A Thesis

For the Degree of Doctor of Philosophy

Knowledge Discovery and Cryptocurrency Price Prediction Based on Blockchain Framework

Zeinab Shahbazi

Department of Computer Engineering

Graduate School

Jeju National University

June 2022

Knowledge Discovery and Cryptocurrency Price Prediction Based on Blockchain Framework

Zeinab Shahbazi
(Supervised by Professor Yung-Cheol Byun)

A thesis submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Engineering

2022.06.02

This thesis has been examined and approved.

Sang Joon Lee

....

Thesis Director

Sang-Joon Lee, Professor, Department of Computer Engineering

Ho-Young Kwak

....

Ho-Young Kwak, Professor, Department of Computer Engineering

Namje Park

....

Namje Park, Professor, Department of Computer Education

Dong Cheol, Lee

....

Dong-Cheol Lee, Professor, Department of Management Information

....

Thesis Supervisor

Yung-Cheol Byun, Professor, Department of Computer Engineering

June 2022

Department of Computer Engineering
GRADUATE SCHOOL
JEJU NATIONAL UNIVERSITY

Acknowledgment

First and foremost, I am thankful to Almighty ALLAH for giving me the strength, knowledge, ability, and opportunity to undertake this study and complete it satisfactorily.

I dedicate this milestone to my beloved mother, Kokab Alizade, and my father, Moharram Ali Shahbazi; I am forever indebted to them for giving me the opportunity and experiences that have made me who I am. I appreciate Factor for becoming my family abroad and always pushing me to do my best in literally every part of my life. I learn all my life lessons, ups and downs, all my dreams, now and future. I wish to thank Sadiqa Jafari, who always encourages me; my sister Zahra Shahbazi and my brother Mahdi Shahbazi for their endless support.

I would like to express my sincere gratitude to my supervisor, Professor Yung-Cheol Byun, for his continuous support of my Ph.D. study and related research and his patience, motivation, and encouraging attitude.

Besides my advisor, I would like to thank my thesis evaluation committee for their insightful comments and valuable suggestions during the thesis defense. Their input helped me in elevating the quality of this thesis.

Finally, I would like to appreciate the undiminishing and unconditional support of Ali Akbar Jalali in every step of this journey.

Zeinab Shahbazi

Dedicated to

My Jooje ❤

Table of Contents

Acknowledgment **iii**

List of Figures **iv**

List of Tables **vii**

Abstract **1**

Chapter 1: Introduction **3**

1.1 Motivation 3

1.2 Background.. 4

1.3 Problem Statement 5

1.4 Research Goal 7

Chapter 2: Related Work **9**

2.1. Concept of Cryptocurrency 9

2.2. Concept of Knowledge Discovery for Cryptocurrency Technology 11

2.2.1 Topic Modeling **16**

2.2.2 Topic Discovery **18**

2.2.3 Information Analysis System **21**

2.3. Concept of Cryptocurrency in Blockchain 23

2.4. Arts and Regulations of Financial Statements 24

Chapter 3: Proposed Cryptocurrency Price Prediction 31

3.1 Cryptocurrency Prediction Architecture.. 31

3.1.1 Data Acquisition for Cryptocurrency 33

3.1.2 System Process 37

3.1.3 Control and Optimization 37

3.2 Topic Modeling Based on LDA in Cryptocurrency Dataset........................... 37

3.2.1 Data pre-processing 39

3.3 General Cryptocurrency Prediction Architecture based on System Security42

3.3.1 Service Layer 44

3.3.2 Predictive Optimal Control Layer 44

3.3.3 Blockchain Layer 45

3.3.4 Physical Layer 46

3.4 Cryptocurrency Prediction Architecture based on Blockchain..................... 47

3.4.1 Difficulty of Mining, Profitability and Hashrate 50

3.4.2 Confirmation Time and Market Capitalization 50

3.5 Cryptocurrency Predictive Analysis 51

3.5.1 Reinforcement Learning Prediction 53

3.5.2 Price Prediction 56

3.5.3 Risk Management 57

Chapter 4: Development of Cryptocurrency Prediction **59**

4.1 Design of Cryptocurrency Prediction Framework 59

4.1.1 Cryptocurrency knowledge Discovery Framework **61**

4.1.2 Cryptocurrency Blockchain Framework **63**

4.2 Implementation of Cryptocurrency Predictive Analysis................................. 64

4.2.1 Development Environment of Cryptocurrency Predictive Analysis **64**

4.2.2 Experimental Setup of Cryptocurrency Prediction **66**

4.2.3 Experimental Results of Cryptocurrency Prediction **70**

4.3 Performance Analysis of Cryptocurrency Prediction 86

4.3.1 Price Management for Prediction **86**

4.3.2 Risk Management for Prediction **89**

4.3.3 Evaluation Metrix of the Cryptocurrency Prediction **97**

Chapter 5: Conclusion and Future Directions **101**

References **103**

List of Figures

Figure 1. Concept of cryptocurrency 10

Figure 2. An overview of the process of predicting the price of cryptocurrencies.......... 15

Figure 3. Topic modeling process.. 18

Figure 4. Visualization of topic discovery content structure 20

Figure 5. Information analysis system scenario.. 22

Figure 6. User wallet growth in years.. 23

Figure 7. Cryptocurrency detail process 24

Figure 8. Money laundering detail process... 27

Figure 9. Overview of cryptocurrency exchange rate prediction 30

Figure 10. Main Architecture of knowledge discovery and cryptocurrency prediction process 32

Figure 11. Detailed data collection and processing steps 33

Figure 12. Detailed data collection and processing steps 34

Figure 13: Data miner platform for data crawling.. 35

Figure 14. LDA topic modeling detail architecture in cryptocurrency dataset 38

Figure 15. Data pre-processing detail architecture 41

Figure 16. General Architecture of knowledge discovery and cryptocurrency prediction process 43

Figure 17. Cryptocurrency service layer general architecture...................................... 44

Figure 18. Cryptocurrency predictive optimal control layer 45

Figure 19. Cryptocurrency blockchain layer 45

Figure 20. Cryptocurrency physical layer..46

Figure 21. Cryptocurrency prediction based on blockchain 48

Figure 22. Detailed cryptocurrency prediction scenario...51

Figure 23. Reinforcement learning framework...54

Figure 24. The Reinforcement learning process in the prediction system.......................55

Figure 25. Price prediction process based on Reinforcement learning............................56

Figure 26. Reinforcement learning-based risk management architecture 58

Figure 27. Predictive optimization flowchart in prediction system.................................60

Figure 28. Knowledge discovery on blockchain framework 62

Figure 29. Sequence diagram of crypto prediction...63

Figure 30. Development Environment of crypto prediction...65

Figure 31. Implementation scenario of LDA topic categorization using pyLDAvis.......69

Figure 32. 30 days actual and predicted values of KuCoin prediction results.................74

Figure 33. 90 days actual and predicted values prediction results of KuCoin.................75

Figure 34. 30 days actual and predicted values prediction results of ChainLink 75

Figure 35. 90 days of actual and predicted value prediction results of ChainLink 76

Figure 36. 30 days actual and predicted values prediction results of Polkadot 76

Figure 37. 90 days actual and predicted value prediction results of Polkadot.................77

Figure 38. MAPE for the classification model 78

Figure 39. Classification model accuracy...78

Figure 40. Factor importance differences 2018-2022...80

Figure 41. Transaction process function for the cryptocurrency blockchain platform....82

Figure 42. Definition of access control in the cryptocurrency system 83

Figure 43. Cryptocurrency transaction portal history 84

Figure 44. Latency of query transaction 85

Figure 45. Response time of different requests 86

Figure 46. Closing price for KuCoin cryptocurrency 88

Figure 47. Closing price for ChainLink cryptocurrency 88

Figure 48. Cryptocurrency closing price for Polkadot 89

Figure 49. Cryptocurrency market risk index regularity 93

Figure 50. Reinforcement learning-based high-risk management 96

Figure 51. Proposed approach MAE comparison 98

Figure 52. Proposed approach MAPE comparison 99

Figure 53. Proposed approach RMSE comparison 99

Figure 54. Proposed approach MSE comparison 100

List of Tables

Table 1: Problems in the cryptocurrency network 6

Table 2. Comparison of cryptocurrencies performance with various features 11

Table 3. The total number of banks related to financial investigation and education attended to interview of cipher trace.. 25

Table 4. Paid penalties of USA for money laundering 28

Table 5. The records of cryptocurrency market capability during 2020 29

Table 6: Overview of collected dataset.. 36

Table 7. LDA parameters for data processing 38

Table 8. Implementation Environment 66

Table 9: NLP Techniques applied for the data pre-processing....................................... 67

Table 10. Cryptocurrency accuracy records statistic prediction..................................... 72

Table 11. Cryptocurrency accuracy records of statistic prediction 73

Table 12. The records of cryptocurrencies breakdown pattern 74

Table 13. The raw feature technical indicators.. 79

Table 14. Cryptocurrencies information 87

Table 15. The return portfolio risk performance = 350 90

Table 16. The return portfolio risk performance = 600 91

Table 17. The return portfolio risk performance = 850 92

Table 18. Risk evaluation of cryptocurrency portfolios 94

Table 19. Risk evaluation of cryptocurrency portfolios for 30 days 95

Abstract

Several types of risks are associated with cryptocurrencies, which affect the intrinsic assessment of risk auditors worldwide. Cryptocurrencies present a wide risk of money laundering from the beginning with their growth in popularity. Financial institutions such as anti-money laundering, banks, and secrecy of banks proceed in various ways, such as specialists of risk, managers of the banks, and officer of compliance whose role involves protecting illegal funds and the use of cryptocurrency. In recent years, research and academic working areas have been paying much attention to cryptocurrency's popularity. Cryptocurrency's uncontrollable and untraceable nature is appealing to humans of this ancestry. Exchange rates can be difficult to predict in financial markets due to their non-linear nature. Research into previous price inflations is used to predict the cryptocurrency's price. Various machine learning algorithms are applied to predict the digital coins exchange rate. In recent years, cryptocurrencies have emerged as a key important factor in financial and business ventures. Despite this, cryptocurrency is not visible due to the market's inconsistency and its volatility. A prior approach to price prediction did not include enough information and solutions for forecasting price changes due to the real-time prediction of costs.

A topic model generates the probability of words to analyze segregated data in available contents. Researchers who use conventional topic modeling on Twitter posts combined posts from multiple users into one document in the last decade. A combination of short text information is needed to conclude the topic relations. Similarly, this procedure may reduce the issue of sparsity, but it cannot identify differences between topics posted by the same user. This thesis proposed knowledge discovery and cryptocurrency price prediction based on the blockchain framework. The general service architecture comprises four layers. Each layer is

separated from the other layer to give the access to developers to replace or add a new module without any effect on the rest of the system. Moreover, the general service architecture contains the sequence diagram and flowchart. The service architecture is mapped into two distinct architectures, i.e., knowledge discovery and blockchain framework.

The knowledge discovery architecture aims to provide detailed information about internet users' shared content related to cryptocurrency. Similarly, the designed platform contributes to the topic modeling approach by using Latent Dirichlet Allocation (LDA) to generate the concept of shared information in different categories and using the prediction and optimization approaches to create the cryptocurrency's future price in terms of allowing the user for selecting the correct option which meets the defined requirement form user.

The blockchain architecture aims to provide a secure and transparent platform for the transactional process between networks. Regarding the defined rules and smart contracts in the system, the process has a visible path to follow the transactions through the network. The outcomes of this study revealed that the proposed model is efficient due to fault-tolerant, scalable, and provides flexibility for the new business models.

Chapter 1: Introduction

1.1 Motivation

In general, cryptocurrencies are used to exchange money. Trading cryptocurrencies has evolved into a popular topic in the past few years. The main meaning of exchange is where people buy and sell the cryptocurrency, and the price of cryptocurrency is the current value of the cryptocurrency. Professional stock market experts usually believe there is no specific place for businesses to grow due to the dependence and volatility of the social moods. Additionally, the ability to predict the value of each cryptocurrency gives the blockchain the ability to process data at full power.

The value of cryptocurrencies is primarily determined by multiple factors, e.g., the previous price trends, social sentiments, the number of trades, and legislative actions. The existing state-of-the-art motivated us to make a reinforcement learning model to predict cryptocurrency price differences. Similarly, the problem of erratic fluctuation in the cryptocurrency price which applied in this process. The model is designed to use a learning-based system to accurately predict the cryptocurrency price market.

Blockchain can transform traditional centralized frameworks. This technology stipulates a system wherein multi-node interaction occurs in peer-to-peer work rather than a single control of the entire network. This multi-node strategy enables decentralized applications to perform their operations among anonymous nodes.

Blockchain technology has paved the way for decentralized markets by providing intelligent contracts, intelligent digital assets, and decentralized cryptocurrency. Furthermore, it mitigates the need of any third party by deploying governance policies that each participant follows through a democratic consensus mechanism.

1.2 Background

Financial markets are among the complex systems that universities have not accepted as complex systems. Consequently, there was an agreement to interact with the elements of complex systems. In 2008, cryptocurrency was developed and designed as a decentralized currency, which can be used to conduct peer-to-peer transactions without involving a bank. In many articles, cryptocurrencies were reported to play a significant role in increasing economic crimes. The Cipher Trace report indicates that about 125 million dollars were lost and stolen due to significant security breaches.

The cryptocurrency agrees to correct the long-standing financial organization trust with a strategy based on the decentralized architecture in the BIS Annual Economic Report for 2018. Internet convenience allows most cryptocurrencies to bypass associated fees in favor of traditional banking systems based on the universal aspect of the internet. A complex method model is equal to the daunting task of defining a hierarchical system design that gathers subsystems related to each other. Hierarchical models extract these resources. Hierarchical models extract these resources.

The matrix is impacted by this issue when there is a high level of covariance. 252.5 trillion dollars have been traded in cryptocurrencies over the past two decades. Reverberation emanates from the cryptocurrency environment. News publishers still focused more attention on the price changes and their effects on the market despite the price changes and the wide range of actions to limit the soaring unabated. Rules exist to protect investors and prevent money laundering. Fiat currency must also be protected from crowds. Based on all the discussed positives, implementation and theories demonstrate that the cryptocurrency market price has moved as planned.

1.3 Problem Statement

Cryptocurrency is one of the prevalent factors in terms of business and finance. There are four recent challenges discussed in this thesis to improve the marketing purpose of this system and comfortable transactions for the users. The problems are security, marketing, inheritance, and regulation, which have many considerations among researchers. Therefore, a cryptocurrency exchange is digital and subject to hackers, operational errors, and malware. Cybercriminals can access thousands of accounts and digital wallets by hacking cryptocurrency exchanges. Cryptocurrencies can fluctuate in value due to any investment when it comes to marketing. In the short time they have been in business, prices fluctuated widely, and they were susceptible to headlines because of the large number of informal and amateur investors. Cryptos are unregulated, which means that without access to a deceased loved one's wallet, a unit of inheritance cannot be obtained. The government has followed a range of aggressive to indifferent policies toward cryptocurrencies. Investors and speculators cautiously monitor international developments regarding the regulation of cryptocurrencies, as governments have reacted aggressively or indifferently.

Table 1: presents the information regarding the problems in the cryptocurrency network and the problem description.

Table 1: Problems in the cryptocurrency network

Issue	Description
Lack of Trust	In terms of various transactions in different networks, there is the possibility of losing money.
Difficult to analyze fake or real information	There is a lot of information related to every crypto type. It's challenging to know which one is real or which is fake.
High price changes	Price changes are very high, and spending decisions on which crypto and selling are difficult.
Transactions between third parties	In terms of third parties transactions, there are lots of money lost and fraud in the crypto network.

1.4 Research Goal

The main objective of this thesis is to present knowledge discovery, cryptocurrency price prediction, and network risk management based on the blockchain framework. This study aims to explore the cryptocurrency framework and predict the prices of digital coins regarding the less expensive digital coin with the highest benefit for the user. Furthermore, we present the knowledge discovery process due to the shared information on social media and internet users' opinions and experiences for the digital coins transaction.

First, we aim to use the topic modeling advantages to discover various topics and the research target. We use topic modeling in terms of knowledge discovery for the user-generated data to extract the recently shared information, trends, and discussions which is an essential aspect of any system. Most service providers apply this approach to improve their system quality of content regarding user requirements. We have used the Latent Dirichlet Allocation (LDA) technique as the topic extraction concept.

Second, we aim to implement the price prediction based on extracted topics and consider the user preferences for predicting the price of digital coins. This system allows overcoming the shared information challenges and data overloading regarding user-generated data in terms of comments, shares, likes, mentions, clicks, etc. We use the LDA topic modeling approach and Reinforcement learning to train and test the network based on LDA results and predict the future digital coins' price.

Third, we used the algorithm for the system credibility estimation before the prediction. The credibility estimation is based on user preferences. Various parameters and metrics are used for the performance measurement, e.g., optimal identified topics, accuracy measurement, etc.

Fourth, we aim to use the blockchain framework for system security and transparency of network transactions. This process can examine professional accounting based on the associated risk of the cryptocurrency and the impact expected from the financial statement. We are finding the intrinsic risk, which is correlated negatively with cryptocurrency. Ranking the exchange level for the trouble controlling regarding the likelihood evaluation and finding the highest likelihood risk of the determined cryptocurrency.

The main contribution of this thesis is as follows:

- Processing the system based on knowledge discovery approaches regarding the multiple data sources for data preparation.
- Applying the Natural Language Processing techniques for all the possible data processing factors.
- Cryptocurrency price prediction due to the high rate of price changes.
- Designing the secure platform for cryptocurrency price prediction keeps the users transactions information records, network information, and personal information.
- Using the private blockchain framework to reduce the system risk and possible attacks from hackers.

This thesis aims to overcome the limitations of topic modeling to find the highest probability based on user preferences. We are finding the similarity between words and shared content. Credibility measurement for the secure prediction result. Result validation is based on optimization: system security estimation, risk control, and price prediction of the following cryptocurrency network.

Chapter 2: Related Work

The following section demonstrates the related literature on Cryptocurrency prediction based on Natural Language Processing (NLP), Knowledge Discovery, and Blockchain. Blockchain mining, cryptography, prediction mechanisms for improving performance based on historical actions and responses, and optimization mechanisms for enhancing the outcomes based on finding the optimal parameters for specific scenarios and optimal solutions to scenario-based issues are some of the key components of cryptocurrency.

2.1. Concept of Cryptocurrency

Due to cryptocurrency's peer-to-peer network design and ungoverned nature, it has become popular due to volume trading, exchange prices, and increased volatility, which have become critical media elements. The area of cryptocurrency for financial covering has been extensively studied, for example, the efficiency of markets [1–4]. The cryptocurrency trading strategy is based on neural networks (NN) [5–8]. According to the various analyses, the neural network system uses the buy-and-hold technique during the bull trend, which causes information incompetence and produces unusual profits. Furthermore, deep learning has been proposed in many recent studies [9–11] to determine the price formulation for trading. Recurrent Neural Networks (RNN) and Long-Short-Term Memory (LSTM) are the first categories of this topic in the state-of-the-art. Lahmiri et al. [12] LSTM and Generalized Regression Neural Networks (GRNN) were proposed to predict digital coin prices. In terms of RMSE, the LSTM performs better than GRNN based on the process between these two algorithms. Tan et al. [13] On the financial market, deep learning models were compared to linear models. A variety of models were tested, including ARIMA, Random Forest, Multi-Layer Perceptron, Prophet, and Regression. Figure 1: shows a detailed description of the cryptocurrency framework [14]. There are five main components in this system: payment system, Protocol, growth, bitcoin, digital currency, and privacy. These protocols are used for the gross settlement and the Ripple transaction protocol.

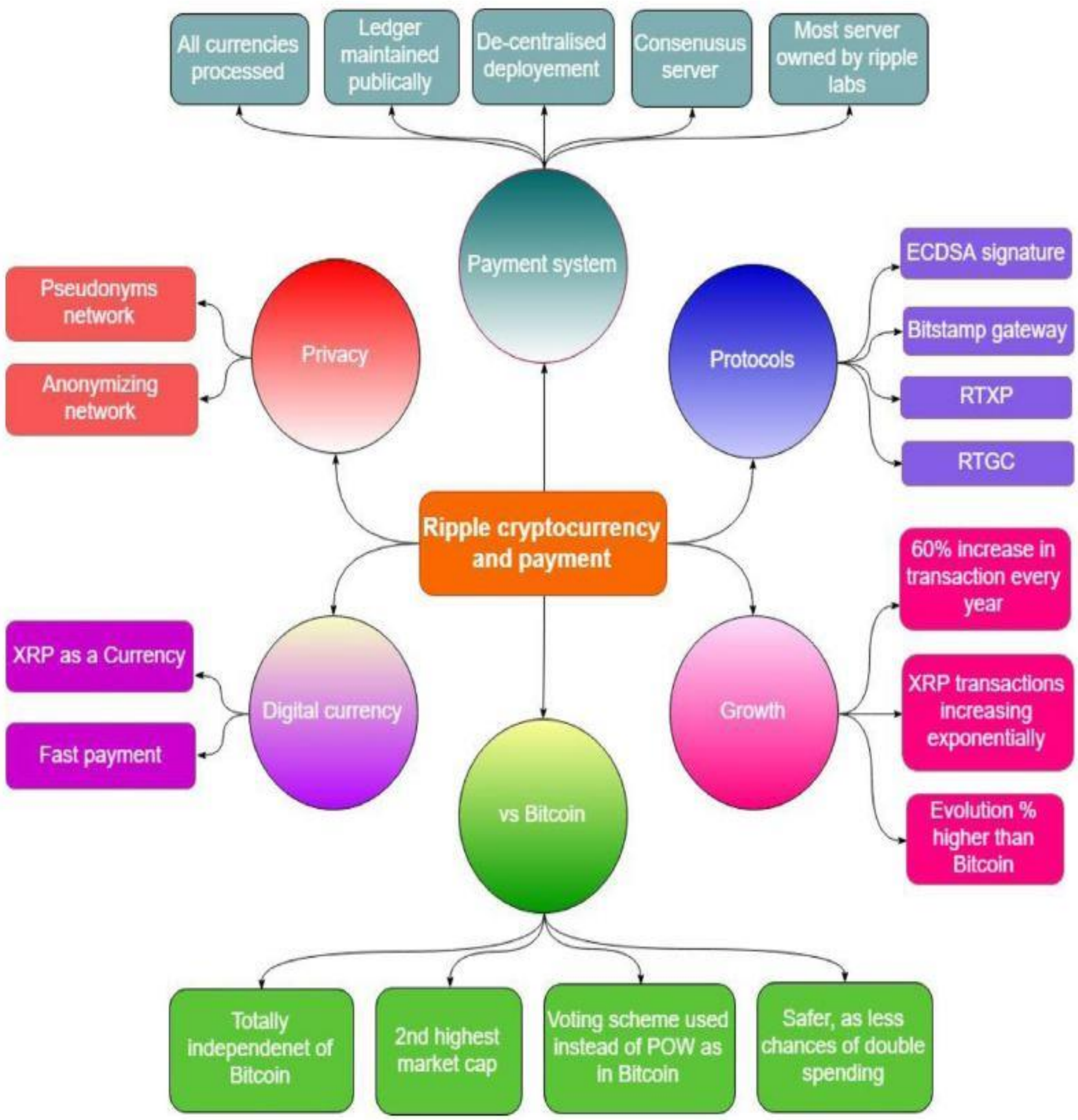


Figure 1. Concept of cryptocurrency

2.2. Concept of Knowledge Discovery for Cryptocurrency Technology

Based on merged information and spread data, the knowledge discovery distributed architecture shows the incompleteness of regional knowledge. Multiple nodes are involved in analyzing big data and choosing the level of computing efficiency [15]. In order to provide detailed answers to the questions, multiple data mining and machine learning methods are used [16]. Three primary groups of techniques are summarized in the study of identifying information, privacy preferences, and data privacy concerns: cryptographic, reconstruction, and heuristic-based techniques. Mendis et al. [17] Envisioned using blockchain nodes for training and aggregation combined with machine learning to provide privacy protection. Statistical analysis is presented in Table 2 for cryptocurrency price prediction. This table is divided into six categories: kind of digital coin, labels, transaction, features, applied algorithm, and performance.

