

**Department of Computer Science**

**MSc Data Science & Analytics**

**Academic Year 2022-2023**

Developing a Transformer-based Deep Learning Model to Investigate the Effect of Sentiment and Emotions in Finance-related News Articles on Cryptocurrency Portfolio Performance

Azadeh Almasi

2207479

A report submitted in partial fulfilment of the requirement for the degree of Master of Science

Brunel University

Department of Computer Science

Uxbridge, Middlesex UB8 3PH

United Kingdom

Tel: +44 (0) 1895 203397

Fax: +44 (0) 1895 251686

**ABSTRACT**

*Give a summary of your dissertation. Try to include the following:*

* *A high-level description of the topic area*
* *An overview of the problem studied and why this is interesting/relevant/ important*
* *A high-level description of the approach taken*
* *A summary of the contribution (what did you find/what was the outcome? What is the key implication of your work?)*

*If possible, try not to exceed 500 words. You must not exceed this one page.*

**ACKNOWLEDGEMENTS**

*Who are you grateful to? Your supervisor?*

Sign in the box below to certify that the work carried out is your own. By signing this box you are certifying that your dissertation is free from plagiarism. Make sure that you are fully aware of the Department guidelines on plagiarism (see the student handbook). The penalties if you are caught are severe. All material from other sources must be properly referenced and direct quotes must appear in quotation marks.

I certify that the work presented in the dissertation is my own unless referenced

Signature..........................................

Date.....................................

1

Insert a word count. This is the sum of the words in all the chapters only. The sum should exclude the words in the title page, abstract, acknowledgements, table of contents, references and any appendices.

**TOTAL NUMBER OF WORDS: 12,000**

[*All the above should not exceed this one page]*

# Table of Contents

Table of Contents v

List of Tables vi

List of Figures vii

CHAPTER 1: Introduction 1

*1.1 Introduction* 1

*1.2 Research aim and objectives* 1

*1.3 Research approach* 1

*1.4 Dissertation outline* 1

CHAPTER 2: Literature Review 2

*2.1 Heading* 2

*2.2 Heading* 2

*2.3 Heading* 2

*2.N Summary* 2

CHAPTER 3: Research Approach 3

*3.1 Heading* 3

CHAPTER 4: Data Analysis 4

*4.1 Heading* 4

CHAPTER 5: Discussion 5

*5.1 Heading* 5

CHAPTER 6: CONCLUsion 6

*6.1 Summary of the dissertation* 6

*6.2 Research contributions* 6

*6.3 Future research and development* 6

REFERENCES 7

*APPENDIX A:* ETHICAL *APPROVAL* 8

*APPENDIX B:* OTHER *APPENDICES* 9

**CHAPTER 1: Introduction**

**1.1 Background**

The market for cryptocurrencies has experienced substantial expansion ever since the first Bitcoin was released in 2009 (Farell, 2015) and has seen a remarkable value increase that has surpassed the most significant historical bubbles over the last three centuries (Naeem, Mbarki and Shahzad, 2021). The number of cryptocurrencies has increased from just one in July 2010 to 2419 by February 2020 (Anamika, Chakraborty and Subramaniam, 2023). As a result, both professional and academic researchers have shown a strong desire to comprehend the behaviour of these newly emerging assets (Naeem, Mbarki and Shahzad, 2021). In the present day, Both investors and speculators are increasingly interested in Bitcoin, which is emerging as the prevailing digital currency(Serafini *et al.*, 2020)

It could be difficult to figure out the worth of cryptocurrencies because there is much debate about their nature, such as whether they are a currency, a financial bubble, or simply a digital asset.

As a result, there is no agreement on which factors drive cryptocurrency prices. Consequently, cryptocurrency pricing is heavily reliant on widely disseminated opinions, sentiments, and emotions in finance related topics (Naeem, Mbarki and Shahzad, 2021) Furthermore, cryptocurrency prices behave differently than traditional currencies, making it extremely difficult to forecast their prices**.** (Abraham *et al.*, 2018)

In order to predict market movements, several methods have been applied, such as statistical analysis, pattern recognition, machine learning, sentiment analysis, and hybrid approaches.

Statistics, as the oldest approach, is employed for data analysis. Traders widely utilize Pattern Recognition, which involves visually identifying patterns and trends in stock data. Machine Learning, particularly with the advancements in deep learning tools, has gained significant popularity for predicting time-series data. The introduction of computer-based recognition through machine learning has further amplified the importance of pattern recognition theory.

Sentiment Analysis takes a different route by analysing crowd-sourced data. It relies on the principle of "wisdom of crowds," considering the collective opinion of individuals as reliable as that of a single expert. This approach leverages news, current events, public releases, and social media to make market forecasts.

In the Hybrid method, a combination of the aforementioned approaches is utilized. This comprehensive approach amalgamates statistical analysis, visual pattern recognition, machine learning techniques, and sentiment analysis to enhance market analysis and prediction.(Serafini *et al.*, 2020)

**1.2 Research aim and objectives**

The aim of this study is to develop a hybrid method by integrating a Transformer-based deep learning model with sentiment analysis methodologies. This hybrid approach investigates the impact of sentiment and emotions expressed in finance-related news articles on cryptocurrency portfolio performance. Through the application of deep learning techniques and sentiment analysis methodologies, this research seeks to explore the correlation between sentiment and emotions found in online forum comments and the performance of cryptocurrency portfolios.

