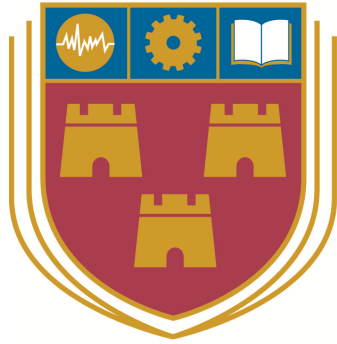Institiúid Teicneolaíochta Cheatharlach

INSTITUTE *of*
TECHNOLOGY
CARLOW

At the Heart of South Leinster

Computer Games Development CW208

Software Functional Specification

Year IV

Josh Tyrrell Browne
C00224660
2020

Section 1 - Project Scope

The initial goal of this project was to investigate the currently available Natural Language Processing (NLP) tools, and test their usability by applying them in a practical use case. NLP refers to a subfield of linguistics, computer science, and artificial intelligence concerned with giving computers the ability to understand text and spoken words in much the same way human beings can. NLP currently has a wide range of use cases such as spam detection, machine translation, text summarization, virtual assistants and chatbots, and last but not least social media sentiment analysis. The area of most interest to me was sentiment analysis. NLP has become an essential tool for uncovering hidden data insights from social media channels. Sentiment analysis can analyze language used in social media posts, responses and reviews to extract attitudes and emotions in response to products, brands, or in my case, stocks and cryptocurrencies.

A personal interest in the cryptocurrency market led me to focus my sentiment analysis implementations on that space. Once this decision was made to perform sentiment analysis in the crypto space, a second research goal became part of the project. That being to investigate a the following questions:

1. How useful is the sentiment data from twitter about cryptocurrency? Can it be used to gain insight to what the future may hold? Or is twitter data too diluted with spam posts and bias speculations?
2. Can accurate scorings of crypto related posts be made with currently available tools?
3. Is there online sentiment trends that relate to upward/downward price movement? Or is online sentiment too random or non related to price movement?

To put these questions to the test, I decided to implement a time series LSTM model, where the input features at a high level are price data and sentiment data. Sentiment data has been used as training data for various ML models in the past, such as Support Vector Machines and various types of neural networks. There has been decent success with these methods in stock market price predictions, but there have been far less attempts at applying these methods to the cryptocurrency space, probably because crypto is a relatively new idea and to many people it is still not a proven success.

It is my hope that this project will further my understanding of NLP tools and their capabilities and Neural Networks, particularly Long-Short Term Memory models. Also, I hope my findings with this work can positively affect the research efforts into these spaces.

Sections 2 - Risks and Assumptions

There are some factors in my design that I would like to consider, as they could impact the reliability of my findings. The technology used here to calculate the public's sentiment is the

NLTK lexicon library. This method may not be the optimal decision for use on tweet data as it does not account for common slang terms and will likely fail to pick up on sarcasm. Also, the corpus of my training data only spans back a period of 40 days which may not be enough to successfully pick up on trends accurately. The Twitter API only allows queries for tweets in the past 30 days so I was unable to recover sentiment data for the distant past and expand my training data.

Section 3 - Product Overview

The end product for this project is a web based application where a user can access sentiment and price statistics about Bitcoin, Ethereum and Cardano cryptocurrencies. Also the user can view the results of my LSTM models for each case. As my models perform short term predictions and I do not plan on continuously updating the websites data, traders will not realistically gain much useful information. Although I do believe there could be useful data generated using these methods with constantly up to date data for use with these models. Another researcher may gain some insight to the type of results that can be found using an LSTM model, and some may find the use of sentiment analysis intriguing. It seems likely that the sentiment analysis tools will improve with time, and in turn yield better results as a feature for use in neural networks. Cryptocurrency traders and market analysts often derive subjective predictions from price charts and sentiment data, so the types of data shown on my website can be useful to those involved in the space.
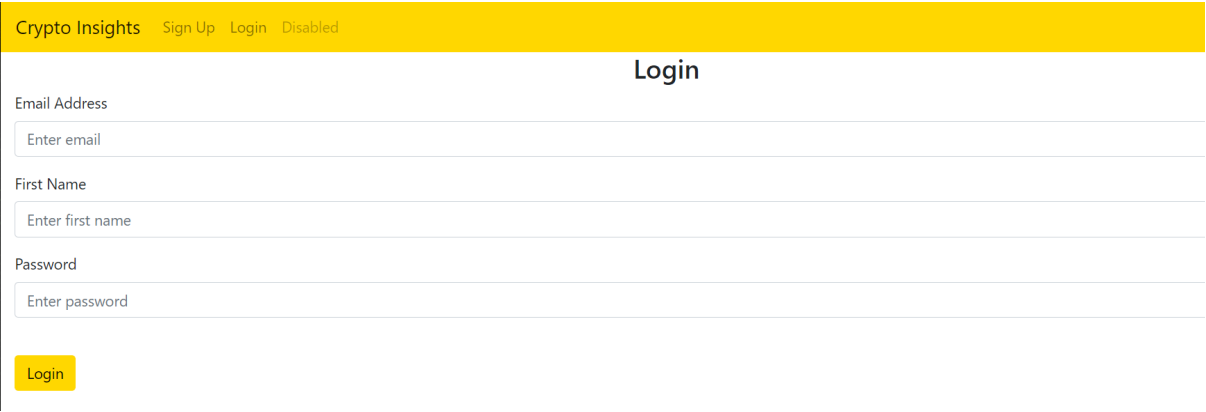
Section 4 - Use Cases

In the following sections I will detail what is involved in the processes with my application, firstly I will identify the sequence of actions that need to be performed to provide a meaningful, observable result to the end user. Secondly from the perspective of an end user and the steps they need to take to reach the final user goal.

