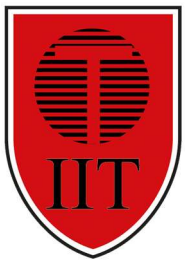


INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

MetaMind : A Meta-Ensemble Approach for Predicting the Future Price of Cryptocurrency

A Product Specification and Prototype Design by

Mr. Ashwinth Ragunathan

W1790169 - 2019713

Supervised by

Ms. Dileeka Alwis

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering degree at the University of Westminster.

May 2023

Declaration

I hereby declare that the dissertation and its components are the product of their own research efforts and have not been submitted to any other institution, while also ensuring that any facts obtained from external sources have been cited appropriately.

Student Name – Ragunathan Ashwinth

Registration number – w1790169 | 2019713

Signature: *Ashwinth* **Date:** 10/05/2023

Abstract

Predicting the price of cryptocurrencies is a challenging task due to their volatile nature. Even a slight shift in market trends can lead to significant price changes, making it difficult to accurately forecast future values. In this research, the author proposes a solution that uses Deep Learning and Meta-Ensemble approach to predict cryptocurrency prices.

The proposed solution utilizes historical Cryptocurrency market data and social media sentiment analysed data to train Deep Learning models, which can then predict future prices. The Meta-Ensemble approach is applied to combine the predictions of multiple models, resulting in a more accurate and robust prediction model. This approach overcomes the limitations of individual models by combining their strengths to produce a more optimum model.

To evaluate the proposed model, the author has conducted experiments on various cryptocurrency datasets, using evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-Squared (R^2) values. The results show that the proposed model outperforms traditional machine learning models in terms of accuracy and performance. This approach has the potential to help investors and traders make informed decisions by accurately predicting cryptocurrency prices.

Keywords : Prediction System, Deep Learning, Meta-Ensemble, Ensemble, Sentiment Analysis, Cryptocurrency, Bitcoin, Data Science

Acknowledgement

I have many people that I need to extend my gratitude towards for successfully completing this research project.

First and foremost, I would like to express my deepest appreciation to my mentor, Ms. Dileeka Alwis, whose encouragement, feedback, and support were instrumental in completing the project despite her very busy schedule. I am extremely grateful for the invaluable experiences gained and the efforts put forth in successfully completing this project. Without her guidance and support, I could not have accomplished this feat.

I also extend my heartfelt gratitude to my parents, my elder brother and sister, for their unwavering support and constant motivation. Their encouragement played a significant role in leading me towards a successful path in life. I would also like to thank my friends and colleagues who provided invaluable assistance during this journey. Finally, I extend my appreciation to the evaluators for taking time of their busy day to give me their valuable feedback regarding my project and the participants of the survey for their valuable feedback which contributed to the successful completion of this project.

Table of Contents

Declaration i

Abstract ii

Acknowledgement iii

1. Introduction 1

1.1 Chapter Overview 1

1.2 Problem Domain 1

1.2.1 Cryptocurrency 1

1.2.2 Social Media Impact and Sentiment Analysis 1

1.3 Problem Definition 2

1.3.1 Problem Statement 2

1.4. Research Motivation 2

1.5 Research Gap 3

1.6 Contribution to the body of knowledge 3

1.6.1 Contribution to the research domain 3

1.6.2 Contribution to the problem domain 4

1.7 Research Challenges 4

1.8 Research Questions 4

1.9 Research Aim 5

1.10 Research Objectives 5

2. Literature Review 8

2.1 Chapter Overview 8

2.2 Concept Map 8

2.3 Problem Domain 8

2.3.1 Blockchain 8

2.3.2 Cryptocurrency 11

2.3.3 Cryptocurrency market overview 12

2.4 Existing Work 12

2.4.1 Existing work for Cryptocurrency price prediction 12

2.4.2 Existing work for other price prediction 18

2.5 Technological Review 18

2.5.1 Sentiment Analysis and Data Processing 18

2.5.2 Deep Learning and Hyper Parameter Tuning 19

2.6 Evaluation and Benchmarking 20

2.6 Chapter Summary 21
3. METHODOLOGY **22**
3.1 Chapter overview 22
3.1 Research Methodology 22
3.2.1 Software Development Life Cycle Model 24
3.2.2 Design Methodology 24
3.2.3 Evaluation Methodology 24
3.3 Project Management Methodology 24
3.4 Schedule 24
3.5 Resource Requirements 24
3.5.1 Software Resources 25
3.5.2 Hardware Resources 25
3.5.3 Data Requirements 25
3.5.4 Skill Requirements 26
3.6. Risks and Mitigations 26
3.6. Chapter summary 28
4. SRS **29**
4.1 Chapter overview 29
4.3 Stake holder analysis 30
4.3.1 Stake holder onion model 30
4.3.2 Stake holder viewpoints 31
4.4 Selection of requirement elicitation methodology 33
4.4.1 Literature Review 33
4.4.2 Survey 33
4.4.3 Brainstorming 34
4.4.4 Prototyping 34
4.5 Discussion of findings 34
4.5.1 Survey 34
4.5.2 Brainstorming 38
4.5.3 Literature 39
4.5.4 Prototyping 39
4.6 Summary of findings 40
4.7 Context diagram 40
4.8 Use case diagram 41
4.9 Use case description 42

4.10 Requirements 43
4.10.1 Functional requirements 43
4.10.2 Non-functional requirements 44
4.11 Chapter Summary 44
5. Social, Legal, Ethical and Professional Issues (SLEP) **45**
5.1 Chapter Overview 45
5.2 SLEP Issues and Mitigation 45
5.3 Chapter Summary 45
6. Design **46**
6.1 Chapter overview 46
6.2 Design Goals 46
6.3 System Architecture Design 47
6.3.1 Architecture diagram 47
6.3.2 Discussion of tiers 48
6.4 Detailed Design 49
6.4.1 Choice of design paradigm 49
6.4.2 Data Flow Diagram 50
6.4.3 UI Design 50
6.5 Design diagram 50
6.5.1 System process flow chart 50
6.6 Chapter summary 51
7. Implementation **53**
7.1 Chapter overview 53
7.2 Technology Selection 53
7.2.1 Technology stack 53
7.2.2 Data-set Selection 54
7.2.3 Development framework 54
7.2.4 Programming languages 56
7.2.5 Libraries 57
7.2.6 IDE 58
7.2.7 Summary of technology selection 58
7.3 Implementation of core functionality 59
7.4 User Interface 64
7.6 Chapter Summary 64
8.0 Testing **65**

8.1 Chapter Overview 65
8.2 Objectives and Goals of Testing 65
8.3 Testing Criteria 65
8.4 Model Testing 65
8.4.1 MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), R^2 (R squared) 65
8.4.2 Cross Model Validation Data Test (Justification for producing the Meta-Ensemble model) 68
8.5 Benchmarking 68
8.6 Functional Testing 69
8.7 Module and Integration Testing 70
8.8 Non-Functional Testing 71
8.8.1 Performance Testing 71
8.9 Limitations of the Testing Process 72
8.10 Chapter Summary 72
9.0 Evaluation **73**
9.1 Chapter Overview 73
9.2 Evaluation Methodology and Approach 73
9.3 Evaluation Criteria 73
9.4 Self - Evaluation 74
9.5 Selection of Evaluators 76
9.6 Evaluation Results 79
9.6.1 Expert Opinion 79
9.6.2 Focus Group Testing 81
9.7 Limitations of Evaluation 81
9.8 Evaluation on Functional Requirements 81
9.9 Evaluation on Non-Functional Requirements 82
9.10 Chapter Summary 83
10.0 Conclusion **84**
10.1 Chapter Overview 84
10.2 Achievements of Research Aims & Objectives 84
10.2.1 Achievement of Research Aims 84
10.2.2 Achievement of Objectives 84
10.3 Utilization of Knowledge from the Course 85
10.4 Use of Existing Skills 86
10.6 Achievement of Learning Outcome 87

10.7 Problems and Challenges Faced 87
10.8 Deviations 88
10.9 Limitations of the Research 89
10.10 Future Enhancements 89
10.11 Achievement of the contribution to the body of knowledge 90
10.11.1 Technical Contribution (Price Prediction System) 90
10.11.2 Domain Contribution (Cryptocurrency) 90
10.11.3 Additional Contribution 90
10.12 Concluding Remarks 90
References I
Appendix A IV
Appendix B V
Appendix C VII
Appendix D IX
Appendix E XIV
Appendix F XVI

List of Figures

Figure 1 Rich picture diagram 29
Figure 2 Stake holder analysis 30
Figure 3 Context diagram 41
Figure 4 Use case diagram 41
Figure 5 Architecture diagram 47
Figure 6 Data Flow Diagram 50
Figure 7 System process flow chart 51
Figure 8 Technology stack 53
Figure 9 Gathering data from Reddit 59
Figure 10 Gathering data from Reddit 59
Figure 11 Gathering data from Reddit 60
Figure 12 Gathering data from Binance 60
Figure 13 Features 61
Figure 14 Data scaling 61
Figure 15 Simple RNN 62
Figure 16 LSTM 62
Figure 17 GRU 62
Figure 18 Bi-LSTM 63
Figure 19 Bi-GRU 63
Figure 20 Ensemble model function 63
Figure 21 Meta-Ensemble model function 64
Figure 22 Concept map IV
Figure 23 Gantt chart V
Figure 24 Home page VII
Figure 25 Prediction page VIII
Figure 26 Home UI - 1 IX
Figure 27 Home UI - 2 IX
Figure 28 Home UI - 3 IX
Figure 29 Home UI - 4 X
Figure 30 Prediction page X
Figure 31 CPU before starting server XIV
Figure 32 GPU before starting the server XIV

Figure 33 CPU after starting the server XV
Figure 34 GPU after starting the server XV
Figure 35 Mail regarding dataset validation XVI
Figure 36 Validation given by domain expert XVI
Figure 37 Private dataset uploaded to Kaggle XVII

List of Tables

Table 1- Research objectives 5
Table 2 The necessity of blockchain 9
Table 3 Evaluation 20
Table 4 Research Methodology 22
Table 5 Risks and Mitigations 26
Table 6 System Layer 31
Table 7 Containing System Layer 31
Table 8 Wider System Layer 32
Table 9 Survey 34
Table 10 Brainstorming 38
Table 11 Discussion of findings 40
Table 12 Use case description 42
Table 13 Requirements 43
Table 14 Functional requirements 43
Table 15 Non-functional requirements 44
Table 16 SLEP Issues and Mitigation 45
Table 17 Design Goals 46
Table 18 Development framework 54
Table 19 Libraries 57
Table 20 IDE 58
Table 21 Summary of technology selection 58
Table 22 RNN base learners 66
Table 23 RNN bas learners 66
Table 24 Ensemble models 66
Table 25 Ensemble models 67
Table 26 Meta-Ensemble models 67
Table 27 Cross Validation Model Data Test 68
Table 28 Benchmarking 68
Table 29 Functional Testing 69
Table 30 Module and Integration Testing 70
Table 31 Evaluation Criteria 73
Table 32 Self - Evaluation 74

Table 33 Selection of Evaluators 76
Table 34 Concept 79
Table 35 Solution 79
Table 36 Scope 79
Table 37 Architecture of the solution 80
Table 38 Implementation of the solution 80
Table 39 Prototype Features 81
Table 40 Usability 81
Table 41 Evaluation on Functional Requirements 81
Table 42 Evaluation on Non-Functional Requirements 82
Table 43 Achievement of Objectives 84
Table 44 Utilization of Knowledge from the Course 85
Table 45 Achievement of Learning Outcome 87
Table 46 Problems and Challenges Faced 87
Table 47 Deliverable dates VI

List of Abbreviations

Abbreviation	Description
ML	Machine Learning
DL	Deep Learning
API	Application Program Interface
IDE	Integrated Development Environment
UI	User Interface
BTC	Bitcoin
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
Bi-LSTM	Bidirectional Long Short-Term Memory
Bi-GRU	Bidirectional Gated Recurrent Unit
NLP	Natural Language Processing
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R^2	R Squared
SDLC	Software Development Life Cycle
GUI	Graphical User Interface
AI	Artificial Intelligence

1. Introduction

1.1 Chapter Overview

Cryptocurrency (a digital currency) is an alternative payment method developed utilizing encryption methods. By utilizing encryption technology, cryptocurrencies can act as both a medium of exchange and a virtual accounting system. In this research, the author tries to find a novel approach for Cryptocurrency price prediction system using social media sentiment analysis and introduce a new Ensemble architecture for prediction models which can be further used in other domains as well. The proposed system will try to automate a number of decision-making steps that a user would ordinarily need to take in order to forecast the future value of cryptocurrencies.

In this document, the topic to be covered, the uniqueness of the research hypothesis, the research technique, the supporting data, and the feasibility of the research questions are all described. The Work Plan concludes with a prediction of the delivery timetable for the project's deliverables.

1.2 Problem Domain

1.2.1 Cryptocurrency

The inception of blockchain technology aided in creating what is currently known as Cryptocurrency and currently there are more than thousands of cryptocurrencies available in the blockchain space (Rizwan, Narejo and Javed, 2019). Cryptocurrency is a decentralized digital currency, and it was designed in a way that it can be used on the internet to make transactions. The first Cryptocurrency which was created in the decentralized network was known as the Bitcoin (BTC), which till date remains as the most influential, highly priced, and well-known Cryptocurrency. After the success of BTC other alt coins like Ethereum (ETH), Binance Coin (BNB), Dogecoin (DOGE) and many more have emerged into the blockchain space. Cryptocurrencies are highly volatile, which means the value of Cryptocurrency can change drastically depending on a range of factors like supply, demand, availability, and investor sentiment as well.

1.2.2 Social Media Impact and Sentiment Analysis

The social media has played a significant influence on the price fluctuations of Cryptocurrency. A prime example for the above-mentioned statement would be the dramatic change in the

Bitcoin price (from $32,000 to $38,000) when the billionaire Elon Musk changed his Twitter account bio to #bitcoin in the year 2021 (Lee et al., 2022). Social medias like Twitter and Reddit are the most common platforms where the topic of Cryptocurrency is highly discussed and by retrieving and processing that information, we can use it to predict the prices of cryptocurrencies.

Sentiment analysis is the best protocol to do the above-mentioned process of retrieving and processing the data from social media platforms. Sentiment analysis (or otherwise known as Opinion Mining) is the process of gathering users' sentiments online (social media) using machine learning techniques and using the data gathered to make proper decisions (Tanwar et al., 2021).

1.3 Problem Definition

Predicting the price of Cryptocurrency has become a major issue for investors and Cryptocurrency holders because the prices of cryptocurrencies are very volatile, and it changes each and every minute. This leads to a problem where the person who plans to buy or sell Cryptocurrency must monitor the price changes happening in the marketplace which is extremely difficult. Many internal and external factors affect the price of cryptocurrencies. Social media trends is one of the major factors which has a direct impact in the prices of Cryptocurrency. The Cryptocurrency which has a favourable trend will go up in value while the opposite will happen if the trends do not favour Cryptocurrency.

To help the user predict the price of Cryptocurrency, it has been identified that several factors must be considered so that the user can have a favourable result in their crypto investment journey by helping them identify high yielding and trending currencies.

1.3.1 Problem Statement

It is extremely difficult to predict the price of Cryptocurrency since the price fluctuates drastically due to the influence of social media trends.

1.4. Research Motivation

The main problem discussed in this proposal applies to both people who are big investors and have ample domain knowledge about cryptocurrencies and people who are planning to invest but have less knowledge in Cryptocurrency prices or are afraid of how volatile it is. It will be hard for a person to predict what the price of Cryptocurrency will be in the future but with the help of machine learning and a vast amount of data the said problem can be tackled by

developing a prediction system. In order to produce an ideal prediction system, it is necessary to understand the factors that will affect the price of cryptocurrencies.

More people have started investing in cryptocurrencies since it is an unbelievably valuable asset, and it shows a promising future for growth as well in the upcoming metaverse space and the proposed solution will contribute to the advancement and accessibility to the domain of crypto, blockchain and prediction system.

1.5 Research Gap

- The past research done on the field of predicting the price of Cryptocurrency most of the researchers have used Deep Learning models to address the issue, but the usage of Deep Learning models are limited to either one or two and in some rare cases three models have been used.
- The usage of Ensemble is popular when building systems that predict the price of Cryptocurrency but over the years the same Ensemble approach (either stacking or weighted average Ensemble) is still being done without taking it to the next level.
- There are many systems which use social media sentiments to predict the price of Cryptocurrency but they either use news and a social media platform (mainly Twitter), or a maximum of two social media platforms. Google Trends has been rarely used in developing a prediction system, but it has not been used with the combination of other social media platforms.

Therefore, this research aims to fill the gap which is mentioned above by creating a novel prediction system that utilises several Deep Learning models and using Meta-Ensemble approach to predict the future price of Cryptocurrency using the sentiment data gathered from different social media platforms.

1.6 Contribution to the body of knowledge

In the following section the author will elaborate on the technical and domain contributions that will be made by doing the proposed project.

1.6.1 Contribution to the research domain

A state-of-the-art prediction system that will use the social media sentiment better than the already existing systems and research to predict the prices of cryptocurrencies will be implemented. The implemented system will also help the users automate the multiple decision-making processes that will be used to predict the prices of cryptocurrencies which will be

beneficial to invest on. It is also expected that the prediction system will be further used on other prediction related applications as well.

1.6.2 Contribution to the problem domain

The type of sentiment which will affect the price of Cryptocurrency will be identified and how that factor will influence the user to buy or sell Cryptocurrency will also be explored when a suitable prediction is provided. By looking at the past prediction systems and it's use cases in the last few years, it is understood that a prediction system will help the users to decide on which Cryptocurrency to buy that would benefit them later and to sell cryptocurrencies which would provide them with the highest yield in profit. This will also pave way to new investors originating in the domain and will eventually help the domain grow even more.

1.7 Research Challenges

Cryptocurrency is a domain which made its inception with the first ever Cryptocurrency Bitcoin (BTC) and much research has been done in this domain and one of that important research is predicting the prices of cryptocurrencies. To predict the price of Cryptocurrency research needs to be done regarding the preferences and factors that are considered when a user wants to invest in a particular Cryptocurrency. Therefore, it is very crucial to pinpoint the key features and factors which affect the prices or values of cryptocurrencies. Due to the high volatility nature of cryptocurrencies, it has been identified as a challenging field to apply the prediction techniques to predict the prices of Cryptocurrency since each and every Cryptocurrency is unique. Also, the social media trends and public opinions directly impact the prices of cryptocurrencies as well (Mai et al., 2018).

Currently existing prediction systems haven't had the necessity to consider social media trends as much as with related to the desirability of investing on cryptocurrencies. Furthermore, the high fluctuation of the Cryptocurrency market opens another challenge of the inability to keep predicting the varying prices of cryptocurrencies. However, given the widespread usage of baseline prediction approaches that have been tried and true across several areas, that alone cannot be considered. A proper prediction architecture has to be constructed using the indicated elements to be taken into account.

1.8 Research Questions

- **RQ1:** What are the features of Cryptocurrency and what factors affect the price of Cryptocurrency?

- **RQ2:** How can a system predict the price of Cryptocurrency which would be useful for investors?
- **RQ3:** What are the recent advancements in the price prediction models and architecture that can be used to develop Ensemble models?

1.9 Research Aim

The aim of this research is to design, develop & evaluate a novel system that will predict the price of a given Cryptocurrency for investing purposes by automating the manual decision-making process.

To further elaborate on the above-mentioned research aim, the author plans to create a system and architecture which can be used to predict the price of a given Cryptocurrency using its relevance and trends in social media. In order to achieve the goals a huge amount of data will be processed and will be streamed into the prediction architecture to automate the numerous decision-making processes a user has to undergo in order to buy or sell cryptocurrencies based on its price. Data Mining techniques will be used in the data gathering process, Natural Language Processing (NLP) techniques, Data analysis, Ensemble modelling and Deep Learning methods will be used to arrive at the best possible predictions.

1.10 Research Objectives

Table 1- Research objectives

Objective	Description	Learning Outcome
Literature Review	Analyse previous work which was done in the research area and critically evaluate them. • RO1: Conduct a preliminary study on the existing prediction systems and architectures. • RO2: Analyse and identify limitations in the existing literature. • RO3: Analyse the user requirements and factors which affect the price of Cryptocurrency. • RO4: Identify the necessary tools required to complete the research solution.	LO1 LO4 LO8

Requirement Analysis	Gather user requirements necessary for the project and critically analyse them. • RO1: Gather information about requirements related to the benefits of investing in cryptocurrencies. • RO2: Gather the requirements for a prediction system and understand the end-user experience. • RO3: Gather feedback from the domain experts regarding the project to build a robust system. • RO4: Gather data regarding Cryptocurrency trends from social media platforms.	LO1 LO3 LO8
Design	Designing the system and architecture of the proposed solution to solve the problems identified in the project. • RO1: Designing the Machine Learning or Deep Learning model necessary for predicting the prices of cryptocurrencies. • RO2: Designing the Natural Language Processing model for social media sentiment analysis. • RO3: Designing a user-friendly Graphical User Interface.	LO1 LO5 LO8
Development	The implementation of the proposed solution to predict the prices of cryptocurrencies. • RO1: Develop a prediction system which can accurately predict the price of a given Cryptocurrency using the relevant machine learning algorithms. • RO2: Developing the multiple models from each social media platform which will be used to predict the prices of cryptocurrencies. • RO3: Creating the Ensemble model for a particular Cryptocurrency from the developed models.	LO1 LO5 LO6 LO7 LO8

Testing and Evaluation	To test and evaluate the prototype model. • RO1: Develop test cases for the model (unit, integration and functionality testing). • RO2: Test the performance of the model. • RO3: Test the performance of the model by bench marking it against existing baseline models.	LO7 LO8

1.11 Chapter Summary

The chapter went through the decisions that were made during the final year project's starting phase. The author described the project's expectations, objectives, goals, and deliverables to ensure a clear implementation. The project requirements, including the problem area, research purpose, research objectives, and the body of knowledge, were also defined. Finally, the challenges regarding the research were also discussed.

2. Literature Review

2.1 Chapter Overview

As discussed in the above chapter Cryptocurrency has become a widespread topic and many investors have a major problem of predicting the price of cryptocurrencies since they are a highly volatile asset, therefore this chapter will highlight about the previous work which was done regarding Cryptocurrency price prediction, critically analysing the work which was done in the past, giving an overview of all the past work and deeply analysing the impact of Cryptocurrency in the world and how social media influences the prices of Cryptocurrency. The past work involves around the work which done in developing the Prediction systems, using Natural Language Processing for opinion mining and various other tools and techniques which were used to predict the price of Cryptocurrency.

2.2 Concept Map

Please refer Appendix A to view the concept map.

2.3 Problem Domain

In the section below the author dives deep into analysing the problem, which was identified in the Introduction chapter.

2.3.1 Blockchain

2.3.1.1 A brief introduction to blockchain

A digital distributed public ledger which is used to store information in blocks which are chained together using cryptographic techniques and algorithmic methods can be defined as a blockchain. The public ledger holds the records of all the transactions which makes blockchain a safe, transparent and decentralized entity (Fallucchi and Gerardi, 2021). In a nutshell blockchain is like a very big database which is widely spread across a network of computers and each computer its own copy of the database. This makes the public ledger safe to store data since the data is immutable and tamper proof.

From its inception into the world blockchain has been evolving and now it's being applied in many fields like real estate, insurance, voting, NFTs, IoT and many other but the primary use case of blockchain was to serve as a public ledger for the first ever Cryptocurrency Bitcoin, which was created by Satoshi Nakamoto in 2008 (Saari, Vimpari and Junnila, 2022).

2.3.1.2 The necessity of blockchain

One of the key important features of blockchain is storing data in new blocks which protects the data from getting corrupted or tampered but the blocks (or nodes) do not form by themselves but are created when the blockchain miners mine for a new block. The miners solve complex mathematical equations to create a new block where the most recent transactions will be recorded and after all the other nodes in the network verify and approve the block, it will then be added to the blockchain network (Ponnakanti and Gonen, 2020). This process is important because the miner will get compensated for their work by receiving Cryptocurrency while the blockchain is getting a new block added to its chain of blocks.

The world keeps evolving each passing day and so does all the existing technologies. This implies that the technologies need to be updated and need to be improved to avoid any malicious activity which can cause harm to the system. Blockchain is a technology which is very well versed in protecting the data stored in it from external threats and data mutations. The following table highlights all necessary features of a blockchain.

Table 2 The necessity of blockchain

Feature	Description of the feature
Decentralization	Blockchain is a distributed ledger technology that allows multiple parties to access and maintain the same data without the need for a central authority. This eliminates the need for intermediaries, such as banks or government institutions, and reduces the risk of corruption or fraud.
Security	Blockchain uses cryptography to ensure the security of data and transactions. Once a transaction is recorded on the blockchain, it cannot be altered or deleted, making it tamper-proof and resistant to hacking.
Transparency	Blockchain offers transparency by providing a public ledger of all transactions. This makes it easy to verify the authenticity of transactions and reduces the risk of fraud.

Efficiency	Blockchain can improve the efficiency of transactions by eliminating the need for intermediaries, reducing the time and cost of transactions, and increasing the speed of settlement.
Trust	Blockchain enables trust between parties by providing a transparent and tamper-proof ledger of transactions. This can help build trust between parties who may not trust each other or may not have a relationship of trust such as healthcare, government and financial institution.

2.3.1.4 The future of Blockchain and AI

In the future everything will be mostly automated, and a large database will be needed to store all the available information. This can be achieved by integrating blockchain and machine learning to design and develop better applications to automate tasks in many sectors like health, business, etc. Blockchain could provide a better platform for understanding AI and help in tracing the decision-making processes in machine learning. On the other hand, AI could help in automating various aspects of blockchain management, such as audit trail monitoring. The advantages of combining AI with blockchain include encryption advantages, ease of blockchain management, improved data protection, and data monetization (Iredale, 2021).

Artificial intelligence (AI) and blockchain technology are increasingly being utilized to help predict Cryptocurrency prices. Machine learning algorithms can analyse vast amounts of historical data and identify patterns and trends that can be used to make more accurate predictions about future prices. Blockchain technology, with its ability to store data securely and transparently, can provide a trustworthy and immutable record of historical prices, which can be used to train AI models (Sebastião and Godinho, 2021). By combining AI and blockchain, Cryptocurrency traders and investors can make more informed decisions about when to buy or sell, potentially increasing their profits and reducing their risks.