An extensive literature review will be performed to evaluate the existing state of research in this field in order to achieve this goal. As a result of this review, valuable insights will be gained and a foundation for the study will be laid. It will also review a variety of data science research methodologies in order to determine which approach is the most appropriate for this project based on the existing literature.

To develop the model, we will begin by searching and finding an open-source dataset that will be used for training and testing the developed model before we proceed with the next steps. To identify a suitable dataset that contains the prices of cryptocurrencies over time, along with the sentiment and emotional comments on online forums associated with them, we will explore a variety of sources and repositories.

It is intended to further investigate the relationship between sentiments, emotions derived from online forum comments, and the performance of cryptocurrency portfolios by developing a Transformer-based deep learning model that will use deep learning to analyse the sentiments, emotions derived from comments made online forums. An analysis of sentiments and emotions and portfolio performance will be conducted through correlations and comprehensive studies.

In its final stage, the investigation will carry out exhaustive testing, evaluation, and discussion of the research findings. This will contribute to the existing body of knowledge that has already been established in this field, while also giving recommendations for potential new lines of investigation.

**1.3 Research approach**

Data science projects can profit from project management and process methodologies. Such methodologies work as a success factor (Schröer, Kruse and Gómez, 2021).

As CRISP-DM (CRoss Industry Standard Process for Data Mining) is one of the project management and process methods that offers a framework for executing big data projects which is independent of both the industry sector and the technology employed, it is one of the project management and process methodologies. This model is intended to make large data mining initiatives less expensive, more trustworthy, repeatable, easier to organize, and quicker(Wirth and Hipp, 2000). Accordingly, we will use CRISP-DM as the process methodology that includes the following steps:1) Business Understanding, 2) Data Understanding, 3) Data 4) Preparation, 5) Modelling, 6) Evaluation, and 7) Deployment.

**1.4 Dissertation outline**

The evaluation of the literature will look at the various models and sentiment analysis methodologies employed in these investigations, as well as their accuracy and potential for boosting the profitability of cryptocurrency portfolios. By combining these insights, we hope to identify crucial insights and knowledge gaps that will help us design our own deep-learning model.

Afterward through the proposed methodology for this research, we will follow the CRISP-DM process model, which provides a framework for carrying out big data projects. The five steps of CRISP-DM methodology will be clarified.

In the part on data analysis, our primary focus is on the creation of a deep learning model that is based on the Transformer. This article investigates the methods that was applied in order to analyse the data and construct the model. The process of data analysis initiates with the collection of relevant datasets, continues with the pre-processing of the data, then moves on to the training of the deep learning model, and finally concludes with the evaluation and testing of the model.

In the discussion section, we will thoroughly analyse and interpret the results gained from the data analysis. This section offers a comprehensive examination and interpretation of the findings, highlighting their relevance within the framework of the research objectives and aims. Moreover, it addresses any limitations or challenges faced during the research and presents an assessment of the implications arising from the findings.

The conclusion acts as a comprehensive summary of the entire dissertation and serves as the last component of the dissertation. This is accomplished by presenting a summary of the most important findings from the data analysis and the debate, as well as by highlighting the most important contributions made by the research. This part offers a comment not just on the research method but also on its implications for the overarching topic of study. In addition to this, it indicates prospective topics for future research and development, providing paths for further exploration and enhancement.

**References**

1. Abraham, J. *et al.* (2018) ‘Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis’, 1(3).
2. Anamika, Chakraborty, M. and Subramaniam, S. (2023) ‘Does Sentiment Impact Cryptocurrency?’, *Journal of Behavioral Finance*, 24(2), pp. 202–218. Available at: https://doi.org/10.1080/15427560.2021.1950723.
3. Farell, R. (2015) ‘An Analysis of the Cryptocurrency Industry’.
4. Mariappan, L.T. *et al.* (2023) ‘A Forecasting Approach to Cryptocurrency Price Index Using Reinforcement Learning’, *Applied Sciences*, 13(4), p. 2692. Available at: https://doi.org/10.3390/app13042692.
5. Naeem, M.A., Mbarki, I. and Shahzad, S.J.H. (2021) ‘Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears’, *International Review of Economics & Finance*, 73, pp. 496–514. Available at: https://doi.org/10.1016/j.iref.2021.01.008.
6. Schröer, C., Kruse, F. and Gómez, J.M. (2021) ‘A Systematic Literature Review on Applying CRISP-DM Process Model’, *Procedia Computer Science*, 181, pp. 526–534. Available at: https://doi.org/10.1016/j.procs.2021.01.199.
7. Serafini, G. *et al.* (2020) ‘Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches’, in *2020 International Joint Conference on Neural Networks (IJCNN)*. *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom: IEEE, pp. 1–8. Available at: https://doi.org/10.1109/IJCNN48605.2020.9206704.
8. Wirth, R. and Hipp, J. (2000) ‘CRISP-DM: Towards a Standard Process Model for Data Mining’.