4.1 Business Use Case
In order to provide the end user with useful information, the website administrator would essentially need to collect up-to-date data and use this data to update website graphs and displayed data. The data needed is cryptocurrency price data and twitter sentiment data. This data can easily be retrieved through running my Python scripts, which make calls to the *Twitter* and *Binance* API's and save the retrieved data to CSV files. They would then use this newer data to update the graphs shown on each webpage, which can also be done easily through running of my scripts. They would then have to retrain LSTM models with the up-to-date data so that the models predictions can be more accurate from processing more data. The data retrieval, data scaling, data batching and model training can all be done easily through running my project scripts.

4.2 System Use Case

The end user of the website may want to gain insightful information on one of the cryptocurrencies my website offers data on, or they may have an interest in the methods I use and they would like to view the results generated from an LSTM model trained on price and sentiment data. When the user first loads up the website, they are directed to the login page, see Fig. 1. Here they will enter the details of their account to login to the website.

Fig. 1 The login page

If they have not yet created an account, they will navigate to the 'sign up' page (Fig. 2) instead. In order to securely store users' passwords, I store a hash value which is generated from the password as a key. I don't actually store the password characters but still am able to ensure an entered password is the correct one.

Fig. 2 The sign up page

Once a user has successfully logged in, they will be navigated to the 'view data' pages (Fig. 3). Here they can view price and sentiment data for the *Bitcoin, Ethereum* and *Cardano* cryptocurrencies. The price and sentiment data is represented and displayed via graphs. Also on these pages, the user can view a word cloud image which is generated from the tweet data of popular tweets regarding those currencies. The final and most important piece of data observation that the user can view on these pages, is the results of my LSTM models. For each of the three currencies listed, there are separate models and databases used for generating the graphs.

This Currencies Price Chart:
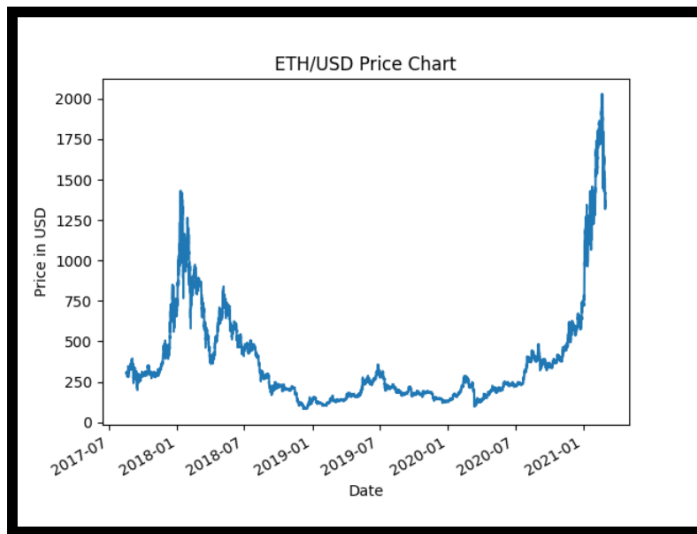
The above graph represents the price data for the currency since its creation date to the present. The data was collected from queries to the Binance API. The data visualisation in graph form was created using the "matplotlib" Python library.

Sentiment Chart:
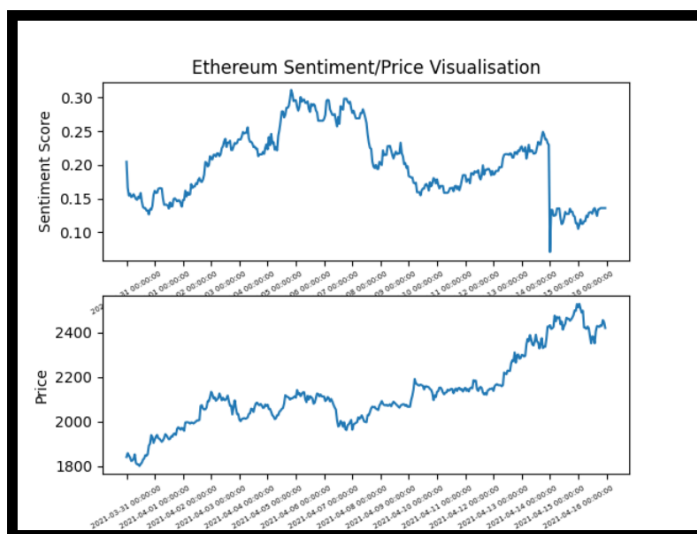
The above chart represents sentiment data plotted on a graph along with price data, it is done this way so we can compare the sentiment trends against the price trends. The two graphs match on the Y axis which represents date times. The sentiment data was generated from collected from collected tweets from the Twitter API and the sentiment values were calculated using the NLTK libraries Sentiment Analyzer.

Fig. 3 The view cryptocurrency data page of my web application (part 1)

Word Cloud From Twitter Data:

The above is a word cloud image. It was generated from a collection of the twitter data retrieved for sentiment analysis.

The results of my LSTM Model trained on data for this currency:

The above represents the predicted values that were generated from my trained LSTM Recurrent Neural Network model. The predicted values are plotted against the true price values to give us an idea of the accuracy in the predictions.

Fig. 3 The view cryptocurrency data page of my web application (part 2)

Once the user is finished observing my data, they will likely like to logout. To do this they will click the 'Logout' button from the navigation bar (shown Fig. 4), upon clicking the user will be logged out of the website and redirected to the login page. Something to note here is that Logout is only an option in the navigation bar once a user has logged in, and the Login and Sign up buttons are only options on the navigation bar if the user has not yet logged in. This functionality is achieved through 'IF' statements in the HTML code.