2.3.2 Cryptocurrency

2.3.2.1 Brief intro to Cryptocurrency

Cryptocurrency is something that you can call as an electronic or digital asset. It is not controlled by any government or bank, but by the people who use it which is also known as being decentralized. Cryptocurrency uses a technique called cryptography to make sure that no one can cheat or fake the transactions. Cryptography, briefly can be defined as a way of making mathematically complex codes and puzzles which makes hacking or manipulation difficult (Menezes, Van Oorschot and Vanstone, 2018).

Cryptocurrency can be stored in a cold wallet or a hot wallet, on your phone or computer, and you can use it to buy things from other people who accept it. The first cryptocurrency was Bitcoin, which started in 2009 (Nakamoto, 2008) and is still exceedingly popular today. There are also many other types of cryptocurrencies, such as Ethereum, Litecoin, Dogecoin, and more. Some people like cryptocurrencies because they are fast, anonymous, and independent. Others think they are risky, volatile, and hard to understand. Cryptocurrencies are still new and developing, so they might change a lot in the future (Poelstra, 2014).

2.3.2.2 Benefits of Cryptocurrency for traders and investors

Cryptocurrencies are a very useful and worthy assets for traders and investors, since they offer a range of benefits which are not offered by traditional financial institutions like banks and financial companies. One major advantage of using Cryptocurrencies as assets is their decentralized nature which means that no financial or government body has control over it but only the public users who use it as digital asset for transaction. This helps in reducing the transaction cost and having a mediator for overlooking the transactions. Cryptocurrencies are a highly liquid type of assets which means that they can be traded easily and quickly making them easy to buy and sell at any time which allows the traders to save a considerable amount of time trading the digital assets. Cryptocurrencies do not share the same regulations and restrictions that are impose on traditional financial institutions like banks or companies, which makes the digital asset more flexible for trading and investing and the Cryptocurrencies are not controlled or governed by any particular country or government, which implies that the same geopolitical risks that affect traditional financial institution doesn't have any effect on them. This makes the digital assets a very safe and useful option for investors who are concerned about political instability or economic uncertainty in their home country or region (Huang et al., 2021).

The unique features mentioned above make Cryptocurrencies a better choice as a liquid asset which can yield a huge amount of profit if used wisely.

2.3.3 Cryptocurrency market overview

A Cryptocurrency market overview involves analysing the cryptocurrency market as a whole, including the various types of Cryptocurrencies, their unique features, and the factors that affect their prices. It provides a broader context for understanding individual Cryptocurrencies and can help identify trends and patterns that may influence their prices. The analysis may involve examining the growth and adoption of Cryptocurrencies, supply and demand dynamics, market sentiment, news and events, and other factors. For instance, the overview may reveal that low market capitalization can lead to rapid price fluctuations or that regulatory announcements and security breaches tend to have a significant impact on prices (Krafft, Della Penna and Pentland, 2018). By gaining insights into the behaviour of cryptocurrency markets, traders and investors can make more informed decisions regarding their investment strategies.

Why is a price prediction system important for Cryptocurrency?

The volatility in the price of Cryptocurrency is very high, which leads to drastic changes in price not only each day but each second which leads to difficulties in predicting the price of a Cryptocurrency. Therefore, a well optimized, better performing and a highly accurate system can help in resolving this major issue.

2.4 Existing Work

2.4.1 Existing work for Cryptocurrency price prediction

2.4.1.1 Cryptocurrency price prediction using Deep Learning

This study (Ji, Kim and Im, 2019) covers the research done by the authors, which predicts the Bitcoin prices using Deep Learning techniques. The authors used the Bitstamp Bitcoin market (USD) time series data from 29th November 2011 to 31st December 2018, which consisted of 2590 days of data. They considered 29 features of the Bitcoin blockchain, including avg-block-size, blockchain-size, cost-per-trans, difficulty, est-trans-vol, hash-rate, market-cap, market-price, miners-revenue, my-wallets, n-trans, n-unique-addr, trade-vol, trans-fees and utxo-count. The authors performed a correlation analysis between these features and the Bitstamp Bitcoin price using Spearman rank correlation coefficients. They excluded features with a correlation coefficient less than 0.75 or greater than 0.95 and used a total of 18 features to develop their prediction models. The authors experimented with several state-of-the-art deep learning methods such as Deep Neural Networks (DNNs), Long Short-Term Memory (LSTM)

models, Convolutional Neural Networks (CNNs), Deep Residual Networks (ResNets), a combination of CNNs and RNNs (CRNN), and their ensemble models for Bitcoin price prediction. They developed both regression and classification models by exploiting the Bitcoin blockchain information and compared their prediction performance under various settings. The results for the experiments showed that although LSTM-based prediction models slightly outperformed the other prediction models for regression problems, DNN-based prediction models performed the best for classification problems. In addition, to determine the applicability of the proposed prediction models to algorithmic trading, they compared the profitability of the proposed models by using a simple trading strategy. The analysis result showed that classification models were more effective than regression models. The authors also acknowledged several limitations of their study, including the fact that their dataset only covered a limited time period and may not be representative of future price movements. They also noted that their models did not take into account news events or market sentiment, which could have a significant impact on Bitcoin prices.

The authors of the paper "Prediction of Bitcoin Price using Deep Learning Model" introduce a new method to forecast Bitcoin prices using a Recurrent Neural Network (RNN) model. This method could potentially offer better accuracy for long-term predictions than previous methods. The authors train their RNN model on time series data and try to enhance the accuracy of their forecasts. However, the paper does not provide a comprehensive evaluation of how their method compares to other methods. The authors only show some results that imply that their method might be effective, but a more rigorous evaluation would be required to fully validate their claims. Moreover, while the application of deep learning techniques such as RNNs may have some benefits for Bitcoin price prediction, these techniques can also be computationally expensive and may require a lot of expertise to apply them correctly. More research would be necessary to determine the feasibility of this method for real-world scenarios and the author has also mentioned that using hourly bitcoin price prediction will be useful for future improvements.(Freeda, Selvan and Hemanandhini, 2021).

The research paper "Bi-LSTM Price Prediction based on Attention Mechanism" by Jiashu Lou, Leyi Cui, and Ye Li explores an innovative method for forecasting the prices of two widely traded assets, gold and bitcoin. The authors employ a bidirectional LSTM neural network with an attention mechanism and a two-layer deep learning network. Their model achieved an accuracy of 71.94% for bitcoin and 73.03% for gold, as measured by AUC. Furthermore, when compared to traditional models, their attention Bi-LSTM model demonstrated superior

performance on their dataset. A key strength of this paper is its incorporation of an attention mechanism with a bidirectional LSTM neural network. The attention mechanism enables the model to concentrate on pertinent information while filtering out irrelevant or noisy data, potentially enhancing its predictive accuracy. Additionally, the bidirectional LSTM neural network allows the model to effectively utilize both forward and backward feature information from the input, further increasing its robustness. However, there are also some limitations to this research. For instance, the paper only focuses on price prediction and does not delve into how to place an asset order or address other issues that may arise in real trading processes such as matching trading books or liquidity. Moreover, while the attention Bi-LSTM model performed well in their experiments, it may not perform as well on other datasets (Lou, Cui and Li, 2022).

In the research paper titled "Comparative Analysis of Cryptocurrency Price Prediction using Deep Learning," the authors Muhammad Zakhwan Mohamed Rafik, Noraisyah Mohamed Shah, and Anis Salwa Mohd Khairuddin investigate the potential of deep learning models in predicting cryptocurrency prices. They suggest using a Recurrent Neural Network (RNN) model and compare its performance to other RNN architectures, including Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). To optimize their models, the authors fine-tune hyperparameters such as the number of epochs, batch size, dropout rate, and optimization algorithm. The methodology employed in this study is comprehensive. The authors carefully selected their input features and experimented with different combinations of these features using datasets for Litecoin (LTC) and Ripple (XRP). They also measure the performance of their models using metrics like Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and execution time. Even though a thorough research has been conducted the authors do mention of some short comings like only using two Cryptocurrencies for testing their models which implies that the research was limited for only those two coins and they acknowledge that adding sentiment analysis to expand the research will be a valuable addition (Zakhwan et al., 2022).

In their research paper, (Hegde, 2023) explores the use of deep learning techniques to predict Bitcoin prices. The authors utilized two datasets for their study. The first dataset was obtained from the Yahoo Finance website and included attributes such as open, high, close, low, volume, and adjacent close. The second dataset was sourced from GitHub and comprised 23 properties related to the Bitcoin Blockchain network, including btc_market_price, btc_total_bitcoins, btc_volume, btc_estimated_transaction_volume, and btc_estimated_transaction_volume_usd.

The author experimented with various deep learning models, including Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN). They found that the ANN model performed the best, with an R square value of 0.9997 and an RMSE value of 65.262. However, the authors acknowledged several limitations to their study. For instance, their dataset only covered a limited time period and may not accurately represent future price movements. Additionally, their models did not account for news events or market sentiment, which could significantly impact Bitcoin prices.

2.4.1.2 Cryptocurrency price prediction using Sentiment Analysis

The author (Aslam et al., 2022) proposes a study that aims to perform sentiment analysis and emotion detection on cryptocurrency-related tweets. The authors propose an ensemble model that combines two recurrent neural networks, namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), to improve the accuracy of sentiment analysis and emotion detection. The study employs a dataset of 40,000 tweets collected from Twitter, which have been preprocessed and annotated for sentiment and emotion using TextBlob and Text2Emotion libraries, respectively. The authors explore the suitability and performance of three feature engineering approaches: term frequency-inverse document frequency (TF-IDF), bag of words (BoW), and Word2Vec. They compare the performance of several machine learning models such as support vector machine (SVM), logistic regression (LR), Gaussian Naive Bayes (GNB), extra tree classifier (ETC), decision tree (DT), and k nearest neighbor (KNN) with their proposed LSTM-GRU ensemble model. According to the study's findings, the proposed LSTM-GRU ensemble model outperforms both machine learning and state-of-the-art models, achieving an accuracy of 0.99 for sentiment analysis and 0.92 for emotion prediction. The authors also discovered that the performance of machine learning models is relatively better when BoW features are used. However, the study has several limitations, such as the use of a dataset consisting of 40,000 tweets collected from Twitter, which may not be representative of all cryptocurrency-related tweets. The study only investigated three feature engineering approaches and did not consider other potential approaches that may enhance performance. Moreover, the study only compared the performance of the proposed LSTM-GRU ensemble model with several machine learning models and did not consider other state-of-the-art deep learning models Future research could address these limitations by employing larger and more diverse datasets, exploring other feature engineering approaches, and comparing the performance of the proposed model with other state-of-the-art deep learning models. Such

research could help to enhance the accuracy of cryptocurrency price prediction, which has significant implications for investors and traders in the cryptocurrency market .

In their paper titled “Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis,” S M Raju and Ali Mohammad Tarif explore an innovative approach to predicting the direction of Bitcoin's price in USD. Their method involves the use of Recurrent Neural Networks (RNN) with long short-term memory cells (LSTM) and sentiment analysis of Twitter and Reddit posts. One notable strength of this paper is its thorough investigation of different machine learning algorithms and their application to cryptocurrency price prediction. The authors compare their LSTM model with the traditional ARIMA method and conclude that LSTM with multi-feature produces more accurate results. This finding highlights the potential for machine learning techniques to enhance the effectiveness of cryptocurrency price prediction. Another strength of the paper is its incorporation of public opinion through sentiment analysis. By analyzing Twitter and Reddit posts, the authors can determine public sentiment towards Bitcoin and utilize this information to enhance their prediction model. This innovative approach could provide valuable insights into the factors that impact cryptocurrency prices. However, the paper also has limitations that should be taken into account. For instance, it only considers Twitter and Reddit posts for sentiment analysis, which may not present a complete picture of public sentiment towards Bitcoin. The authors could broaden their analysis by incorporating other social media platforms, news articles, and forums to capture a more comprehensive view of public opinion. Additionally, while the authors compare their LSTM model with ARIMA and conclude that LSTM with multi-feature performs better, they do not provide a detailed analysis of their model's performance. More comprehensive metrics such as accuracy, precision, recall, and F1-score could be beneficial in assessing the performance of the model (Raju and Tarif, 2020).

2.4.1.3 Cryptocurrency price prediction using Ensemble models

In their research, (Livieris et al., 2020) proposed a novel approach to forecasting major cryptocurrency hourly prices. By combining ensemble learning strategies with advanced deep learning models, the authors aimed to improve upon traditional methods for predicting cryptocurrency prices. One significant benefit of this approach is the use of ensemble learning strategies that can effectively combine the predictions of multiple deep learning models to produce more accurate and reliable forecasts. The authors evaluated three commonly used ensemble learning strategies and found that they could be effectively combined with advanced

deep learning models such as LSTM, Bi-directional LSTM, and convolutional layers. These models can capture complex patterns in time-series data and produce accurate forecasts. However, the study is limited to major cryptocurrencies and does not consider other cryptocurrencies with different price dynamics. Furthermore, the research is based on historical data and may not accurately predict future price movements in a rapidly changing market.

The paper authored by Ahmed Ibrahim from the University of Waterloo titled "Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-based Prediction Model" introduces a novel approach to predicting early market movements of cryptocurrency by utilizing sentiment analysis of Twitter data. The proposed Composite Ensemble Prediction Model (CEPM) outperforms other state-of-the-art techniques using Twitter datasets collected during the Era of COVID-19. The CEPM framework comprises of five stages, including text preprocessing, sentiment scoring, individual XGBoost classifications, composite ensemble aggregation, and model validation. The use of an ensemble model that combines multiple individual models is an improvement over previous approaches that rely on a single model. However, the study is based on Twitter data collected during the Era of COVID-19, and it is not clear how well the model would perform outside of this specific time period. The author suggests future research directions to adjust the number of incremental iterations of each XGBoost and incorporate various sentiment scoring schemes compared to VADER, indicating that there is still room for improvement in the proposed model (Ibrahim, 2021).

The paper by authors Zi Ye, Yinxu Wu, Hui Chen, Yi Pan, and Q. Jiang titled "A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin" presents a novel method to forecast Bitcoin prices using an ensemble deep learning model that incorporates price data, technical indicators, and sentiment indexes extracted from Twitter comments on Bitcoin. The authors show that their model achieves higher prediction accuracy than conventional models, with a mean absolute error (MAE) 88.74% lower than the daily prediction. A key contribution of this study is the use of a stacking ensemble technique that combines two types of neural networks, LSTM and GRU, to enhance decision accuracy. This technique enables the model to benefit from the advantages of both LSTM and GRU while reducing their individual drawbacks. Furthermore, the use of social media texts as a source of public opinion data is an innovative choice that sets this paper apart from other studies that depend on news websites as their main source of information. However, this research also has some limitations. For example, the authors only use data from September 2017 to January 2021

to train and test their model. It would be interesting to see how their model performs on more recent data or on data from different time periods. Also, while the authors provide a comprehensive methodology section, they do not discuss potential limitations or challenges associated with their approach, such as the effect of fake or misleading information on social media. Therefore, further research could be done to explore these limitations (Ye et al., 2022).

2.4.2 Existing work for other price prediction

2.4.2.1 Copper price prediction using RNN and Ensemble models

In their paper "Copper price movement prediction using recurrent neural networks and ensemble averaging," Jian Ni, Yue Xu, Zhi Li, and Jun Zhao explore the effectiveness of using recurrent neural networks (RNNs) for predicting copper price movements. Specifically, they investigate the performance of three types of RNNs - long short-term memory (LSTM), bidirectional LSTM (BiLSTM), and gated recurrent unit (GRU) - and evaluate them against a memory-free artificial neural network (ANN) model. To optimize prediction accuracy, the authors apply data slicing and augmentation procedures to extract useful information from both domestic and international copper-related market indices. Their results show that RNN models with memory units outperform the memory-free ANN model, and the combination of two RNNs (LSTM and BiLSTM) with a shorter input window length through ensemble averaging approaches (AVG or OLS) produces the best prediction model. The authors suggest that simplicity is key when deploying RNN models for copper price prediction. However, it is important to acknowledge that the study is limited to a specific dataset and the performance of the approach may differ when applied to other datasets or markets. Additionally, the study only considered a few types of RNNs and ensemble averaging approaches, leaving room for exploration of other methods (Ni et al., 2022).

2.5 Technological Review

The price prediction system will be using Sentiment Analysis, Deep Learning and Ensemble approach to build the final prediction model. The models will be trained on the gathered and pre-processed dataset.

2.5.1 Sentiment Analysis and Data Processing

The data social media sentiment data gathered using the appropriate data gathering techniques will be cleaned and processed for sentiment analysis since the raw data cannot be used to train the system. For sentiment analysis popular libraries like NlTK and TextBlob will be used to do

the analysis in order to gather the sentiment data. After the sentiment analysis is performed the data will then be merged with historical Bitcoin data.

2.5.2 Deep Learning and Hyper Parameter Tuning

The Deep Learning models will be used as base learners to develop the proposed system. After analysing the previous works, especially the ones done using multiple Deep Learning models (Ni et al, 2022) the author decided to use only RNN models, specifically a Simple RNN, LSTM, GRU, Bi-LSTM and Bi-GRU were chosen as the base learners. The reason for choosing the RNN models as base learners are listed below.

- Recurrent neural networks (RNNs) have emerged as a powerful tool for cryptocurrency price prediction due to their ability to capture the temporal dependencies in the data.
- RNNs can learn from historical price data and identify patterns and trends that can be useful in predicting future prices.
- RNNs can handle the high volatility and non-linear nature of cryptocurrency prices, making them well-suited for this task.
- The large amounts of data involved in cryptocurrency price prediction can be efficiently processed by RNNs, providing a diverse range of information sources for the model.
- RNNs can also adapt to changing market conditions, which enables the models to adapt allow for latest trends and patterns in the data.
- RNNs can be used to forecast different types of cryptocurrency prices, such as the high, low, or closing prices, providing flexibility in the types of predictions that can be made.

In a nut shell RNNs are super powerful Deep Learning models that are well suited for time series analysis and capturing trends in data easily, which are necessary for this project since the data that will be used for the project has sentiment data included in it and a time series of data will be used in the dataset to predict the future price of Cryptocurrency.

The hyper parameters for the RNN base models will be tuned by introducing a validation split in the dataset that will be used to train and test the models. The validation score of each model will be tracked and the models will be tuned for optimum performance.

2.5.3 Ensemble approach and Evaluation

The Deep Learning models that are trained will be joined together using a stacking Ensemble approach to increase the overall performance and will help in producing highly accurate predictions (Ye et al., 2022). The Ensemble approach will be followed by using Regression

models like Random forest regressor, Linear regressor, Elastic net regressor, Support vector regressor, Decision tree regressor, Gradient boosting regressor and Ada boosting regressor.

After training all the models the models performance will be evaluated using the popular evaluation metrics like Mean Absolute Error, Root Mean Squared Error and R Squared error.

2.6 Evaluation and Benchmarking

2.6.1 Evaluation

To evaluate the performance of the proposed system a comprehensive evaluation approach is necessary. The evaluations are necessary to check whether the system has the ability to accurately predict the price of Cryptocurrency to provide meaningful insight to the traders and investors. To achieve the above-mentioned goal the system should be tested using the adequate evaluation metrics. The table below shows the evaluations metrics that are necessary for evaluating the system.

Table 3 Evaluation

Evaluation metric	**Description of the metric used**	**Purpose of the evaluation**
MAE	Mean Absolute Error is used to measure the difference between the prediction and the true value. Lower the score better the prediction.	To check the accuracy of the prediction.
RMSE	Root Mean Squared Error is used to measure the square root of the mean square error of the prediction and true value. Lower the score better the prediction.	Useful in measuring the overall error of the prediction.
R^2	R Squared error is used to measure the proportion of the variance in the dependant	Useful to check if the model fits well with the provided data.

	variable and the independent variable. The R^2 value lies between 0 and 1, where higher the score better the prediction.	

2.6.1 Benchmarking

Since the proposed system is novel system in the domain proper benchmarking against other existing system will be difficult to perform. Therefore, existing system which can be considered as similar to the proposed system will be used to perform benchmarking.

After all the evaluation and benchmarking are performed the data gathered will be made publicly available so that in the future it can enable other researchers to evaluate or compare new models or system that will emerge in the domain of Cryptocurrency price prediction.

2.6 Chapter Summary

In this chapter the author discussed about the concept map showing how the decisions were made to do the project, then the problem domain of the proposed system was addressed. After that the past work done in the domain of Cryptocurrency price prediction was explored and an analysis of the past research was given. Next a review about the technology selected to build the proposed system was addressed. Finally, the evaluation and benchmarking of the proposed system were discussed.

3. METHODOLOGY

3.1 Chapter overview

This chapter outlines the study methods and systematic design for the project, including research, development, and project management methodologies. The author discusses the selection of each methodology with justifications and provides a project plan with a Gantt chart and deliverables for each research component. Additionally, the chapter covers the data collection process, stages of research, and a detailed explanation of the selected method for data collection. Finally, the author discusses potential project risks and ways to mitigate them.

3.1 Research Methodology

The methodologies of a project are required to maintain the quality of a project and to manage the project throughout its development and implementation. The Saunders Research Onion Model has been used to discuss about the methodologies that are necessary for the project.

Table 4 Research Methodology

Research Philosophy	The ideologies that served as the foundation for the majority of earlier studies were constructivism, interpretivism, and positivism. The **positivism** method will be used in this study since the author intends to gather data and interpret it in order to come up with a useful solution.
Research Approach	Research approaches are the various ways that a researcher could approach an issue when conducting research to attempt and find a solution. Since a methodological investigation must be conducted in order to prove the hypotheses, the author intends to use the **deductive technique**.
Research Strategy	**Surveys**, **previous research**, and **interviews** will all be used as research methods during the project. The main methods will be surveys and

	previous research, with interviews conducted to assess the planned system once it has been completed.
Research choice	There are three different sorts of research choices: mono, mixed, and multi-methods. Since only quantitative techniques are utilized in the study, the author intends to use the **mono** method.
Time Horizon	Since data will be obtained and used periodically over time while the project is being implemented, the time horizon for the research will be **longitudinal**.
Technique & Procedures	Web scraping will be utilized to collect the essential data from social media sites, and surveys, studies, academic papers, open-source datasets, and historical bitcoin data will also be utilised.

Research Hypothesis: The research hypothesizes by gathering user sentiment from social media platforms and creating individual Machine Learning models using each platform. After creating the models, Ensemble them and create a new model which will be able to perform better than the existing models to predict the prices of cryptocurrencies.

Research Process: Optimize the existing prediction models using Machine learning and Ensemble modelling.

Prototype Input: The Cryptocurrency for which the user wants to predict the price.

Research output: The system will run through all the processes to finally display the future price of the Cryptocurrency selected by the user.

3.2 Development Methodology

3.2.1 Software Development Life Cycle Model

There are different types of Software Development Life Cycle models available and from the list of all the available models, the author will be using the Agile methodology as the research development model for the following reasons,

- Agile methodology allows the project to be broken down into iterations.
- During the development stage it enables the addition and removal of requirements if necessary.

3.2.2 Design Methodology

For the design methodology the author has chosen to use Object Oriented Analysis and Design (OOAD). The OOAD design pattern was chosen due to its reusability nature and flexibility, which is necessary for a research project where requirements will change until the final system is developed.

3.2.3 Evaluation Methodology

As an evaluation statistic, the author has chosen to use Mean Absolute Error (MAE) and Root Mean Square error (RMSE). After applying the absolute value to the difference and averaging it across the entire dataset, MAE calculates the difference between the forecast value from the ML model and the data that is accurate.

3.3 Project Management Methodology

The Agile methodology is useful for the iterative, flexibility and responsiveness of a system while Prince2 is a commonly used methodology in project management. Therefore, combining both Prince2 and Agile methodologies, the author has chosen to use the Prince2 Agile methodology as the project management methodology.

3.4 Schedule

The gantt chart and deliverable dates can be found in Appendix B.

3.5 Resource Requirements

The following are the resources that are necessary for the author to complete the project successfully and deliver a proper working prototype.

3.5.1 Software Resources

- **Operating System:** An operating system which can handle heavy computational tasks without any issues (Preferred OS systems would be Windows, Linux and Mac).
- **Python or R:** The programming language that will be used to develop the Machine Learning models.
- **Tensorflow or Scikit Learn Python packages:** The libraries which will be used to develop the models.
- **NodeJS or Flask or Django:** Web framework which will be used to develop the prototype.
- **PyCharm or Jupyter Notebook:** Integrated Development Environment for writing the codes necessary for the program.
- **GitHub or GitLab:** Version controlling and creating backups of code.
- **Zotero or Mendeley:** Tool to manage and store research articles, papers and journals, and to manage references.
- **Google Colaboratory:** Cloud development platform provided by Google to train and develop Machine Learning models.
- **Microsoft Office or Google Docs:** To create documents related to the project.
- **Google Drive or One Drive or Drop Box:** To save and backup documents on the cloud.

3.5.2 Hardware Resources

- **An Intel Core i5 processor or above:** A machine with a processor which can handle heavy resource consuming computational tasks.
- **Minimum 45 GB ROM or more:** A disk space of more than 45GB of data is necessary to store the datasets and Machine Learning models.
- **Minimum 8GB RAM or more:** A machine with a RAM capacity which can manage datasets and run development environments.

3.5.3 Data Requirements

- **Twitter tweet data:** Sentiment data regarding Cryptocurrency will be gathered from the Kaggle open-source datasets, which will be combined with historical Cryptocurrency data to develop the proposed system.

- **Reddit data:** Sentiment data regarding Cryptocurrency will be gathered from the subreddits which mainly talk about cryptocurrencies and will be combined with historical Cryptocurrency data to develop the proposed system.
- **Historical data of Cryptocurrency:** The historical data related to Cryptocurrency will be gathered using the Binance Exchange API, which will be necessary for developing the proposed system.

3.5.4 Skill Requirements

- Deep understanding about the prediction systems.
- An overall understanding of creating the pipeline and model architecture.
- An overall understanding of performing sentiment analysis on text data.
- Deep understanding about the algorithms and Deep Learning models.
- A deep understanding about model Ensemble.
- Ability to create the proposed prediction system.