Table 2. Comparison of cryptocurrencies performance with various features

#	Crypto	Considered Labels	Time of Transaction	Selected Features	Applied Algorithm	Final Performance
1 [18]	Bitcoin	Address exchange, Service of gambling,	Nov 2018	Embedding	HDDT+ ECOC	0.91%
2 [19]	Bitcoin	Exchange entity, Gambling entity,	0-561 Blocks	Network, Volume	GBDT, Cascading	0.99%

		Mining entity, Mining pool,				
3 [20]	Bitcoin	Mining address pool, Services for mixology, Minors, Interchange	520.850 to 520.950 Blocks	Embedding, Temporal, Volume	RF	0.96%
4 [21]	Bitcoin	Faucet openings, Exchange, Gambling, HYIP	2009-2017	Network, Volume, Temporal	RF	0.70%
5 [22]	Bitcoin	Address interchange, Faucet, HYIP, Mining Pool	2009-2018	Network, Volume, Temporal	Light GBM	0.86%
6 [23]	Bitcoin	Enterprises that hold exchanges, Mining pool, Darknet	0-514.971	Network, Volume,	Temporal	0.91%

		market				
7 [24]	Bitcoin	Exchange entity, Hosted wallet, Darknet market, Merchant service	Not disclosed	Network, Volume, Temporal	Extra trees	0.96%
8 [25]	Ethereum	Smart contract authors	-	Stylometric	RF	0.91%
9 [26]	Litecoin	Daily price	2009-2018	Market information, Network	RF	Prediction contribution with origin feature
10 [27]	Ethereum	Daily Price	2016-2018	Mining difficulty, Volume, Market information	LR	0.99%
11 [28]	Litecoin	Daily price	2017-2019	Mining node, Volume,	SNN	Lowest MAPE

				Market document		
12 [29]	Bitcoin	Daily price	2015-2017	Mining, Volume, Market information	GASEN	0.64%
13 [30]	Bitcoin	Five min price orientation in one day	2017-2019	Mining hardness, Capacity, Market information	LR, LSTM	0.66%
14 [31]	Bitcoin	30th, 90th and next day prediction	2013-2019	Mining difficulty, Volume, Market information	LSTM	MAE, RMSE, MAPE, 0.62% to 0.65%
15 [32]	Bitcoin	Direction and daily price	2017	Network, volume	PDE	0.82%

Researchers around the world have an interest in cryptocurrency prices. Prices fluctuate based on various factors, including transaction costs, difficulty in mining, market trends, popularity, alternative coins, and others. Cryptocurrency prices can change over time, making prediction difficult due to the named factors. A look at the blockchain and machine learning-based prediction of cryptocurrency prices is provided in Figure 2. Bitcoin, litecoin, ripple, monero, tether, and IOTA are some of the categories of cryptocurrency price data available. They used Litecoin and Monero as data sources for this process. It then organizes the data for the blockchain network and feature extraction by processing it and slipping it into a training and test set. There are several blocks and transaction IDs included in the blockchain framework. Blockchain features include cryptocurrency tokens, digital wallets, smart contracts, and distributed apps.

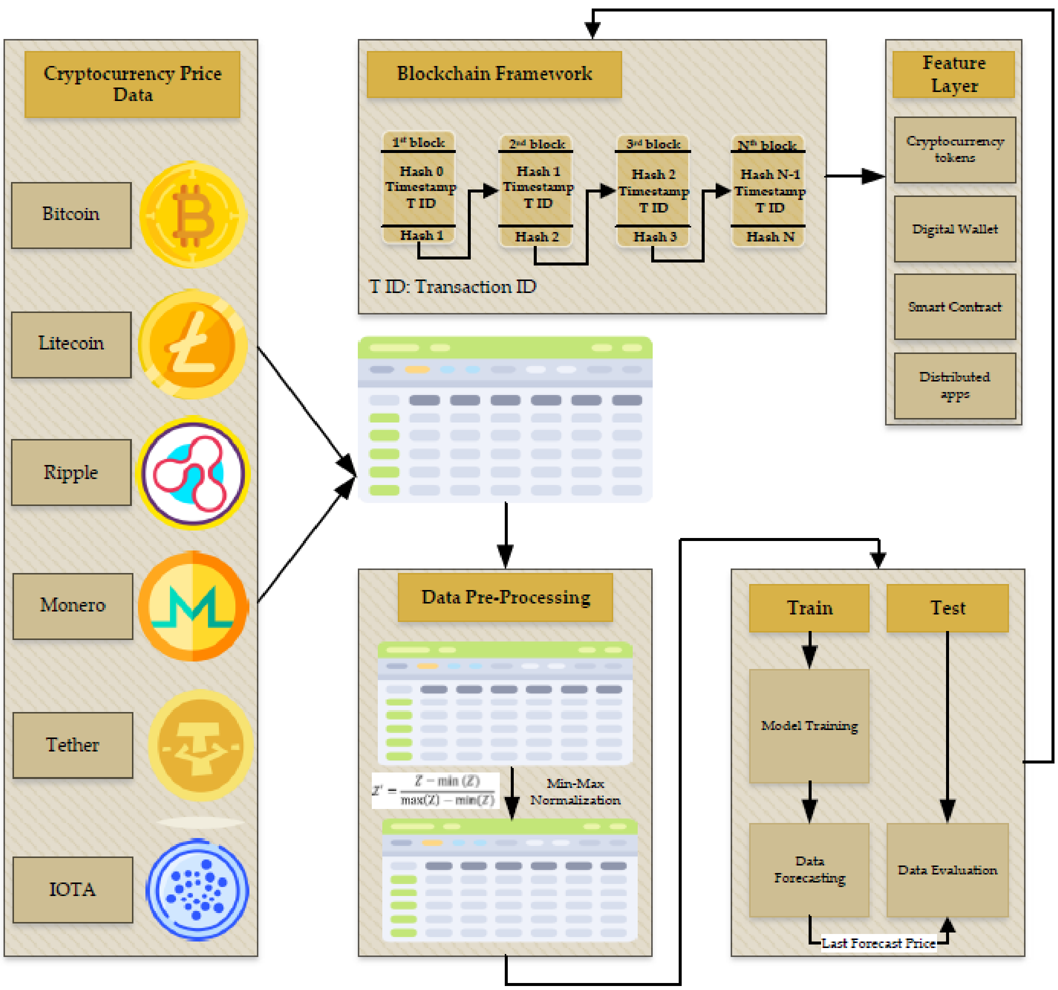


Figure 2. An overview of the process of predicting the price of cryptocurrencies.

2.2.1 Topic Modeling

Topic identification and text segmentation are the two primary tasks of subject analysis. Most text processing algorithms employ such a strategy as automatic text indexing in knowledge recovery approaches. By specifying topics and sub-topics in a document, topic identification and segmentation are intended to find differences. Modeling topics provides a method for identifying confidential information in a document. Word clusters are called topic clusters [33].

An essential challenge in large corpora of documents utilizing data and technology is analyzing the amount of data in each document. The cutting-edge technique from LDA is used to study latent subjects in documents, which can be measured as a pattern of comments [34]. Text segmentation based on topical passages, also known as boundary detection, is one of the areas of study related to topical courses. By operating text segments, the research aims to compare the before and after of each boundary. The process initially breaks the text into three to five sentences, resulting in vectors. Potential boundaries between topics can be identified based on the cosine similarity between vectors [35].

Salton and Singhal [36] Automated text overview is based on passage extraction employing knowledge of text frame. The inter-document link generation approach generates intra-document links within this method. However, spoken dialogue can be segmented using text segmentation. Researchers in [37] proposed lexical chaining using CNN transcripts segmented coarsely. Based on the Maximum Entropy model for lecture and topic segments, Christensen and Kolluru introduced an automatic speech recognition system (ASR) [38]. Huzeh and Moore [39] have looked at how lexical cohesion can predict segment boundaries in spoken multi-part dialogs using a verbal cohesion model. Research in topic segmentation relevant to the dialog domain applies most of the same techniques as text segmentation, but they're reconciled with spoken language features. By calculating the sentence distance matrix presented by Ji and Zha in [40], the authors calculate the cohesion of information contained in sentences. In addition to determining the optimal boundary topics, each entry is based on the similarity between a pair of sentences. As a final step, the matrices were transformed against grayscale images, and the topic boundaries were sharpened using anisotropic diffusion.

QA's passage retrieval is the intermediate step for identifying text regions. Additionally, it provides answers to user questions [41], which aims to improve quality assurance. QA passages are either fixed length or variable length. Another way of participating or fragmenting sentence passages is to use word parts [42] and provide semantic clues such as similarities between sentences, the relevance of questions, and conjunctions [43].

In addition to the word-based segmentation system, there are many different processes for text segmentation, including algorithms based on identified topics. It requires latent semantic analysis to map high dimensional material into low dimensional semantic space to calculate topic distributions based on latent Dirichlet allocation (LDA). The latent semantic analysis must map the high-dimensional content to a low-dimensional semantic space. A text segmentation method would also need to identify relationships between sub-topics and map topics to segments in a subsequent step [44,45]. Document groups are found operating LDA in many existing works. The LDA model does not have document design knowledge and cannot extract more information from text.

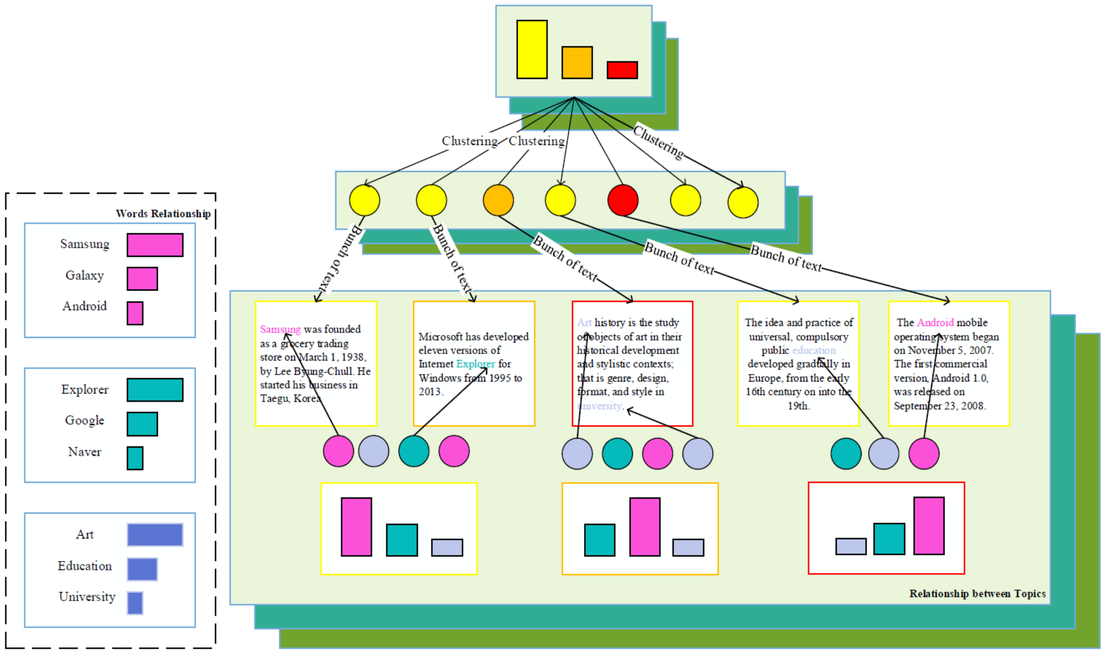


Figure 3. Topic modeling process

Each cluster is associated with one document in Figure 3. Groups are related to topics in a specific way. One collection contributes to one document and is not associated with the information of other users. A topic model is developed using the LDA system, which already evaluates the clustering of each document. Word relationships are used to choose topics per category.

2.2.2 Topic Discovery

LDA topic modeling: The Latent Dirichlet Allocation (LDA) is a popular model in terms of generative statistics for topic generalization, illustrating unobserved or hidden parts of perceptions [46]. The similarity between subtopics is determined using LDA. Generalizing the different subtopics into one main topic causes the word embedding process and LDA approach for further processing. Figure. 4 shows how LDA works in detail. A case-by-case approach used in LDA means the separation of documents and the topics generated with the LDA process has the highest performance and correct segmentation in a large corpus of data. The segmentation process is also unknown for short types of text data, which can contain only one text or a limited document length. By optimizing the value and adjusting the parameters, LDA topic modeling provides optimal options. The similarity between subtopics is determined by operating LDA to generalize the subtopics to one main topic and further process it based on the word embedding output file.

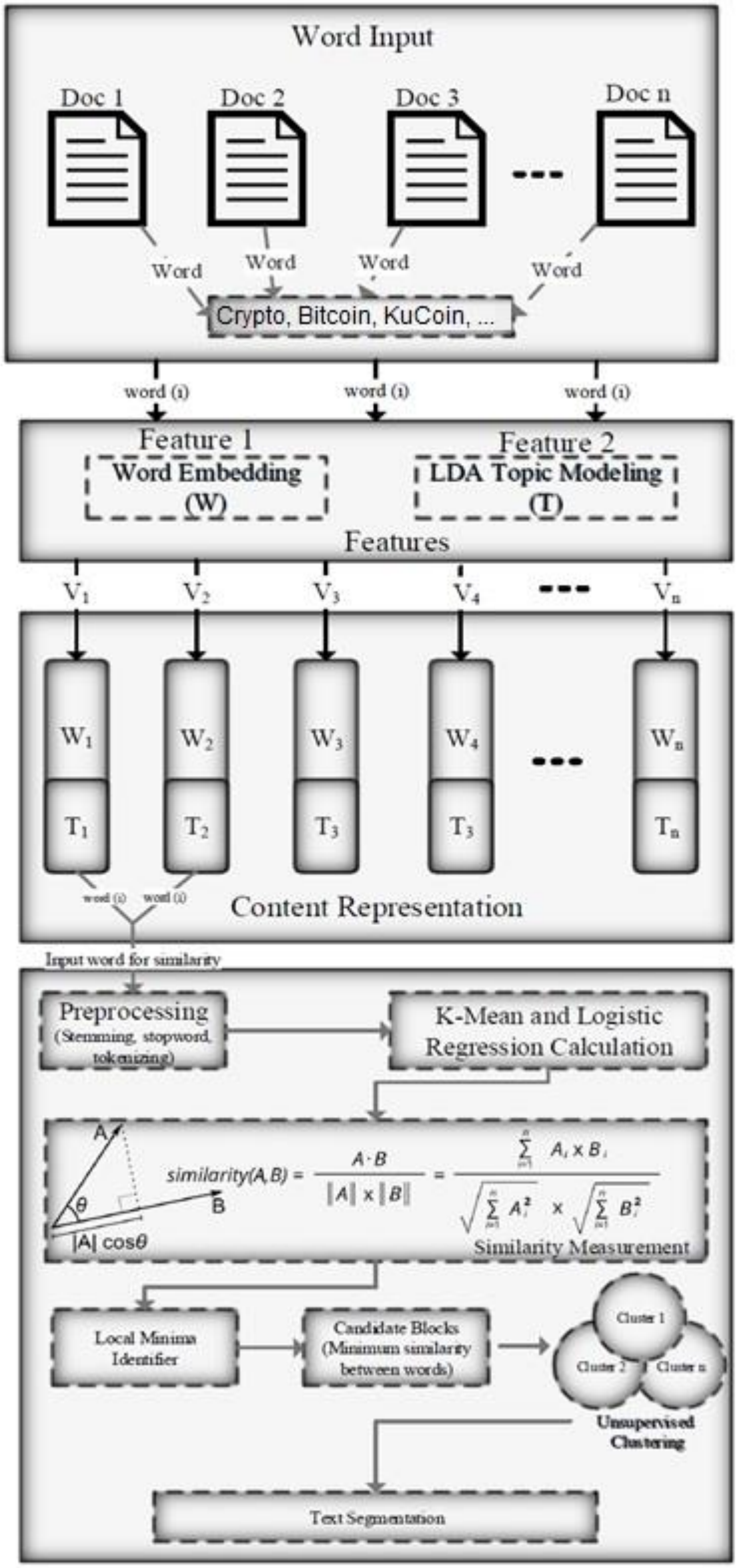


Figure 4. Visualization of topic discovery content structure

2.2.3 Information Analysis System

This information analysis system analyzes user conduct on the available dataset and predicts user preferences based on the analysis. User behavior includes click information, ratings on bought things, statements, sharing, cart information, etc. are all included in user behavior [47–52].

Its primary purpose is to assist users who lack information on specific items to make informed purchasing decisions. The scientific community has presented various RS since the 1990s, including articles, movies, and products [53-60]. There are five main types of information analysis systems: (1) collaborative filtering, (2) content-based, (3) knowledge-based, (4) hybrid, and (5) demographic. Burke [61] categorized these in this manner. Goldberg et al. [62] are the authors of the collaborative filtering (CF) concept. Users or products can be grouped at the same point in the CF predicting process [63].

The assumption is that if two users have a similar interest in one effect, they will also have a similar interest in other developments. Researchers have found that neighborhood-based methods are more adaptable and efficient than different techniques in adaptability, efficiency, etc. A majority of studies have used machine learning algorithms and strategies to enhance the performance of existing approaches. CF contains several potential aspects, but it has a sparsity issue due to a lack of interest in products [64]. There is a significant discrepancy between the number of items bought and the number of ratings customers give. Many of the items purchased have few or no ratings.

The online shopping system has captured somebody's attention throughout the world during the last few years. A lot of people have access to online shopping websites like Amazon. The primary focus of some CF strategies is the items purchased by a user only during a specific period; Health care is one example of an IoT-based platform [65-67]. In terms of accuracy, one of the most promising uses of IoT is indoor localization [68-69], which is highly accurate are added. Using two attributes, i.e., the dataset and the user behavior changes, Song et al. [73] analyzed changes in user behavior.

Examination results show no significant purchasing patterns changes observed using the defined rules. A few studies examined temporal dynamics [74–81]. Time spent on various web pages is associated with time spent purchasing.

As shown in Figure 5, the suggested system predicts neighbor cryptos. Upon clicking on certain items, the neighbors are generated. Another critical factor is the number of crypto directions. For one crypto, it's common for the user to click several times in various approaches. A more detailed explanation of this crypto may interest the user. Each user follows the exact procedure. Therefore, this model seeks to determine the complicated relationship between product and user.

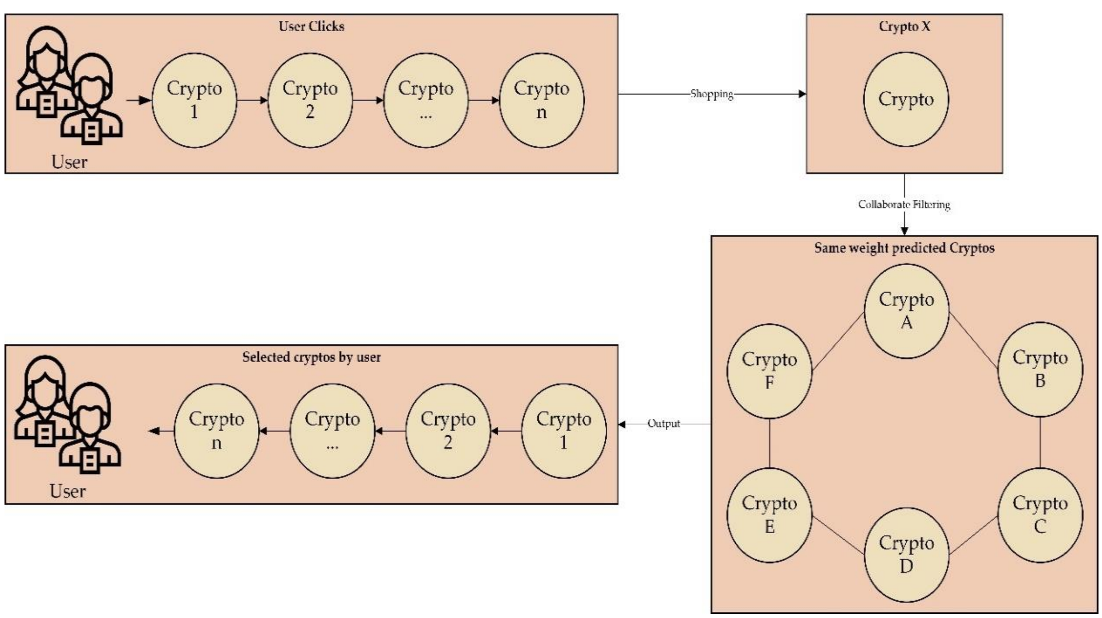


Figure 5. Information analysis system scenario

2.3. Concept of Cryptocurrency in Blockchain

Peer-to-peer payment systems with blockchain technology take things a step further. This enables the application of the Internet of Things for distributed storage applications, giving the system trust, privacy, and security [82]. Many blockchains and cryptocurrencies are designed to cater to specific application ranges. The connection between cryptocurrency and Blockchain arises because cryptocurrency provides an incentive for machines to validate blockchains. Blockchain-based currencies use cryptography to increase transparency, immutability, and decentralization [83]. Likewise, cryptocurrency usage has risen in correlation with the increased use of Blockchain. Blockchain has an inherent value influenced by a wide range of factors. The process saves the values and allows for a better understanding of price changes based on the deals. Figure 6 shows cryptocurrency coin prices over time. Almost 19 billion US dollars' worth of market capitalization was recorded for 2017. The market capitalization of 90% of the market is based on seven currencies.

