MSc Project - Reflective Essay

Project Title:	Bitcoin price trend prediction using tweet sentiment analysis
Student Name:	Krishna Sameer Kummetha
Student Number:	210470013
Supervisor Name:	Prof. Eranjan Udayanga Padumadasa
Programme of Study:	MSc Big Data Science

1. Strengths/Weaknesses

1.1 Strengths

- Using twitter over other social media for sentiment analysis is better due to the ease with which we can extract data and due to it's use by prominent members as compared to other social media.
- Pre-processing using NLTK is better as it does not require you to create your own stop words list for every for all the NLP tasks. It saves time as we do not to focus on rewriting functions.
- Labelling tweets positive, negative or neutral using VADER[1] is better than manually doing it as through this way we can get more training data as compared to manual labelling.
- Sequential model in CNN is good for text classification as it contains various filters which are able to detect useful features.
- The chosen timeline between 01/2/2021 to 01/03/2022 for the analysis purpose is ideal as we can see during this time there are a lot of fluctuations in bitcoin price in this time interval.

1.2 Weaknesses

- Not enough data to get proper prediction
- Can only attempt to predict upward or downward trend but we will not be able to extent to which it decreases using the classification models, which will give the investor a better idea.
- Not enough features considered.

2. Presentation of possibilities for further work

The work can be extended to further improve the model proposed in the following ways:

- Predicting price movements hourly as compared to daily could have given better results as we would get larger training data.
- Time series forecasting using hourly data could have given more accurate results as we could predict more accurately the extent to which the price increases or decreases.
- More features such as number of retweets for a particular tweet, number of followers for the user, considering if user is verified or not could add weightage to the scores and hence giving us more accurate results.

3. Critical analysis of the relationship between theory and practical work produced

In this project the final result of what is achieved is close to what was said theoretically. I was able to predict the trend of bitcoin price movement using twitter sentiment analysis to a certain extent. Although the results could have been better with more feature extractions such as adding weightage to the scores of the tweets based on the number of likes the tweets has got, rehsares of the tweets and the region in which it was tweeted, etc among other features. We could have also combined sentiments of various other sources such as from reddit and other official forums to get more accurate results. More advanced Recurrent neural network models such as LSTMs(Long short term memory)[2] and GRUs(Gated Recurrent Unit)[2] could improve the performance. LSTMs are a kind of recurrent neural networks which can learn dependence of order in predicting sequence problems. So in this way it could help us predict future prices based on the previous prices. GRUs are similar to LSTMs but a simplified version. So all these combined we could have got results closer to what was thought.

4. Awareness of Legal, Social Ethical Issues and Sustainability

Real time prediction of bitcoin price using tweet sentiment analysis is ethical as we are just scrapping data from a public platform which are visible to anyone. The data collected doesn't involve any personal information about the person which could harm the privacy of the person involved. Our project is compliant with the GDPR guidelines doesn't violate any rights. The crypto price prediction is considered legitimate as it helps the investors understand the crypto market better and enables them to make an informed decision. Using this could prevent investor from incurring heavy losses. In future we just need to make sure that while collecting any data from any social media we don't collect any of the private information of any person that could invade their privacy.

5.References

[1] Hutto, C.J. & Gilbert, E.E. (2014). “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14)”. Ann Arbor, MI, June 2014.

[2] Charu C. Aggarwal “Neural Networks and Deep Learning”