3.6. Risks and Mitigations

Table 5 Risks and Mitigations

Risk	Probability of occurrence	Magnitude of the loss	Mitigation Plan
Lack of knowledge in the selected domain.	5	5	Seek the help of domain experts and developers who work or have abundance of knowledge in the domain. Try to read research papers, articles which are available online and following an online course or tutorial can help further understand the technical aspects of the domain.
Lack of resources to train the models	4	5	The training process of the Machine Learning models require very high

			resource consuming computational process, and the personal computers processing power will not be enough to do the heavy-duty task. Therefore, using online notebook such as Google Colabaratory can help mitigate the problem.
Loss of development code due to unforeseen circumstances	3	5	Losing the development code which is necessary to implement the project is a very high risk. Therefore, backing up the codes to a cloud storage, external hardware or on GitHub can help prevent this risk.
Loss or corruption of documents.	3	4	The documents are crucial for the project since it explains everything that has been followed to develop the project. Therefore, to avoid the corruption or loss of document it is necessary to have back ups of the document in a cloud or external storage and backup the document in a daily basis.
Unpredictable risks (getting sick)	2	5	There are some risks which can not be

			predicted and do not know when it might happen. Therefore, it is better to work on a planned schedule to finish the necessary tasks and to always keep in touch with the mentor and update them with the progress of the research.

3.6. Chapter summary

In this chapter the author discussed about selecting the methodologies and justifications for completing the project goals within the time frame, as well as a detailed project plan with Gantt chart and deliverables. It also discusses the possible risks that could occur during the project period and their mitigation strategies.

4. SRS

4.1 Chapter overview

This chapter discusses about the overall approach of identifying the stakeholders who will be important for the system, collecting the necessary system requirements, and analysing the project data. The underlying system flow and the main stakeholders who are involved with the system are depicted using a rich picture diagram. To explain about the intended system functionality, a use case diagram and a use case description are provided. The chapter concludes with a list of functional and non-functional requirements necessary for the prototype of the proposed system.

4.2 Rich picture diagram

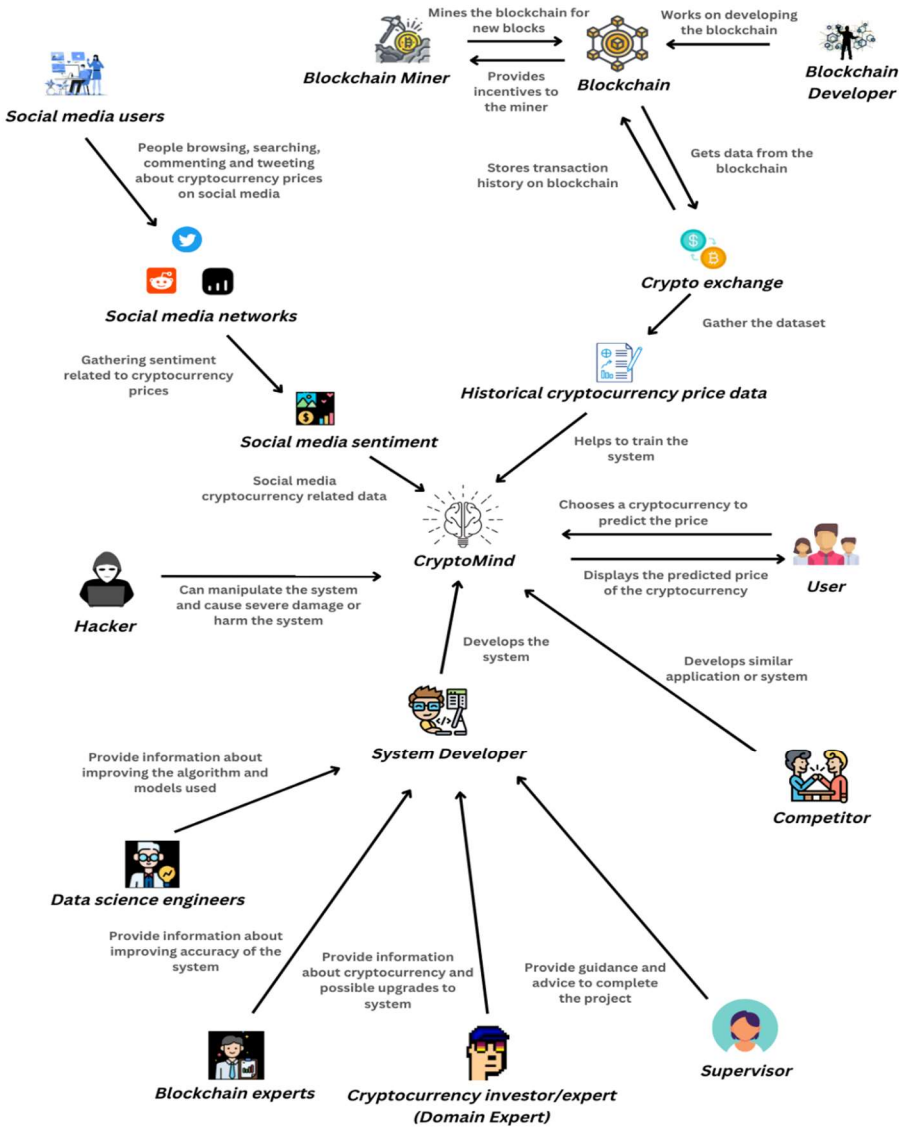


Figure 1 Rich picture diagram

4.3 Stake holder analysis

The section below highlights about the stake holders of the proposed system. The onion model given below has 3 layers in which different stakeholders are identified and further explanation about the stakeholders are given in the stakeholder viewpoint section.

4.3.1 Stake holder onion model

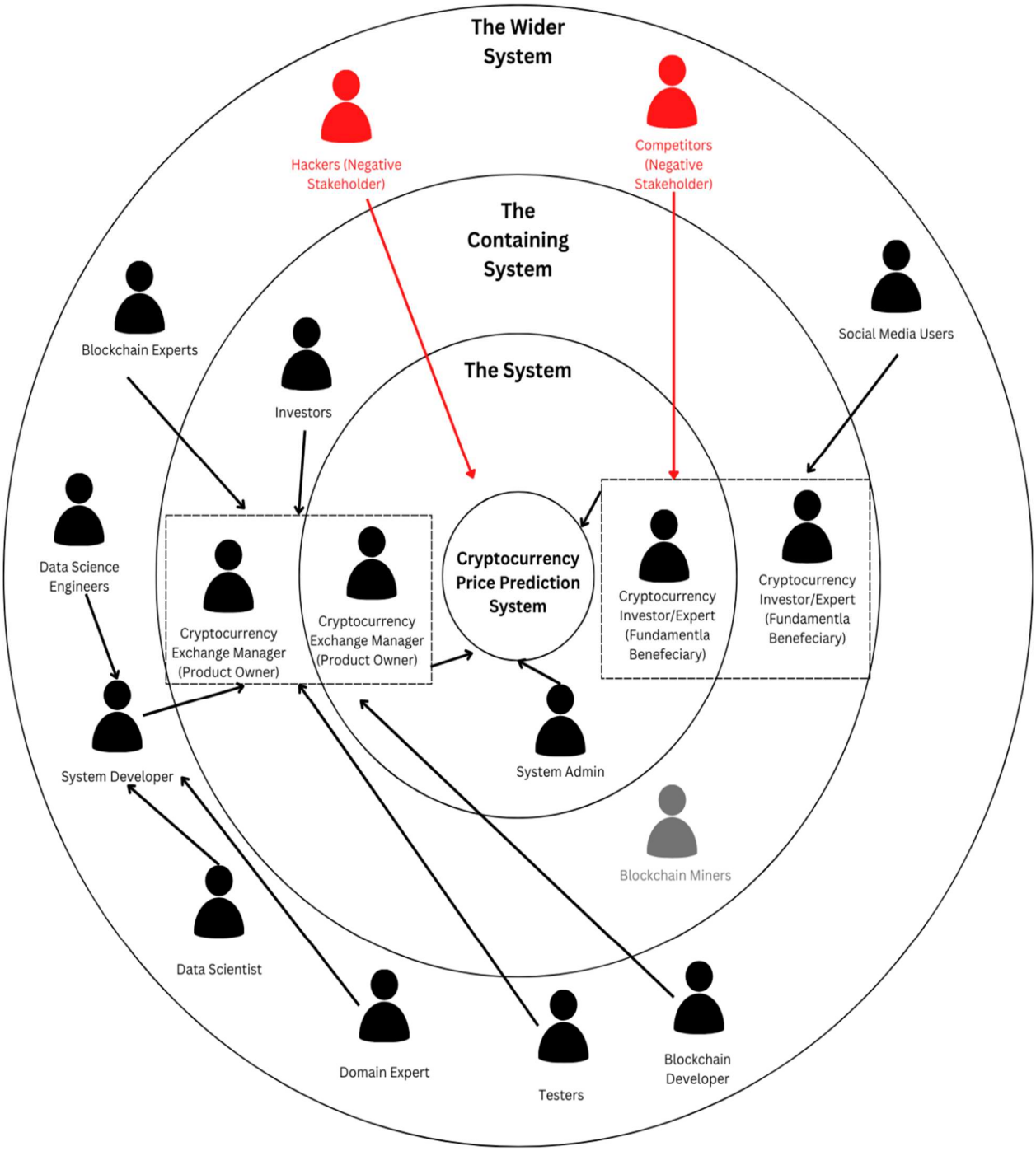


Figure 2 Stake holder analysis

4.3.2 Stake holder viewpoints

4.3.2.1 System Layer

In this section the stakeholders of the System layer are analysed.

Table 6 System Layer

Stake Holder	Contribution	Description
Cryptocurrency Exchange Manager	The owner of the system and manages all the operations	Makes sure the system functions without any errors or problems and overlooks all the operations.
Cryptocurrency Investor/Expert	The main beneficiary of the system	Uses the system to predict the prices of Cryptocurrency which aids them in deciding whether invest or sell a Cryptocurrency.
System Admin	Administration	Responsible for configuring and handling the overall system for optimum performance.

4.3.2.2 Containing System Layer

In this section the stakeholders of the Containing System layer are analysed (excluding Cryptocurrency Exchange Manager and Cryptocurrency Investor/Expert since they were explained in the previous layer).

Table 7 Containing System Layer

Stakeholder	Contribution	Description
Investors	Financial beneficiaries	Invest in the system during the development phase and makes profit from the system after deploying it.
Blockchain Miners	Mine the blockchain for new blocks	The miner mine for new blocks which help in storing the transaction history of crypto trades and help in maintaining the security and integrity of the blockchain network.

4.3.2.3 Wider System Layer

In this section the stakeholders of the Wider System layer are analysed.

Table 8 Wider System Layer

Stakeholder	Contribution	Description
Social Media User	Influence crypto prices	By tweeting, commenting and sharing post social media users can influence the price changes in cryptocurrencies.
Blockchain Developer	Developer	Helps in creating the decentralized applications necessary for the system.
Testers	Quality inspection	Help in testing the system for bugs and errors so that the system can perform without any issues.
Domain Experts	Experts	Has abundant knowledge about the domain and helps the developer create the system.
Data Scientist		Has abundant knowledge in the field of Data Science and helps the developer in developing the machine learning models and algorithms.
Blockchain Experts		Has abundant knowledge in the field of Blockchain technology. Provides the information necessary to improve system performance
Data Science Engineer		Has abundant knowledge in the field of Data Science and helps in manipulating data to make a better prediction system.

System Developer	Develops the system	The developer develops the overall system with the aid of the experts.
Hackers	Negative stakeholders	An anonymous user who can cause harm or potentially take over the entire system.
Competitors		Builds similar applications which performs similar or better than the proposed system.

4.4 Selection of requirement elicitation methodology

The below section discusses about the methodologies chosen to gather requirements. Requirement gathering is a fundamental aspect of a research project which help in gathering the information and data necessary to build a proper working application or system which is robust and fool proof. Literature review, conducting survey, brainstorming and prototyping are the requirement elicitation techniques used by the author to gather requirements for this project.

4.4.1 Literature Review

The important aspect of the final year project is research, and the research is conducted through the literature review the author develops for the project. The literature review is a culmination of the past research that was done on the topic and critically analysing the research that was done. By analysing the past research, the author was able to identify the contributions made to the topic and the possible gaps that were left by the past research.

4.4.2 Survey

A survey is a means of gathering data and information from a targeted audience in order to obtain insights and draw conclusions about a particular problem. The survey was conducted using an online questionnaire (google form) to gather necessary data regarding the requirements for the proposed system. This survey served as a tool to identify the expectations and thinking process of the target audience which will be greatly useful when developing the prototype.

4.4.3 Brainstorming

Brainstorming was chosen as a requirement elicitation tool because the author needed to freely think of new ideas and solutions to fill the gap which was identified through the literature review process. Brainstorming allowed not only the author but other individuals such as the supervisor, peers, and other technical and domain experts to share their ideas as well, which greatly helped in the requirement gathering process.

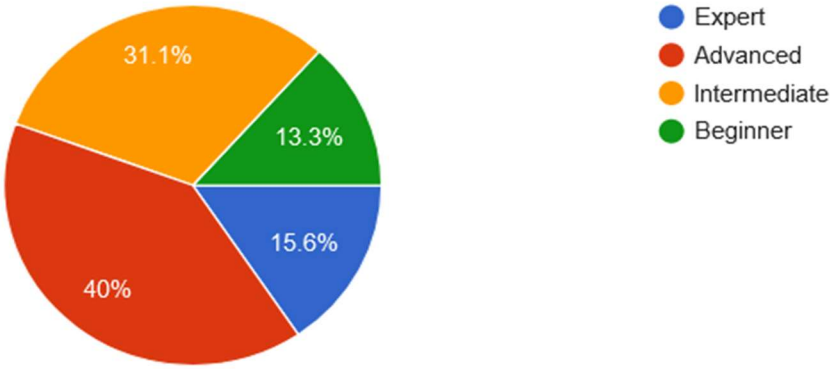
4.4.4 Prototyping

Agile Software Development was the life cycle process chosen initially by the author to develop the proposed system, therefore using the above-mentioned life cycle process for prototyping, the author would be able to repetitively try out different implementations and testing methods to develop the optimum performing model when completing the project.

4.5 Discussion of findings

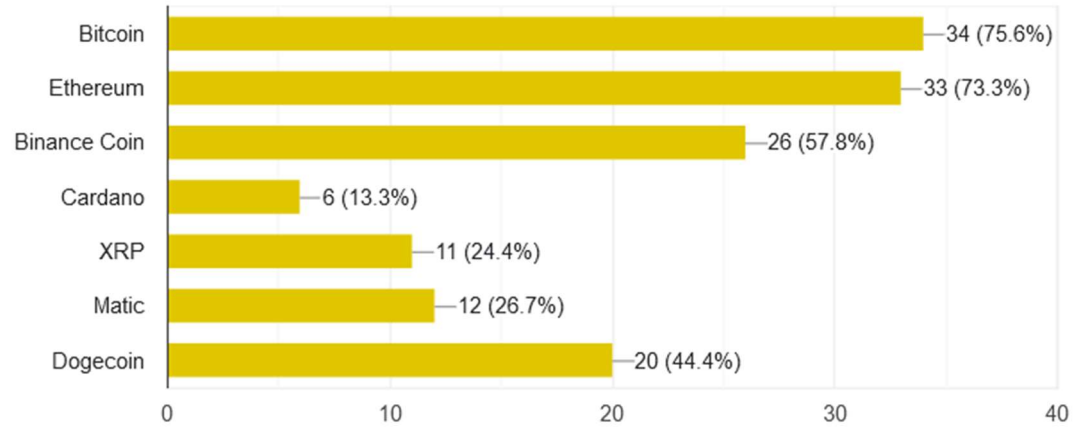
4.5.1 Survey

Table 9 Survey

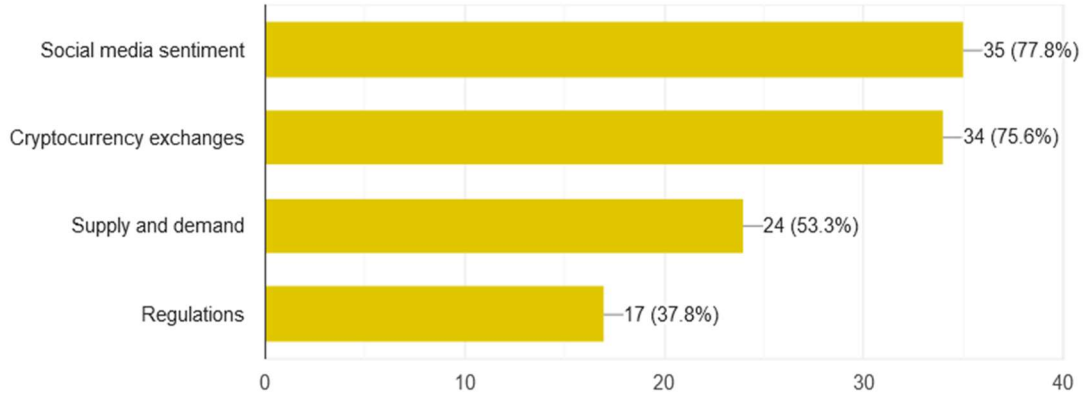
Question	**How will you identify your expertise level in trading or investing cryptocurrencies?**
Aim of question	To understand the knowledge of the person answering the question.
Outcome	

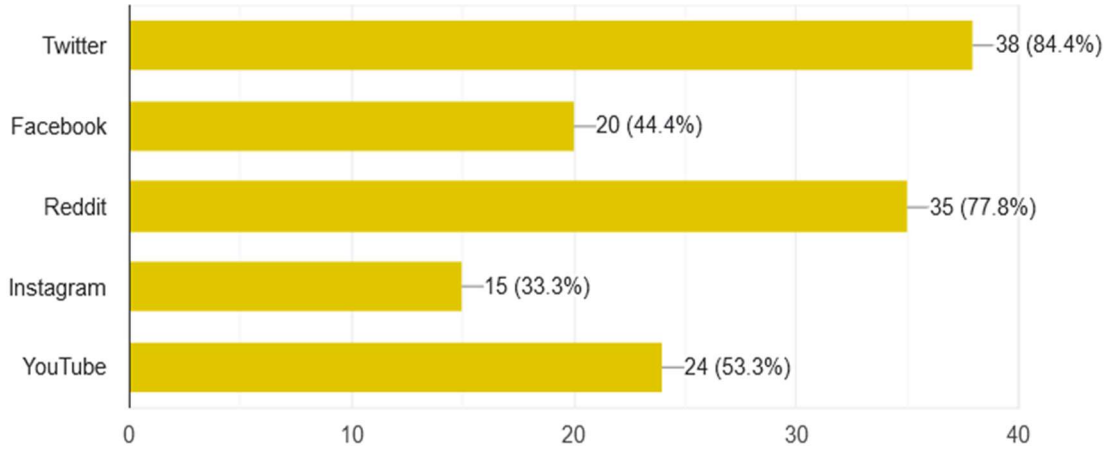
Findings	This questionnaire was conducted using a targeted audience who had knowledge about cryptocurrencies and majority of the users who filled the form have an advanced or intermediate level knowledge in trading cryptocurrencies while the others have just started trading or are experienced veteran Cryptocurrency traders.
Concluding remark	From this analysis the author has understood that most of the users who filled the form are considerably experienced crypto traders and have sound knowledge regarding cryptocurrencies.
Question	**What are the cryptocurrencies that you have tried to trade or invest?**
Aim of question	To know the type of cryptocurrencies the person has tried to trade or invest and to find out the trending coins in the market.

Outcome



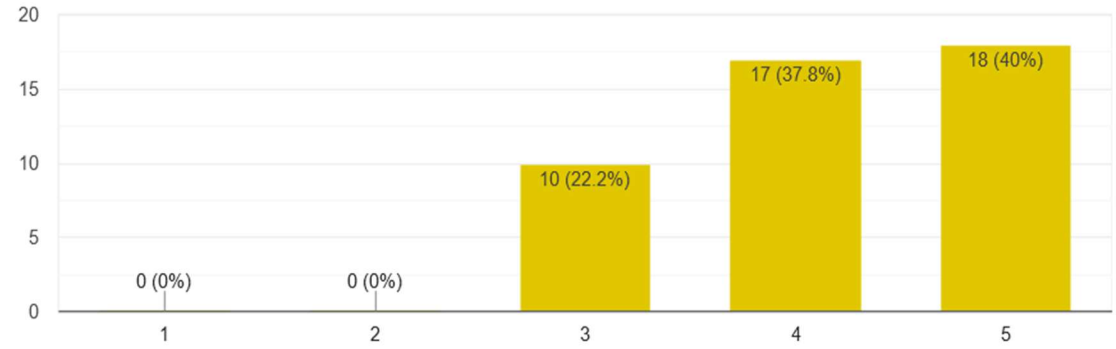
Findings	Majority of the traders have traded Bitcoin and Ethereum more than the other alt coins. Binance coin and Dogecoin are the other best traded coins after the above-mentioned two coins and the rest of the coins Matic, XRP and Cardano aren't traded much compared to the rest of the coins.
Concluding remark	The author understood that the most popularly traded coins are Bitcoin and Ethereum.

Question	**Since the prices of cryptocurrencies are very volatile and change rapidly, what do you think has a major influence for the above-mentioned phenomena?**
Aim of question	To know what factors, affect the price of a Cryptocurrency.
Outcome	
Findings	As mentioned in the aim this question was used to find what factor had the biggest impact in affecting the price of Cryptocurrency and most of the users consider the social media sentiment to be the main factor which affects the prices of cryptocurrencies closely followed by the crypt exchanges, which also played a major role in drastic price differences in the crypto world in the last few months.
Concluding remark	It's satisfying to see that most of the users consider social media to be a major cause since the project plans to use the social media sentiment to develop the model.
Question	**What type of social media platforms do you use to get information about cryptocurrencies and cryptocurrencies prices?**
Aim of question	To identify which social media the person uses to gather information regarding cryptocurrencies.
Outcome	



Findings	Majority of the users have used Twitter or Reddit as their primary social media platform to gather information regarding Cryptocurrency. While YouTube is used by more than 50% of the users the other two social medias (Facebook and Instagram) aren't used by a considerable number of users to gather information.
Concluding remark	From this the author understood that the best place to gather data regarding Cryptocurrency is either Twitter or Reddit.
Question	**Do you think a prediction system which uses social media sentiment, will help in overcoming the problem of predicting the price of cryptocurrencies?**
Aim of question	To identify the persons interest in this project and the importance of the technical contribution of the project

Outcome



Findings	The users consider the prediction system to be very much useful in predicting the price of cryptocurrencies.
Concluding remark	The overall prediction system which will be developed and the technical contribution which will be achieved by completing the project is considered to be extremely beneficial.
Question	**What other functionalities or features would you like to see in the Cryptocurrency price prediction system?**
Aim of question	Additional non-functional requirements that can make the system more user friendly
Findings	Most of the additional requirements mentioned by the users are to include historical price of cryptocurrencies, develop a robust and safe system, the user interface must be user friendly, and the system needs to show prediction graphs.
Concluding remark	The additional requirements will be a valuable addition to the system and the author plans to implement those requirements if time permits.

4.5.2 Brainstorming

Table 10 Brainstorming

Finding	**Description**
Using more than one social media platforms to gather data	When brainstorming to develop different ideas to incorporate into the proposed system the author with the help of the domain experts found out that most of the previous research models used one social media platform either Twitter or Reddit to gather data regarding cryptocurrencies instead of using both or more than two social media platforms. By using both Twitter and Reddit and if possible, by incorporating a third social media platform, the data gathered form the platforms will be more which will then be useful for creating a more optimised model to predict the prices of cryptocurrencies.

Predicting multiple coins	Most of the previous work done regarding Cryptocurrency price prediction involve only predicting the price of Bitcoin or Ethereum and very rarely the application allows to predict two different cryptocurrencies. Therefore, by creating an application which allows the user to predict more than one Cryptocurrency price will lead to the application will appeal to a wider audience of Cryptocurrency users.

4.5.3 Literature

Finding	**Citation**
After reading the literature it was found that creating a highly accurate prediction model can add tremendous value to a Cryptocurrency price prediction system	(Mohapatra, Ahmed and Alencar, 2019)
Using Deep Learning models instead of using Machine Learning models can vastly improve the overall performance of the system.	(Livieris et al., 2020)
Predicting the price of Cryptocurrency one hour into the future has not been tried or tested	(Ider, 2022)
Using several Deep Learning models can Benefit the system and increase the overall prediction accuracy of the system	(Falcon and Lyu, 2021)
Using 1 hour interval of Cryptocurrency price data can be done as a future work	(Mittal et al., 2019)

4.5.4 Prototyping

Prototyping is the method of creating a system which behaves similar to the final application but having only the core aspects of the proposed system. The prototype creation process was an iterative process which involved tuning the features and parameters, pre-processing the data

and training the machine learning model to get the desired outcome. By developing the prototype the author found out there were many challenges and requirements which emerged that needed to be addressed. The author also found that the accuracy of the system can be improved by incorporating more datasets which are available as open source on the internet. By overcoming the challenges and fulfilling the requirements the performance of the final proposed system performance can be drastically boosted to optimum level and the predictions made will be highly accurate to the actual prices.

4.6 Summary of findings

Table 11 Discussion of findings

No.	Findings	Literature Review	Survey	Brainstorming	Prototyping
1	The model should be able to predict the prices of Cryptocurrency	✓	✓		
2	Use two or more social media platforms to gather the social sentiment	✓	✓	✓	
3	The system should be able to predict more than two crypto coins		✓	✓	
4	The overall system should perform without any bugs or issues		✓		✓
5	The UI of the system must be user friendly		✓		✓

4.7 Context diagram

Prior to development, the system's boundaries and interactions with internal and external components should be established. This is depicted in the diagram below.