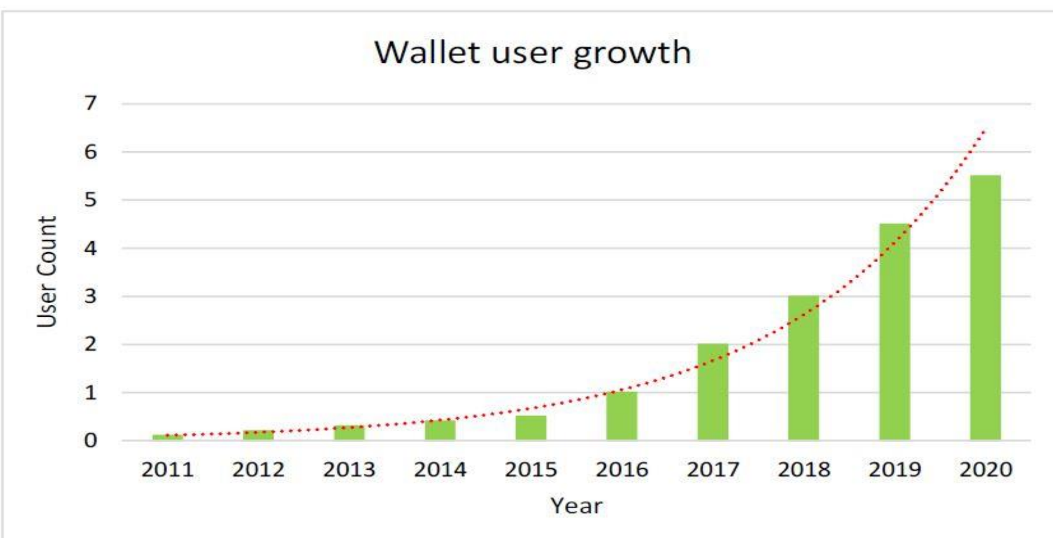


Figure 6. User wallet growth in years

2.4. Arts and Regulations of Financial Statements

Banks and financial institutions made certain adjustments to identify money laundering and suspicious activity. Lack of oversight regarding appropriate regulations led to increased money laundering in the cryptocurrency system [84]. Cryptocurrency schemes pose a security risk due to a lack of regulation [85]. Governments should not restrict Blockchain and cryptocurrencies because of their popularity in developing regulations to govern financial institutions, including safeguards to prevent losses [86]. According to one of the case studies from the Financial Action Task Force [87], the Altaf Khanami was laundering a large amount of money illegally (Billions of dollars) for drugs, weapons, and various terrorist groups. Money from Nigeria has also been used in support of criminal activities. Figure 7 explains the cryptocurrency components and verification method.

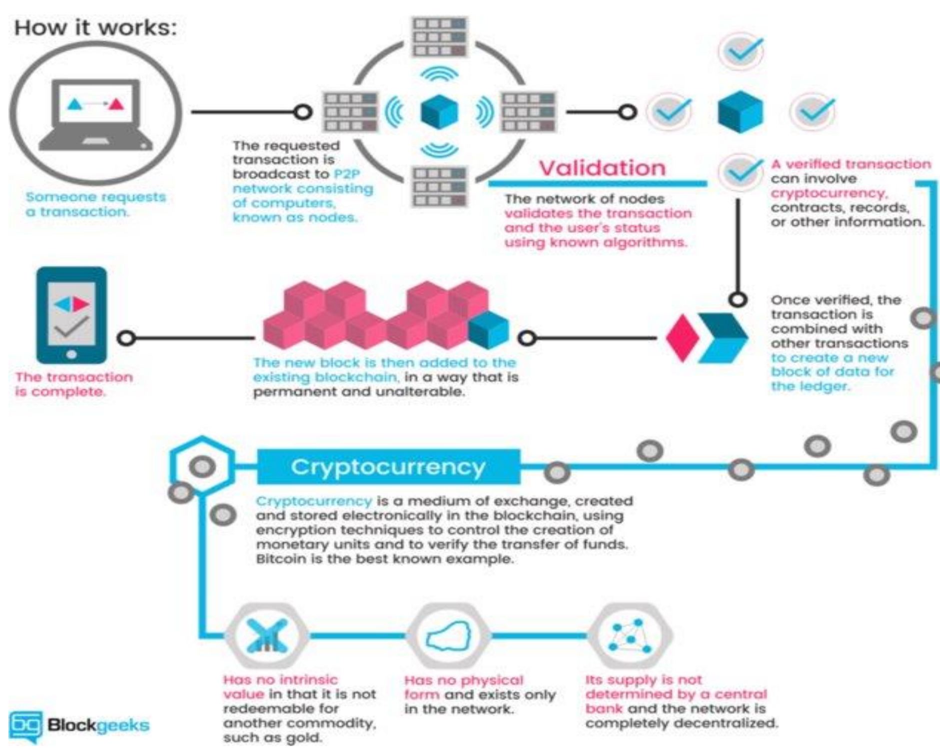


Figure 7. Cryptocurrency detail process

The following Table 3 provides information about the cipher trace interviews attended by banks involved in the financial investigation.

Table 3. The total number of banks related to financial investigation and education attended to interview of cipher trace.

Participants of Research	Job Title	Number of Bank	Location	Years in current position
1	Risk specialist BSA/AML	2	San Francisco	32 Years
2	Risk specialist BSA/AML	1	San Francisco	3 Years
3	Risk specialist BSA/AML	3	San Francisco	1 Years
4	Risk specialist BSA/AML	2	San Francisco	3 Years
5	Risk specialist BSA/AML	2	San Francisco	1 Year
6	Risk specialist BSA/AML	3	San Francisco	7 Years
7	Financial crimes investigator	2	San Francisco	10 Years
8	Financial spiffy unit administrator	1	San Francisco	3 Years

9	Director of cipher trace	N/A	Texas	2 Years
10	Director of due diligence	1	San Francisco	4 Years
11	Compliance manager	2	San Francisco	12 Years
12	Compliance manager	1	San Francisco	6 Years
13	Policy execution supervisor	3	San Francisco	3 Years
14	Investigator	2	San Francisco	16 Years
15	Manager	1	San Francisco	2 Years
16	leader	2	San Francisco	1.5 Years
17	Manager	1	San Francisco	4 Years

2.5 Financial Institutions Efficiency in Money Laundering

As the poles of finance flow, the financial industry and banks are prime money laundering targets. Money laundering tarnishes banks' reputation, which loses customers' trust and value and makes them fear for the safety of financial transactions [88]. Zali et al. [89] used the legal trick for criminals by successfully moving money into another account with the help of the financial employees to avoid paying fines due to the lack of detailed anti-money laundering policies. Figure 8 shows how cryptocurrency architecture facilitates money laundering.

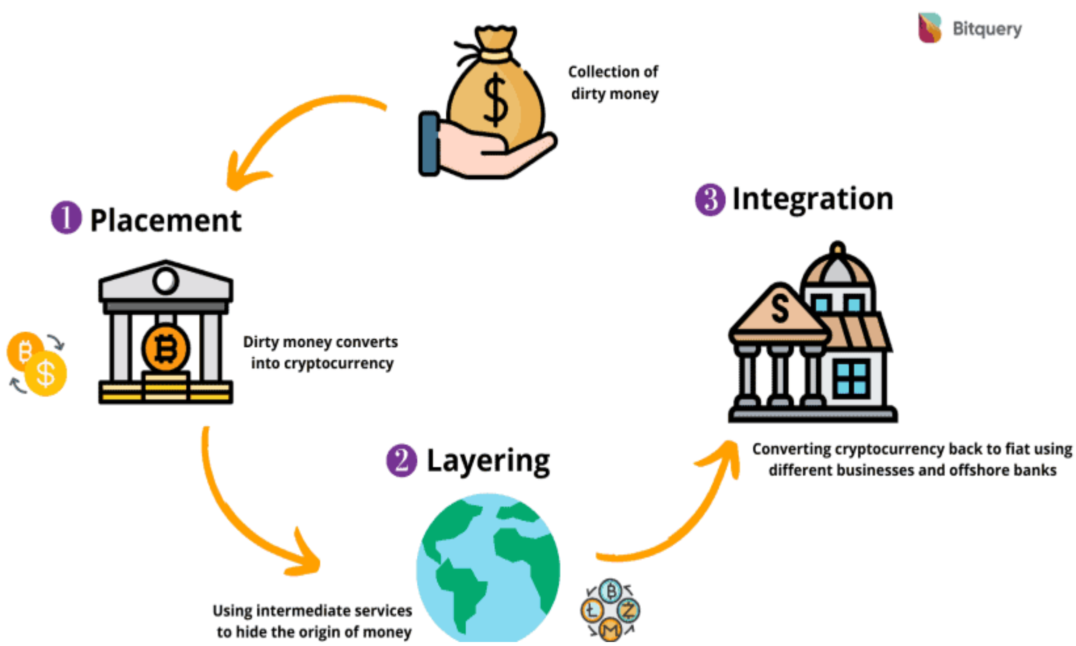


Figure 8. Money laundering detail process

Table 4 shows the penalties that USA banks paid from 2012 to 2019.

Table 4. Paid penalties of USA for money laundering

Bank Title	**Time Period**	**Paid Price of Penalty**
California Pacific Bank	2019	225.000$
Citibank NA	2018	70M
U.S. Bank NA	2018	613 M
Banamex (Citigroup)	2017	97 M
JPMorgan Chase	2014	2.05 B
HSBS	2012	1.92 B

2.6 Money Laundering in Cryptocurrency System

Financial crimes increased sixfold between 2015 and 2018 in various cryptocurrency types [90]. Since the advent of cryptocurrency, tax evasion, money laundering, and terrorism have become standard financial practices. Criminals do not need to be identified, making it difficult for banks to identify them and track their activities [91]. Cipher-Trace [92] states the misuse of funds, which appears to be associated with cryptocurrency holders with almost 4.3 billion dollars in 2019. Because of the developed level of cryptocurrency technology, financial sector institutions struggle to deal with cryptocurrency crimes due to the lack of knowledge regarding the bank systems and the ability to prevent money laundering [93]. Technology advances such as computer-based cryptography have enhanced cryptocurrency enforcement entities, which have not been wholly adapted yet [94]. Table 5 explains the market capabilities of different cryptocurrencies in 2020. That Five cryptocurrencies have been documented and compared.

Table 5. The records of cryptocurrency market capability during 2020

#	Type of Crypto	Market Capability	Established Date
1	Tether	5 B	2017
2	Bitcoin Cash	7 B	2011
3	XPR	12 B	2012
4	Ether	24 B	2013
5	Bitcoin	177 B	2009

2.7 Cryptocurrency Prediction Models

Digital coin prices have been predicted with machine learning algorithms in recent years. The linear regression model (LR) and support vector machine (SVM) has been used by Sin et al. [95] to predict the bitcoin price based on time-series data information from 2012 and 2018. A prediction model based on low error rates matches multiple parameters.

The window lengths and coefficient weights are used to construct filters. The price of bitcoin can be predicted with various window sizes through filters. The model developed by Azari et al. [96] utilized ARIMA to forecast the future value of bitcoin based on data from 2015 to 2018. The results report requires a minimum of 0.02 Residual Sum Square. Figure 9 summarizes the proposed approach. Machine learning regression and classification models were used to estimate digital coin prices.

The training and testing phases are both parts of the process. Further crypto-exchange rate prediction is conducted using blockchain network validation and verification. Azari et al. [96]

proposed the prediction model based on ARIMA for predicting the future value of bitcoin based on the available datasets between 2015 and 2018. A minimum of 0.02 residual sum of squares is shown in the results. Hans et al. [97] compared ARIMA and Prophet in an experiment to predict cash flow by applying LSTM and multilayer perceptrons. A diagram of the proposed method is shown in Figure 9. Regression and classification models were used to predict the price of digital coins.

The train and test phases both involve several tasks. The algorithms involved in these phases determine which features to select and which URLs to classify. Blockchain networks further process the exchange rate prediction by verifying transactions and validating their validity. During the prediction phase, the mentioned database shows data about the 30-day and 90-day predictions of the transactions using Regression and Classification models.

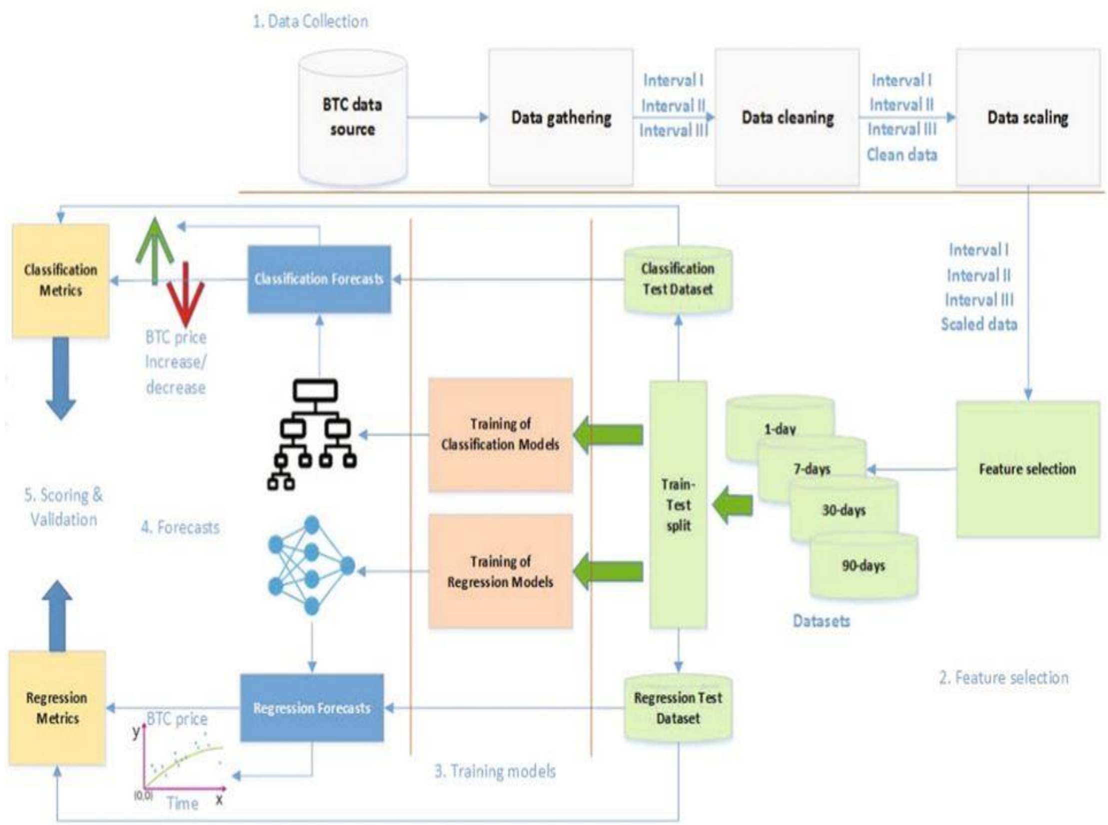


Figure 9. Overview of cryptocurrency exchange rate prediction

Chapter 3: Proposed Cryptocurrency Price Prediction

This chapter presents our proposed Cryptocurrency price prediction based on knowledge discovery on the blockchain framework. The proposal aims to provide the highest-ranking predicted crypto for delivering the highly related crypto to the user request from the system. Furthermore, the proposed mechanism predicts the crypto price and operates the risk management module to provide fault detection and recovery based on the blockchain framework.

In section 3.1, we present the proposed cryptocurrency prediction architecture. Section 3.2 offers the general architecture of cryptocurrency prediction. Section 3.3 presents the process of prediction and optimization. Finally, section 3.4 presents the proposed prediction system for risk management using the blockchain framework.

3.1 Cryptocurrency Prediction Architecture

This section presents the main overview of the proposed cryptocurrency prediction. Figure 10: shows the components applied in this system. There are four main layers in the designed framework which are defined: (1) Data Acquisition for cryptocurrency, (2) System Process (Training and Testing), (3) Control and Optimization, and (4) prediction. Figure 10 shows that we have collected the dataset from various social media websites and analyzed them based on different factors. The raw data is stored in the blockchain database for training and testing purposes. The control and optimization layer is processed by applying a reinforcement learning algorithm to improve the system performance.

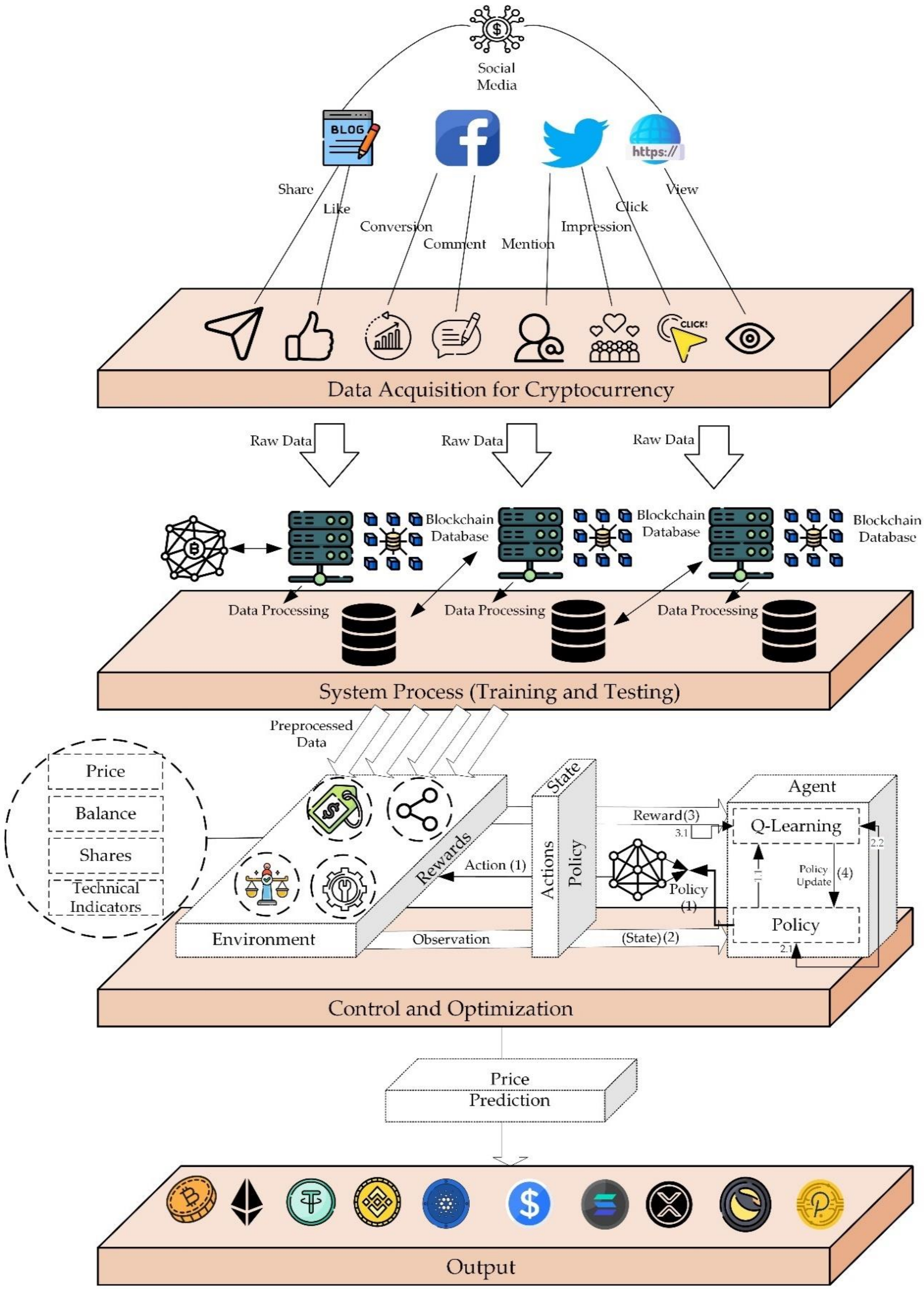


Figure 10. Main Architecture of knowledge discovery and cryptocurrency prediction process

3.1.1 Data Acquisition for Cryptocurrency

This layer presents the details of data collection from social media websites. To do this, we have collected the data from various social media websites, and our target for this purpose was Twitter, Facebook, Blogs, and other websites related to crypto information. The second data source for this system is from the https://coinmarketcap.com/all/views/all/ website, which gives the hourly information on crypto price changes, volume, etc. To crawl the data, we have used the data miner feature, one of the prominent features of google chrome. The data is collected based on the user's information sharing, likes, conversion, comments, mentions, impressions, clicks, and views. Figure 11: Shows the details of data collection and processing for the prediction.

