypto Forecast: AI predictions - link

8 – Binance -link

9 – WazirX -link

10 – Google Firebase Documentation - link

11 – Microsoft Azure Documentation - link