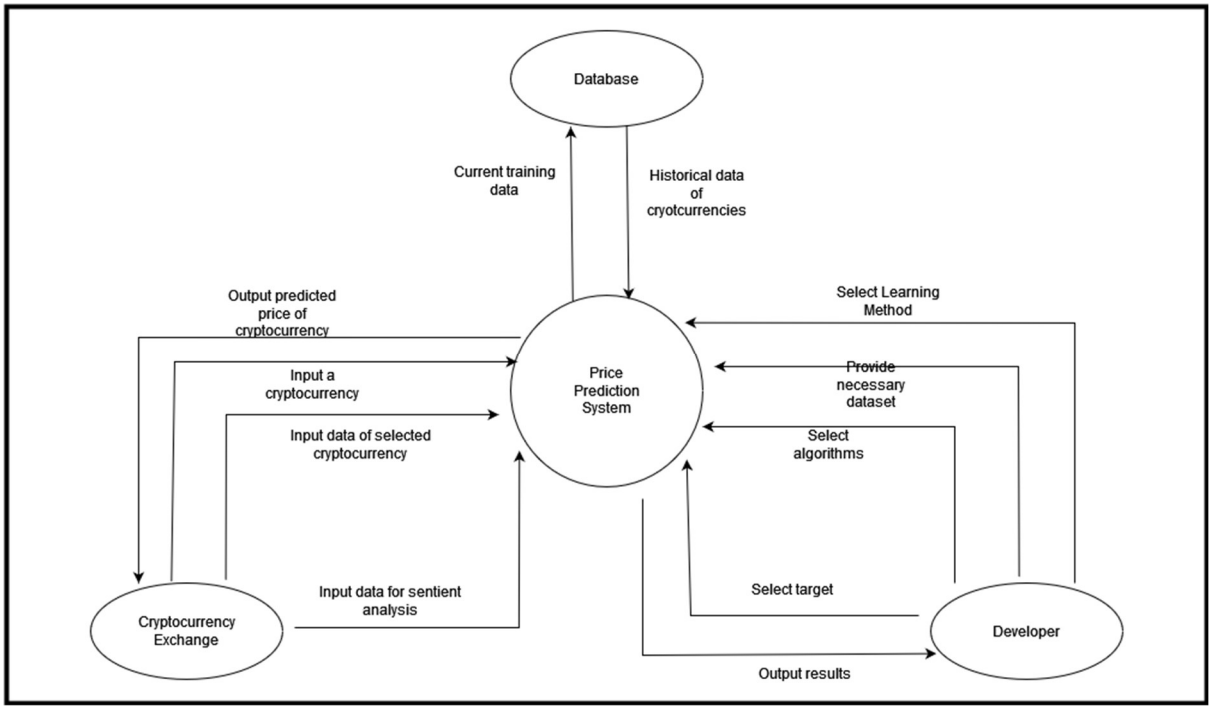


Figure 3 Context diagram

4.8 Use case diagram

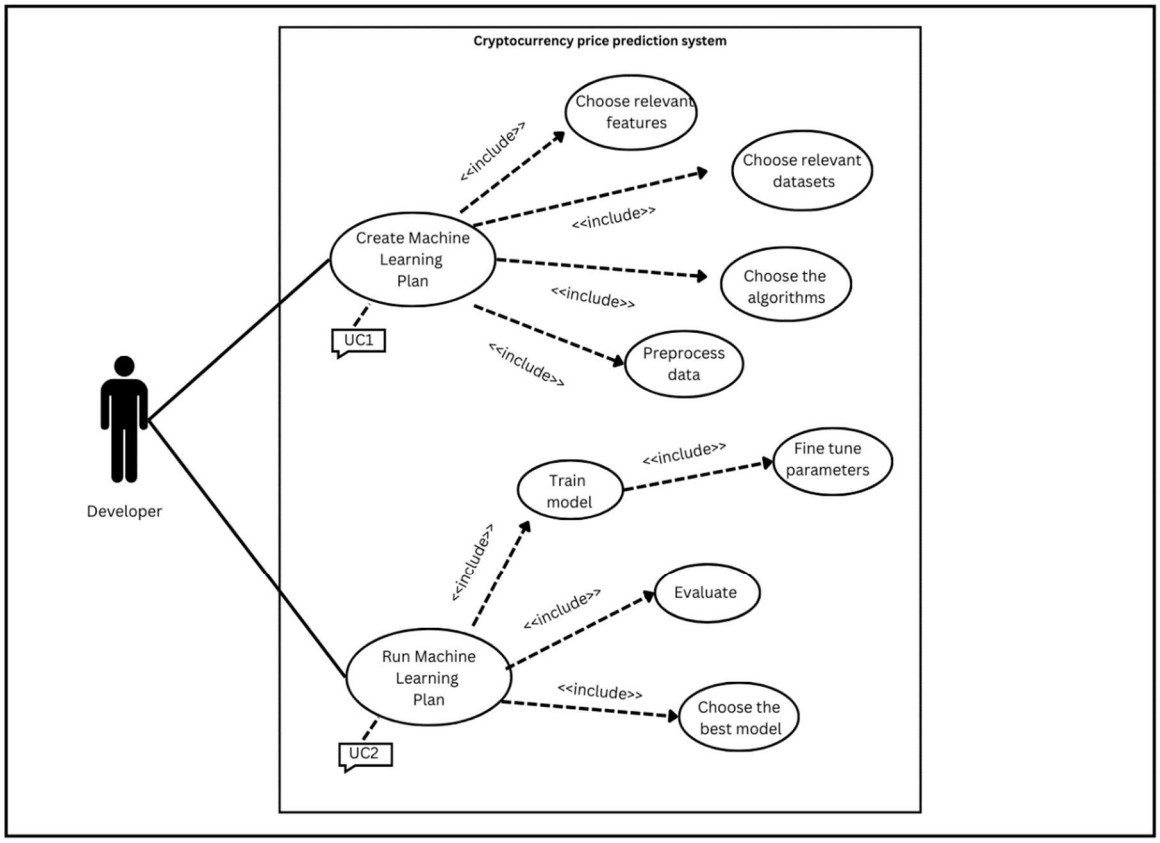


Figure 4 Use case diagram

4.9 Use case description

Table 12 Use case description

Use Case ID	01
Use Case Name	Create Deep Learning Plan
Description	Creating the plan for the machine learning advancements that is required for the building the prediction model.
Primary actor	Developer
Supporting actor (if any)	None
Preconditions	The target variables necessary for the prediction should be decided, the datasets must be in the correct formats for pre-processing and the algorithms for the prediction model should be decided.
Main flow	• Developer devises the plan for creating the machine learning model. • Inputting the necessary datasets to train the model. • Choosing the target features necessary for the prediction to work.
Alternate flow	None
Exception flow	The application failing or crashing due to errors
Post condition	The final plan for the advancement of the machine learning model will be created.

Use Case ID	02
Use Case Name	Run Deep Learning Plan
Description	Executing the plan created for the machine learning advancements.
Primary actor	Developer
Supporting actor (if any)	None
Preconditions	A fool proof and well-developed machine learning plan must be available.
Main flow	Executing the developed machine learning plan.

Alternate flow	Developer pausing or stopping the execution of the machine learning process.
Exception flow	The application failing or crashing due to errors
Post condition	The overall report of the execution and performance metrics will be created regarding the executed process.

4.10 Requirements

This section explains about the functional and non-functional requirements selected for the project by the author. The author has used the MoSCoW technique to set the priority for the requirements and the priority table is given below.

Table 13 Requirements

Priority Level	**Description**
Must have (M)	Core functionality of the system.
Should have (S)	Requirements that are not necessary for the system to work properly but if added will add value to the system.
Could have (C)	Luxury requirements which are optional and do not add much value to the system.
Will not have (W)	The requirement that will not be available at the moment and not considered to be available in the system as well.

4.10.1 Functional requirements

Table 14 Functional requirements

Requirement ID	**Requirement Description**	**Priority Level**
FR01	The user must be able to input the Cryptocurrency and sentiment analysed data to the system	M
FR02	The pre-processing should be done correctly for the data	M
FR03	The processed data must be passed to the model for prediction	M
FR04	The system must be able to predict the price of the given Cryptocurrency without any errors	M

FR05	The system should predict the price of Cryptocurrency one hour into the future	M
FR06	The predicted results must be accurate and should be displayed to the user	M
FR07	The predicted results must be displayed in tabular format in the GUI	S
FR08	The graph of prices vs. time should be displayed in the GUI	S

4.10.2 Non-functional requirements

Table 15 Non-functional requirements

Requirement ID	Requirement Description	Priority Level
NFR01	The accuracy of the proposed system being higher than existing system.	C
NFR02	The overall performance of the system should be better optimised compared to the existing systems	C
NFR03	The user interface of the system must be user friendly and easy to use without any difficulties	C
NFR04	The system should be able to adapt to new technology and frequently be updated	W
NFR05	The system should be prone to security breaches and cyber attacks	W

4.11 Chapter Summary

This chapter gives a thorough evaluation of the system's needs. The Rich Picture Diagram is used by the author to depict the system's interaction with society and the stakeholders' effect on the system, which is represented by the Saunders' Onion model. To gather ideas and opinions from stakeholders, the author additionally applies requirement gathering strategies such as literature research, survey, brainstorming and prototyping. Finally, the gathered information is then used to define the system's use cases, functional and non-functional requirements.

5. Social, Legal, Ethical and Professional Issues (SLEP)

5.1 Chapter Overview

The purpose of this section is to provide a clear understanding of the potential social, legal, ethical, and professional challenges that could occur and to ensure that the research was conducted in accordance with ethical principles and professional standards.

5.2 SLEP Issues and Mitigation

Table 16 SLEP Issues and Mitigation

Social	**Legal**
• The survey conducted was anonymous and personal data such as name, age and gender were not collected. • .The interviewers were informed of the reason for the interview before hand and their consent was taken to add their names to the evaluation.	• The tools, programming languages and frameworks used were open source. • The system does not require personal data to make predictions.
Ethical	**Professional**
• The reference and citations are documented in a proper way. • The reason for the survey was conveyed to the participants before conducting it	• All the tools used throughout the development of the system are legal. • No pirated software were used to develop the system. • None of the data used in the project was altered and high level of research standards were followed while doing the research.

5.3 Chapter Summary

This chapter outlines the SLEP (Social, Legal, Ethical and Professional) issues encountered during the research and the corresponding mitigation strategies, highlighting the focus on identifying and addressing potential challenges throughout the various stages of the study.

6. Design

6.1 Chapter overview

This chapter describes the design choices and solutions for the proposed model that were based on the requirements obtained. High-level and low-level concept diagrams, such as component diagrams, data flow diagrams, system process flowcharts, and UI mock-ups, are included in the design. These design aspects paint a clear image of how the design goals will be met, as well as the reasons behind the design decisions made.

6.2 Design Goals

The design goals for the proposed system are mentioned in the table below.

Table 17 Design Goals

Design goal	**Description**
Correctness	The main purpose of the application should be to give the required output with very few or no errors at all. To achieve this the program correctness must be satisfied and it should be of the highest possible level. The system predicting the price of the Cryptocurrency to be accurate as possible will help the user make informed decisions.
Performance	Any application's performance is critical, and it is important to reduce mistakes or delays at both the logical and presentation ends of the system. Building numerous machine learning models at the same time requires a substantial amount of time and computer power; thus, minimizing the use of computational resources will improve the performance of the suggested solution.
Scalability	The world of crypto is always evolving and the prices of cryptocurrencies will always change with extreme volatility. Therefore, the system should be able to accommodate more new data to improve its performance in the long run.
Reusability	It is important that the system should be a reusable system and not a system which cannot be reused at all. A reusable system allows other developers to build on top of existing systems to improve the overall system.

Adaptability	The proposed system must be able to adapt to new data, technology and requirements since everything keeps evolving day by day. Therefore, adaptability should be an essential aspect of the system so it stays relevant even in the future.

6.3 System Architecture Design

6.3.1 Architecture diagram

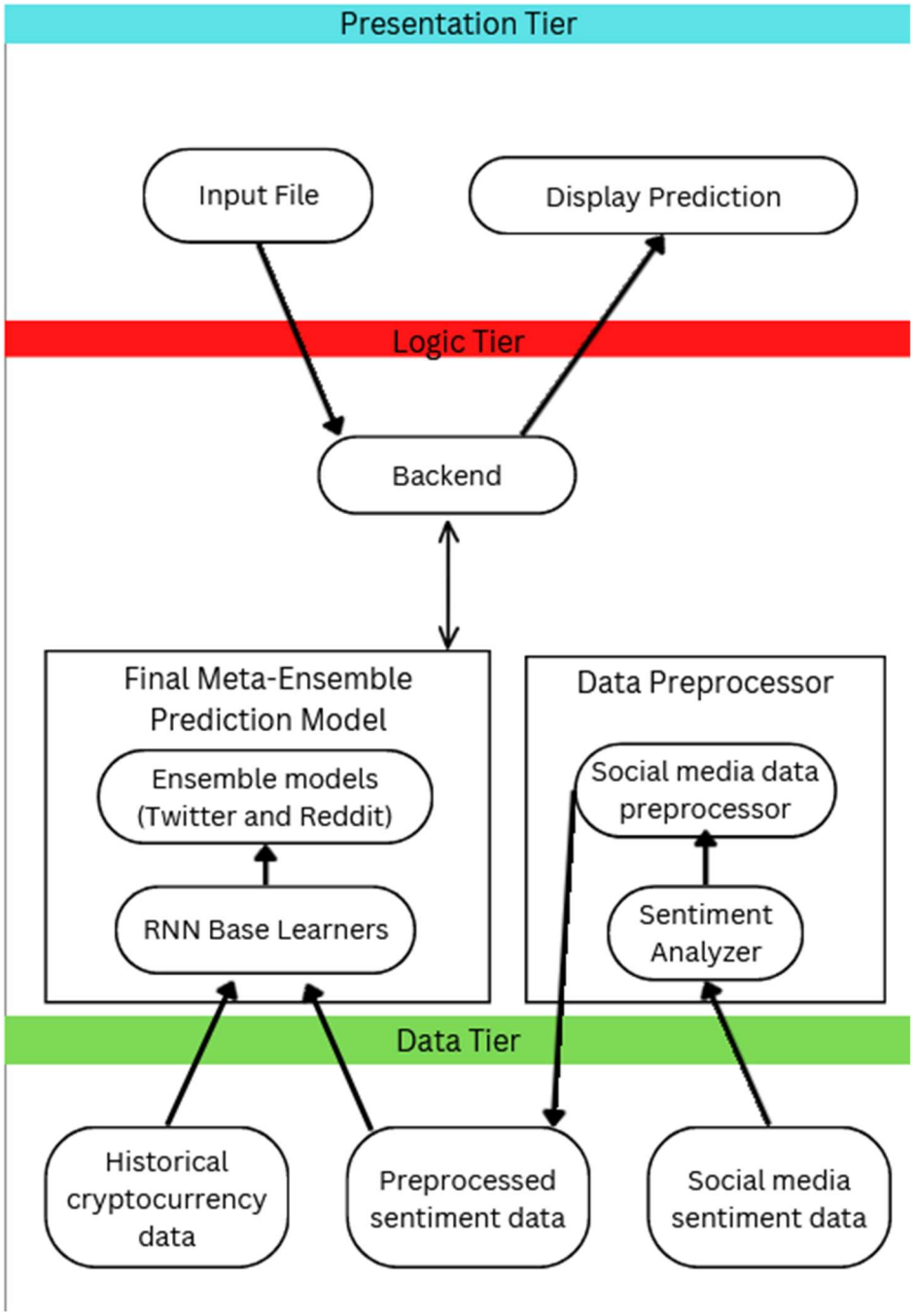


Figure 5 Architecture diagram

6.3.2 Discussion of tiers

Presentation Tier

The presentation tier is what the user will be interacting with when using the application.

1. Input File: The UI will prompt the user to input the csv file to predict the future price.
2. Display Prediction: The UI will display the predicted price of the Cryptocurrency.

Logical Tier

The logical tier is where all the process and the core aspect of the application is located.

1. Backend: The interface which helps in the communication between backend services and frontend.
2. Data Pre-processor: The interface where the sentiment gathered from the social media will be used to create the pre-processed sentiment data.
    a. Sentiment Analyzer: Important in extracting the necessary data from the gathered social sentiment data.
    b. Social Media Data Pre-processor: Responsible for separating and classifying data which will be needed to create the sentiment data.
3. Final Meta-Ensemble Prediction Model: The final model which is built by combining the best Ensemble models in a stacking Ensemble.
    a. RNN Base Learners: The Deep Learning prediction models that will be used to learn from the processed data.
    b. Ensemble models: The model which will be built using the RNN Base Learners by stacking Ensemble.

Data Tier

The data tier is where all the data required for the proposed system is located.

1. Historical Cryptocurrency Data: The data that is required to train the recommender models which will be collected from open-source platforms and crypto exchanges.
2. Pre-processed Sentiment Data: The data that will be gathered by using NLP techniques which will be used to create the social media sentiment-based recommender.
3. Social Media Sentiment Data: The raw data which will be collected from social media platforms.

6.4 Detailed Design

The author has stated below the design paradigm that will be followed throughout the entire development process of creating the proposed system.

6.4.1 Choice of design paradigm

The author has decided to follow the Structured System Analysis and Design Method (SSADM) to develop the proposed system. The SSADM is a development methodology that follows a structured approach to analyse and design a proposed system from scratch. The development methodology also follows six structured phases such as gathering requirements, feasibility study, analysis, logical design, prototyping and finally implementation which helps in developing a properly structured and organized design. Previously in the Methodology chapter the author has mentioned to use Object Oriented Design Method (OOAD) to develop the proposed system, but the design paradigm has been changed to the new methodology due to the following reasons mentioned below,

1. The time period given for the research is short therefore following the OOAD method will be redundant due to the time strain.
2. Since the project involves the usage of machine learning and data science aspects more, the usage of SSADM methodology seems more optimised rather than using OOAD.

6.4.2 Data Flow Diagram

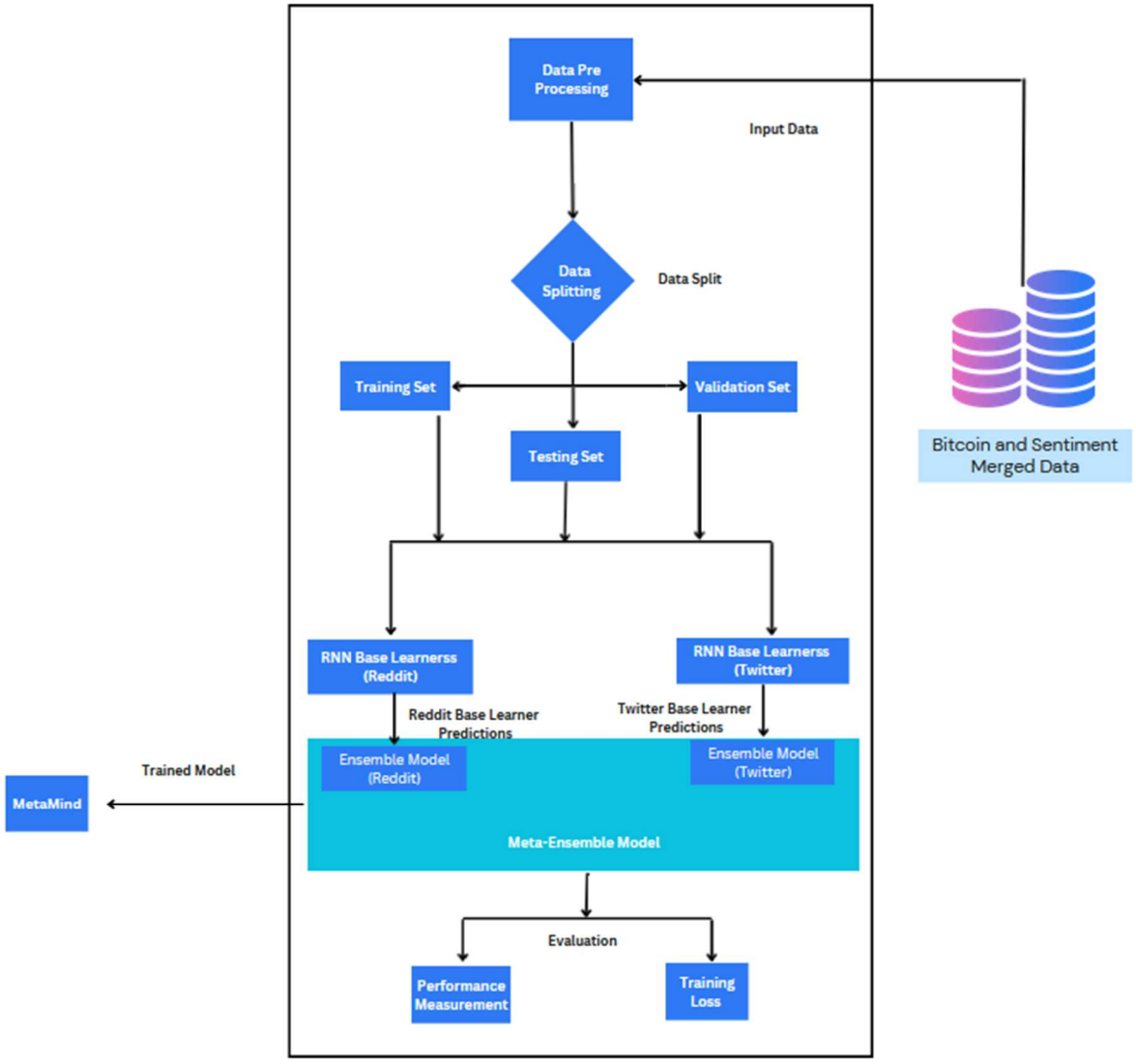


Figure 6 Data Flow Diagram

6.4.3 UI Design

The UI design can be found in the Appendix C.

6.5 Design diagram

6.5.1 System process flow chart

The overall process flow of the system and the selection procedure are depicted in the diagram given below. The diagram explains how the implemented system is expected to work.

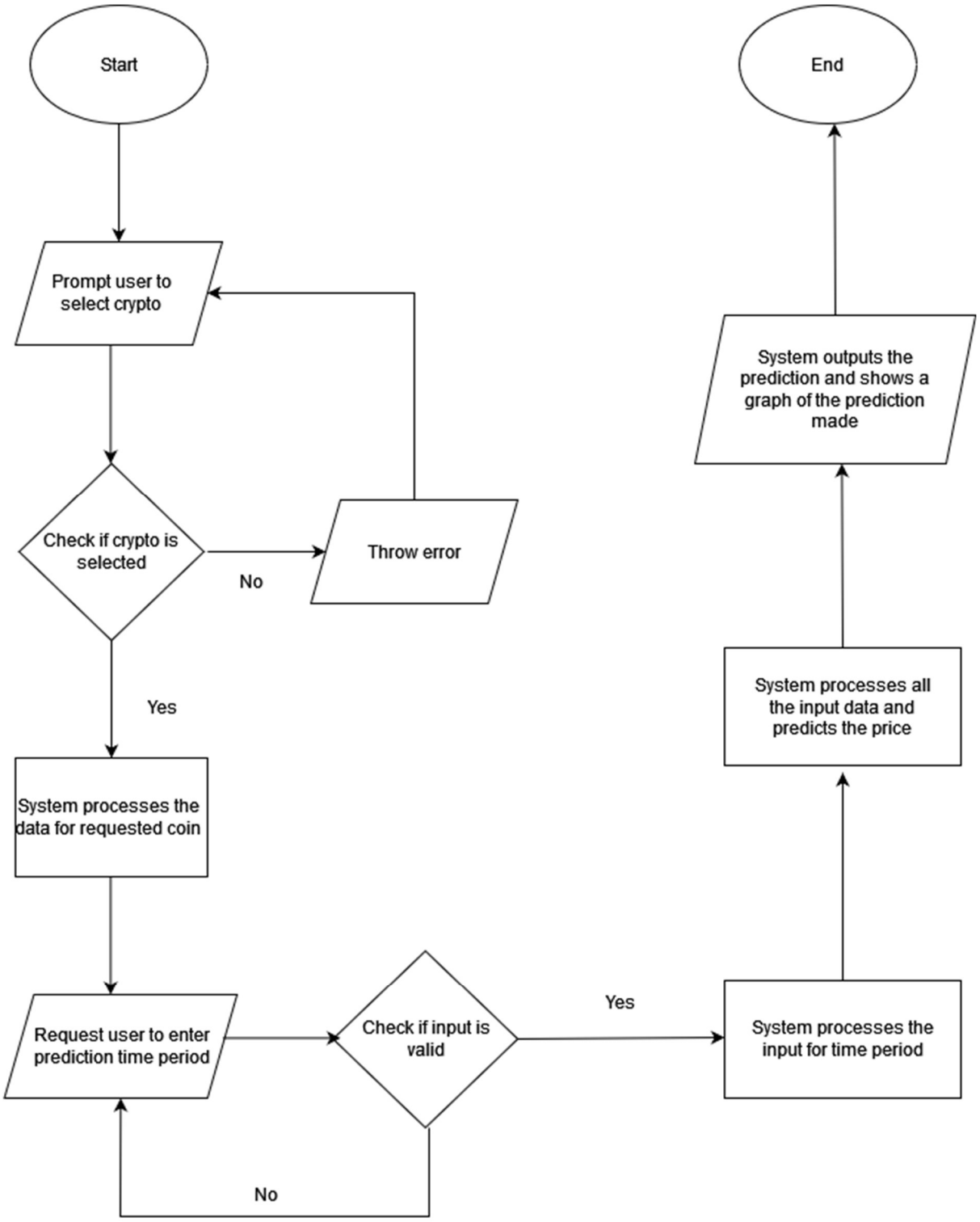


Figure 7 System process flow chart

6.6 Chapter summary

The chapter discusses about the design goals of accuracy, scalability, usability, performance, and flexibility. The system architect is then discussed by the author explaining the purpose of

each tier and its overall process. The author has also discussed about the reason for choosing the SSADM design paradigm. The Data flow diagrams were also used to show data flow across components, while flowcharts were used to create the entire workflow. Finally, the chapter concludes with the display of mock - ups for the UI design.

7. Implementation

7.1 Chapter overview

This chapter focuses on the intended system development, comprising of the information on the technological stack, programming languages, and development tools needed to create the prototype. The author will explain their choices and also talks about transforming design ideas into code, using libraries and datasets, as well as essential code samples and their explanations.

7.2 Technology Selection

The section below discusses about the technologies, frameworks, programming languages and IDE's chosen for the development of the proposed system.

7.2.1 Technology stack

Figure 8 Technology stack

The author has chosen Windows operating system as the default development platform since it is widely used by many people, user friendly and easy to use as well. Other operating systems such as MacOS and Linux will be used for research purposes.