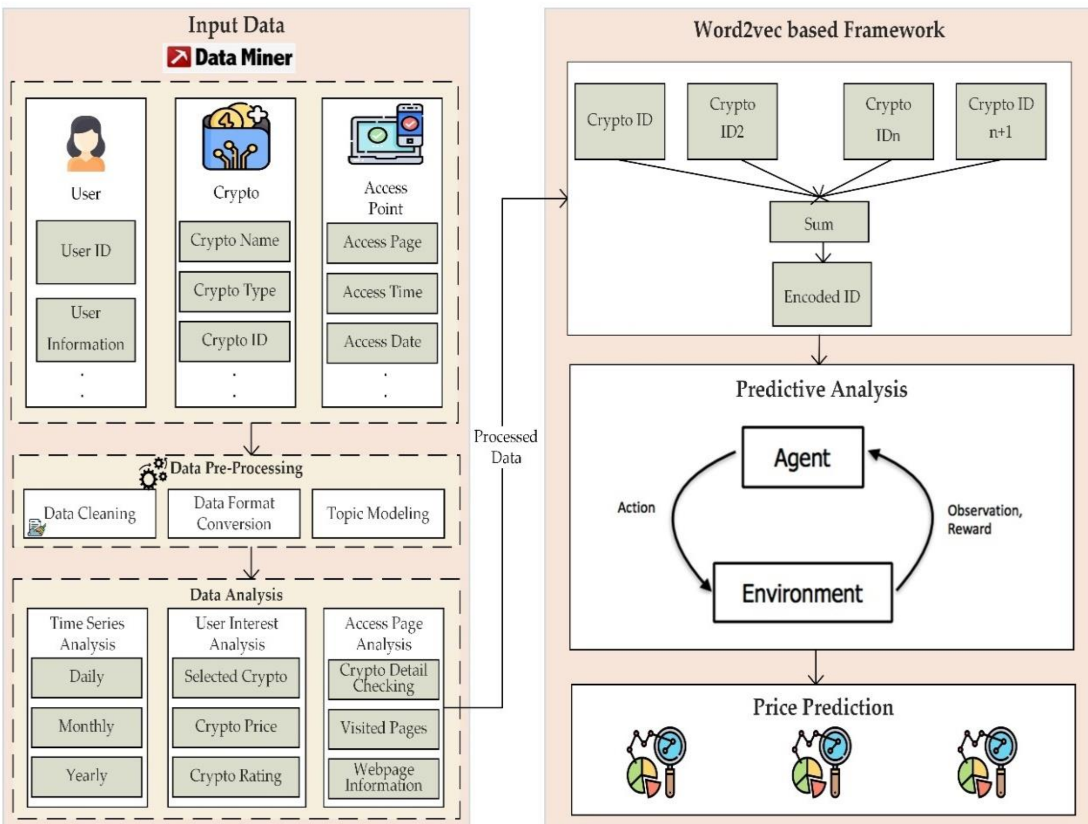
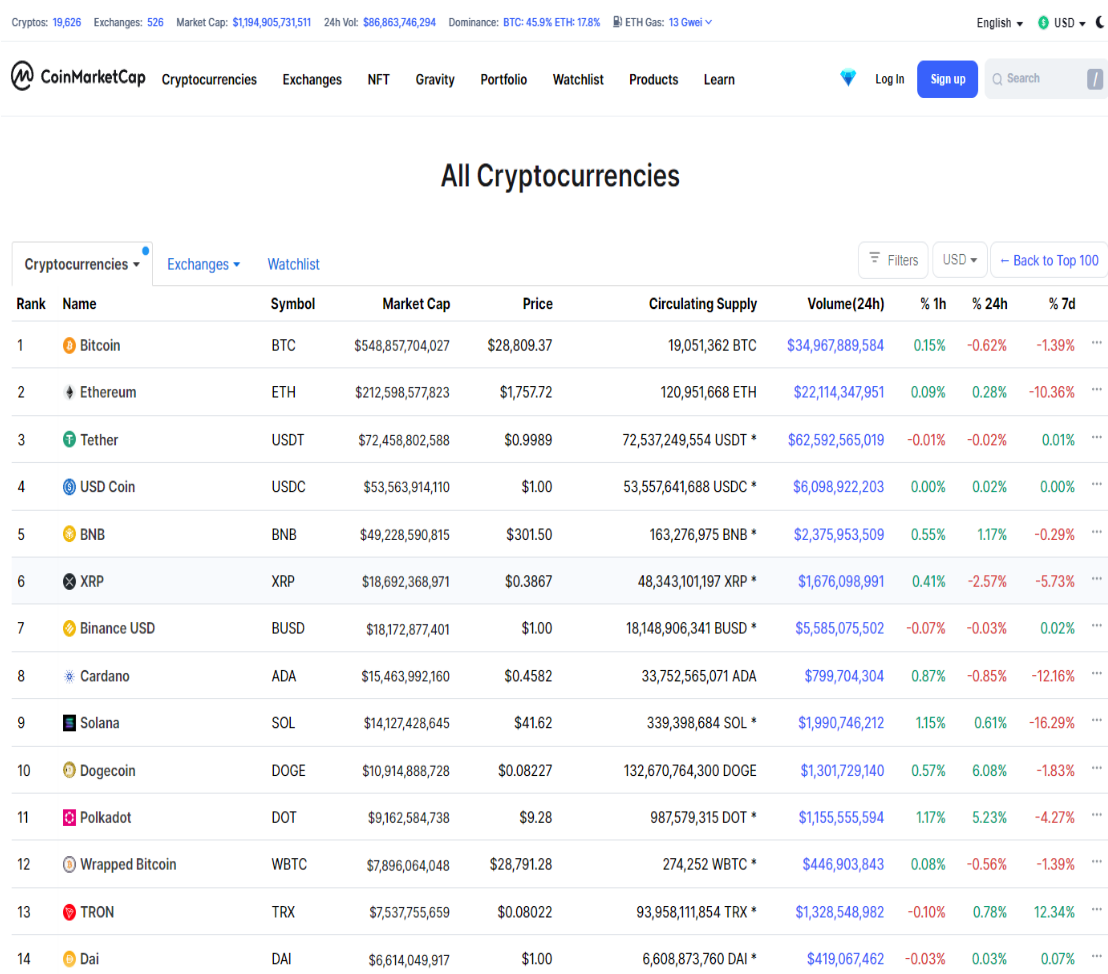


Figure 11. Detailed data collection and processing steps

One of the data collection sources is https://coinmarketcap.com/all/views/all/. As shown in Figure 12, the data collection environment is as follows: the complete version of digital coins hourly changes records. This website presents the cryptocurrency types, price, 24-hour percentage, seven days percentage, market cap differences, etc., are presented on this website. The data time duration is from 2022 for 3-months records. The total records of all the collected data are 100.000 records which took almost one week to collect due to collecting from multiple sources.


Cryptos: 19,626 Exchanges: 526 Market Cap: $1,194,905,731,511 24h Vol: $86,863,746,294 Dominance: BTC: 45.9% ETH: 17.8% ETH Gas: 13 Gwei English USD

CoinMarketCap Cryptocurrencies Exchanges NFT Gravity Portfolio Watchlist Products Learn Log In Sign up Search

All Cryptocurrencies

Cryptocurrencies Exchanges Watchlist Filters USD ← Back to Top 100

Rank	Name	Symbol	Market Cap	Price	Circulating Supply	Volume(24h)	% 1h	% 24h	% 7d
1	Bitcoin	BTC	$548,857,704,027	$28,809.37	19,051,362 BTC	$34,967,889,584	0.15%	-0.62%	-1.39%
2	Ethereum	ETH	$212,598,577,823	$1,757.72	120,951,668 ETH	$22,114,347,951	0.09%	0.28%	-10.36%
3	Tether	USDT	$72,458,802,588	$0.9989	72,537,249,554 USDT *	$62,592,565,019	-0.01%	-0.02%	0.01%
4	USD Coin	USDC	$53,563,914,110	$1.00	53,557,641,688 USDC *	$6,098,922,203	0.00%	0.02%	0.00%
5	BNB	BNB	$49,228,590,815	$301.50	163,276,975 BNB *	$2,375,953,509	0.55%	1.17%	-0.29%
6	XRP	XRP	$18,692,368,971	$0.3867	48,343,101,197 XRP *	$1,676,098,991	0.41%	-2.57%	-5.73%
7	Binance USD	BUSD	$18,172,877,401	$1.00	18,148,906,341 BUSD *	$5,585,075,502	-0.07%	-0.03%	0.02%
8	Cardano	ADA	$15,463,992,160	$0.4582	33,752,565,071 ADA	$799,704,304	0.87%	-0.85%	-12.16%
9	Solana	SOL	$14,127,428,645	$41.62	339,398,684 SOL *	$1,990,746,212	1.15%	0.61%	-16.29%
10	Dogecoin	DOGE	$10,914,888,728	$0.08227	132,670,764,300 DOGE	$1,301,729,140	0.57%	6.08%	-1.83%
11	Polkadot	DOT	$9,162,584,738	$9.28	987,579,315 DOT *	$1,155,555,594	1.17%	5.23%	-4.27%
12	Wrapped Bitcoin	WBTC	$7,896,064,048	$28,791.28	274,252 WBTC *	$446,903,843	0.08%	-0.56%	-1.39%
13	TRON	TRX	$7,537,755,659	$0.08022	93,958,111,854 TRX *	$1,328,548,982	-0.10%	0.78%	12.34%
14	Dai	DAI	$6,614,049,917	$1.00	6,608,873,760 DAI *	$419,067,462	-0.03%	0.03%	0.07%

Figure 12. Detailed data collection and processing steps

The data collection and data crawling process is from the data miner data crawler system. We focus on Facebook, Twitter, Blogs, users comments on these platforms, likes, dislikes, tweets, and retweets. Figure 13: shows the data miner platform and the proposed system process's data collection.

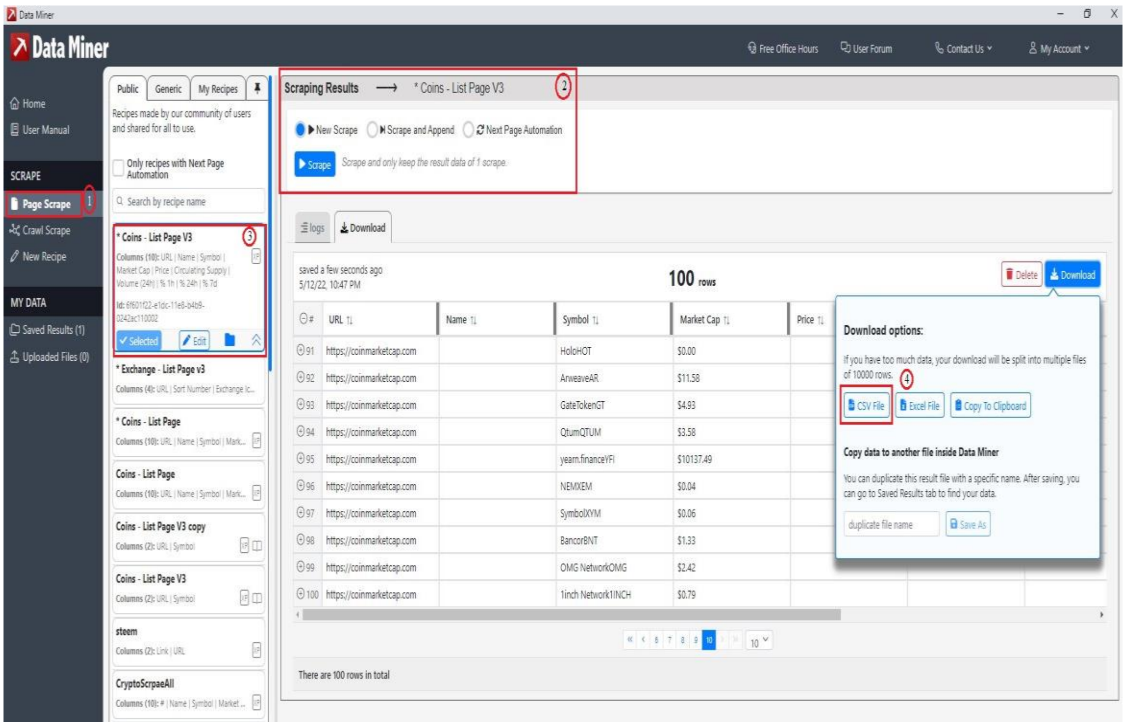


Figure 13: Data miner platform for data crawling

Table 6: shows the details of the collected dataset from the social media platforms using the data miner web crawler platform.

Table 6: Overview of collected dataset

Coin Name	User Location	Hashtags	Price	Sources
Bitcoin	Atlanta, GA	#bitcoin	2686.6	Twitter Web App
Litecoin	London, England	#cryptocurrency trading	2702.45	Buffer Finance
Bitcoin	Florida, USA	#bitcoin #crypto	2699.76	Btc_p_bot
Ethereum	Atlanta, GA	#BTC #ETH	2697.45	Twitter Web App

3.1.2 System Process

The system processing layer is designed for the raw data train and test purposes. The collected data from the previous layer is stored in the blockchain database for security reasons not to be accessible to everyone to avoid any attack from hackers. The crypto and user dataset is a type of sensitive information that is not allowed to be accessed in public and requires storing in a blockchain framework. The data processing, training, and test set are processed in this section.

3.1.3 Control and Optimization

This layer mainly focuses on the system performance. In this section, we have applied the reinforcement learning algorithm due to the learning-based aspect. Reinforcement learning input is the collected data mentioned above, and the output is the predicted cryptos to the user. In the case of a suitable prediction, the user rating allows Reinforcement learning to process the prediction correctly. In case of a wrong prediction, the system will change the process and predict the results until getting the correct output.

3.2 Topic Modeling Based on LDA in Cryptocurrency Dataset

In this section, the LDA-based topic modeling process is presented. LDA is one of the popular techniques for data pre-processing and clustering. Figure 14: Shows the LDA topic modeling and clustering architecture in the proposed system.

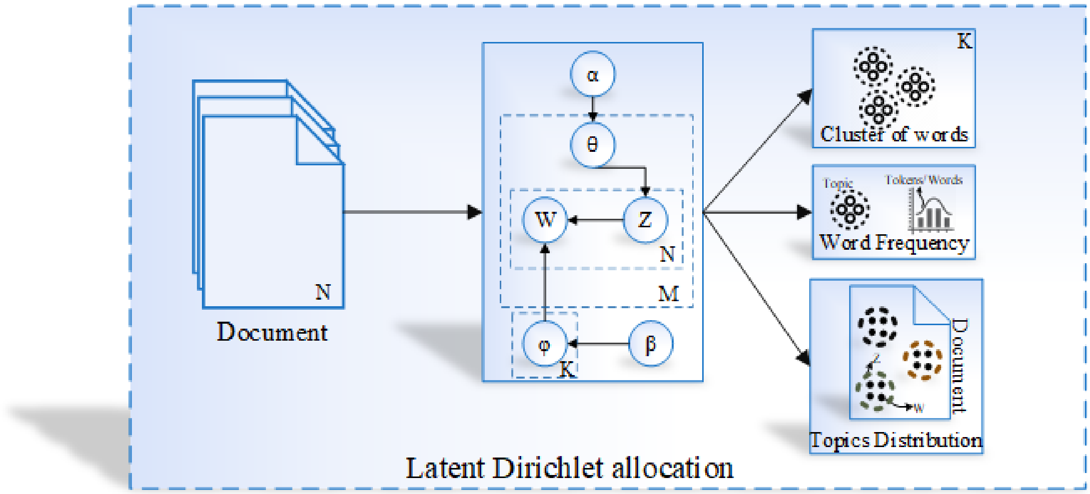


Figure 14. LDA topic modeling detail architecture in cryptocurrency dataset

Due to various types of cryptocurrencies, topic modeling is one of the required processes for cryptocurrency prediction. The highest probable topic regarding user request's defined factors will be categorized and extracted from the input dataset in this process. Several steps are required for topic modeling. (1) data pre-processing, (2) applying LDA, and (3) topic probability distribution. Table 7: presents the details of LDA parameters.

Table 7. LDA parameters for data processing

LDA Parameters	Description	Type
α	k-dimensional	Topic weight
β	v-dimensional	Word weight
θ	Float	Probability
K	Integer	Topics quantity
W	Input words	Word
N	INT type	Words quantity
Z	n-dimensional	Topic value representation

3.2.1 Data pre-processing

The pre-processing data section contains the steps of stop word removal, stemming dataset, tokenization, removing quotes, removing URLs, etc. The main focus here is to extract the information of every crypto for the future perdition steps. The word distribution and topic distribution were evaluated by applying the LDA technique. The output of this step is the highest topic probability of the cryptocurrency dataset, which is suitable for further application. The cosine similarity between segments is calculated based on the vectors denoted by Y_i . This is useful for locating possible breakpoints between segments. Domain-independent text segmentation algorithm also used this process to realize the approximate performance. Cosine similarity between the two parts was evaluated as:

$$\boldsymbol{Similarity} = \mathbf{cos}(\boldsymbol{\theta}) \frac{\boldsymbol{X.Y}}{||\boldsymbol{X}||||\boldsymbol{Y}||} = \frac{\sum_{n}^{i=1} \boldsymbol{X_i Y_i}}{\sqrt{\sum_{n}^{i=1} \boldsymbol{X_i^2}} \sqrt{\sum_{n}^{i=1} \boldsymbol{Y_i^2}}} \tag{1}$$

$$\boldsymbol{C_{Distance}} = \frac{\boldsymbol{Cos^{-1}(cosinesimilarity)}}{\boldsymbol{\pi}} \tag{2}$$

$$\boldsymbol{C_{Similarity}} = \{\mathbf{1} - \boldsymbol{C_{Distance}}\} \tag{3}$$

A visual representation of the relationship between two adjacent segments and the local minima identifier is provided. According to the similarity score, there is a substantial reduction in similarity at the borders between two contiguous pieces that differ in knowledge regarding the textual relationships across them. A predetermined sentence segmentation has been applied to the dataset. A comparison between the proposed algorithm and the existing segmentation method is undertaken. The actual segmentation (Pd) and WindowDiff (WD) metrics measure the difference between accurate and involved segmentation methods. Hyp and eval are hypothesized segmentations, while reference segmentations are based on the evaluated. Segmentation (Ref).

P_d d is the size of words at half the average segmentation size, and if the exact segments aren't reached, as a result, it collects the penalties among the terms. Below Equation specify P_d as:

$$\boldsymbol{P_d} = (\boldsymbol{Hyp}, \boldsymbol{Ref}) = \sum_{X} \leq \boldsymbol{y} \leq \boldsymbol{n}\, \boldsymbol{D_m}(\boldsymbol{x}, \boldsymbol{y}) \left(\boldsymbol{S_{Ref}}\, (\boldsymbol{x}, \boldsymbol{y}) \oplus \boldsymbol{S_{Hyp}}(\boldsymbol{x}, \boldsymbol{y})\right) \tag{4}$$

Both SRef and SHyp are indicators in the case of x and y sentences belonging to the same segment when d is defined as the average of true segmentation. Distance Probability defines as Dm (x, y) for all possible distances between randomly chosen documents. WindowDiff (WD) means the terms fixed-size through the document and disciplines the number of limitations based on Hyp segmentation when the words are not matched in proper segmentation below Equation specify WD as:

$$\boldsymbol{windowDiff_d}(\mathbf{Hyp}, \mathbf{Ref}) = \frac{\sum_{x=1}^{M-d} |\, \boldsymbol{r}(\boldsymbol{x}, \boldsymbol{d}) - \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{d})}{\mathbf{M-d}} \tag{5}$$

Where r(x,d) presents the boundary number of Ref segmentation, which contains x and x+d sentences, h(x,d) illustrates the boundary number of Hyp segmentation, including the x and x+d sentences. M demonstrates the number of penalties, and d offers several word sizes.

Figure 15: Shows the overview of data pre-processing steps. There are four major parts of sorting datasets, data statistics, filtering data, and generating from data patterns.

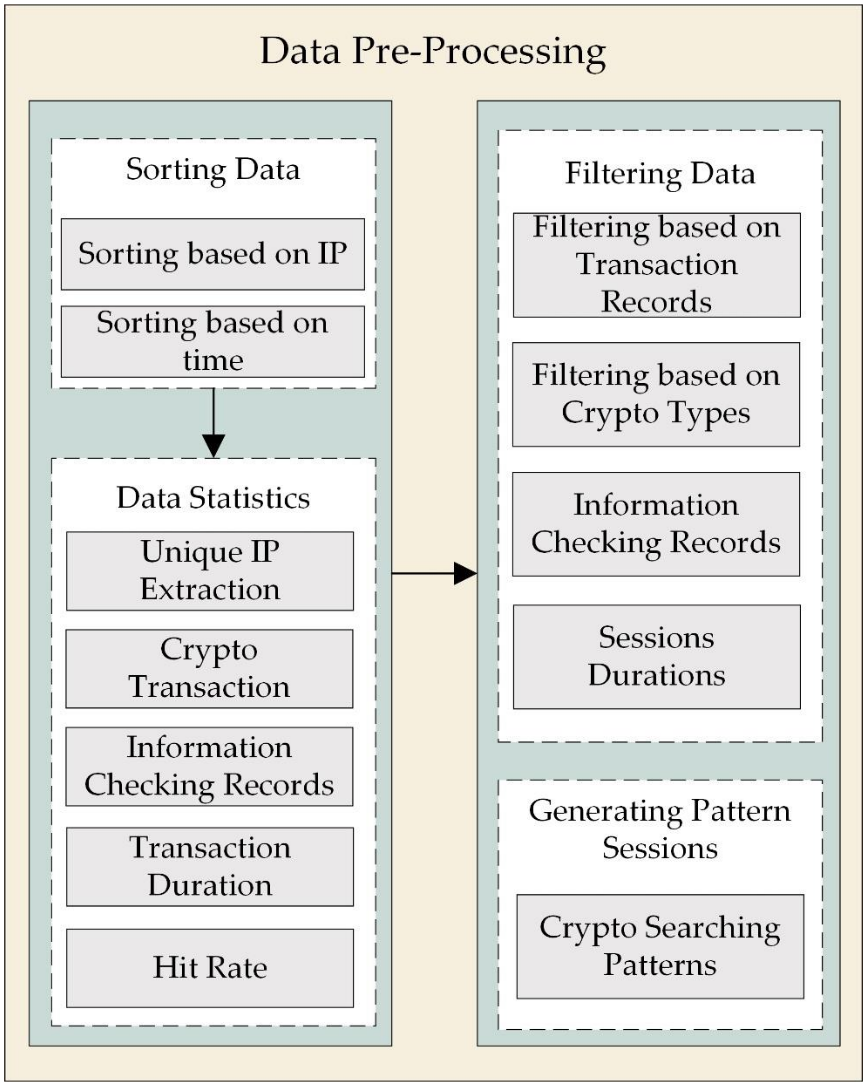


Figure 15. Data pre-processing detail architecture

3.3 General Cryptocurrency Prediction Architecture based on System Security

This section presents the general architecture of cryptocurrency price prediction and the knowledge discovery process. Figure 14: provides an overview of the layer-based system with details of each layer. The presented model contains four layers (1) Service Layer, (2) Predictive Control Layer, (3) Blockchain Layer, and (4) Physical Layer.