7.2.2 Data-set Selection

The project chosen by the author for development is a data science project, therefore data is an important aspect of the project, and a fair amount of data is necessary to build the machine learning models which will be used by the users to predict the prices of cryptocurrencies.

The data which were necessary are given below,

1. Historical price data of the cryptocurrencies
2. Sentiment data from social media

The historical price data of the cryptocurrencies is required to train the machine learning model. The author has chosen to predict the price of Bitcoin one hour into the future, therefore it was necessary to gather the required dataset for the above-mentioned Cryptocurrency but after browsing the internet and searching the open-source databases like Kaggle, the author found out that the required data is not available online. Therefore, the author plans on collecting the data from scratch.

The sentiment data from social media is necessary for the testing of the machine learning model which also effects the technical contribution as well. The author plans to use web scraping techniques and python libraries to gather the necessary data from one of the most popular social media platforms which is Reddit since the data necessary for Twitter is readily available online in open-source dataset websites like Kaggle. These two popular social media platforms were chosen since the topic of Cryptocurrency is highly discussed in a daily basis.

7.2.3 Development framework

Table 18 Development framework

Framework	Description about framework	Pros	Cons

Flask	The flask framework is a lightweight web framework which is used to create medium sized web application using the Python language.	• Flask is a lightweight framework which is easy to understand and develop. • It is a very flexible framework allowing developers to create small and simple or large and complex web applications. • The framework is easy to learn to compared to other complex frameworks	• A built-in admin interface is not present in the framework. • The framework is less comprehensive compared to other frameworks like Django
Django	Django is a high-level Python framework which follows the MVT (Model View Template) architecture pattern, and it is used to develop secure and robust websites.	• Django supports rapid development which allows it to create web applications faster. • Django is a more secure framework since it has cross-site scripting and	• People who are new to programming or have never used Django before will find it overwhelming to learn the concepts of the framework.

		protects the web applications from SQL injection attacks. • Django has a large community of builders.	• The performance offered by the web applications created by the framework maybe slower compared to other frameworks

Since the author is using Python as the main programming language he had to choose between Flask or Django to implement the prototype and after carefully comparing the pros and cons of both frameworks, Flask was chosen as it is easier to learn and lightweight compared to Django.

7.2.4 Programming languages

The two main programming languages used for machine learning are Python and R. Both languages are good in their own way, but the author has chosen to use Python language due to the following reasons,

- Python has a vast expansion of libraries which supports and helps in creating machine learning models (TensorFlow, PyTorch, Scikit-learn).
- Python has several libraries for data analysis which allows the developer to easily pre-process data and build features for machine learning models (Pandas, NumPy).
- Python works well with a variety of tools and technologies, making it easy for developers to integrate Python with other technologies to build machine learning models.
- Python is an open-source programming language, therefore it has many resources which are available online.
- Since Flask is being used as the framework for building the web application, the author has chosen to use Python as the main programming language which makes the development process simple and less difficult.

From the vast selection of front-end technologies available in the tech stack, the author planned to use JavaScript for the system's frontend to provide an interactive and appealing user experience. This decision was taken to create a dynamic user experience and to make system interaction more engaging and user friendly.

7.2.5 Libraries

Table 19 Libraries

Library	Justification for choosing the library
TensorFlow	The TensorFlow library enables the author to build the machine learning models as well as it also offers additional benefits like visualizing and debugging the training process while building the models.
Pandas	The Pandas library enables to use a wide rang of functionalities like cleaning, sorting, filtering, transforming and manipulating data, which is necessary for analysing the data which is used to train the machine learning model.
NumPy	NumPy is a library which is used for numerical computing. The library was chosen because it works well with the other libraries which were chosen to do scientific computing and data analysis.
Matplotlib	Matplotlib is a library which is used to plot graphs and create graphics. The author chose this library since it can be integrated with the other libraries for data visualization which will help in analysing the data that is being used in this project.
Scikit-learn	Scikit-learn is an open-source machine learning library used in the development of machine learning application and models. Classification, regression, clustering, dimensionality reduction and model selection are some of the algorithms provided by this library and the author plans to use this library for the range of algorithm it offers for training the machine learning model.

NLTK	NLTK (Natural Language Toolkit) is an open-source library for building Python programs to work with human language data. It provides a wide variety of language processing libraries, tools, and data sets, making it a valuable resource for natural language processing.

7.2.6 IDE

Table 20 IDE

IDE	**Reason for choosing the IDE**
VSCode	It is an open-source IDE which is commonly used by all the developers since it free and light weight. VSCode also provide extensions for multiple programming languages which makes it a more versatile IDE compared to other IDEs.
Google Colab	Google Colab is a versatile and user-friendly tool for developing and testing machine learning models. Its cloud-based software which allows users to work remotely online, can access the platform form different devices and provides space to store the code base as well. Furthermore, the online IDE provides access to GPUs and TPUs, allowing users to quickly and efficiently train models on big datasets.

7.2.7 Summary of technology selection

Table 21 Summary of technology selection

Component	**Tools**
Programming language	Python
Libraries	TensorFlow Pandas NumPy Matplotlib Scikit-learn React
Framework	Flask
IDE	VSCode

	Google colab
Version control	Git GitHub

7.3 Implementation of core functionality

The main goal of the project is to build a Prediction system than can give the best results to the user, therefore the gathered raw data cannot be used to train the system. So, the gathered data needs to be pre-processed in a better to get the best features that can be used to build an optimised system.

Gathering data from Reddit

```
def get_pushshift_data(query, after, before, sub):
    url = 'https://api.pushshift.io/reddit/search/submission/?q='+str(query)+'&size=1000&after='+str(after)+'&before='+str(before)+'&subreddit='+str(sub)
    print(url)
    sleep(1) # pushshift has a rate limit, so sleep for 1 second before making the request
    try:
        r = Request(url)
        r = urlopen(r)
        data = json.loads(r.read().decode())
        data = data['data']
    except JSONDecodeError:
        print("JSONDecodeError occurred, retrying...")
        data = []
    return data
```

Figure 9 Gathering data from Reddit

```
def write_subs_to_file(filename):
    upload_count = 0
    if os.path.exists(filename):
        keep_header = False
    else:
        keep_header = True

    with open(filename, 'a', newline='') as file:
        a = csv.writer(file, delimiter=',')
        headers = ['post_id', 'title', 'selftext', 'url', 'author', 'score', 'publish_date', 'num_of_comments',
                   'permalink', 'flair']
        if keep_header:
            a.writerow(headers)
        for sub in sub_stats:
            a.writerow(sub_stats[sub][0])
            upload_count += 1
        # print(str(upload_count) + ' submissions have been uploaded')
```

Figure 10 Gathering data from Reddit

```
# Download reddit posts from sub_reddit with keywords given by key_word
sub_reddit = 'bitcoin'
key_word = 'bitcoin'

output_filename = '/content/drive/MyDrive/FYP/dataset/test_data/reddit_data_test.csv'
# Search all the posts from start_date to end_date overall
start_date = datetime.datetime(2023, 4, 29, 0)
end_date = datetime.datetime(2023, 5, 2, 0)

# In each itration get reddit posts for one day, to avoid getting blocked by server
one_day = datetime.timedelta(hours=24)
after_date = start_date
after = str(int(after_date.timestamp()))
before_date = start_date + one_day
before = str(int(before_date.timestamp()))
```

Figure 11 Gathering data from Reddit

The above images represent the codes used to extract data from Reddit using the 'Pushshift' API to gather necessary data from the social media platform. The date range was set to 2021 1st of January to 2021 31st of December and the Bitcoin subreddit was used to gather the data.

Gathering data from Binance

```
import pandas as pd
from binance.client import Client

# Replace with your API keys
api_key = 'API_KEY'
api_secret = 'API_SECRET'

client = Client(api_key, api_secret)

# Define the start and end dates for the data
start_date = '2023-04-29'
end_date = '2023-05-02'

# Get the data from the Binance API
klines = []
interval = Client.KLINE_INTERVAL_1HOUR
symbols = client.get_all_tickers()
symbols = [symbol['symbol'] for symbol in symbols]
if 'BTCUSDT' not in symbols:
    print('BTCUSDT symbol not found')
else:
    symbol = 'BTCUSDT'
    while start_date < end_date:
        temp_date = pd.to_datetime(start_date) + pd.Timedelta(days=1)
        temp_date = temp_date.strftime('%Y-%m-%d')
        temp_klines = client.get_historical_klines(symbol, interval, start_date, temp_date)
        klines += temp_klines
        start_date = temp_date

# Convert the data to a pandas dataframe
columns = ['timestamp', 'open', 'high', 'low', 'close', 'volume', 'close_time', 'quote_asset_volume', 'num_trades', 'taker_buy_base_asset_volume', '
df = pd.DataFrame(klines, columns=columns)
df['timestamp'] = pd.to_datetime(df['timestamp'], unit='ms')
df.set_index('timestamp', inplace=True)

# Save the data to a CSV file
df.to_csv('bitcoin_data_2023.csv')
```

Figure 12 Gathering data from Binance

The code segment above is used to gather data from the Binance Exchange. The API was provided by Binance, and the data was gathered from the timeline between January 1st 2021 and December 31st 2021.

Cleaning, Pre-processing, and Performing Sentiment Analysis on the Gathered data

The cleaning and pre-processing of the gathered data were done using by taking out unnecessary noise words by creating an array of the noise words to be removed and removing links, numbers, special characters, url, hashtags, numbers and emojis.

Sentiment analysis was performed on the cleaned and pre-processed social media sentiment data using the NLTK library, TextBlob and Flair to gather all the necessary sentiment analysed data which will be incorporated when building the individual datasets.

The images of the code segments to carry out the above processes can be found in the Appendix D.

Implementing the Base Learners (RNN)

Initially after creating the datasets to train the models the features needed to be extracted and then the hyperparameters needed to be tuned so that the data doesn't overfit and ruin the training, therefore the data was split into a train, test and validation split of 70%, 20% and 10%. The images of the code segments that perform the above-mentioned processes are given below.

```
[ ]  # Create the target variable by shifting the 'close' column one hour into the future
     df['target'] = df['close'].shift(-1)

     # Drop missing values
     df = df.dropna()

     # Visualise the dataframe
     df
```

```
[ ]  features = ['pos', 'neg', 'neu', 'close', 'volume']
     df[features]
```

```
[ ]  # Split into features and target
     X = df.drop('target', axis=1).values
     y = df['target'].values.reshape(-1, 1)
```

Figure 13 Features

```
# Split into train, validation, and test sets
train_size = int(0.7 * len(X))
val_size = int(0.1 * len(X))
test_size = len(X) - train_size - val_size
X_train, X_val, X_test = X[:train_size], X[train_size:train_size+val_size], X[train_size+val_size:]
y_train, y_val, y_test = y[:train_size], y[train_size:train_size+val_size], y[train_size+val_size:]

# Scale the data
scaler_X = MinMaxScaler()
X_train_scaled = scaler_X.fit_transform(X_train)
X_val_scaled = scaler_X.transform(X_val)
X_test_scaled = scaler_X.transform(X_test)

scaler_y = MinMaxScaler()
y_train_scaled = scaler_y.fit_transform(y_train)
y_val_scaled = scaler_y.transform(y_val)
y_test_scaled = scaler_y.transform(y_test)

# Reshape input to be 3D [samples, timesteps, features]
n_features = X.shape[1]
X_train_reshaped = X_train_scaled.reshape((X_train_scaled.shape[0], 1, n_features))
X_val_reshaped = X_val_scaled.reshape((X_val_scaled.shape[0], 1, n_features))
X_test_reshaped = X_test_scaled.reshape((X_test_scaled.shape[0], 1, n_features))
```

Figure 14 Data scaling

The Deep Learning models were used as base learners for the project. Five individual base learners such as Simple RNN, LSTM, GRU, Bi-LSTM, Bi-GRU were used to learn from the features extracted from the two datasets that were created. The images of the code segments for the base learners are given below.

LSTM base learner

```
# Define the model
model_lstm = Sequential()
model_lstm.add(LSTM(80, activation='relu',kernel_regularizer=regularizers.l2(0.001), input_shape=(1, n_features), return_sequences=True))
model_lstm.add(Dropout(0.2))
model_lstm.add(LSTM(60, activation='relu', kernel_regularizer=regularizers.l2(0.001)))
model_lstm.add(Dropout(0.2))
model_lstm.add(Dense(1))
model_lstm.compile(optimizer='adam', loss='mse')

# Fit the model
history = model_lstm.fit(X_train_reshaped, y_train_scaled, epochs=20, batch_size=50, validation_data=(X_val_reshaped, y_val_scaled), verbose=2, shuffle=False)


# Plot the loss and validation loss for lstm model
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Figure 15 Simple RNN

Simple RNN base learner

```
# Define the model
model_rnn = Sequential()
model_rnn.add(SimpleRNN(100, activation='relu', kernel_regularizer=regularizers.l2(0.001), input_shape=(1, n_features), return_sequences=True))
model_rnn.add(Dropout(0.25))
model_rnn.add(SimpleRNN(80, activation='relu', kernel_regularizer=regularizers.l2(0.001), return_sequences=False))
model_rnn.add(Dropout(0.25))
model_rnn.add(Dense(1))
model_rnn.compile(optimizer='adam', loss='mse')

# Fit the model
history = model_rnn.fit(X_train_reshaped, y_train_scaled, epochs=15, batch_size=65, validation_data=(X_val_reshaped, y_val_scaled), verbose=2, shuffle=False)


# Plot the loss and validation loss for simple rnn model
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Figure 16 LSTM

GRU base learner

```
# Define the model
model_gru = Sequential()
model_gru.add(GRU(100, activation='relu', kernel_regularizer=regularizers.l2(0.001), input_shape=(1, n_features), return_sequences=True))
model_gru.add(Dropout(0.2))
model_gru.add(GRU(75, activation='relu', kernel_regularizer=regularizers.l2(0.001)))
model_gru.add(Dropout(0.2))
model_gru.add(Dense(1))
model_gru.compile(optimizer='adam', loss='mse')

# Fit the model
history = model_gru.fit(X_train_reshaped, y_train_scaled, epochs=20, batch_size=60, validation_data=(X_val_reshaped, y_val_scaled), verbose=2, shuffle=False)


# Plot the loss and validation loss for gru model
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Figure 17 GRU

Bidirectional LSTM base learner

```
# Define the model
model_bilstm = Sequential()
model_bilstm.add(Bidirectional(LSTM(100, activation='relu', kernel_regularizer=regularizers.l2(0.001), input_shape=(1, n_features), return_sequences=False)))
model_bilstm.add(Dropout(0.2))
model_bilstm.add(Dense(1))
model_bilstm.compile(optimizer='adam', loss='mse')

# Fit the model
history = model_bilstm.fit(X_train_reshaped, y_train_scaled, epochs=20, batch_size=100, validation_data=(X_val_reshaped, y_val_scaled), verbose=2, shuffle=False)


# Plot the loss and validation loss for bilstm model
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Figure 18 Bi-LSTM

Bidirectional GRU base learner

```
# Define the model
model_bigru = Sequential()
model_bigru.add(Bidirectional(GRU(100, activation='relu', kernel_regularizer=regularizers.l2(0.001), input_shape=(1, n_features), return_sequences=False)))
model_bigru.add(Dropout(0.2))
model_bigru.add(Dense(1))
model_bigru.compile(optimizer='adam', loss='mse')

# Fit the model
history = model_bigru.fit(X_train_reshaped, y_train_scaled, epochs=15, batch_size=100, validation_data=(X_val_reshaped, y_val_scaled), verbose=2, shuffle=False)


# Plot the loss and validation loss for bigru model
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.xlabel('Epoch')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Figure 19 Bi-GRU

Creating the Ensemble Models (Reddit and Twitter)

```
def stacked_ensemble(meta_model, base_preds_train, y_train_scaled, base_preds_test, y_test_scaled, save_path=None):
    # Fit the meta-model
    meta_model.fit(base_preds_train, y_train_scaled)

    # Generate predictions from the meta-model
    meta_preds_test = meta_model.predict(base_preds_test)

    # Reshape the predictions to be two-dimensional
    meta_preds_test = meta_preds_test.reshape(-1, 1)

    # Inverse transforming the scaled data
    meta_predict_test_inv = scaler_y.inverse_transform(meta_preds_test)
    y_test_actual = scaler_y.inverse_transform(y_test_scaled)

    # Calculate the mean squared error, mean absolute error, and r2 score
    mse = np.sqrt(mean_squared_error(y_test_actual, meta_predict_test_inv, squared=False))
    mae = mean_absolute_error(y_test_actual, meta_predict_test_inv)
    r2 = r2_score(y_test_actual, meta_predict_test_inv)

    # Save the model to a file
    if save_path is not None:
        with open(save_path, 'wb') as file:
            pickle.dump(meta_model, file)

    # Return a dictionary containing the model, model predictions, and evaluation metrics
    results = {
        'model': meta_model,
        'mse': mse,
        'mae': mae,
        'r2': r2
    }
    return results
```

Figure 20 Ensemble model function

The Ensemble models were created individually using the five base learners as inputs for the relevant dataset (Bitcoin with Reddit and Bitcoin with Twitter). The regression learners like

Random Forest, Linear Regression, Elastic Net Regression, Support Vector Regression, Decision Tree Regression, Gradient Boosting Regression and Ada Boosting Regression were used to create the Ensemble models. A common function was used to train and test the models and the image for the code segment for the function is given above.

Creating the Final Meta-Ensemble Model (Reddit and Twitter combined)

```
def evaluate_meta_model(final_meta_model, meta_model_train_preds, meta_model_test_preds, y_train_scaled, y_test_scaled):
    # Train the meta-model
    meta_model = final_meta_model
    meta_model.fit(meta_model_train_preds, y_train_scaled)

    # Generate predictions from the meta-model
    meta_model_test_preds = meta_model.predict(meta_model_test_preds)

    # Reshape the predictions to be two-dimensional
    meta_model_test_preds = meta_model_test_preds.reshape(-1, 1)

    # Inverse transforming the scaled data
    meta_model_test_preds_inv = scaler_y.inverse_transform(meta_model_test_preds)
    y_test_actual = scaler_y.inverse_transform(y_test_scaled)

    # Calculate the mean squared error, mean absolute error, and r2 score
    rmse = np.sqrt(mean_squared_error(y_test_actual, meta_model_test_preds_inv, squared=False))
    mae = mean_absolute_error(y_test_actual, meta_model_test_preds_inv)
    r2 = r2_score(y_test_actual, meta_model_test_preds_inv)

    print(f"RMSE: {rmse:.4f}")
    print(f"MAE: {mae:.4f}")
    print(f"R^2 Score: {r2:.4f}")
    print("Prediction of bitcoin closing price: ", meta_model_test_preds_inv)
```

Figure 21 Meta-Ensemble model function

The above image shows the function that was used to create the different Meta-Ensemble models. The meta model takes the predictions made by the Ensemble models that were created using the different datasets (Reddit and Twitter) as inputs to train and make predictions. After the model makes the predictions, evaluations are conducted to check whether the model is a well performing model.

7.4 User Interface

The user interface is an important part of the project since it will be utilised by the end users. The author developed a web-based application to be used by the end users and the images of the web application is shared in the Appendix D.

7.6 Chapter Summary

In this chapter the author has mentioned the technological aspect of the project that is required to develop the proposed system. Firstly, the author talks about the technology selection that was made for this project like the technology stack, the datasets, the frameworks, the programming languages, libraries and IDEs that were chosen with the explanation of why it was chosen. Finally, the core functionality of the project is discussed which includes the codes which were used to create the prototype.

8.0 Testing

8.1 Chapter Overview

The chapter focuses on the different testing methods followed by the author to validate the performance and accuracy of the prototype. The goals and objectives of conducting the testing, the different testing criteria, model testing, functional testing and non-functional testing are further discussed in this chapter with the scores gathered while training and testing the models.

8.2 Objectives and Goals of Testing

The main purpose of testing a software system is to ensure that it works as expected based on the requirements that were gathered. These requirements include both functional and non-functional aspects of the system and are listed as follows,

- Make sure that the implemented system meets the MoSCoW technique's mandatory and important functional requirements.
- Helps in improving the performance and usability of the system.
- Ensures that the system is absence of bugs and errors.
- Helps the system to adhere to good coding practices.

Testing also helps to identify possible areas of improvement or enhancement for the system in the future.

8.3 Testing Criteria

The system was tested under two criteria to ensure that it matches the intended design and follows the best practices.

- Functional testing – This is the first criterion, which was used to check whether the system performed all the functions as specified by the functional requirements.
- Structural testing – This is the second criterion, which was used to examine the quality of the code and its compliance with the software engineering standards.

8.4 Model Testing

8.4.1 MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), R^2 (R squared)

MAE was used as evaluation matric since it can evaluate how well a prediction system forecasts future values based on the historical data. Lower the MAE score for a model, the better it performs.

RMSE measures the average squared difference between the actual and predicted values of a model. Achieving low RMSE scores determines that the model performs better.

R^2 is a way of measuring how well a machine learning model can predict the outcome based on the input. It's a number that tells us how good the model is at fitting the data. R^2 ranges from 0 to 1 where 0 means the model is not good while 1 means the model is perfect

The best performing models are bolded.

8.4.1.1 RNN base learners (Twitter)

Table 22 RNN base learners

Base Leaner	MAE	RMSE	R^2
Simple RNN	2628.98	55.70	0.7700
LSTM	2010.90	48.64	0.8663
GRU	2146.28	50.22	0.8480
Bi-LSTM	**1180.95**	**39.92**	**0.9393**
Bi-GRU	2065.38	50.80	0.8409

8.4.1.2 RNN bas learners (Reddit)

Table 23 RNN bas learners

Base Leaner	MAE	RMSE	R^2
Simple RNN	2595.28	56.19	0.7618
LSTM	1991.14	49.51	0.8565
GRU	2198.56	50.20	0.8482
Bi-LSTM	**1242.21**	**40.31**	**0.9369**
Bi-GRU	1758.52	47.75	0.8758

8.4.1.3 Ensemble models (Twitter)

Table 24 Ensemble models

Ensemble Model	MAE	RMSE	R^2
Random forest regressor	448.7563	27.5556	0.9862
Linear regressor	**386.7846**	**22.5477**	**0.9939**
Elastic net regressor	2247.4281	52.6150	0.8168
Support vector regressor	1339.0530	45.8028	0.8948

Decision tree regressor	555.6305	28.4381	0.9843
Gradient boosting regressor	455.0534	27.5945	0.9861
Ada boosting regressor	858.8580	36.1743	0.9590

8.4.1.4 Ensemble models (Reddit)

Table 25 Ensemble models

Ensemble Model	**MAE**	**RMSE**	**R^2**
Random forest regressor	448.8232	27.2960	0.9867
Linear regressor	**376.9415**	**22.9562**	**0.9933**
Elastic net regressor	2166.3083	52.9717	0.8119
Support vector regressor	1460.0957	45.5765	0.8969
Decision tree regressor	537.9431	28.8580	0.9834
Gradient boosting regressor	441.4857	27.9509	0.9854
Ada boosting regressor	812.8985	37.2513	0.9540

8.4.1.5 Meta-Ensemble models (Reddit and Twitter Combined)

Table 26 Meta-Ensemble models

Meta-Ensemble Model	**MAE**	**RMSE**	**R^2**
Random forest regressor	400.2379	22.5142	0.9922
Linear regressor	363.3302	21.9440	0.9930
Elastic net regressor	290.634	20.0939	0.9951
Support vector regressor	2204.0325	50.5283	0.8024

Decision tree regressor	1667.2215	43.6854	0.8896
Gradient boosting regressor	**274.1641**	**19.6632**	**0.9955**
Ada boosting regressor	654.5694	29.5107	0.9770

8.4.2 Cross Model Validation Data Test (Justification for producing the Meta-Ensemble model)

The Ensemble models from Twitter and Reddit produced two Linear Regression models that were very similar and in performance and had very similar evaluation scores as well which implied that one of the models should be chosen. Therefore, the author requested the advice of an experienced technical expert in the field of Artificial Intelligence and had extensive knowledge of Machine Learning and Deep Learning. The expert suggested that a cross model data validation test should be performed to determine whether the models fit for their counterpart datasets and perform evaluations on the test.

Table 27 Cross Validation Model Data Test

Ensemble model	Dataset	MAE	RMSE	R^2
Linear Regression (Twitter)	Reddit sentiment data merged with Bitcoin	33.6621	980.3275	0.9693
Linear Regression (Reddit)	Twitter sentiment data merged with Bitcoin	30.1011	707.0960	0.9828

If the evaluation scores were the same one model had to be dropped but the evaluation scores varied by a very big degree which led to the author combining both datasets and Ensemble models to create the final Meta-Ensemble model.

8.5 Benchmarking

Table 28 Benchmarking

Study	**MAE**	**RMSE**	**R^2**	**Models Used**
(Hegde, 2023)	N/A	66.262	0.9997	ANN
(Aslam et al., 2022)	N/A	N/A	0.99	LSTM-GRU

((Ibrahim, 2021))	88.74	173.40	N/A	Stacking of 2 base model (LSTM and GRU)
Proposed Prediction System	274.1641	19.6632	0.9955	Meta-Ensemble model

8.6 Functional Testing

Functional testing was performed to check whether the system met the functional requirements that were identified in the requirement gathering phase.

Table 29 Functional Testing

Test Case	Test Description	Expected results	Actual results	Result
1	Viewing and scrolling through the web application	Can view and scroll through the web application	Can view and scroll through the web application	Passed
2	Uploading a file to generate predictions	File being uploaded and the name of the file being displayed	File being uploaded and the name of the file being displayed	Passed
3	Restricting the upload of any files other than a CSV file	The system only allows the uploading of CSV files	The system only allows the uploading of CSV files	Passed
4	Clicking the prediction button after file upload generating predictions	After uploading the files when the predict button is clicked the predictions are generated	After uploading the files when the predict button is clicked the predictions are generated	Passed

5	While the predictions are being generated a loading animation is displayed	The loading animation is displayed while the predictions are being processed and generated	The loading animation is displayed while the predictions are being processed and generated	Passed
6	Displaying the generated predictions	The generated predictions are displayed in a tabular format	The generated predictions are displayed in a tabular format	Passed
7	Going back to the home page by clicking the back button	The back button in the predictions page when clicked redirects back to the home page	The back button in the predictions page when clicked redirects back to the home page	Passed

8.7 Module and Integration Testing

Table 30 Module and Integration Testing

Module	Input	Expected output	Actual output	Status
Retrieve data	User inputs sentiment analysed Cryptocurrency data	The system gathers the data for pre processing	The system gathers the data for pre processing	Pass
Pre-processing the input data	Input data from the user	The system processes the data for the base models to make predictions	The system processes the data for the base models to make predictions	Pass

Base RNN learners	Pre-processed data	The base RNN learners of make predictions using the pre-processed data	The base RNN learners make predictions using the pre-processed data	Pass
Ensemble models	Predictions of Base RNN learners	The Ensemble models make predictions using the predictions of the base RNN learners	The Ensemble models make predictions using the predictions of the base RNN learners	Pass
Meta-Ensemble model	Predictions of the Ensemble models	The meta model outputs the final prediction of the close price 1 hour into the future	The meta model outputs the final prediction of the close price 1 hour into the future	Pass

8.8 Non-Functional Testing

8.8.1 Performance Testing

The efficiency and usage of resources in training and fitting the model were evaluated to ensure that the system is lightweight and has low resource consumption. The author used their own personal computer to conduct the performance testing, and the specifications of the computer are as follows:

- Operating System: Microsoft Windows 11
- RAM – 8GB
- Processor – Intel® Core™ i7 – 1165G7 @ 2.80GHz

The images for the performance of the application before and after the local server was run can be found in the Appendix E.

8.9 Limitations of the Testing Process

The main limitations faced while conducting the testing was lack of resources and time since testing is a long and very crucial process when developing a Deep Learning-based system and finding relevant research to benchmark against was a difficult task.

8.10 Chapter Summary

This chapter presents the test plan and results of the proposed solution. It starts with the testing objectives and criteria, then describes the model testing, which was conducted using different evaluation metrics, as well as the benchmarking against past research. It also evaluates the functional and non-functional requirements of the prototype and discusses about the module and integration testing as well. Finally, the limitations of the testing methodology is addressed by the author.

9.0 Evaluation

9.1 Chapter Overview

This chapter is dedicated to evaluating the system prototype designed, implemented, and optimized for performance through extensive testing. The evaluation methodology and approach are discussed, followed by the evaluation criteria, which include assessments from technical, domain, and industry experts, as well as self-evaluation. Feedback was received from the target audience, technical and domain experts, and the author's self-evaluation. The evaluation covers both functional and non-functional requirements, and limitations of the evaluation are also discussed.

9.2 Evaluation Methodology and Approach

For the evaluation of the prototype, a qualitative method was employed to gather feedback from domain experts, industry experts, and fellow researchers who are also end-users. The evaluations were conducted online through video calls. Thematic analysis was used to summarize the feedback received from evaluators. While the system was previously tested and analysed quantitatively, this chapter focuses on the qualitative evaluation. The chapter provides an overview of the evaluation methodology and approach, including criteria, self-evaluation, evaluation results, and limitations. The evaluation involved assessments from technical, domain, and industry experts, as well as self-evaluation. Overall, the evaluation was conducted with respect to the requirements gathered in the SRS chapter.

9.3 Evaluation Criteria

The evaluation criteria for the system were determined through thematic analysis of interviews with experts and other research aspects that needed to be assessed. The table describes the evaluation process's criteria, and the deployed system is evaluated using a quantitative technique. The critical elements for analysis are listed in the table to ensure the system is thoroughly evaluated.

Table 31 Evaluation Criteria

Criteria	Evaluation Purpose
The overall concept of the project	To assess the concept and novelty of the research project and obtain feedback on the

	implemented system's ability to address the problem in the domain.
The scope and depth of the project	To ensure that the research has sufficient depth and scope for a final year project in the field of Prediction Systems, specifically Cryptocurrency and Blockchain.
Research contribution	To assess the research contribution made to the domain, technical side of the project and any additional contributions made to other areas as well.
High level architecture and design of the project	To assess the success of the system architecture and ensure that it meets the necessary design standards.
Prototype and solution	The purpose of the prototype evaluation is to assess the proof of concept, confirm the suitability of the development approach, and determine if the implemented prototype is capable of answering the proposed question
Quantitative analysis of the results	To ensure the validation and accuracy of the metrics used for evaluating and analysing the results produced by the research is relevant.
Limitations and possible improvements	To identify areas for future development and improvements, potential avenues for future work related to the research as well as identifying additional features that could enhance the system's reliability.

9.4 Self - Evaluation

The table below is the authors own evaluation regarding the project.

Table 32 Self - Evaluation

Criteria	Self - Evaluation

The overall concept of the project	The overall concept of the research is to develop a novel Cryptocurrency price prediction system using a combination of Deep Learning and Ensemble models, which can accurately forecast the future price of a Cryptocurrency. The system aims to provide a reliable and efficient tool for investors and traders to make informed decisions
The scope and depth of the project	The scope and depth of the developed price prediction system cover the implementation of advanced algorithms for accurate prediction, a comprehensive analysis of Cryptocurrency market trends, and the incorporation of multiple data sources to achieve reliable results.
Research contribution	The technical contribution to the research done by the author is building a novel price prediction system which uses Deep Learning and Meta-Ensemble approach to predict the future price of Cryptocurrency. The model significantly increasing the accuracy of the prediction which can potentially provide valuable insight for traders and investors can be highlighted as a contribution to the domain.
High level architecture and design of the project	The system architecture and design were planned out carefully and developed to adhere to necessary development standards.
Prototype and solution	A significant amount of effort has been put into developing the proposed price prediction protype, starting from the data collection to data pre-processing, using powerful Deep Learning models as base learners to learn

	using the processed data and finally building the Ensemble models, which were further enhanced into a Meta-Ensemble model that showed promising results when evaluations were conducted.
Quantitative analysis of the results	Looking at the past research which were done on the similar domain and the models that were used to build the overall system the relevant evaluation metrics (RMSE, MAE, R^2) were used to evaluate and analyse the overall performance of the system.
Limitations and possible improvements	The implemented Cryptocurrency price prediction system has potential for future growth by expanding its features and incorporating other platforms which share information about Cryptocurrency like discord and telegram, and adding Cryptocurrency news sentiment can further enhance its accuracy.

9.5 Selection of Evaluators

The evaluators for the project were chosen based on the following categories.

a) Technical Experts – Experts who have sound knowledge in the field of research and have extensive knowledge in the field of Prediction systems, Data Science, Data Engineering, Machine Learning, Deep Learning and Artificial Intelligence.
b) Domain Experts – Experts who have experience in the field of Blockchain, Cryptocurrencies and Crypto Exchanges.
c) Target Audience – The end users of the system who have knowledge in trading and investing Cryptocurrency.

Table 33 Selection of Evaluators

Type	Expert ID	Details	Reason

Technical Experts	EID1	**Porf. Nihal Kodikara** Senior Professor	Choosing Professor Nihal as an evaluator is the best option because of his abundant knowledge in the fields of machine learning and artificial intelligence. His familiarity with the latest advancements and academic background guarantees valuable insights and useful criticism.
	EID2	**Dr. Abeegithan Jeyasothy** PhD in Artificial Intelligence, NTU, Singapore (Senior Quantitative Researcher, WordQuant)	Dr. Abeegithan has a deep understanding of machine learning, artificial intelligence and quantitative analysis and his evaluation will add tremendous value to the technical aspects of this project. His feedback will be useful in identifying potential technical challenges and recommending effective solutions, ultimately helping the author to improve the accuracy and efficiency of this project.
Domain Experts	EID3	**Mr. Amri Shafeek** Co-Founder, Director Neurosens Ltd Director and investor at Numuinn Venture Capital	Mr. Amri has experience in both the practical and investment aspects of the field. He also has very good knowledge in the field of blockchain and has been giving financial advice for a long time. His insights will be valuable since the domain of this project falls into the financial side as well.
	EID4	**Mr. Aslam Shafeek** Managing Partner	Mr. Aslam's experience as a managing partner at Numuinn

		Numuinn Ventures, Blockchain and Cryptocurrency Advisor	Ventures, combined with his expertise in blockchain and cryptocurrency, make him an ideal domain expert for this project. His knowledge of the industry and current market trends can provide valuable insights to the project, particularly in relation to blockchain and cryptocurrency integration, and he is also a long time Cryptocurrency trader as well as an investor, therefore his evaluation will not only provide feedback as domain expert but also a valuable target audience as well.
Target Audience	EID5	**Mr. Sharangan Balathayalan**	Mr. Sharangan has been investing and trading Cryptocurrency for the past one and a half years. He has very good knowledge of how the market works and therefore he has the required knowledge and experience to give valuable insights and constructive criticism regarding the project.
	EID6	**Mr. Nirahulan Suvendaran**	Mr. Nirahulan is a beginner in the field of trading and investing in Cryptocurrency. He has been trading for the past few months and has made decent portfolio. As a beginner, his evaluation regarding the project will be very useful in understanding the complexity of using the system.

9.6 Evaluation Results

9.6.1 Expert Opinion

9.6.1.1 Domain Experts

9.6.1.1.1 Concept

Table 34 Concept

Question	
What do you think of the concept?	
Evaluator ID	**Feedback**
EID3	The concept is great since you're predicting the prices of Cryptocurrency and the fact that you went with Bitcoin as the main currency is commendable.
EID4	Predicting the price of Cryptocurrency, especially Bitcoin is a great idea and the fact that you have chosen Binance, Twitter and Reddit to gather data is brilliant.

9.6.1.1.2 Solution

Table 35 Solution

Question	
What do you think of the solution?	
Evaluator ID	**Feedback**
EID3	The prototype is remarkable and works very well, good job.
EID4	A simple and straightforward prototype but produces accurate and informative results. You have nailed your objective very well.

9.6.1.2 Technical Experts

9.6.1.2.1 Scope

Table 36 Scope

Question
What do you think of the scope?

Evaluator ID	Feedback
EID1	The scope of the project is interesting and technically challenging as well.
EID2	The project has a good scope and considerable amount of depth.

9.6.1.2.2 Architecture of the solution

Table 37 Architecture of the solution

Question	
What do you think of the architecture of the solution?	
Evaluator ID	**Feedback**
EID1	The architecture and the design of the solution are very well defined.
EID2	The architecture is a bit too complex and the overall design for the project is good.

9.6.1.2.3 Implementation of the solution

Table 38 Implementation of the solution

Question	
What do you think of the implementation of the solution?	
Evaluator ID	**Feedback**
EID1	The novel approach of using Deep Learning with Meta-Ensemble is commendable. Even though simple learners are used as Ensemble models the evaluation scores look promising and the model accuracy also looks good. Incorporating visual graphs to show the prediction will be a good addition.
EID2	Using multiple Deep Learning models, Ensemble models to produce a Meta-Ensemble model is a great implementation and the final model provides very good evaluation scores as well. The researcher has conducted cross-model validation to justify the creation of the Meta-Ensemble model as well.

9.6.2 Focus Group Testing

9.6.2.1 Prototype Features

Table 39 Prototype Features

Question	
What do you think of the prototype features of the application?	
Evaluator ID	**Feedback**
EID4	This prototype predicts the price of Cryptocurrency taking all factors. Good project. Very helpful for traders and investors. As an improvement, showing a graph which corresponds to the prediction can add more value to the application.
EID5	This is going to be very useful for us. The system predicts the price of Cryptocurrency with a single button click.
EID6	A very useful product for traders like me.

9.6.2.2 Usability

Table 40 Usability

Question	
What do you think of the usability of the application?	
Evaluator ID	**Feedback**
EID4	A very user-friendly and easy to use interface.
EID5	It's easy to use and very efficient.
EID6	A simple and clean application.

9.7 Limitations of Evaluation

The limited availability of domain experts, the scarcity of evaluators and potential end users in the cryptocurrency space posed a significant challenge in obtaining comprehensive feedback and evaluating the performance of the developed system.

9.8 Evaluation on Functional Requirements

Table 41 Evaluation on Functional Requirements

Requirement ID	**Requirement Description**	**Status**	**Priority Level**

FR01	The user must be able to input the Cryptocurrency and sentiment analysed data to the system	Implemented	M
FR02	The pre-processing should be done correctly for the data	Implemented	M
FR03	The processed data must be passed to the model for prediction	Implemented	M
FR04	The system must be able to predict the price of the given Cryptocurrency without any errors	Implemented	M
FR05	The system should predict the price of Cryptocurrency one hour into the future	Implemented	M
FR06	The predicted results must be accurate and should be displayed to the user	Implemented	M
FR07	The predicted results must be displayed in a tabular format in the GUI	Implemented	S
FR08	The graph of prices vs. time should be displayed in the GUI	Not Considered	S

9.9 Evaluation on Non-Functional Requirements

Table 42 Evaluation on Non-Functional Requirements

Requirement ID	Requirement Description	Status	Priority Level
NFR01	The accuracy of the proposed system being higher than existing system.	Implemented	C
NFR02	The overall performance of the system should be better optimised compared to the existing systems	Implemented	C
NFR03	The user interface of the system must be user friendly and easy to use without any difficulties	Implemented	C
NFR04	The system should be able to adapt to new technology and frequently be updated	Partially Implemented	W

NFR05	The system should be prone to security breaches and cyber attacks	Not Considered	W

9.10 Chapter Summary

This chapter details the evaluation methodologies used in the research project, including the criteria and approach taken. The author critically self-evaluates the project and discusses the feedback received from domain experts, technical experts, and end-users. The evaluators' choices and conclusions are summarized based on pre-defined criteria and discusses the limitations of the evaluation. Finally, the functional and non-functional requirements of the project are examined.

10.0 Conclusion

10.1 Chapter Overview