The Service layer represents the provided services of the presented approach. Two primary services are considered: type of services, user management, and crypto management. The provided services are exchange cost optimization, exchange management, and transaction management. This information's used to define the transaction records through the blockchain framework. The user management section contains the user profile, location, identity, user exchange, and request. This information is collected from users using the cryptocurrency framework to secure system security and transparency during network transactions. The second layer is the predictive optimal control layer, which contains the prediction process and optimal system control. This process includes cryptocurrency data collection, modeling and analysis, aggregation, prediction and optimization, price mapping, and crypto control mechanism. When the tasks are generated in the analysis layer, the suitable data tasks are used to run the prediction algorithm and find the target features. For example, the price prediction can be named for this section in cryptocurrency. The control algorithm is trying to find the duration of the operation. The third layer is the Blockchain, which contains data integrity and access control sections. The consensus manager, distributed architecture, auditability, traceability, and transparency are the considered factors in the data integrity section. Authentication, authorization, and identity management are the selected factors in the access control layer. The main reason for using these three factors is to control the users connected to the cryptocurrency framework and the network transaction details to improve the system's safety. The last layer is the physical layer which contains the cryptocurrency sources and points of users buying and selling cryptos. This thesis focuses on the cryptocurrency prediction regarding the highest network security and benefit based on user preferences.

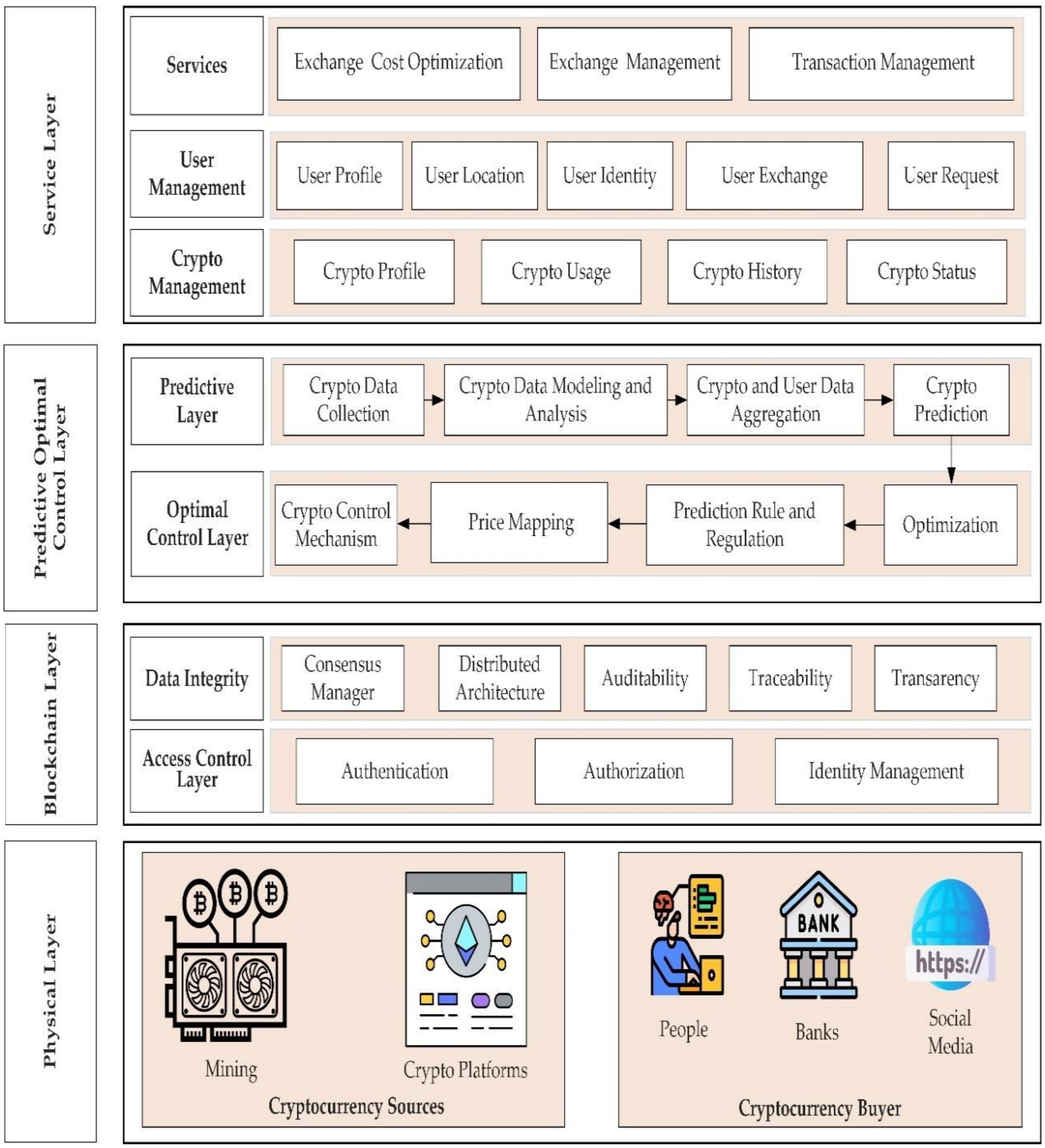


Figure 16. General Architecture of knowledge discovery and cryptocurrency prediction process

3.3.1 Service Layer

Figure 17: The cryptocurrency service layer architecture includes exchange cost optimization, exchange management, and transaction management services. These features are processing the secure transaction into the network regarding the user data management. The information related to user profile, location, identity, exchange, and user request is collected from the social media and their crypto account information. The crypto usage, history, and status details are analyzed to improve system performance and security. Every user considers different factors for crypto investment and plans to get a high benefit based on the money spent on buying cryptos. The user profile has the complete information for the crypto price prediction due to having rich user data.

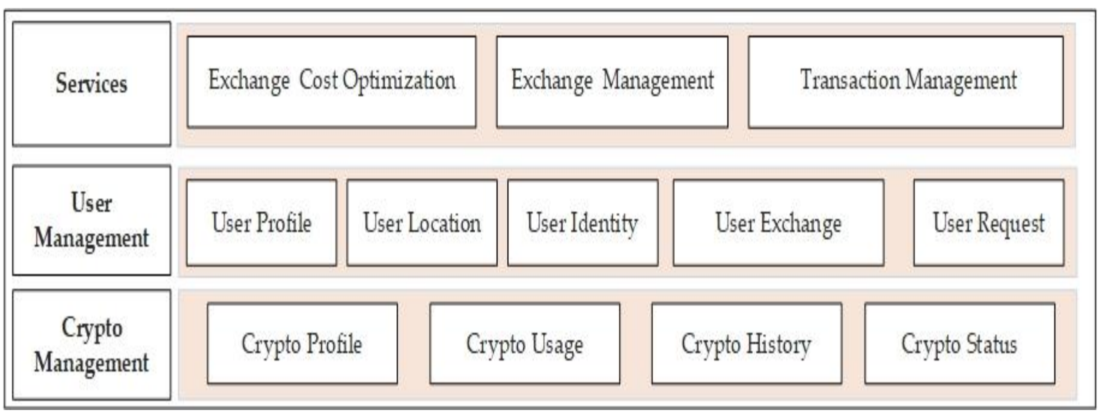


Figure 17. Cryptocurrency service layer general architecture

3.3.2 Predictive Optimal Control Layer

Figure 18: The cryptocurrency predictive optimal control layer contains two main sections: the predictive and optimal control layers. This layer was designed and processed to collect the user and crypto information for suitable data analysis in the proposed system. We mentioned the data-processing analysis and topic modeling using the LDA approach for the highly related

crypto price prediction. The optimal control processing is responsible for system optimization, the crypto price mapping, and controlling the proposed mechanism.

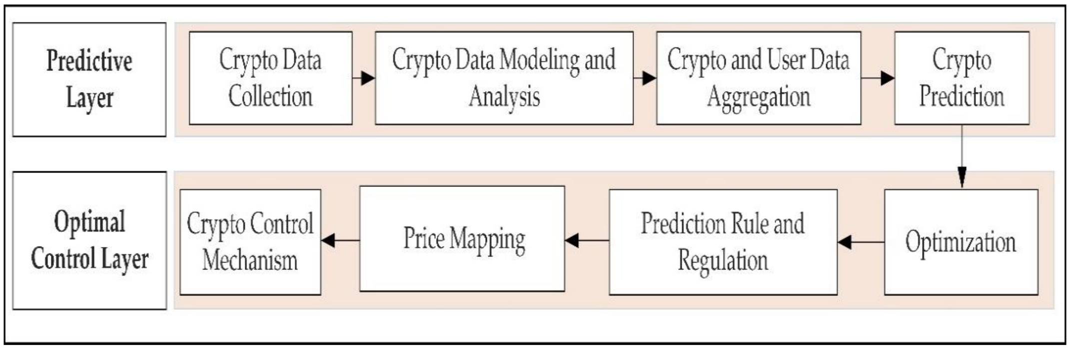


Figure 18. Cryptocurrency predictive optimal control layer

3.3.3 Blockchain Layer

Figure 19: The cryptocurrency blockchain layer on which the basis of this design is system security. The blockchain layer contains the data integrity and access control sections to validate the network transactions, user identification, and system traceability and transparency. To avoid hacking the transactions, applying the blockchain technique to secure the information shared in the network is required.

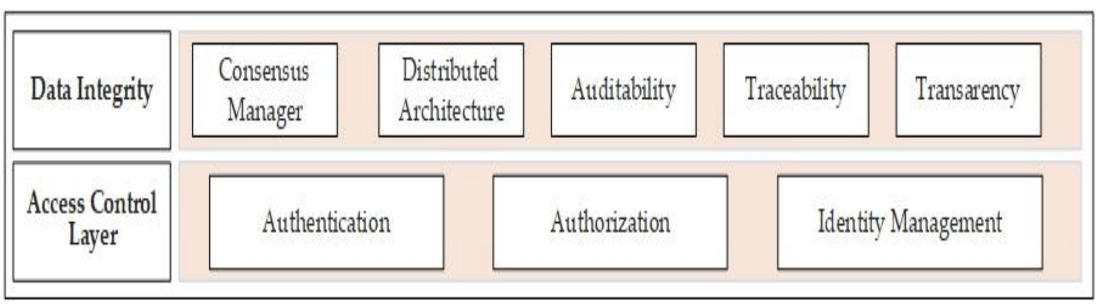


Figure 19. Cryptocurrency blockchain layer

3.3.4 Physical Layer

Figure 20: The cryptocurrency physical layer contains two critical sections: cryptocurrency sources and buyers due to various types of cryptocurrencies, which is essential to find the best and most accurate information. Crypto mining generates new cryptocurrencies and verifies recent transactions, which is a very costly process, and not everyone can do it. Regularly, cryptocurrency buyers are categorized into three main groups: people, banks, and social media. People who are saving money buy the cryptos and sell them with the highest benefit. The exact process is for the other two categories.

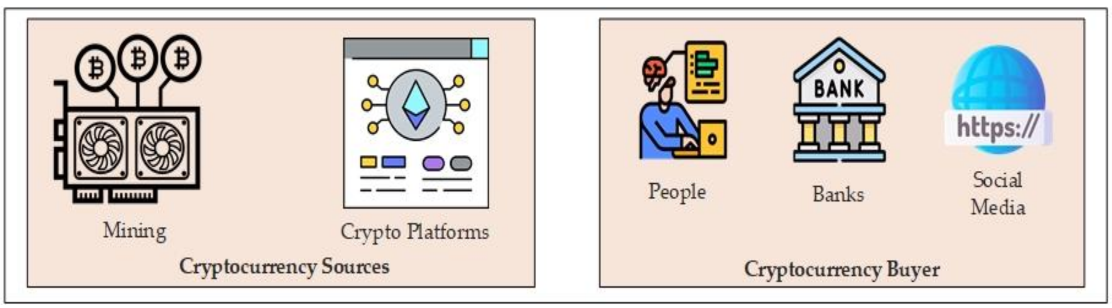


Figure 20. Cryptocurrency physical layer

3.4 Cryptocurrency Prediction Architecture based on Blockchain

Blockchain is a secure platform to store sensitive data to avoid attack, harm, data format changes, etc. In this thesis, we applied the blockchain framework based on the following reasons: (1) to secure the network transactions, (2) to store the user data in the safety database, (3) to easily trace the network transactions, and (4) to track the user's behavior and records. Figure 21: Shows the detail of the blockchain network in the proposed system. The blockchain framework in this system contains three main layers: The service layer, the Data integrity layer, and the Access control layer. The responsibility of each layer starts from the service layer, which contains the user management components and crypto transaction management.

The followings are the key points that demonstrate the importance of blockchain in the proposed cryptocurrency price prediction framework. The presented system progresses regarding the enabled data in smart contracts and modules of predictive analysis by utilizing the permissioned blockchain framework, Hyperledger Fabric, which gives permission only to the authenticated participants interested in linking with each other. This process causes transaction efficiency in terms of throughput and minimizes the latency of transactions in the whole blockchain network. The main goal of the enabled data analysis module based on smart contract is to explore the stored data in blockchain-related to smart contract and extract the hidden knowledge and underlying pattern that is significant for the management of cryptocurrency to use the social media-related contents effectively. The aim of the predicted analysis module based on smart contract is to use the extracted knowledge to build the prediction model based on a machine learning algorithm for cryptocurrency price prediction. The prediction model allows cryptocurrency users to plan the business in the future and various strategies such as cryptocurrency future demand, top selling, top buying cryptocurrency, etc. The presented approach is secure due to avoiding the access of unauthorized users to the transaction data and history and any other information related to the digital transactions. The secure cryptocurrency management system is defined to be lightweight due to the communication of frontend applications and RESTful API for the blockchain development system. The RESTful API server avoids the system offload computation. Moreover, Hyperledger Clipper is one of the standard tools of blockchain used to confirm the cryptocurrency price prediction framework. Furthermore, measurement of performance, e.g., MAE, RMSE, and MAPE scores used for estimating the predictive analysis model performance.

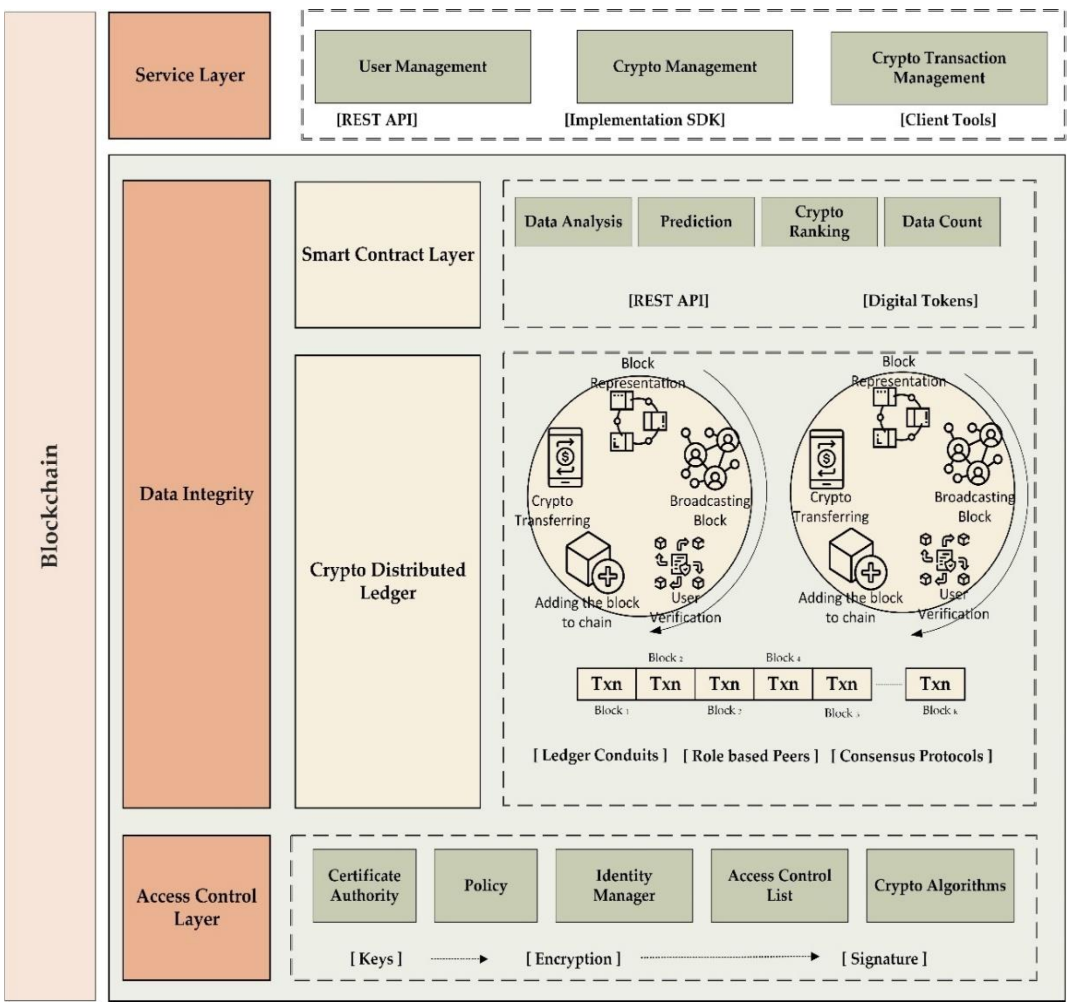


Figure 21. Cryptocurrency prediction based on blockchain

The user management is supposed to track all the information regarding the crypto users and store them in Blockchain to avoid any changes to improve the system security. The crypto management is responsible for the processing phase of suitable crypto prediction to users, and crypto transaction management is supposed to track the transaction records. Generally, this process follows the REST API, Implementation SDK, and client tool.

The Data integrity layer is divided into smart contracts and crypto distributed ledger. The smart contract is an essential part of any blockchain framework due to its rule-based behavior. The smart contract contains sentiment analysis, predicted sentiments, crypto ranking, and sentiment count. The importance of this approach is to extract the essential views for the crypto prediction, which is a susceptible process. System safety is in the first rank during this process due to working in cryptos and network transactions. The sentiment analysis, count, and prediction process the collected data to keep the necessary parts for further processing.

The distributed ledger is responsible for block representation, broadcasting block, user verification, adding the league in the chain, and crypto transferring. This process summarizes into ledger conduits, role-based peers, and consensus protocols. These five factors are thoroughly analyzing the crypto network in detail.

The access control layer is the last step of showing the information to the user in terms of having the certificate authority, policy, identity manager, access control list, and crypto algorithms. In the last step to have the secure prediction of cryptos and being sued about the presenting result to the user, the user identification is a significant step in case the user is joint to the network and similarly is the authorized user and has the verification. Getting the output of the blockchain network is required first having the keys, subsequent encryption, and last, by getting a signature from the network user will be able to access the crypto details.

3.4.1 Difficulty of Mining, Profitability and Hashrate

Single coin mining and coin block required the mining difficulty. Mining difficulty means solving the problems in blockchain networks which are complex. The mining difficulty based on the period decrease or increase depends on the number of network miners. The mining difficulty is automatically connected to the amount of power in the hashrate or network to stabilize the block mining. As the hash rate increases, the problem will be more significant and opposite. Mining a block needs more hash by verifying transactions with the necessary hardware. The trade-off between power consumption and hash rate exists. Mining rate and difficulty have an exchange in between. Using power-consuming resources and time to generate the most negligible valuable income is inefficient. Minors receive minimal rewards as the number of children grows.

3.4.2 Confirmation Time and Market Capitalization

Confirming transactions is required by logging into a block table and calculating the average time. Activated users and their location about the table block update need around ten minutes to complete the transaction logging method. Each currency amount is derived based on USD values. The market capitalization shows the value of cryptocurrency. This process means the evaluation of stock market capitalization based on the price, time, shares, etc., factors. The details of every cryptocurrency and hourly record are mentioned at https://coinmarketcap.com/. Regarding the need of users and crypto selections, this website shows the changes between various cryptos and the differences from one week before in terms of numbers and figures. The amount of market cap and changes in volume are 24 hours, and circulating supply is discussed in detail.

3.5 Cryptocurrency Predictive Analysis

The cryptocurrency prediction service based on predictive optimization comprises the following components, i.e., Centralized crypto data collection, decentralized crypto data collection, and hybrid crypto data collection. All three components are stored in the blockchain network. The input data type process is crypto analysis and prediction, which gets the input data from the stored dataset in the blockchain network. After finalizing the prediction, the details of predicted crypto, crypto price, and indicated data are further processed for the optimization and the optimal crypto data stored in the blockchain network. The optimal crypto type and price are also stored in the details in the blockchain network. The further process of this network is the crypto prediction to users based on the mentioned process. This architecture process works in different types of crypto datasets. Figure 22: Shows the detailed process of cryptocurrency optimization.