This chapter provides a summary of the research project and its accomplishments. The author explains the achievement of the research aim and objectives, followed by the utilization of knowledge and new skills acquired during the project. The chapter also addresses the challenges faced during the project period and how they were solved. Additionally, the chapter discusses deviations from the original plan, limitations, and future enhancements. The author also highlights the contribution of the research project to the existing body of knowledge. Overall, this chapter provides concluding statements for the research project and its unique contribution to the research community.

10.2 Achievements of Research Aims & Objectives

10.2.1 Achievement of Research Aims

The aim of this research is to design, develop & evaluate a novel system that will predict the price of a given cryptocurrency for investing purposes by automating the manual decision-making process.

The aim of the research was successfully achieved by designing, developing, and evaluating a novel prediction system which can predict the price of cryptocurrency 1 hour into the future. Many Deep Learning models were used as base learners which were finally combined to Ensemble models which were further combined to the final Meta-Ensemble model.

10.2.2 Achievement of Objectives

The completion status of the research objectives mentioned in the research objectives has been summarized in the table below, marking the achievement of each objective.

Table 43 Achievement of Objectives

Objective	Description	Status
Literature Review	Analyse previous work which was done in the research area and critically evaluate them.	Completed
Requirement Analysis	Gather user requirements necessary for the project and critically analyse them.	Completed

Design	Designing the system and architecture of the proposed solution to solve the problems identified in the project.	Completed
Development	The implementation of the proposed solution to predict the prices of cryptocurrencies.	Completed
Testing and Evaluation	To test and evaluate the prototype model.	Completed

10.3 Utilization of Knowledge from the Course

Table 44 Utilization of Knowledge from the Course

Module	**Knowledge Utilized**
Programming Principles 1	The module which laid the foundation for basic Python programming and introduced all the basic concepts necessary to program using the language.
Programming Principles 2 and OOP	The modules which introduced Java and the OOP paradigm. The modules taught the basics of Java programming and all the concepts in Java programming which was later advanced by incorporating the OOP concepts which is a fundamental skill for any programmers.
Algorithms: Theory Design and Implementation	The module covered the basics of data structures and algorithms which provided the author with the knowledge necessary to design high-performing algorithms and develop data science models, with a focus on algorithm selection based on difficulty, search techniques, and graphs.
Mathematics for Computing	Mathematics is a subject which has its presence in all the fields. The module was very useful for understanding the mathematical notations and computations that were done in the research papers and it was also useful for understanding and building algorithms.
Web Design and Development, Server-side Web Development and Advanced	The module provided a comprehensive understanding of web designing and development, covering the fundamentals of HTML, CSS, JavaScript, and other relevant technologies. This knowledge was instrumental in designing and developing the front-end of the system, enabling the author to design and develop the interaction parts of websites and fulfil user requests.

Server-side Web Development	
Software Development Group Project	This module served as the initial inspiration for the author to work on research by covering all the essential steps from recognizing a problem to designing, developing, and testing a prototype. It also provided a foundation for good research by teaching fundamental aspects such as project management, documentation, and reporting. Additionally, the knowledge gained from this module aided the author in managing the development project for the research-based prototype.

10.4 Use of Existing Skills

Python – The author learned the basics of Python programming through the Programming Principles 1 module. This was helpful in understanding and writing all the codes for the core functionality of the proposed system and the building the APIs.

Machine Learning – The author had an interest in Machine Learning and watched a few courses online to understand the basics of machine learning and the maths related to it. The skill was useful when implementing the proposed system.

Blockchain and Cryptocurrency – During the Industrial placement the author worked as a Blockchain developer and from that was introduced to the technology which sparked the interest in creating the proposed system. The Blockchain technology taught the author how the Cryptocurrency system works and the usage of cryptocurrency. This was very useful in understanding the features that were necessary to build a price prediction system for cryptocurrency.

Web development – During the Industrial placement apart from working as a Blockchain developer the author also worked as a Web development developer which gave the necessary skills to build the frontend of the proposed system.

10.5 Use of New Skills

Deep Learning – The author had no prior knowledge in the field of Deep Learning therefore he had to learn everything from scratch by reading articles, watching videos, and doing courses which taught the basics about Deep Learning and the maths related to it.

Prediction System – The author has never created nor had any experience working with Prediction systems. So, the author had to search the internet to learn about building and using Prediction systems from scratch and many free online resources like Google ML courses, Medium articles and YouTube helped him do that.

Natural Language Processing – The idea of using Natural Language Processing to extract sentiment data from tweets and comments was very new to the author and many Medium articles were referred to finally apply the processing techniques to the gathered data.

Data scraping/ Data gathering – The author had to refer many articles regarding data gathering using APIs to legally gather data that was necessary for building the proposed system.

10.6 Achievement of Learning Outcome

Table 45 Achievement of Learning Outcome

Learning Outcome	Description
LO1 LO2 LO3	The research thought how to conduct the proper research by critically analysing the work done on the area of interest and gathering requirements necessary to conduct the research.
LO4 LO5 LO6	The author had to learn many technical skills from scratch do to not having prior knowledge on them. The author also improved his communication skills by conducting interviews and meeting domain and technical experts
LO8	The entire research process was filled with trials and errors and the author had to adapt to solve the problems which arise while conducting the research. This helped the author develop his problem solving and self-learning skills.
LO7	Managed to complete the research project in the given time frames, deadlines and changing requirements.

10.7 Problems and Challenges Faced

Table 46 Problems and Challenges Faced

Problem/ Challenge	Solution
Cryptocurrency is a relatively new and complex domain that involves the use of digital assets as a medium of exchange.	Reading past research papers and journal articles regarding the domain and the research that was done, the author was able

	to understand the overall complexity of the domain.
Lack of data regarding the historical prices of Bitcoin and sentiment data of Reddit in open-source databases like Kaggle hindered the author from building a proper dataset.	The author had to write codes to gather data online and since the data needed to be legally gathered the APIs from Binance Exchange and Pushshift API were used to gather data regarding historical Bitcoin prices and Reddit comments respectively.
The author found it difficult to benchmark the proposed system since there was no previous research done regarding Meta-Ensemble.	To overcome this problem the author had to search for past research papers which were similar to the research conducted by the author so that proper benchmarking could be done.
Processing data required a vast amount of computational resources (mainly the RAM) and the author didn't have the necessary resources to process the data.	To mitigate this circumstance the author used Google Collaboratory to do all the pre-processing, training, and testing tasks which were required to build the model.
Random power outages leading to low battery and loss of internet connection.	Having a backup UPS to charge the laptop and using the mobile data during power outages helped the author overcome this problem.
Since Google Collaboratory was used to run all the codes and scripts, the extraction of historical Bitcoin data could not be performed since Binance API doesn't support the US based apps or server.	To overcome the challenge the author had to run the scripts locally on the PC to extract the relevant data.

10.8 Deviations

The initial plan of the author was to incorporate multiple cryptocurrencies and Google trends data into the system, but it was heavily underestimated since the time and effort necessary to gather the required data (since access needs to be granted and it can take a long time to get access to required APIs) and pre-process it will not be sufficient compared to the time frame

given to complete the project. Therefore, it became evident that a separate project solely focused on these topics would be necessary.

10.9 Limitations of the Research

Even though the project was completed successfully there were several limiting factors that effected the progress of the project. Some of the major limiting factors are mentioned below.

- Due to the limited time allocated for the research, the author had to manage time effectively, which made it challenging to work and had to learn several technical skills from scratch, which further added to the time constraints in completing the system implementation.
- The current system only accepts csv files with the necessary features to make predictions
- Finding the proper datasets was a difficult task since only Twitter data was available online as an open-source dataset in Kaggle but the Reddit data and historical Bitcoin data needed to be built from scratch.
- Cryptocurrency is not a very popular domain in Sri Lanka and finding the relevant domain experts was a very difficult task.
- The current system was trained using Twitter and Reddit sentiment data only but has the possibility of expanding to use other social media platforms sentiment data.

10.10 Future Enhancements

- Introducing multiple new features from either the historical cryptocurrency data or sentiment data might have a chance of improving the overall performance of the model.
- Testing out other major Deep Learning models other RNN, like CNN and GANs has a chance of improving the accuracy of the model.
- Incorporating more cryptocurrencies into the system can enable the system to predict the price of multiple cryptocurrencies.
- Adding more sentiment data from other platforms like news and articles can help improve the model.
- Creating a cryptocurrency Recommendation system that can be combined with the existing model to create a hybrid model that can predict as well as recommend cryptocurrency will be a huge addition to the system.

10.11 Achievement of the contribution to the body of knowledge

After successfully completing the research, the contributions made by the author for the different domains are discussed below.

10.11.1 Technical Contribution (Price Prediction System)

The author, after countless days and hours of research finally created a novel Price Prediction System that uses Deep Learning and Meta-Ensemble approach to predict the price of Bitcoin 1 hour into the future.

10.11.2 Domain Contribution (Cryptocurrency)

Significantly improving the accuracy of predicting cryptocurrency prices while answering the research gap and paving way for future research on the domain for developing improved cryptocurrency trading strategies.

10.11.3 Additional Contribution

1. Scripts involving the gathering and processing of historical Bitcoin data was created by the author to gather data necessary for the project using the Binance API.
2. A dataset containing the historical data of Bitcoin which was validated by a domain expert is available as a private dataset in Kaggle (see Appendix F) and will be made publicly available after the project viva.

10.12 Concluding Remarks

In conclusion, this project was challenging and complex, but the use of Deep Learning and Meta-Ensemble approaches along with social sentiment data showed promising results in predicting cryptocurrency prices. Extensive testing was carried out to ensure the reliability and accuracy of the model, and positive feedback was received from the evaluators. This novel approach provides a valuable contribution to the cryptocurrency domain and can be useful for investors and traders in making informed decisions. Throughout the project, the highest standards of research were followed, and the resulting model reflects this. Overall, this project represents a significant achievement in the field of cryptocurrency price prediction and paves the way for future research in this area.

References

Tanwar, S. et al. (2021). Deep Learning-Based Cryptocurrency Price Prediction Scheme With Inter-Dependent Relations. *IEEE Access*. Available from https://doi.org/10.1109/ACCESS.2021.3117848 [Accessed 10 October 2022].

Mittal, A. et al. (2019). Short-Term Bitcoin Price Fluctuation Prediction Using Social Media and Web Search Data. *2019 Twelfth International Conference on Contemporary Computing (IC3)*. August 2019. Noida, India: IEEE, 1–6. Available from https://doi.org/10.1109/IC3.2019.8844899 [Accessed 11 October 2022].

Mai, F. et al. (2018). How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Journal of Management Information Systems*, 35 (1), 19–52. Available from https://doi.org/10.1080/07421222.2018.1440774 [Accessed 11 October 2022].

Falcon, A. and Lyu, T. (2021). Daily Cryptocurrency Returns Forecasting and Trading via Machine Learning. *undefined*. Available from https://www.semanticscholar.org/paper/Daily-Cryptocurrency-Returns-Forecasting-and-via-Falcon-Lyu/cafff6f95805706e59ab50a0b55c5f06c8bffc53 [Accessed 14 October 2022].

Ider, D. (2022). Cryptocurrency Return Prediction Using Investor Sentiment Extracted by BERT-Based Classifiers from News Articles, Reddit Posts and Tweets. *undefined*. Available from https://www.semanticscholar.org/paper/Cryptocurrency-Return-Prediction-Using-Investor-by-Ider/37a123f00f93858c763da4dac3b99e528ac81160 [Accessed 16 October 2022].

Mohapatra, S., Ahmed, N. and Alencar, P. (2019). Krypto Oracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments. *2019 IEEE International Conference on Big Data (Big Data)*. December 2019. 5544–5551. Available from https://doi.org/10.1109/BigData47090.2019.9006554 [Accessed 18 October 2022].

Livieris, I.E. et al. (2020). Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms*, 13 (5), 121. Available from https://doi.org/10.3390/a13050121 [Accessed 18 October 2022].

Poelstra, A. (2014). A Treatise on Altcoins. 2014. Available from https://www.semanticscholar.org/paper/A-Treatise-on-Altcoins-Poelstra/e246abb304521341a58642e062e3cc3906bdc8fe [Accessed 21 October 2022].

Lee, H. et al. (2022). The Credibility Cryptocurrency Valuation: Statistical Learning Analysis for Influencer Tweets. *2022 International Conference on Information Networking (ICOIN)*. 12 January 2022. Jeju-si, Korea, Republic of: IEEE, 58–61. Available from https://doi.org/10.1109/ICOIN53446.2022.9687194 [Accessed 23 October 2022].

Rizwan, M., Narejo, S. and Javed, M. (2019). Bitcoin price prediction using Deep Learning Algorithm. *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*. December 2019. Karachi, Pakistan: IEEE, 1–7. Available from https://doi.org/10.1109/MACS48846.2019.9024772 [Accessed 15 November 2022].

Huang, X. et al. (2021). The Diversification Benefits of Cryptocurrency Asset Categories and Estimation Risk: Pre and Post COVID-19. *SSRN Electronic Journal*. Available from https://doi.org/10.2139/ssrn.3894874 [Accessed 18 November 2022].

Krafft, P.M., Della Penna, N. and Pentland, A.S. (2018). An Experimental Study of Cryptocurrency Market Dynamics. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 21 April 2018. Montreal QC Canada: ACM, 1–13. Available from https://doi.org/10.1145/3173574.3174179 [Accessed 26 November 2022].

Sebastião, H. and Godinho, P. (2021). Forecasting and trading cryptocurrencies with machine learning under changing market conditions. *Financial Innovation*, 7 (1), 3. Available from https://doi.org/10.1186/s40854-020-00217-x [Accessed 5 December 2022].

Saari, A., Vimpari, J. and Junnila, S. (2022). Blockchain in real estate: Recent developments and empirical applications. *Land Use Policy*, 121, 106334. Available from https://doi.org/10.1016/j.landusepol.2022.106334 [Accessed 9 December 2022].

Menezes, A.J., Van Oorschot, P.C. and Vanstone, S.A. (2018). *Handbook of Applied Cryptography*, 1st ed. CRC Press. Available from https://doi.org/10.1201/9781439821916 [Accessed 15 December 2022].

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. Available from https://www.semanticscholar.org/paper/Bitcoin%3A-A-Peer-to-Peer-Electronic-Cash-System-Nakamoto/4e9ec92a90c5d571d2f1d496f8df01f0a8f38596 [Accessed 16 December 2022].

Fallucchi, F. and Gerardi, M. (2021). Blockchain, State-of-the-Art and Future Trends World. In: Fazzolari, R. and Jaber, A.A. (eds.). *Proceedings of the Scholar's Yearly Symposium of Technology, Engineering and Mathematics (SYSTEM 2021)*. CEUR Workshop Proceedings. 27 July 2021. Catania, Italy: CEUR, 80–88. Available from https://ceur-ws.org/Vol-3092/#p13 [Accessed 19 December 2022].

Ponnakanti, H.P. and Gonen, B. (2020). Relative Importance of Blockchain in Cryptocurrency, Banking & Health Care. *Archives of Business Research*, 8 (12), 6–25. Available from https://doi.org/10.14738/abr.812.9428 [Accessed 21 February 2023].

Iredale, G. (2021). Are Blockchain and AI a Perfect Combination? *101 Blockchains*. Available from https://101blockchains.com/blockchain-and-ai/ [Accessed 22 February 2023].

Ji, S., Kim, J. and Im, H. (2019). A Comparative Study of Bitcoin Price Prediction Using Deep Learning. *Mathematics*, 7 (10), 898. Available from https://doi.org/10.3390/math7100898 [Accessed 22 February 2023].

Freeda, S.E., Selvan, T.C.E. and Hemanandhini, I.G. (2021). Prediction of Bitcoin Price using Deep Learning Model. *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2 December 2021. Coimbatore, India: IEEE, 1702–1706. Available from https://doi.org/10.1109/ICECA52323.2021.9676048 [Accessed 26 February 2023].

Zakhwan, M. et al. (2022). Comparative Analysis of Cryptocurrency Price Prediction Using Deep Learning. 28 December 2022. 63–74. Available from https://doi.org/10.21467/proceedings.141.10 [Accessed 28 February 2023].

Lou, J., Cui, L. and Li, Y. (2022). Bi-LSTM Price Prediction based on Attention Mechanism. Available from https://doi.org/10.48550/ARXIV.2212.03443 [Accessed 3 March 2023].

Aslam, N. et al. (2022). Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model. *IEEE Access*, 10, 39313–39324. Available from https://doi.org/10.1109/ACCESS.2022.3165621 [Accessed 4 March 2023].

Raju, S.M. and Tarif, A.M. (2020). Real-Time Prediction of BITCOIN Price using Machine Learning Techniques and Public Sentiment Analysis. *ArXiv*. Available from https://www.semanticscholar.org/paper/Real-Time-Prediction-of-BITCOIN-Price-using-Machine-Raju-Tarif/e35b295febcdaf3551d3b062356cdafff82ad554 [Accessed 10 March 2023].

Livieris, I.E. et al. (2020). Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms*, 13 (5), 121. Available from https://doi.org/10.3390/a13050121 [Accessed 14 March 2023].

Ibrahim, A. (2021). Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-based Prediction Model. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. 21 April 2021. Toronto, ON, Canada: IEEE, 1–5. Available from https://doi.org/10.1109/IEMTRONICS52119.2021.9422647 [Accessed 8 April 2023].

Ye, Z. et al. (2022). A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin. *Mathematics*, 10 (8), 1307. Available from https://doi.org/10.3390/math10081307 [Accessed 15 April 2023].

Ni, J. et al. (2022). Copper price movement prediction using recurrent neural networks and ensemble averaging. *Soft Computing*, 26 (17), 8145–8161. Available from https://doi.org/10.1007/s00500-022-07201-w [Accessed 23 April 2023].

Hegde, Dr.G. (2023). Bitcoin Price Prediction using Deep Learning. *International Journal for Research in Applied Science and Engineering Technology*, 11 (4), 1780–1788. Available from https://doi.org/10.22214/ijraset.2023.50481 [Accessed 2 May 2023].

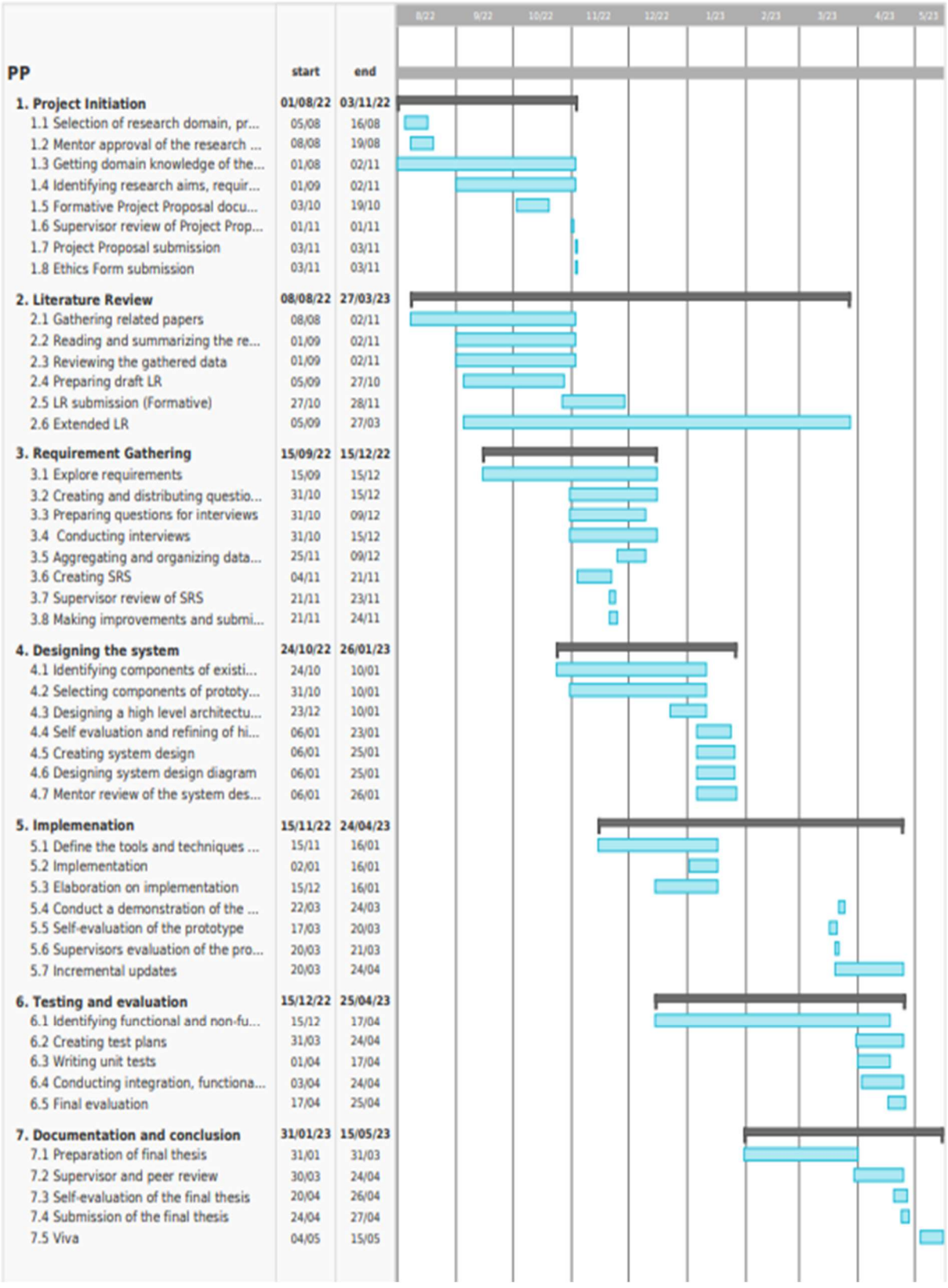
Appendix A

Concept Map

Figure 22 Concept map

Appendix B

Gantt chart



PP	start	end
1. Project Initiation	**01/08/22**	**03/11/22**
1.1 Selection of research domain, pr...	05/08	16/08
1.2 Mentor approval of the research ...	08/08	19/08
1.3 Getting domain knowledge of the...	01/08	02/11
1.4 Identifying research aims, requir...	01/09	02/11
1.5 Formative Project Proposal docu...	03/10	19/10
1.6 Supervisor review of Project Prop...	01/11	01/11
1.7 Project Proposal submission	03/11	03/11
1.8 Ethics Form submission	03/11	03/11
2. Literature Review	**08/08/22**	**27/03/23**
2.1 Gathering related papers	08/08	02/11
2.2 Reading and summarizing the re...	01/09	02/11
2.3 Reviewing the gathered data	01/09	02/11
2.4 Preparing draft LR	05/09	27/10
2.5 LR submission (Formative)	27/10	28/11
2.6 Extended LR	05/09	27/03
3. Requirement Gathering	**15/09/22**	**15/12/22**
3.1 Explore requirements	15/09	15/12
3.2 Creating and distributing questio...	31/10	15/12
3.3 Preparing questions for interviews	31/10	09/12
3.4 Conducting interviews	31/10	15/12
3.5 Aggregating and organizing data...	25/11	09/12
3.6 Creating SRS	04/11	21/11
3.7 Supervisor review of SRS	21/11	23/11
3.8 Making improvements and submi...	21/11	24/11
4. Designing the system	**24/10/22**	**26/01/23**
4.1 Identifying components of existi...	24/10	10/01
4.2 Selecting components of prototy...	31/10	10/01
4.3 Designing a high level architectu...	23/12	10/01
4.4 Self evaluation and refining of hi...	06/01	23/01
4.5 Creating system design	06/01	25/01
4.6 Designing system design diagram	06/01	25/01
4.7 Mentor review of the system des...	06/01	26/01
5. Implemenation	**15/11/22**	**24/04/23**
5.1 Define the tools and techniques ...	15/11	16/01
5.2 Implementation	02/01	16/01
5.3 Elaboration on implementation	15/12	16/01
5.4 Conduct a demonstration of the ...	22/03	24/03
5.5 Self-evaluation of the prototype	17/03	20/03
5.6 Supervisors evaluation of the pro...	20/03	21/03
5.7 Incremental updates	20/03	24/04
6. Testing and evaluation	**15/12/22**	**25/04/23**
6.1 Identifying functional and non-fu...	15/12	17/04
6.2 Creating test plans	31/03	24/04
6.3 Writing unit tests	01/04	17/04
6.4 Conducting integration, functiona...	03/04	24/04
6.5 Final evaluation	17/04	25/04
7. Documentation and conclusion	**31/01/23**	**15/05/23**
7.1 Preparation of final thesis	31/01	31/03
7.2 Supervisor and peer review	30/03	24/04
7.3 Self-evaluation of the final thesis	20/04	26/04
7.4 Submission of the final thesis	24/04	27/04
7.5 Viva	04/05	15/05

Figure 23 Gantt chart

Deliverable dates

Table 47 Deliverable dates

Deliverable	Date
Project Proposal Document and Ethics Form The initial proposal document regarding the proposed project and ethical form.	3rd November 2022
Literature Review Document The document will assess the solutions and current system critically.	27th October 2022
Software Requirement Specification The document that will outline the specifications required to create the prototype.	24th November 2022
System Design Document The document outlining the suggested prediction system's design.	5th December 2022
Test and Evaluation Report The document that will outline the testing and assessments that were done.	23rd March 2023
Draft Project Report The thesis' preliminary draft, which will be given to the supervisor for comments and feedback.	30th March 2023
Prototype The primary functional prototype of the suggested system.	24th April 2023
Final Thesis The final report that will go through all the conclusions, results, and choices.	24th April 2023

Appendix C

UI Wireframes

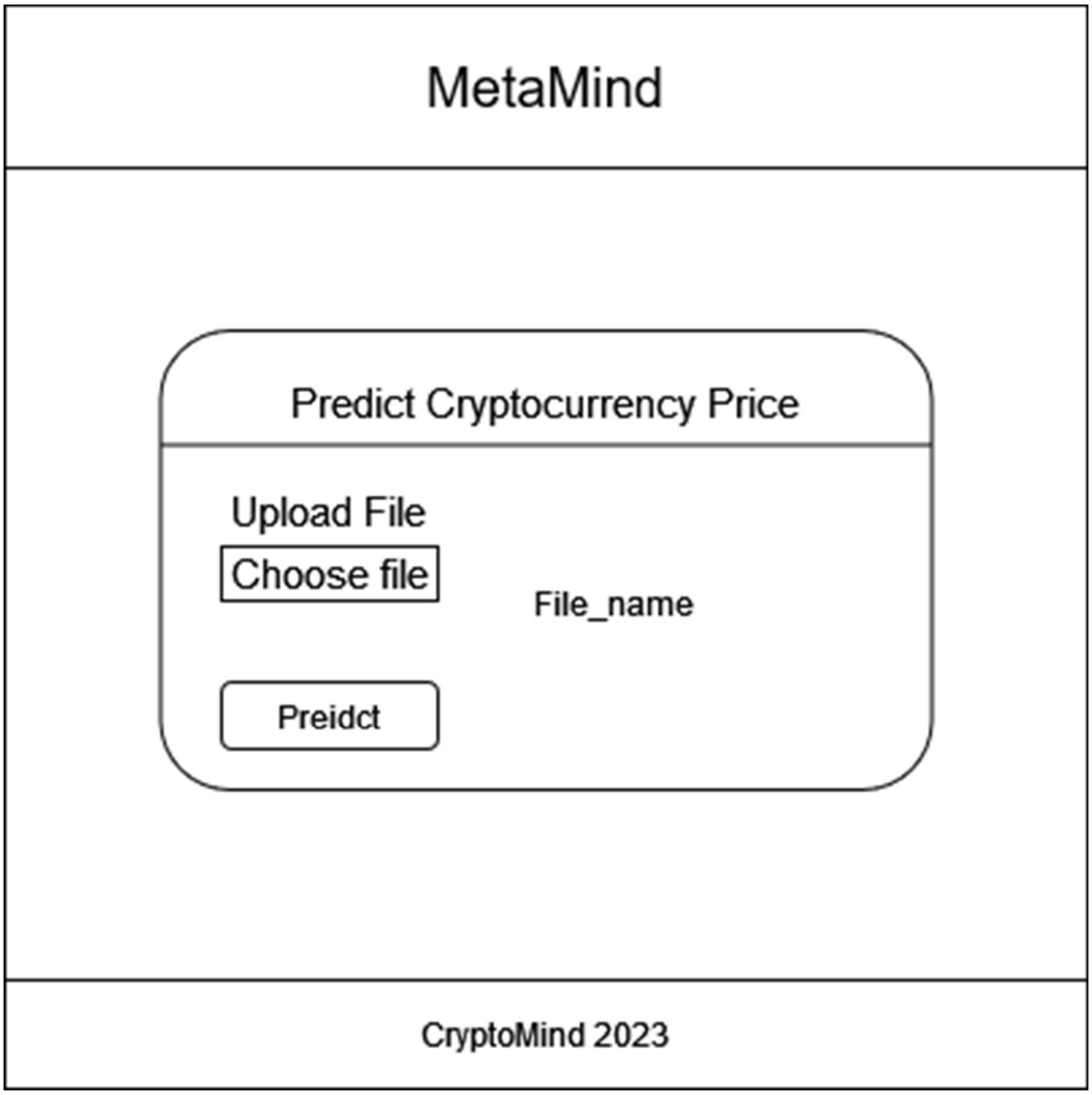


Figure 24 Home page

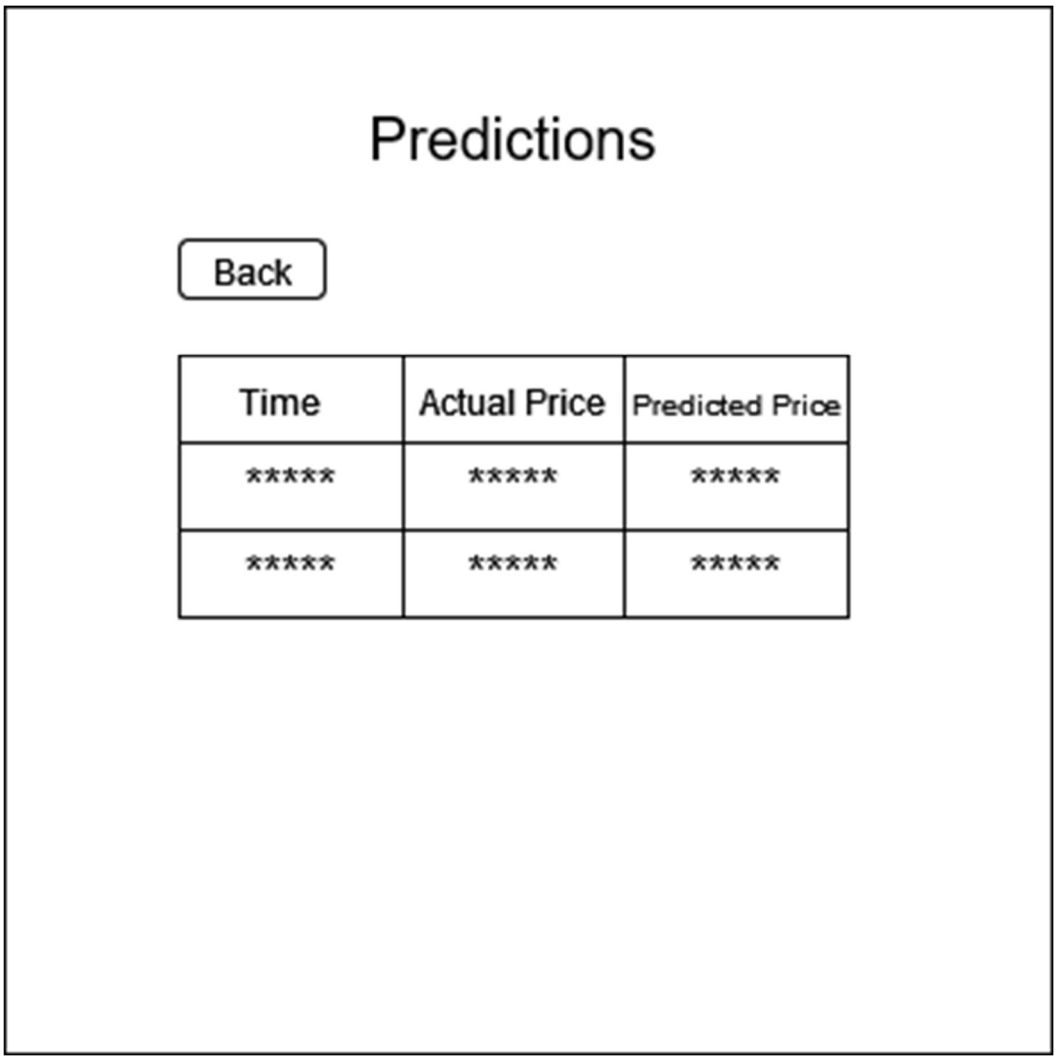


Figure 25 Prediction page

Appendix D

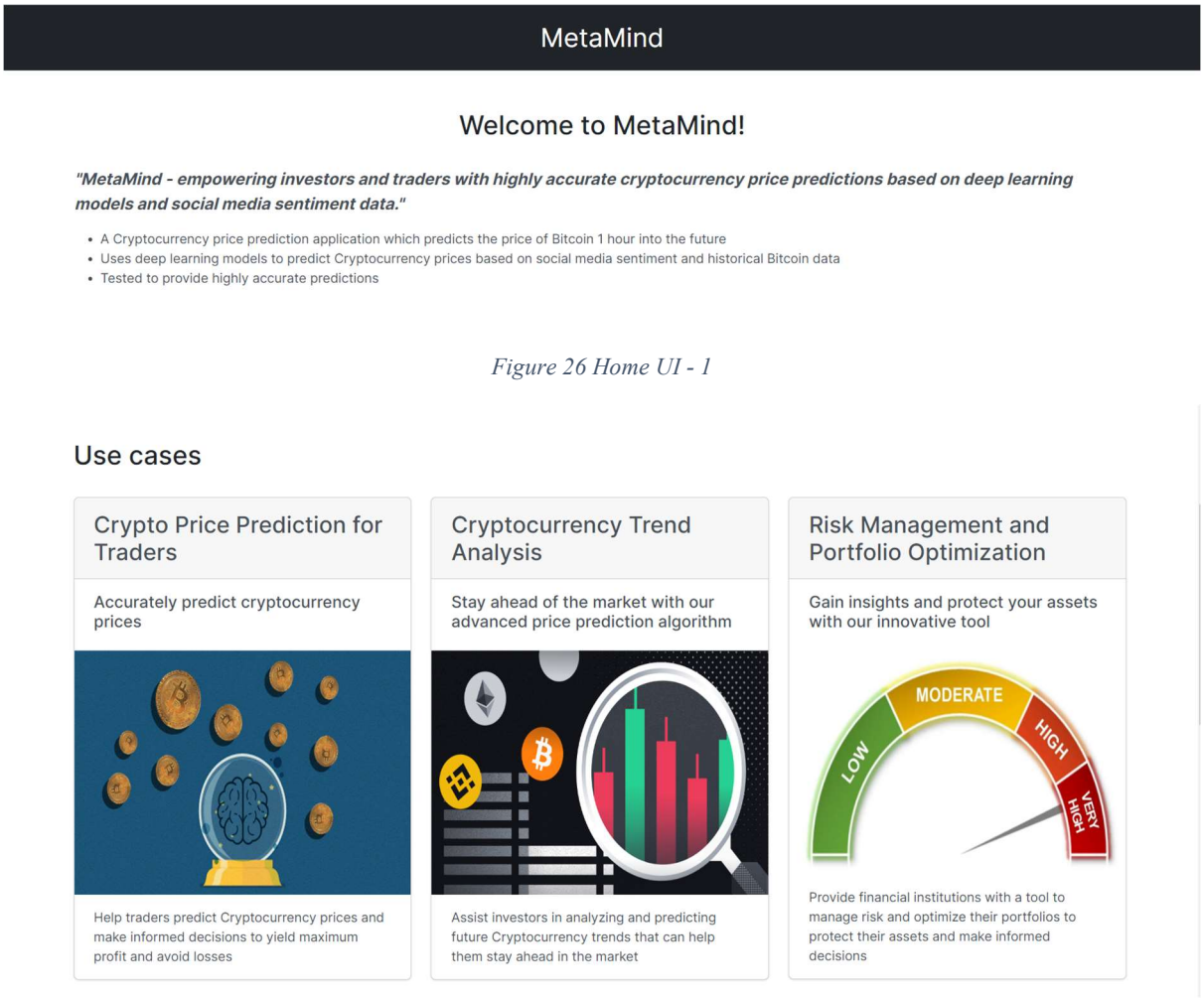


Figure 26 Home UI - 1

Figure 27 Home UI - 2

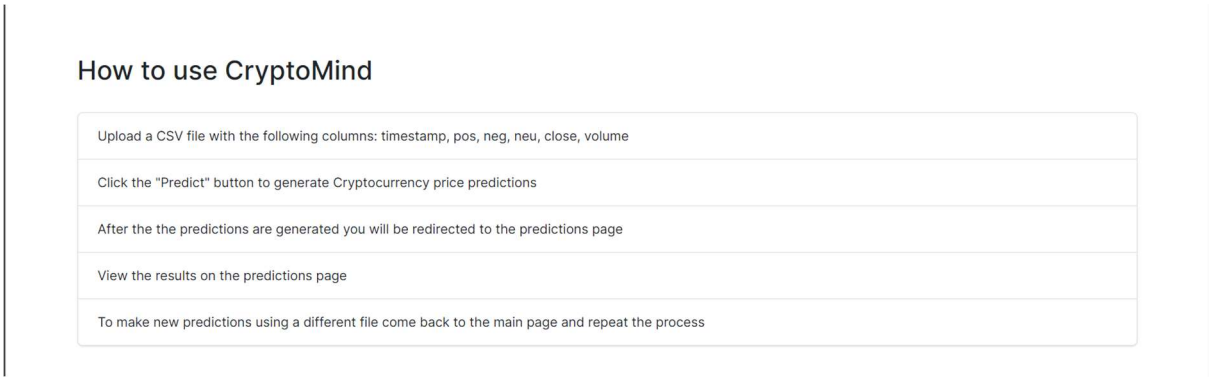


Figure 28 Home UI - 3

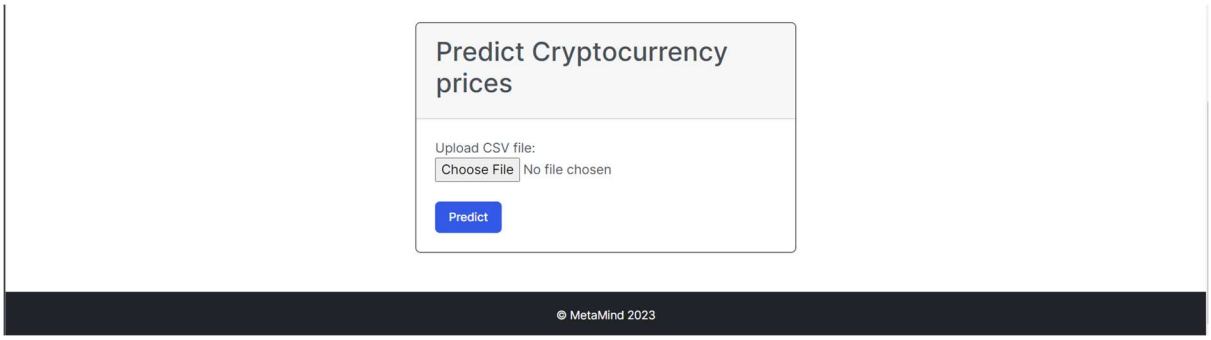


Figure 29 Home UI - 4

The images above show the main page of the application where the user is given information regarding the application like what the application is, how it is useful for the end user and the instructions to use the system. The user can upload a csv file by clicking 'choose file' button and after uploading the file the user can click the 'predict' button to view the predictions.

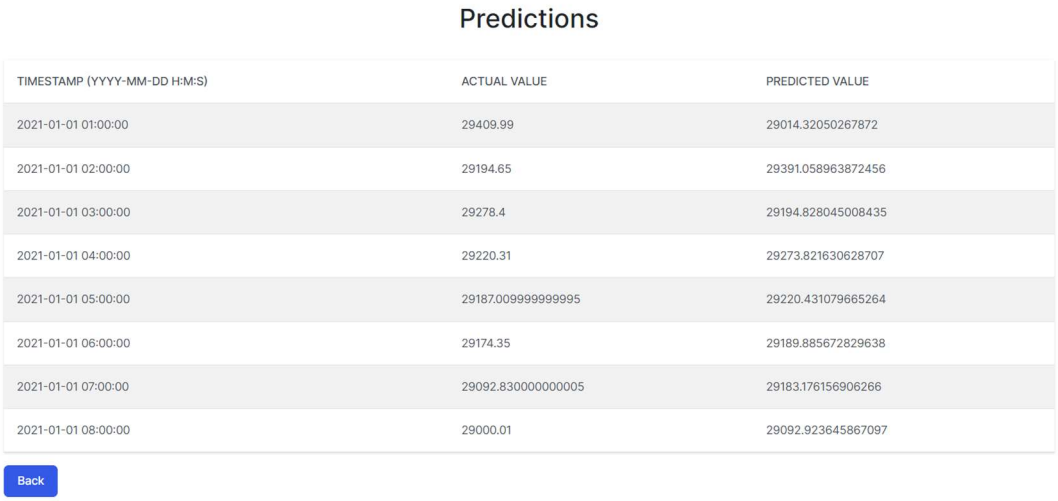


Figure 30 Prediction page

The image above shows the output of the final predictions made by the system for the file uploaded by the user.