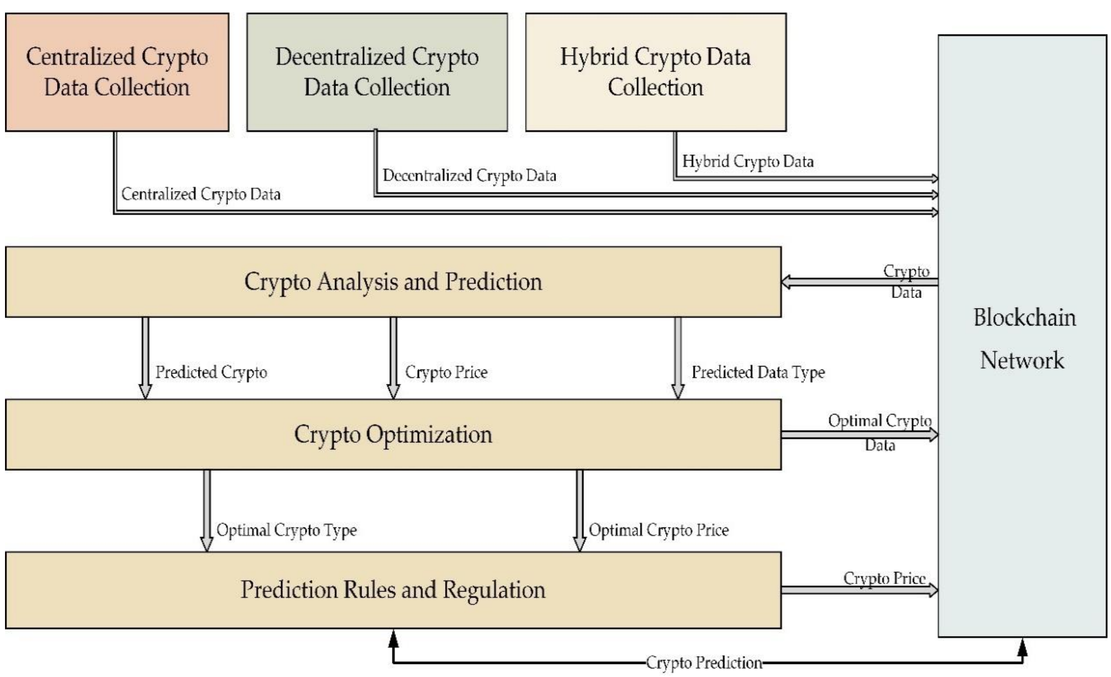


Figure 22. Detailed cryptocurrency prediction scenario

In this process, the applied algorithm for prediction and optimization is reinforcement learning which directly optimizes the user interest based on the FeedRec optimization algorithm using the simulation process. Engaging your customers virtually needs "ground truth." Based on these parameters, the algorithm predicts the optimal policy and optimal delay in user concentration. A simulation is defined by S (z_t, i_t; β_z); the prediction dataset is applied by employing a mini-batch of SGD. This dataset is immediately used with the manufacturing simulator, which utilizes prediction policy pb. Prediction policy pb is considered based on how to minimize the significant weight loss as demonstrated in Equations down:

$$\gamma(\boldsymbol{\beta_z}) = \sum_{t=0}^{T-1} \frac{1}{m} \sum_{j=1}^{m} (\boldsymbol{v_{0:t}}, \boldsymbol{D})\, \boldsymbol{\delta_t}(\boldsymbol{\beta_z}) \tag{6}$$

$$\boldsymbol{\delta_t}(\boldsymbol{\beta_z}) = \boldsymbol{\lambda_g} * \boldsymbol{\psi}\left(\boldsymbol{g_t}, \boldsymbol{g^{\wedge}}_{\boldsymbol{t}}\right) + \boldsymbol{\lambda_x} * \left(\boldsymbol{x_t} - \boldsymbol{x^{\wedge}}_{\boldsymbol{t}}\right)\boldsymbol{2} + \boldsymbol{\lambda_y} * \boldsymbol{\psi}\left(\boldsymbol{y_t}, \boldsymbol{y^{\wedge}}_{\boldsymbol{t}}\right) + \boldsymbol{\lambda_w} * (\boldsymbol{w^r}\, \boldsymbol{w^{\wedge r}})^{\boldsymbol{2}} \tag{7}$$

Generally, N represents the total number of directions in the forecast data. The significant ratio π between v0:t is used to reduce the disparity. π refers to the policies from the Q-network, such as E-greedy. Ψ presents the cross-entropy, that is, the representation of loss function. To avoid a large ratio, C is defined as the hyper-parameter. Using the loss function as the loss function, we define cross-entropy as ψ, in which hyperparameter C is a means of preventing large balances. λ is a hyper-parameter controller to evaluate the various tasks using the multi-task loss function to compare regression loss and multi-task loss $\delta_t(\beta_z)$.

As reported by β_zThe extracted π from Adapting to the Q-Network requires continuous effort. To observe the formal policies adaptive, as well as π, the S network likewise keeps changes to provide optimal accuracy. Research aimed at improving user satisfaction and diversity was mentioned as an effective process for the prediction. Previous research has identified diversity as a valuable tool for improving user satisfaction. The system optimizes the engagement of users unintentionally as well. As a result of the FeedRec framework discussed above, it is possible to optimize user engagement through various means in diversity immediately; an engagement list and a prediction list are both types of lists.

1. Linear Style: The results with more increased entropy demonstrate the most satisfying linear relationship. This can often result in the user getting more details and making better use of the design. The Equation below calculates the probability of a user looking for articles on the system.

   $$P\ (\mathrm{Use} \mid \boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_n}) = x\alpha\ (\boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_n}) + y,\ x > 0 \tag{8}$$

   The article prediction system described as $(\phi_1, \ldots, \phi_n)$, and the meaning of entropy is defined as xα $(\phi_1, \ldots, \phi_n)$. x and y are used in the range of 0, 1.

2. Quadratic Style: The high user satisfaction made by moderate entropy. The probability of a user using a system and searching for articles is defined as the below Equation.

   $$P\ (\mathrm{Use} \mid \boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_n}) = \exp\left(-\frac{(\boldsymbol{\alpha}(\boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_n} - \boldsymbol{\mu})\mathbf{2})}{\boldsymbol{\theta}}\right)) \tag{9}$$

Evaluations of the user-system agent relationship are shown above. This method indicates that FeedRec can adjust entropy differences among predicted lists and user attention based on various types of dispensation.

3.5.1 Reinforcement Learning Prediction

Reinforcement Learning (RL) may be a valuable tool for evaluating the reward selected from the action. Learning-based systems are designed using a model-based and model-free approach. Figure 23 shows the detailed design of the Reinforcement learning framework.

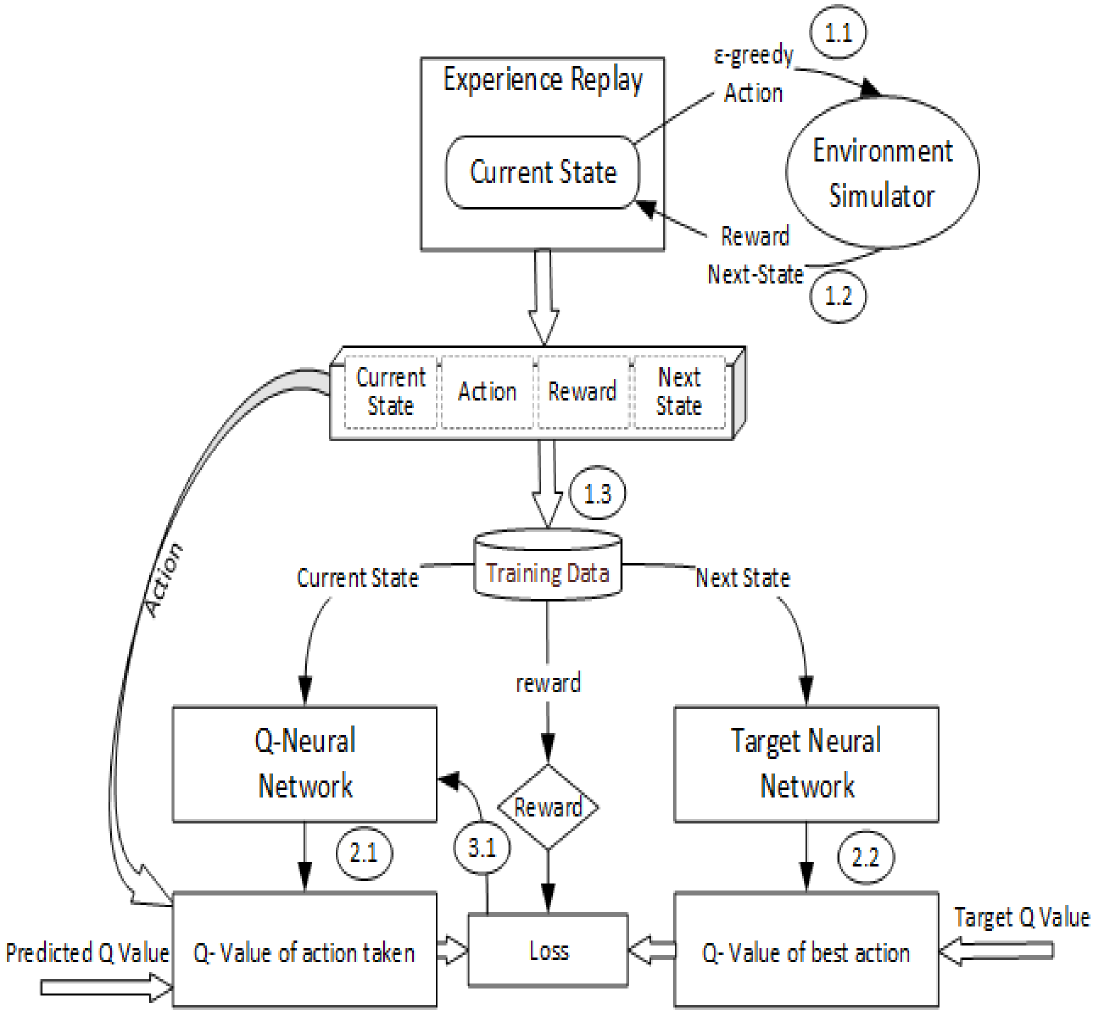


Figure 23. Reinforcement learning framework

In addition to improving the designed system structure, training the learning-based model can improve system performance accuracy. A primary objective of RL is to determine what activities to take in different states to raise the rewards in the future. Occasionally, the RL algorithm takes effort based on premium and specifies a policy for valuable actions based on the tip. A price prediction is a process of predicting prices, Figure 24: A summary of the RL algorithm's prediction steps and the collected data is presented. A series of steps are involved in pre-processing the raw data, engineering the features, transforming the elements, and selecting the components. After splitting the dataset, training and testing the price information was done using the RL procedure.

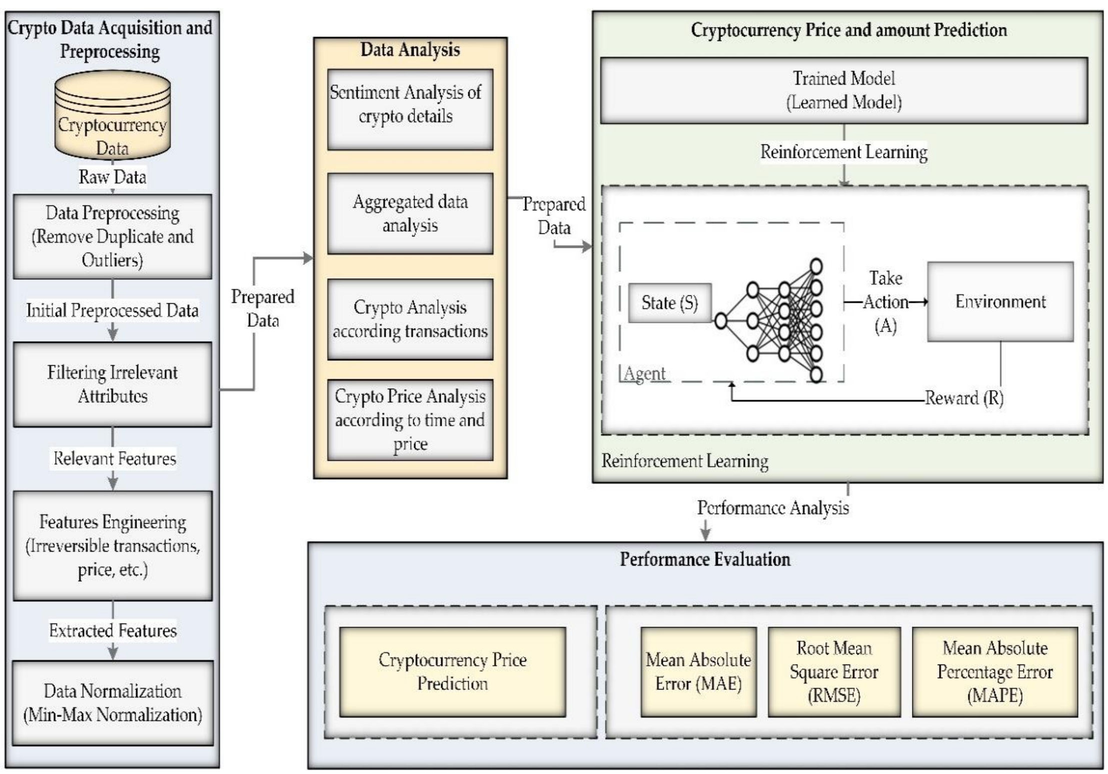


Figure 24. The Reinforcement learning process in the prediction system

3.5.2 Price Prediction

Three necessary authorities of data are used in cryptocurrency price prediction. Market statistics are the first. Another is blockchain network information such as transaction count, transaction fee, hash rate, etc. The last is usage statistics from Google Trends and Twitter. All three data sources were aggregated before being loaded into the data loader. The next step is to normalize the data and create the data stacks. For predicting the price for the next N days, reinforcement learning is used. Figure 25: Describes the process of predicting price in detail.

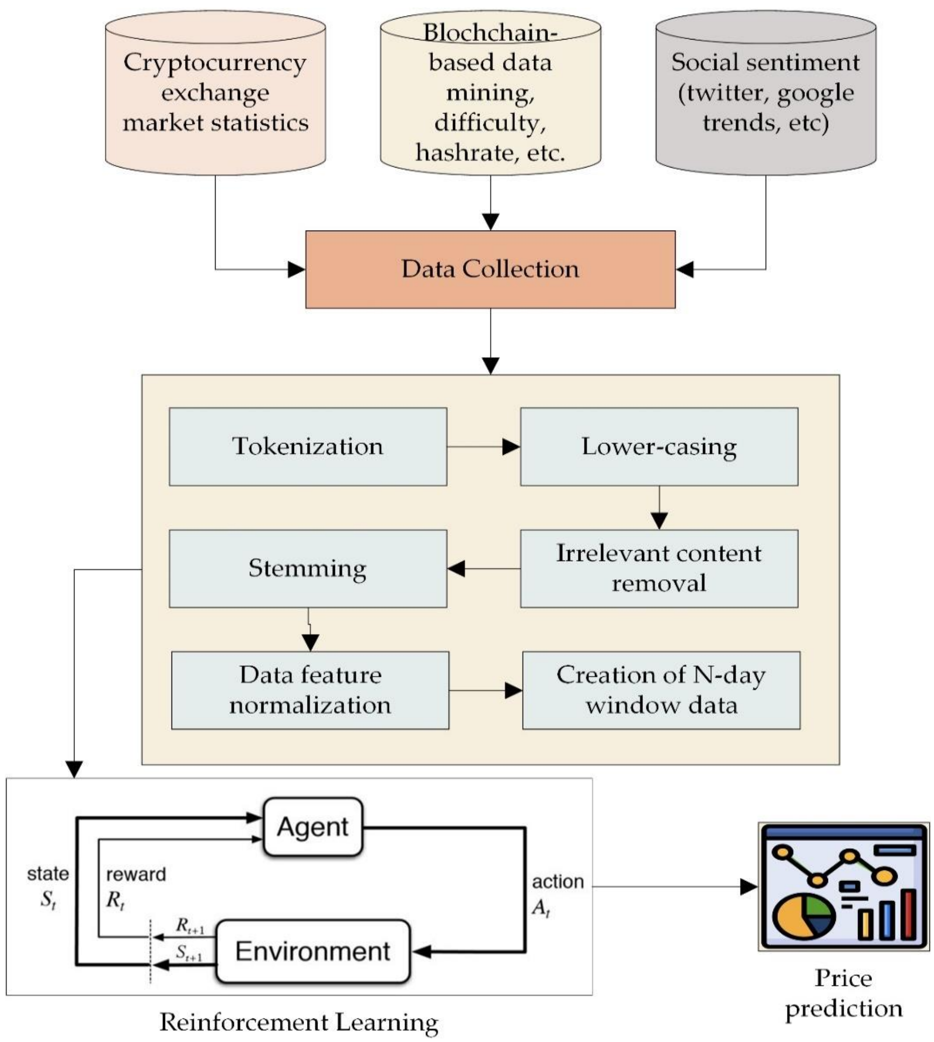


Figure 25. Price prediction process based on Reinforcement learning

3.5.3 Risk Management

The risk assessment implication is analyzed to process further the system regarding the generated risk performance portfolio results. The risk assessment implications are examined for the generated risk performance portfolio results to process the system further. Reference portfolios, risk-minimum portfolios, and determined portfolios have been added that compare the five cryptocurrencies. Portfolios are first compared with benchmark portfolios to determine which risk reduction assets are used. The cryptocurrency portfolio asset was protected from downside risk by four risk metrics. Each portfolio was evaluated on the following metrics: Regret (RE), Semi-Variance (SV), Expected Shortfall (ES), and Value-at-Risk (VaR). In reinforcement learning, the correct input improves the system's performance. Figure 26: Demonstrates the RL process for risk management. System risks are meant to be identified, evaluated, and prioritized within the suggested risk management system. By analyzing the risks and profits of the management issues, the portfolio's management problem can be expressed as an RL-based trading system with specifications.

To express portfolio management in the RL architecture, the design agent determines strategies for trading assets based on the current state of the capital markets. Asset trading information is linked to the environment. The agent provides a trading strategy. Based on the evaluation of this trading strategy, the agent is equipped with the following state and is rewarded.

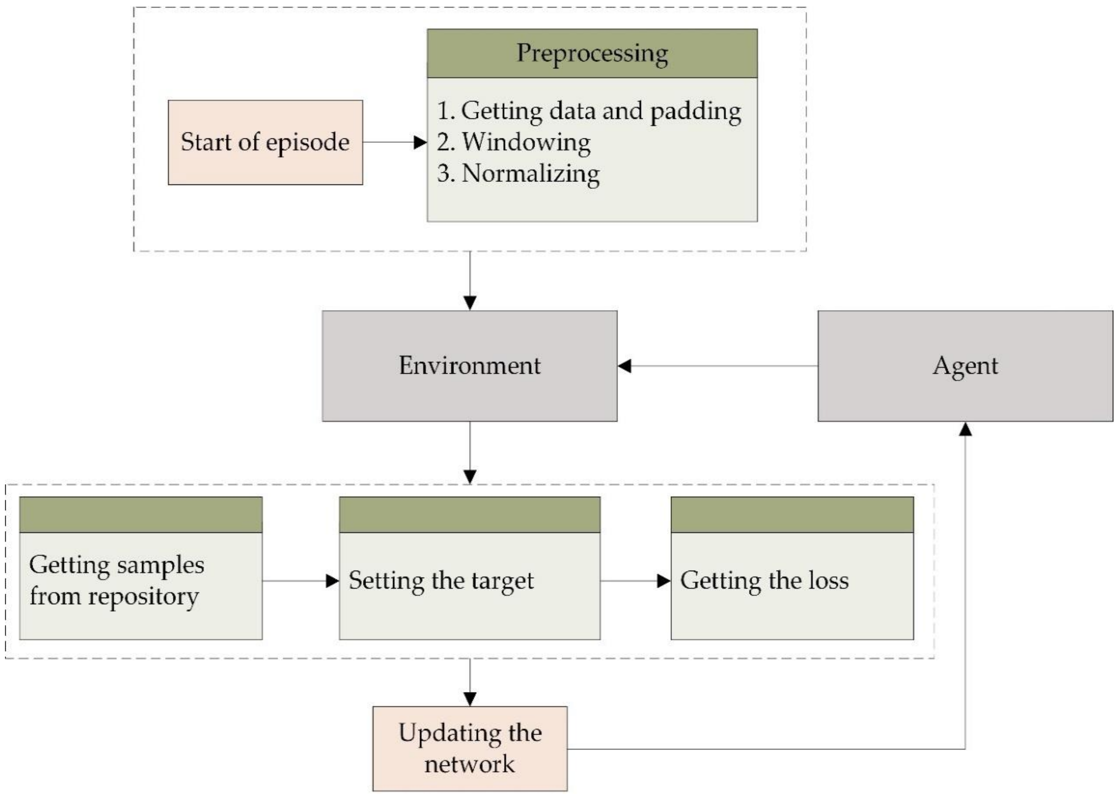


Figure 26. Reinforcement learning-based risk management architecture

Chapter 4: Development of Cryptocurrency Prediction

This chapter presents the design and implementation of cryptocurrency prediction based on predictive optimization. The structure contains the flowchart of prediction service architecture based on predictive optimization, a sequence diagram of prediction service architecture, system modeling, and objective function.

4.1 Design of Cryptocurrency Prediction Framework

The prediction data is gathered from social media in the presented predictive optimized model. The dataset includes crypto details, sentiments, price, date, time, comments, likes, dislikes, shares, and clicks processed with the various data processing techniques. The user data contain individual information such as user IP, user click information, and shared information, to name a few. The processed data is applied for the feature selection to reduce the number of variables used for input and develop the predictive model. In the presented system, we evaluated some features such as cryptocurrency price, cryptocurrency amount, and cryptocurrency risk management. The selected features are used for the prediction based on Reinforcement learning and optimizing the predicted value. Figure 27: shows the detailed process of the predictive optimization process of the proposed system. The input data is from internet sources, Facebook, Twitter, blogs, and crypto coin details websites. The details of data acquisition focus on the sentiments, price, date, time, etc. The user data collection and crypto data analysis go to the data aggregation process and feature selection. In terms of feature selection, crypto ranking, crypto prices, sentiment count, etc., considers, and by applying the machine learning and deterministic models, the prediction is completed.

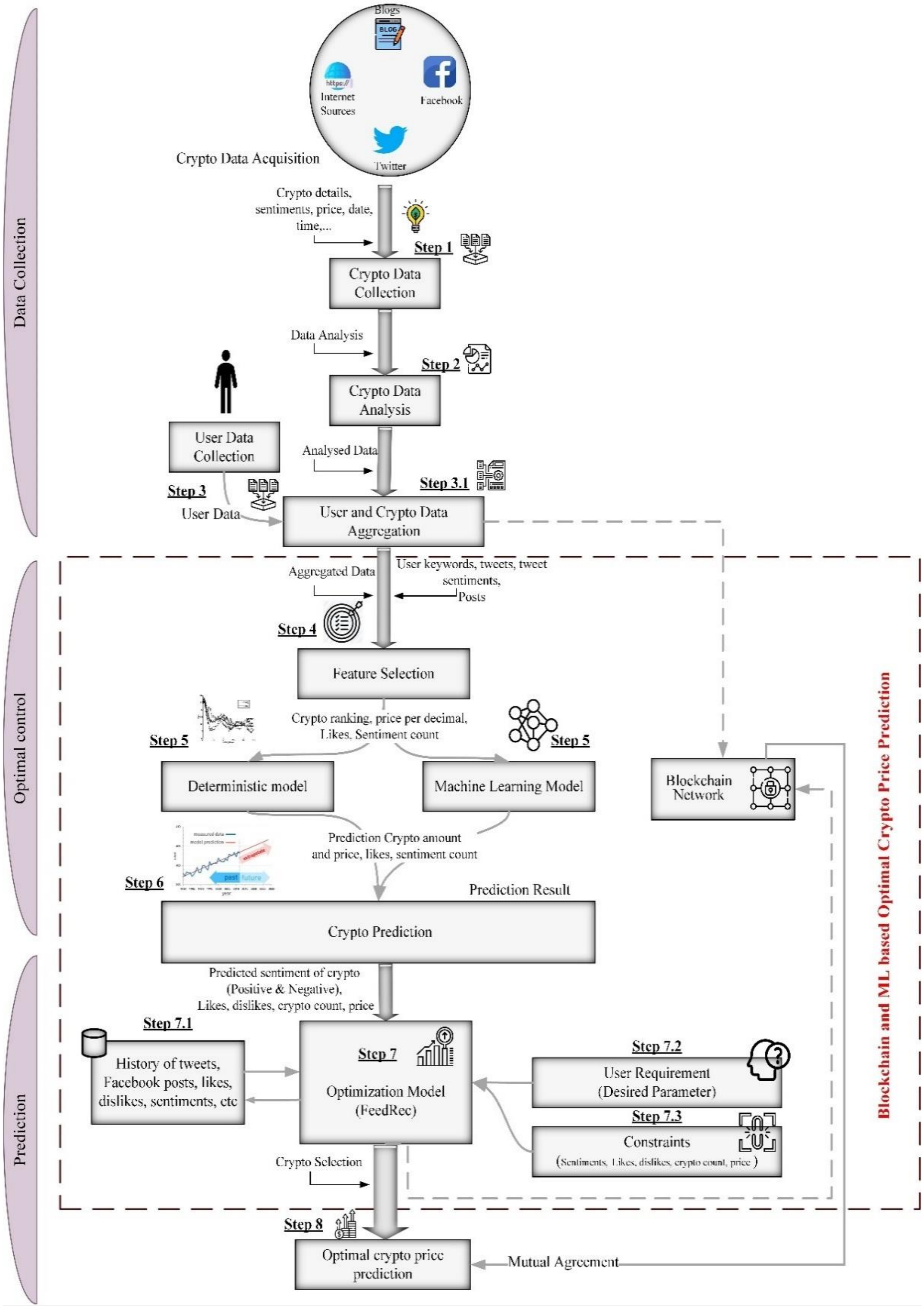


Figure 27. Predictive optimization flowchart in prediction system

The prediction layer gives the highest possible results regarding the user preferences and predicted sentiments. Moving forward to optimization, the applied optimization algorithm is FeedRec. The requirements are desired parameters from the user and the constraints information such as price, sentiments, etc. The history of tweets, Facebook posts, likes, dislikes, etc., is also required. After optimization, the final result is the optimal crypto prediction to users with the highest possibility regarding user interest and preferences. The optimal value defines the rules of payments and the prediction system price mechanism. Finally, all the transactions are recorded in the blockchain system.

4.1.1 Cryptocurrency knowledge Discovery Framework

Knowledge discovery is one of the critical frameworks for the proposed system, which is defined in Figure 28. The primary meaning of knowledge discovery is to extract vital substances and information from the provided contents. This process shows the main topic of the content that focuses on. In this approach, we defined two phases of prediction and exchange rate. The system database, as explained before, is the contents collected from the social media websites and coin market cap. There are two training and testing phases which focus on applying feature extraction techniques and machine learning algorithms for prediction. The mentioned machine learning algorithm in this system is Reinforcement learning. The data splitting process divided the data into 80% training and 20% testing sets. The feature extraction components are divided into memory, CPU, file, network, process, and registry. The classified URLs out of the testing set save into the database. Moving forward to the exchange rate phase, a transaction request from the user goes through the p2p network and requires transaction validation. The verified transaction contains the cryptocurrency records and defined smart contract rules that are supposed to add a new block to the system in terms of successful transactions. Successful marketing entirely depends on applying encryption technique, not allowed to another commodity, network-based and decentralized network.

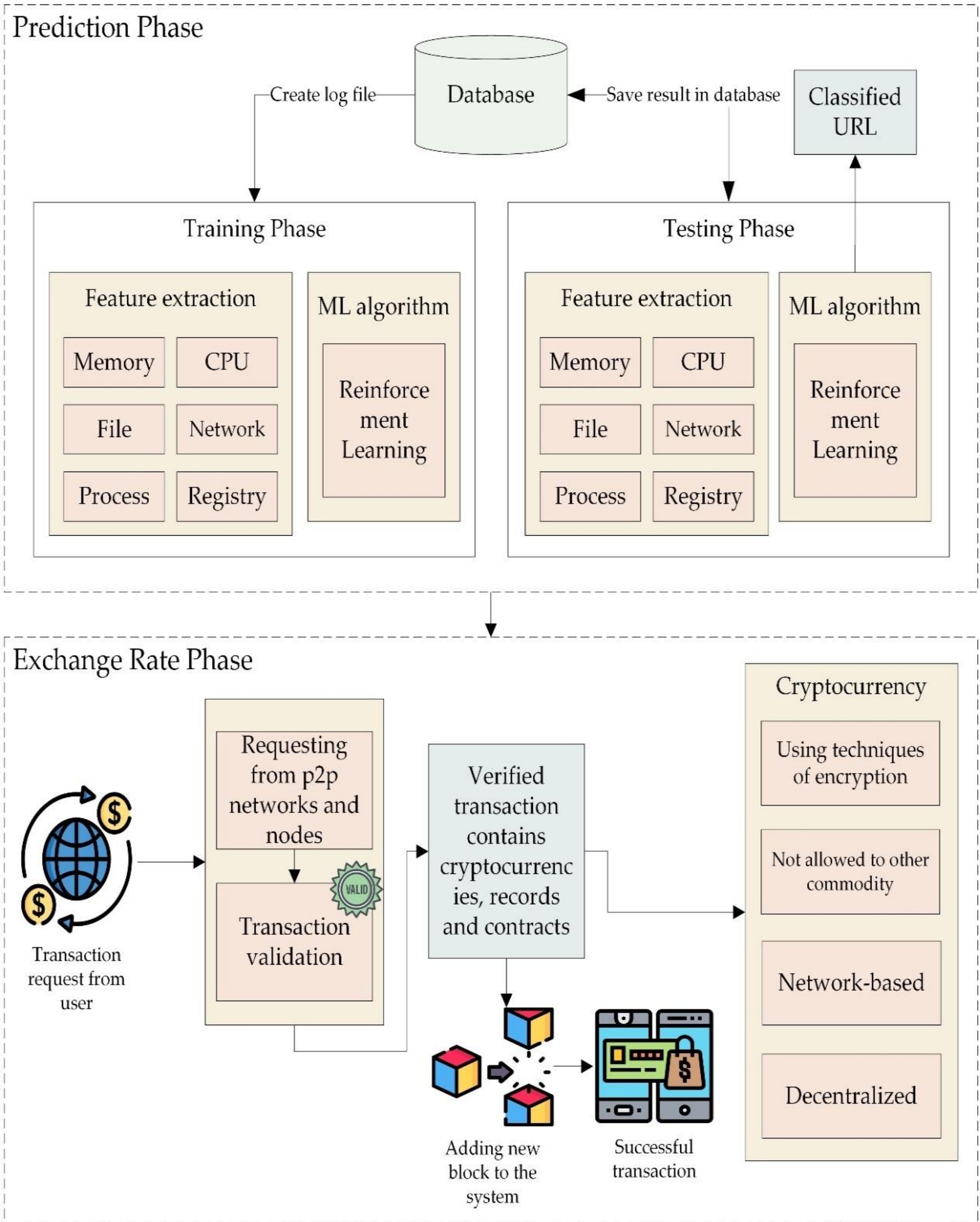


Figure 28. Knowledge discovery on blockchain framework

4.1.2 Cryptocurrency Blockchain Framework

The blockchain framework was designed and implemented in the following process. The cryptocurrency information contains the coins details sent to the user known as the seller of digital currencies. This process moves forward to managing the crypto information and crypto market until getting the request from the user who wants to buy the digital coins. The detailed procedure is shown in Figure 29: Sequence diagram of cryptocurrency prediction architecture regarding predictive optimization.

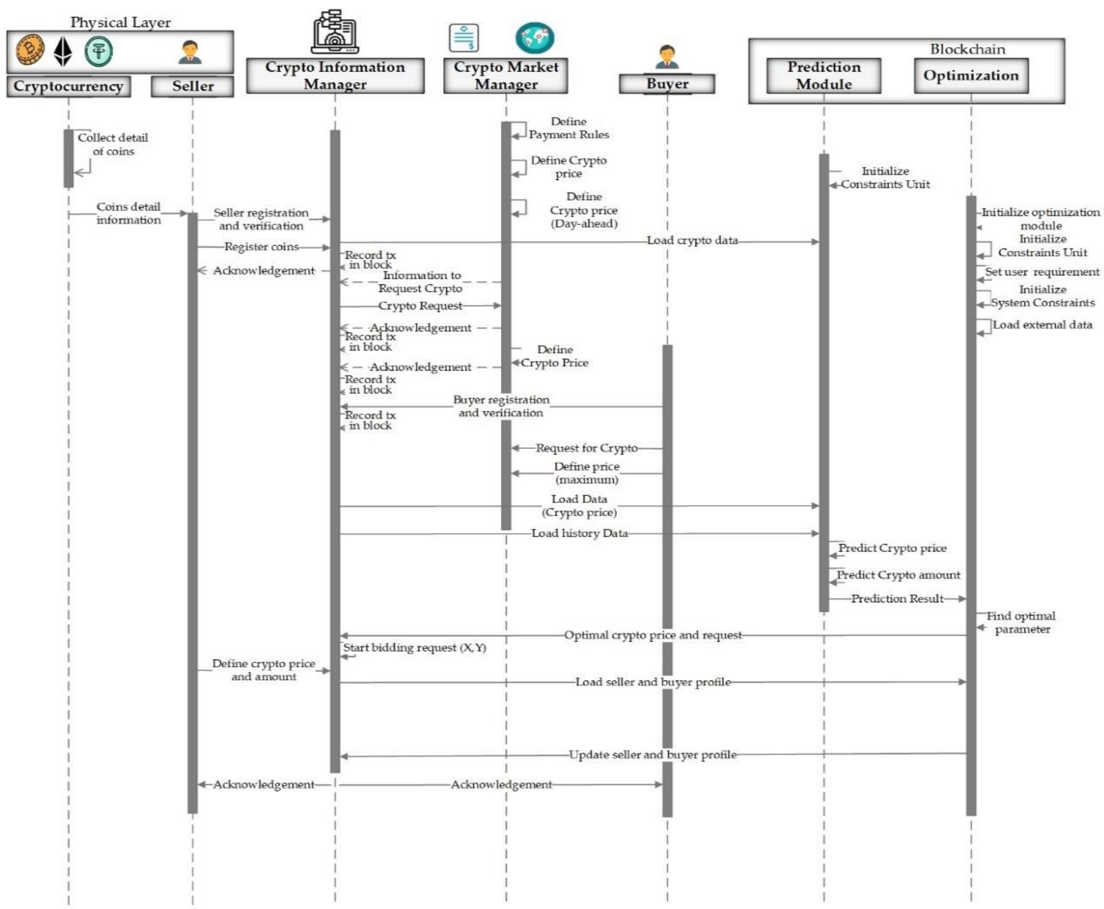


Figure 29. Sequence diagram of crypto prediction

4.2 Implementation of Cryptocurrency Predictive Analysis

In this section, we present the implementation of prediction service architecture. This section contains two subsections of development environment and execution results.

4.2.1 Development Environment of Cryptocurrency Predictive Analysis

In this section, we discuss the proposed architecture development environment. Figure 30: Shows the summary of the system implementation requirements. In total, there are five important sections to the stary implementation process. The first layer is data collection which, as explained before, the internet is available, and we mainly focus on a few websites. The first one, which contains the hourly information of cryptocurrencies, is https://coinmarketcap.com/all/views/all/. The others are from Twitter, Facebook, blogs, internet websites, user comments, shares, likes, dislikes, etc. The process of data crawling is by a data miner web scraping tool. Moving forward to the second layer, libraries. This contains the machine learning and optimization libraries. As mentioned in previous sections, machine learning and optimization libraries contain the applied Reinforcement learning algorithm and FeedRec optimization. The frameworks used in this process are Notify.js, bootstrap, jQuery, and Net framework. Moving to the programming language, we have defined two modules based on python and Java. The python module contains the prediction, control, and optimization. The Java module contains access control and security. The requirements for implementation are python, java, docker, and visual studio. The visualization section contains the desktop and web frameworks. The desktop framework works with the windows forms and SQL. The web framework works with HTML and SQL.

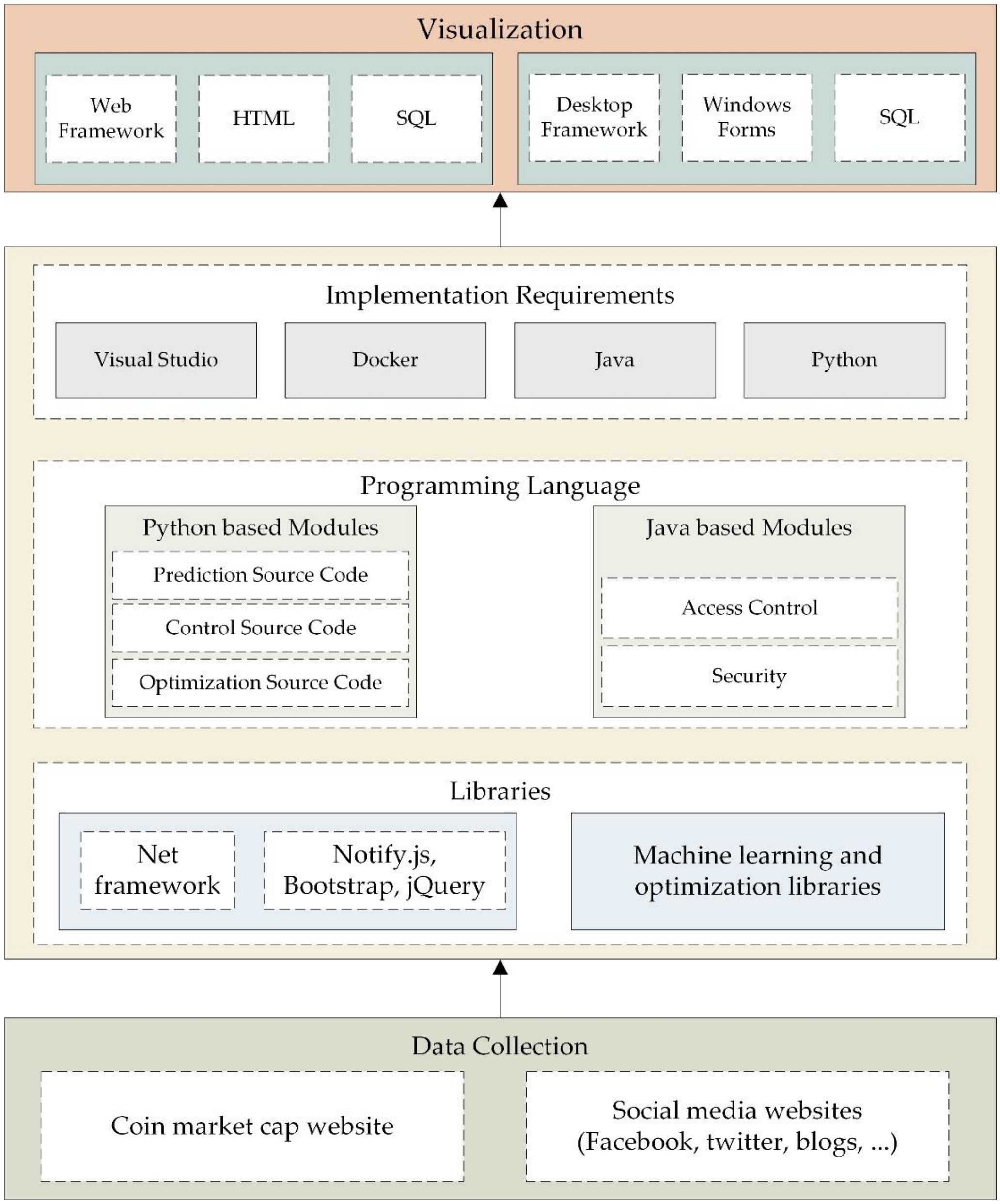


Figure 30. Development Environment of crypto prediction

4.2.2 Experimental Setup of Cryptocurrency Prediction

This section details the price prediction system's implementation process and development area. Table 8 summarizes the machine learning and blockchain framework components in this system. The operating system of the machine learning environment is windows 10. Browser is Internet Explorer, Firefox, and Chrome. Python and IDE are the programming languages for train and testing the developed models. The blockchain framework was designed based on the Ubuntu Linux 1804 LTS operating system and Node.js programming language. The CPU model is Intel(R) Core (TM) i7-8700 @3.20 GHz, and the docker engine version is 18.06.1-ce. Docker composer version is 1.13.0, and the IDE is Composer Playground. The memory usage in this system is 12GB.

Table 8. Implementation Environment

Techniques	Elements	Statement
Machine Learning	Operating System	Windows 10
	Browser	IE, Firefox, Chrome
	Programming Language	Python, IDE
Blockchain	Operating System	Ubuntu Linux 1804 LTS
	Programming Language	Node.js
	CPU	Intel (R) Core (TM) i7-8700 @3.20 GHz
	Docker Engine	V18.06.1 - ce
	Docker Composer	V1.13.0
	IDE	Composer Playground
	Memory	12GB

4.2.2.1 Natural Language Processing Techniques

We have applied the NLP techniques for detail processing the collected dataset in this process. Due to various data sources in the proposed approach, there are many contextual datasets regarding the users comments, tweets, posts, etc., which require content-based analysis. Table 9: shows the details and description of the NLP technique for the proposed approach.

Table 9: NLP Techniques applied for the data pre-processing

Process	Description
Expand Contractions	The short form of the words that are not suitable for analysis, such as don’t, it’s, and aren’t, requires changing to the complete version using expanded contractions.
Lower Case	Checking the words lower case and upper case are very different during the analysis process and treated in different directions for further analysis.
Remove Punctuations	Removing punctuations is another aspect of text analysis. In total, 32 punctuations require to take care of it. The string module helps to replace the punctuations with an empty string.
Remove words and digits	Combining digits and words in the text makes the process difficult in terms of text combination for analysis.

containing digits	
Remove Stop words	One of the common statutes of text analysis is removing stop words that don't have any positive point for the dataset. The stop words have no meaning, such as this, there, were, etc. The famous library for stop word removal is NLTK which contains 180 stop words list.
Rephrase Text	Text modification and pattern changes to the specific string for easier identification to match the email patterns and to change them to the string file like email address.
Stemming	Reducing the words to reach the main root of the word. Such as: running to run, doing to do, etc.
Remove White Spaces	The extra space into the context file while doing the pre-processing. This problem solves by removing the extra white spaces.

4.2.2.2 Training Set

In data training and testing, we consider 80% of the dataset for the training section. The extracted information details are in terms of:

1) Social media related links to cryptocurrency information
2) Comments which show the satisfaction or not satisfying information.
3) Users like and dislike, share, tweet, and retweet.

This scenario extracts words and evaluates their similarity and topic probability distribution by applying the LDA technique. All of the extracted helpful information is the input for the Reinforcement learning model. The output generates the topic distribution based on positive and negative information utilized to create the proper prediction. In this process, we reached over 95% accuracy in information classification. Figure 31: Shows the LDA topic categorization in the proposed approach.

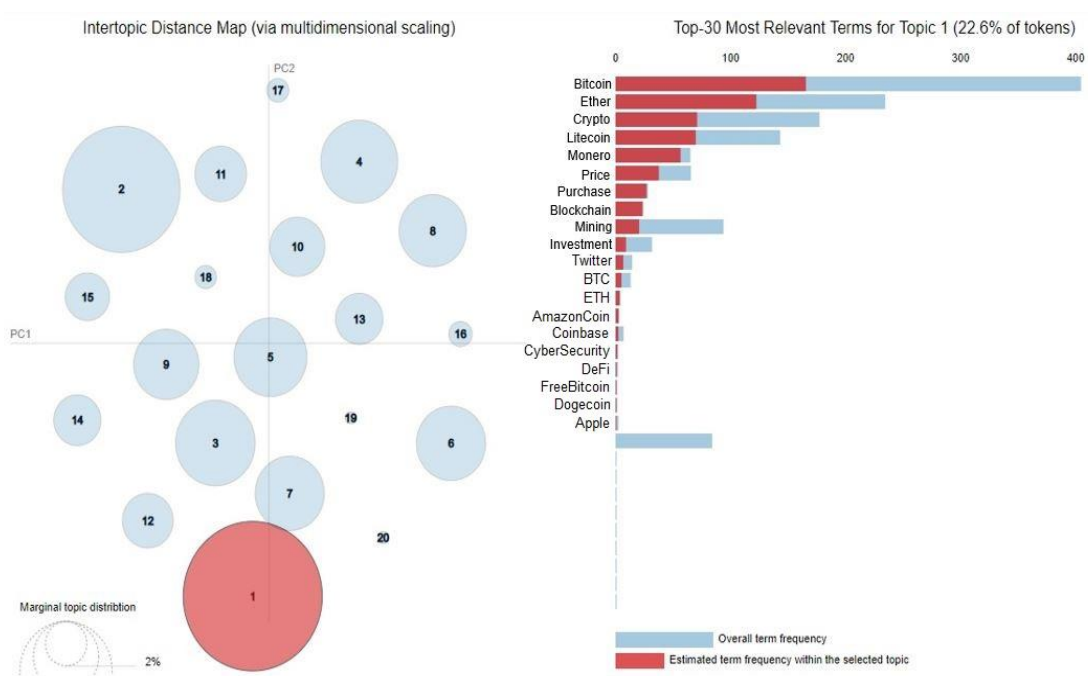


Figure 31. Implementation scenario of LDA topic categorization using pyLDAvis

As shown in Figure 31, each circle represents a topic in this process. Distance between circles proposes similarity between topics. The right-side flowchart shows the 30 top words in each category. The blue color represents the total number of words in the dataset. The red color shows the number of words that happen in each topic.