```
def get_sentiment_report(input_filename, output_filename):
    df = pd.read_csv(input_filename)
    df = df[['title', 'selftext', 'publish_date']]
    df = df.fillna('')

    df['text'] = df['title'] + ' ' + df['selftext']
    df.set_index('publish_date', inplace=True)
    df.drop(['title', 'selftext'], axis=1, inplace=True)

    for row_i, row in df.iterrows():
        tb_sentiment_polarity_dict = dict()
        tb_sentiment_subjectivity_dict = dict()
        flair_sentiment_dict = dict()

        sid_pos_dict = dict()
        sid_neg_dict = dict()
        sid_neu_dict = dict()
        sid_com_dict = dict()

        data = row['text']
        print(row_i)
        print(data[0:15])
        flair_s = flair.data.Sentence(data)
        flair_sentiment.predict(flair_s)
        flair_total_sentiment = flair_s.labels
        flair_val = get_sentiment_val_for_flair(flair_total_sentiment)

        flair_sentiment_dict[str(row_i)] = flair_val
        tb_sentiment_polarity_dict[str(row_i)] = TextBlob(data).sentiment[0]
        tb_sentiment_subjectivity_dict[str(row_i)] = TextBlob(data).sentiment[1]

        ss = sid.polarity_scores(data)
        sid_pos_dict[str(row_i)] = ss['pos']
        sid_neg_dict[str(row_i)] = ss['neg']
        sid_neu_dict[str(row_i)] = ss['neu']
        sid_com_dict[str(row_i)] = ss['compound']

        flair_df = pd.DataFrame.from_dict(flair_sentiment_dict, orient='index', columns=['reddit_flair'])
        flair_df.index.name = 'timestamp'

        tb_polarity_df = pd.DataFrame.from_dict(tb_sentiment_polarity_dict, orient='index',
                                                columns=['reddit_tb_polarity'])
        tb_polarity_df.index.name = 'timestamp'
```

Figure 31 Reddit sentiment analysing

```
    tb_polarity_df = pd.DataFrame.from_dict(tb_sentiment_polarity_dict, orient='index',
                                            columns=['reddit_tb_polarity'])
    tb_polarity_df.index.name = 'timestamp'

    tb_subjectivity_df = pd.DataFrame.from_dict(tb_sentiment_subjectivity_dict, orient='index',
                                                columns=['reddit_tb_subjectivity'])
    tb_subjectivity_df.index.name = 'timestamp'

    sid_pos_df = pd.DataFrame.from_dict(sid_pos_dict, orient='index',
                                        columns=['reddit_sid_pos'])
    sid_pos_df.index.name = 'timestamp'

    sid_neg_df = pd.DataFrame.from_dict(sid_neg_dict, orient='index',
                                        columns=['reddit_sid_neg'])
    sid_neg_df.index.name = 'timestamp'

    sid_neu_df = pd.DataFrame.from_dict(sid_neu_dict, orient='index',
                                        columns=['reddit_sid_neu'])
    sid_neu_df.index.name = 'timestamp'

    sid_com_df = pd.DataFrame.from_dict(sid_com_dict, orient='index',
                                        columns=['reddit_sid_com'])
    sid_com_df.index.name = 'timestamp'

    final_senti_df = pd.concat([flair_df, tb_polarity_df, tb_subjectivity_df, sid_pos_df, sid_neg_df, sid_neu_df,
                  sid_com_df], axis=1)

    if os.path.exists(output_filename):
        keep_header = False
    else:
        keep_header = True

    final_senti_df.to_csv(output_filename, mode='a', header=keep_header)

return
```

Figure 32 Reddit sentiment analysing

```
import os
import pandas as pd
import flair
from textblob import TextBlob
from nltk.sentiment import SentimentIntensityAnalyzer

flair_sentiment = flair.models.TextClassifier.load('en-sentiment')
sid = SentimentIntensityAnalyzer()

def get_sentiment_report(input_filename, output_filename):
    df = pd.read_csv(input_filename, usecols=['title', 'selftext', 'publish_date'], na_values='')
    df['text'] = df['title'].fillna('') + ' ' + df['selftext'].fillna('')
    df.set_index('publish_date', inplace=True)
    df.drop(['title', 'selftext'], axis=1, inplace=True)

    sentiment_columns = ['reddit_flair', 'reddit_tb_polarity', 'reddit_tb_subjectivity',
                         'reddit_sid_pos', 'reddit_sid_neg', 'reddit_sid_neu', 'reddit_sid_com']
    sentiment_dicts = [dict() for _ in range(len(sentiment_columns))]

    for row_i, row in df.iterrows():
        data = row['text']
        flair_s = flair.data.Sentence(data)
        flair_sentiment.predict(flair_s)
        flair_total_sentiment = flair_s.labels
        flair_val = get_sentiment_val_for_flair(flair_total_sentiment)

        sentiment_dicts[0][str(row_i)] = flair_val
        sentiment_dicts[1][str(row_i)] = TextBlob(data).sentiment[0]
        sentiment_dicts[2][str(row_i)] = TextBlob(data).sentiment[1]

        ss = sid.polarity_scores(data)
        sentiment_dicts[3][str(row_i)] = ss['pos']
        sentiment_dicts[4][str(row_i)] = ss['neg']
        sentiment_dicts[5][str(row_i)] = ss['neu']
        sentiment_dicts[6][str(row_i)] = ss['compound']

    final_senti_df = pd.concat([pd.DataFrame.from_dict(sentiment_dict, orient='index', columns=[column_name])
                                for sentiment_dict, column_name in zip(sentiment_dicts, sentiment_columns)],
                               axis=1)

    with open(output_filename, 'a') as f:
        final_senti_df.to_csv(f, header=f.tell()==0)

    return
```

Figure 33Reddit sentiment analysing

```
df_short

df_short.to_csv('shorted_tweets')

df_short[['neg', 'neu', 'pos', 'compound']] = df_short['text_clean'].apply(lambda tweet: pd.Series(sid.polarity_scores(tweet)))

df_short

df_short.set_index('datetime', inplace=True)
df_short

df_short.to_csv('cleaned_tweets.csv')
```

Figure 34 Twitter sentiment analysing

Appendix E

The below images show the GPU and CPU usage before starting the local server,

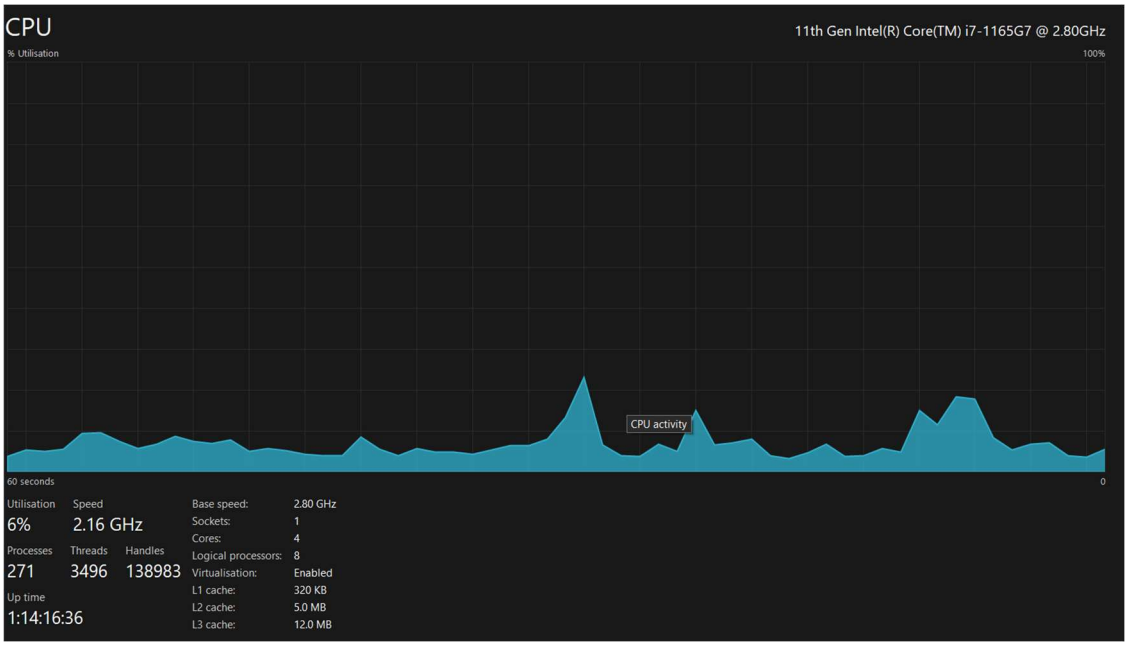


Figure 35 CPU before starting server

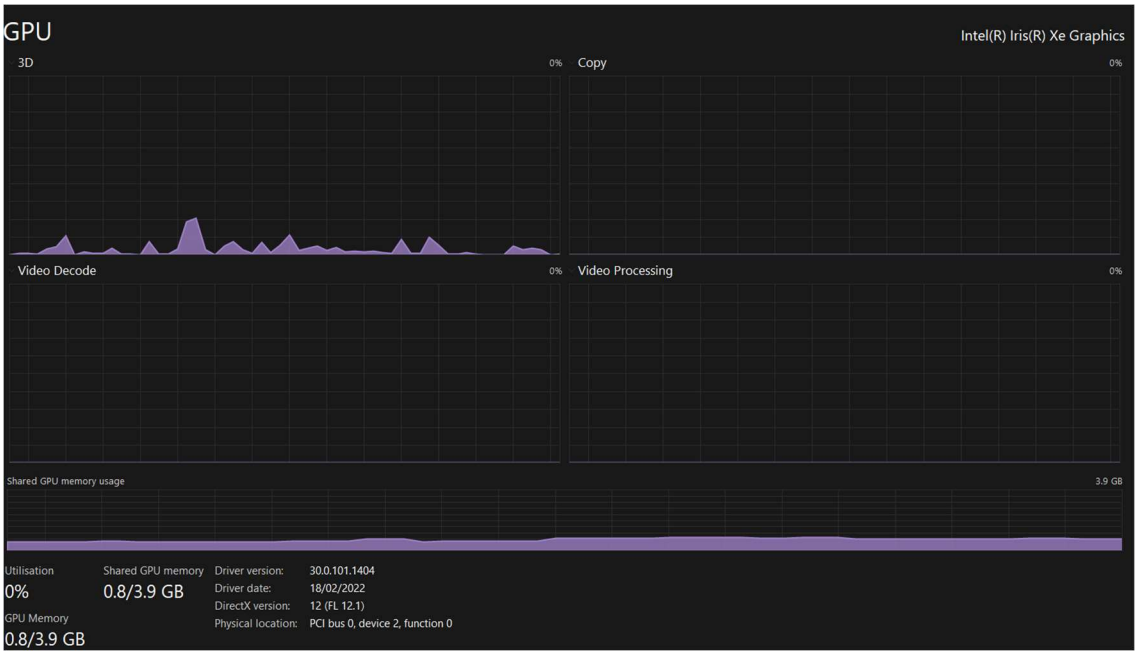


Figure 36 GPU before starting the server

The below images represent the GPU and CPU usage after running the local server,

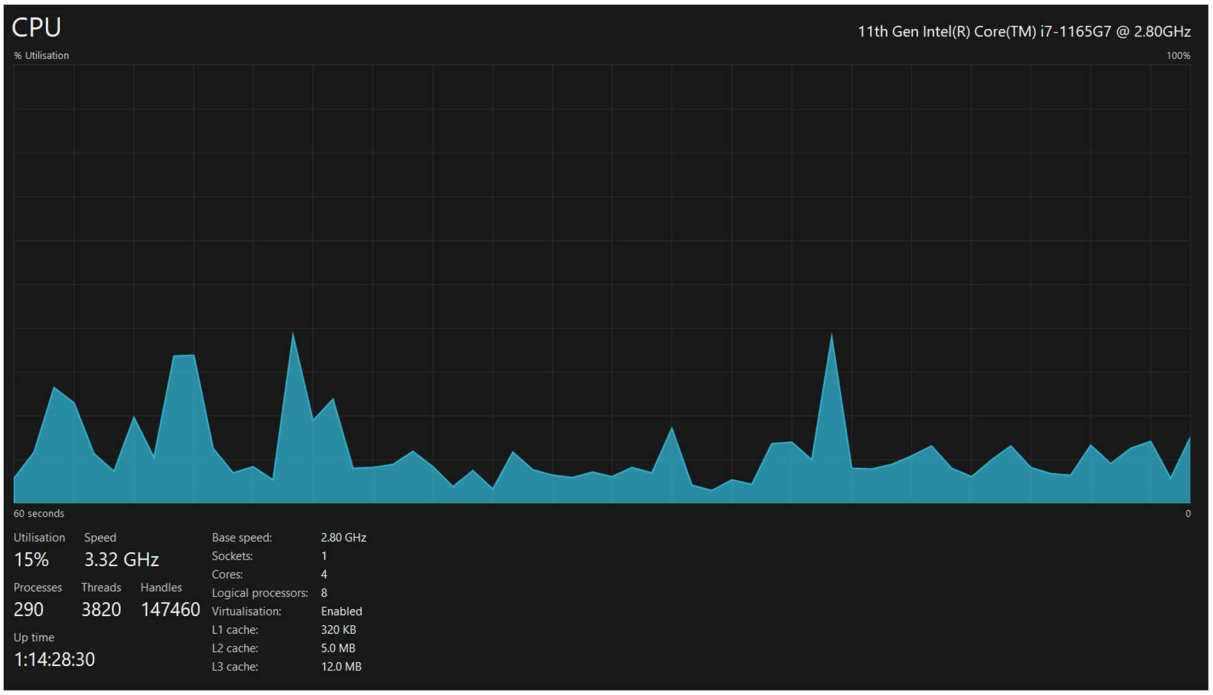


Figure 37 CPU after starting the server

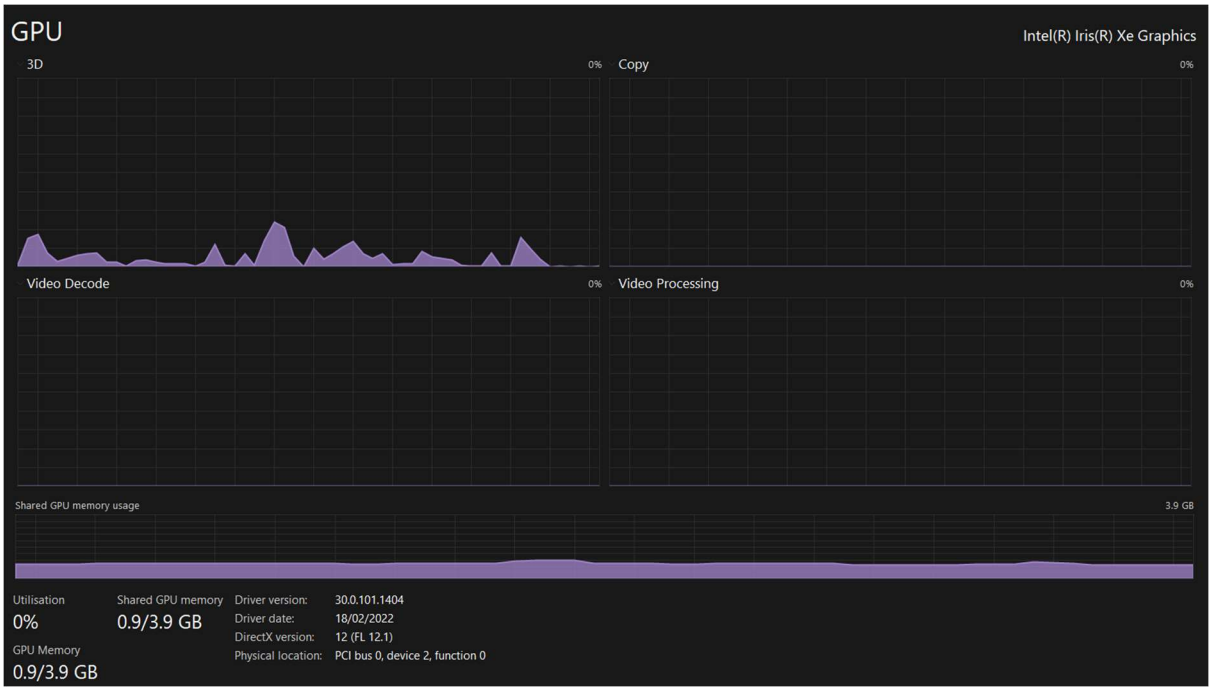


Figure 38 GPU after starting the server

Appendix F

Validation from the domain expert

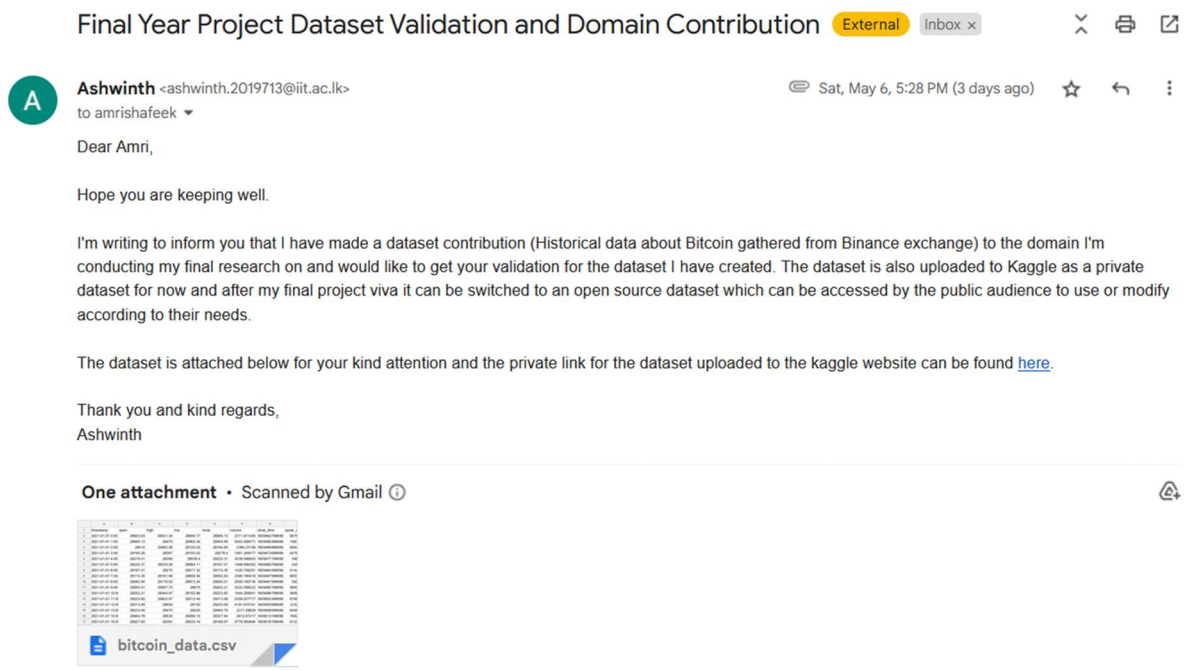


Figure 39 Mail regarding dataset validation

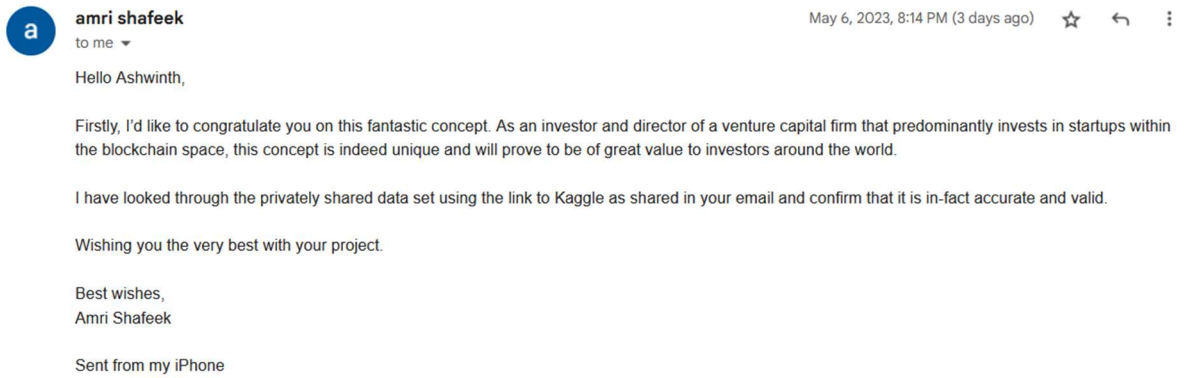


Figure 40 Validation given by domain expert

Kaggle Private Dataset

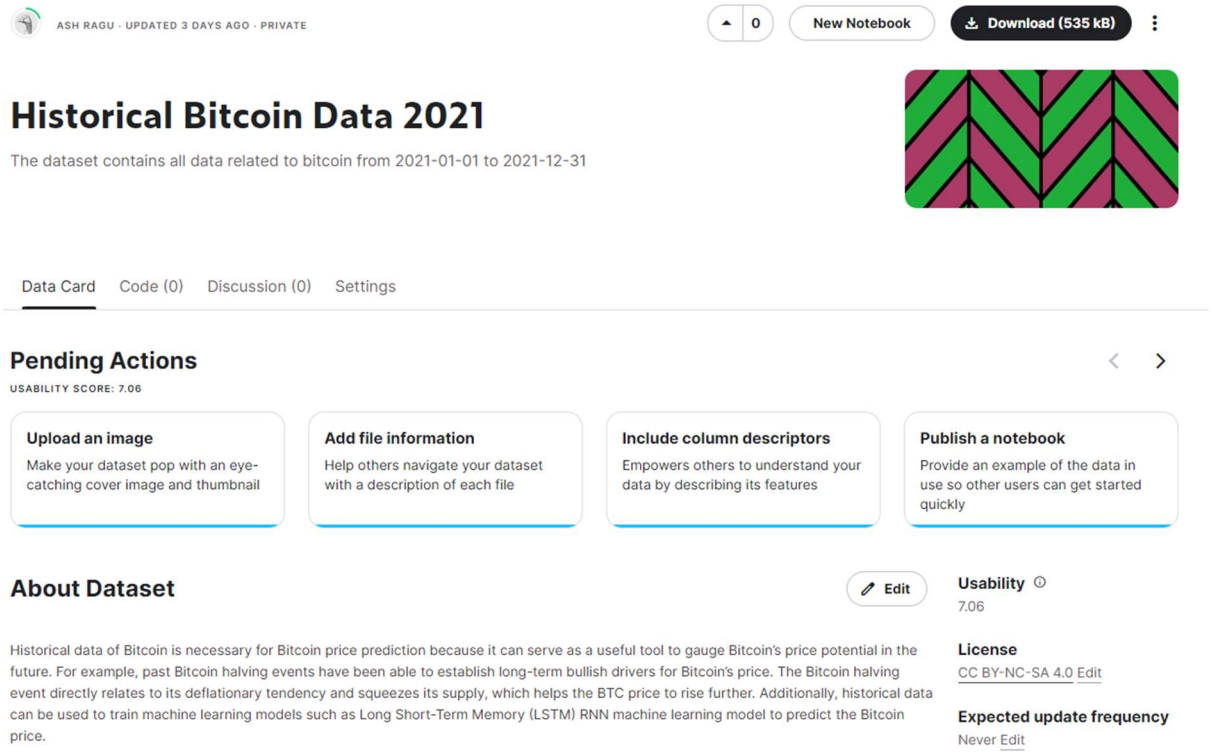


Figure 41 Private dataset uploaded to Kaggle