4.2.2.3 Testing Set

The proposed system's testing set follows the same process as the training set. In the case of testing, 20% of the dataset was considered for analysis. The processing is based on the data pre-processing and topic extraction to extract the probability of the topic. The testing set results are the input of the Reinforcement learning system for predicting the ground truth of collected information. The Reinforcement learning algorithm is a learning-based system that focuses on the user ratings regarding the results and based on the ratings. The system improves the performance of prediction. The topic probability extraction authorizes the prediction using the authentication module.

4.2.3 Experimental Results of Cryptocurrency Prediction

This section dedicates the discussed details regarding the experiments and the corresponding results. We divide this section into three topics: reactive optimal control layer results, cryptocurrency price prediction based on knowledge discovery, and Blockchain.

4.2.3.1 Experimental Results of Predictive Optimal Control Layer

In this process, we selected three cryptocurrencies named KuCoin, ChainLink, and Polkadot for further process. The main reason for this section is regarding the shared information of users and recent research records on these digital currencies, which have recently become very famous among researchers worldwide.

HRP is a graph-based theory and employs machine learning approaches in three main phases:

- Clustering
- recursive bisection
- quasi-diagonalization

After applying the Hierarchical Tree Clustering algorithm to the assets, the first step is to divide them into various clusters. A correlation distance matrix A is extracted from two asset correlation matrices x and y as follows:

$$\boldsymbol{A(x,y)}\sqrt{\boldsymbol{0.5*(1-p(x,y))}} \quad (10)$$

Step two displays the Euclidean distance processing applied to all pair-wise manner columns, which gives us the augmentation matrix distance Aˆ as the following Equation:

$$\boldsymbol{A^{\wedge}(x,y)}=\sqrt{\boldsymbol{\sum_{m=1}^{i}(A(m,x)-A(m,y))^2}} \quad (11)$$

Clusters are generated by recursive application of the Equation above. By defining the set of clusters as C and the first cluster as (x^*, y^*) evaluated as Equation below:

$$\boldsymbol{C[1]=argmin_{x,y}A^{\wedge}(x,y)} \quad (12)$$

Aˆ is the matrix distance for the evaluation process, and all the assets use the C [1] single clustering linkage. Therefore, for every asset x out of the cluster, the distance of the new cluster is evaluated as Equation below:

$$\boldsymbol{A^{\wedge}(x,C[1])=\min(A^{\wedge}(x,x^*),A^{\wedge}(x,j^*))} \quad (13)$$

Table 10 shows the result of the KuCoin, ChainLink, and Polkadot cryptocurrencies statistic prediction accuracy. Results from the baseline show that most of the time, the classifier predicts the same class for every currency. Table 10 shows the Mean, Median, Var, Min, and Max records.

Table 10. Cryptocurrency accuracy records statistic prediction

#		Mean	Median	Var	Min	Max
KuCoin	RL	0.7899	0.8022	<0.003	0.7874	0.8048
	CNN	0.7989	0.7989	<0.003	0.7958	0.7877
	MLP	0.7734	0.7734	<0.002	0.7648	0.7736
	LSTM	0.7498	0.7698	<0.003	0.7448	0.7736
	Baseline	0.7689				
ChainLink	RL	0.8985	0.8988	<0.003	0.8946	0.887
	CNN	0.8785	0.8785	<0.003	0.8727	0.8828
	MLP	0.8286	0.8288	<0.003	0.8269	0.8346
	LSTM	0.8289	0.8285	<0.003	0.8188	0.8359
	Baseline	0.8288				
Polkadot	RL	0.9886	0.9886	<0.003	0.9878	0.8248
	CNN	0.9789	0.9722	<0.003	0.9758	0.9855
	MLP	0.9685	0.9622	<0.003	0.9768	0.9669
	LSTM	0.9696	0.9698	<0.003	0.8725	0.9673
	Baseline	0.9688				

RL algorithm predictions of digital coin exchange rates for the next ten days are displayed in Table 11. Detailed information about the dates, actual coin rates, predicted rates, and error rates are shown in four columns. The Equation below evaluates the error rate:

$$\boldsymbol{Error\ Rate} = \frac{\boldsymbol{Actual - Predicted}}{\mathbf{Actual}} * \mathbf{100} \tag{14}$$

Table 11. Cryptocurrency accuracy records of statistic prediction

Date	Actual Rate	Predicted Rate	Error (%)
1 Mar 22	25853	25831.85	0.21
2 Mar 22	25885	25826.37	0.19
3 Mar 22	25863	25992.86	0.10
4 Mar 22	25863	25992.86	0.15
5 Mar 22	25863	28992.86	0.30
6 Mar 22	25883	25889.79	0.08
7 Mar 22	25871	25889.29	0.07
8 Mar 22	25886	25896.78	0.17
9 Mar 22	25983	25985.28	0.20
10 Mar 22	25865	25873.78	0.08

A breakdown of the pattern classes for cryptocurrencies is shown in Table 12. There are two types, zero and one, shown in Table 12. The value in class 1 is based on USD, and the rest is in class 0.

Table 12. The records of cryptocurrencies breakdown pattern

#		KuCoin	ChainLink	Polkadot
Train	0	76.88%	82.69%	96.66%
	1	67.54%	58.75%	47.78%
Test	0	76.89%	82.88%	96.88%
	1	67.55%	58.34%	47.53%

Figure 32 and Figure 33 demonstrate the actual and predicted value differences using RL for exchange rate prediction for 30 and 90 days of KuCoin.

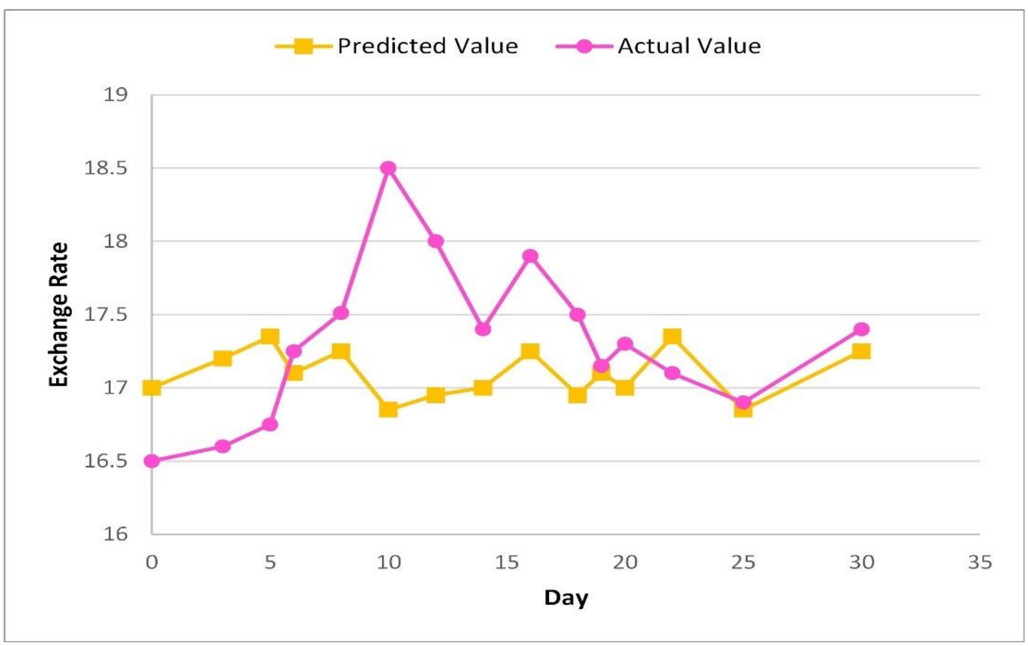


Figure 32. 30 days actual and predicted values of KuCoin prediction results

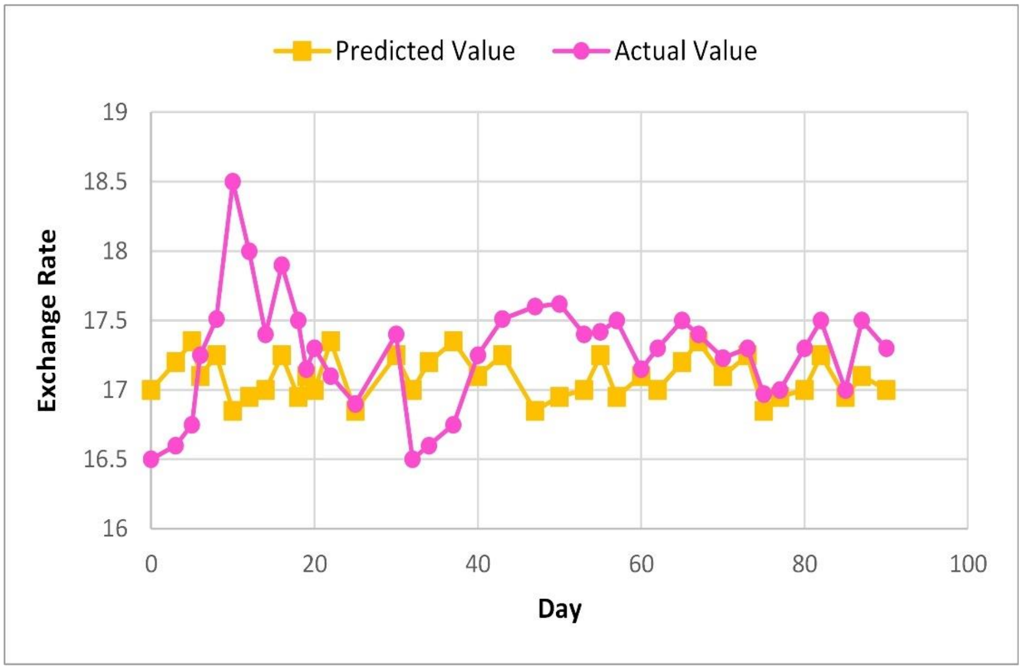


Figure 33. 90 days actual and predicted values prediction results of KuCoin

Figure 34 and Figure 35 show the actual and predicted value of exchange rate prediction for 30 and 90 days of ChainLink.

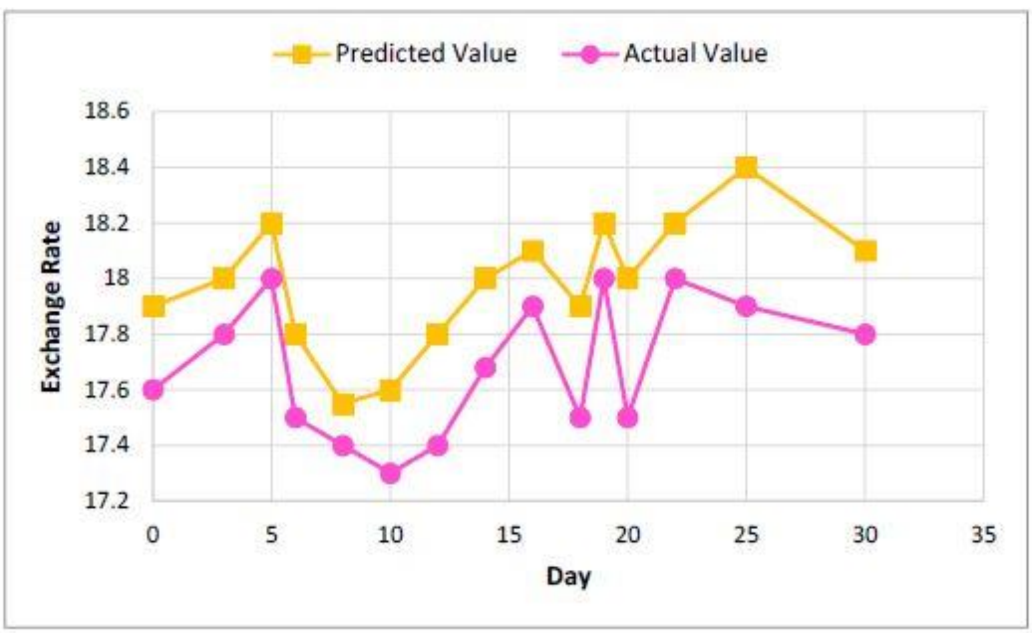


Figure 34. 30 days actual and predicted values prediction results of ChainLink

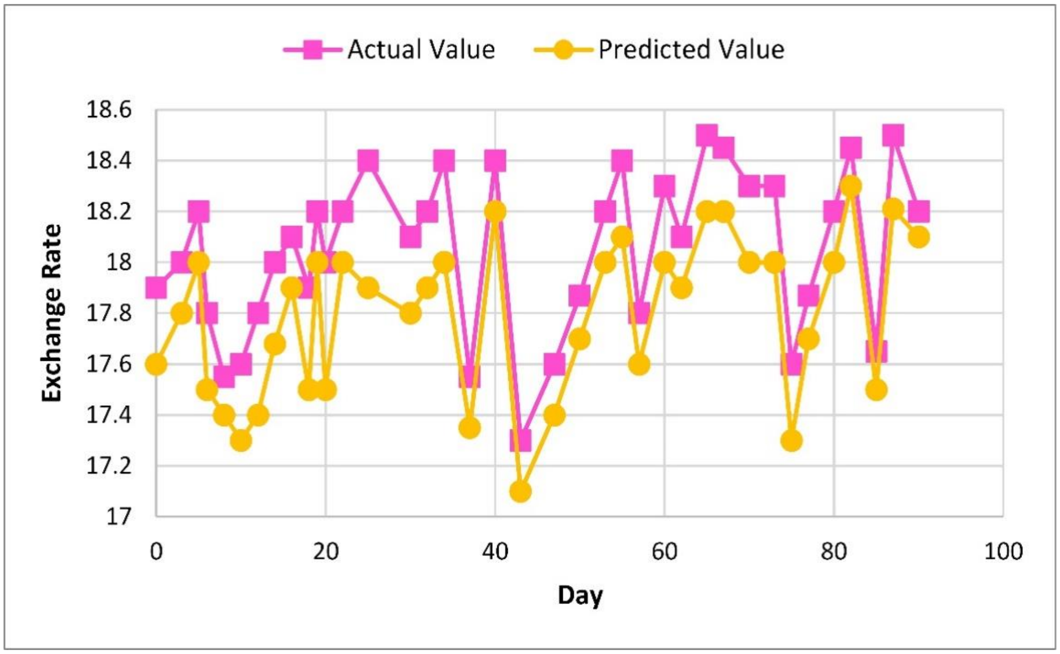


Figure 35. 90 days of actual and predicted value prediction results of ChainLink

Figure 36 and Figure 37 show the actual and predicted value of exchange rate prediction for 30 and 90 days of Polkadot.

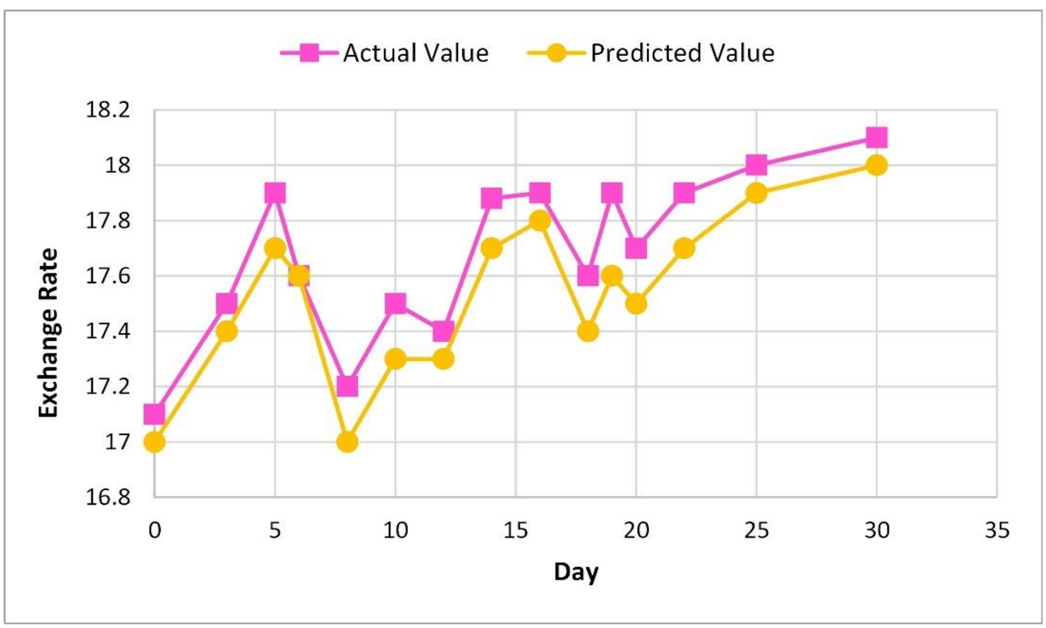


Figure 36. 30 days actual and predicted values prediction results of Polkadot

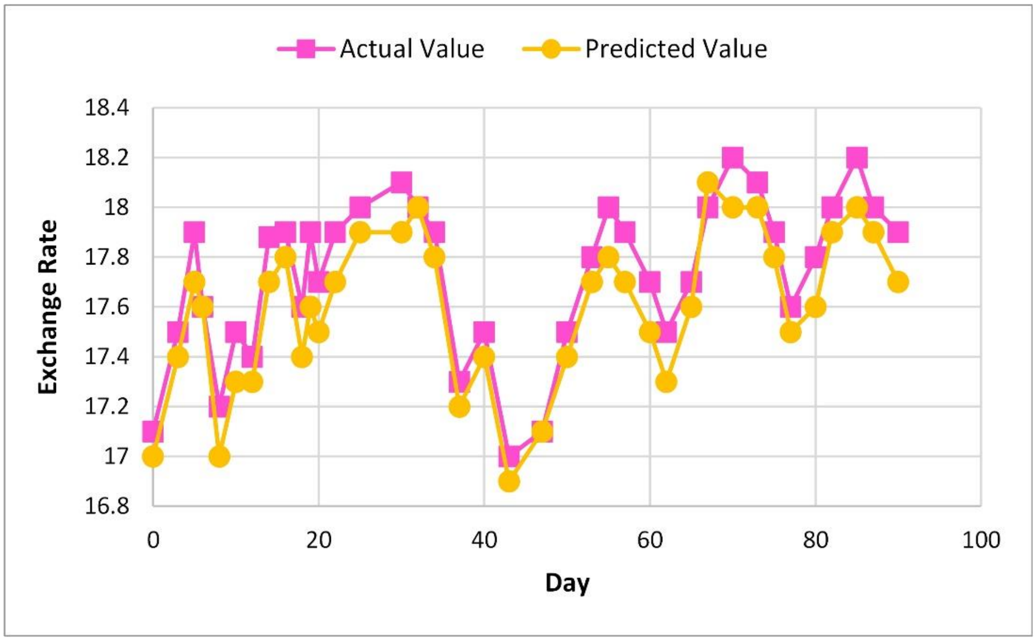


Figure 37. 90 days actual and predicted value prediction results of Polkadot

MAPE results for machine learning algorithms in 7 days, 30 days, and 90 days are shown in Figure 38. This figure compares RL performance with CNN, MLP, and LSTM. The MAPE calculation determines the overall average error of the real values as they change over time. Predictions that are better on some days are preferred over those not.

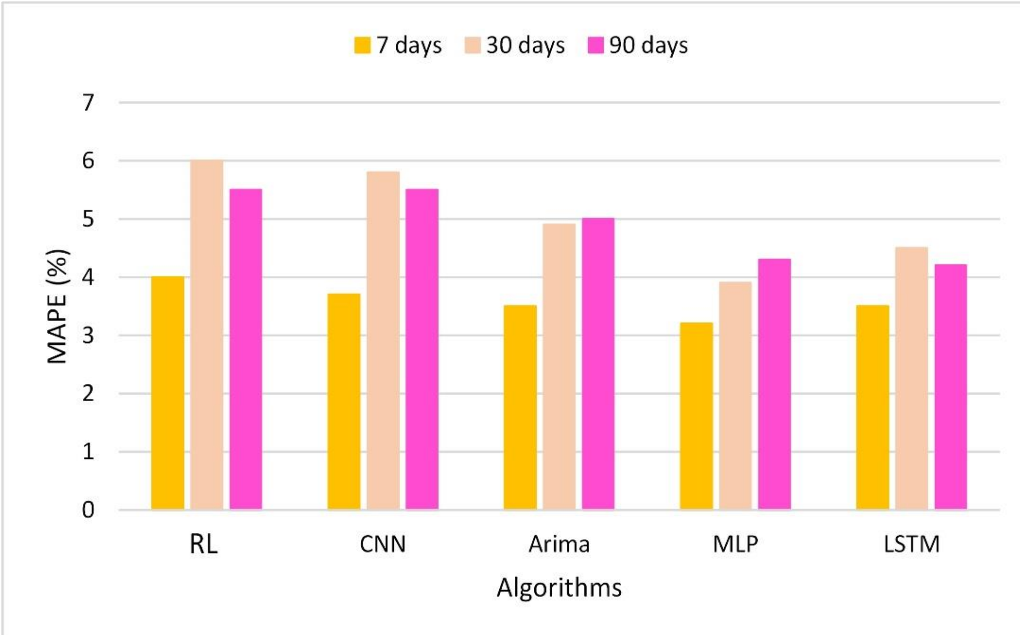


Figure 38. MAPE for the classification model

Compared to the proposed approach, Figure 39 illustrates the accuracy of the machine learning algorithms. The applied RL algorithm had the highest accuracy score for the seven days, 30 days, and 90 days. During this process, CNN produces the lowest score.

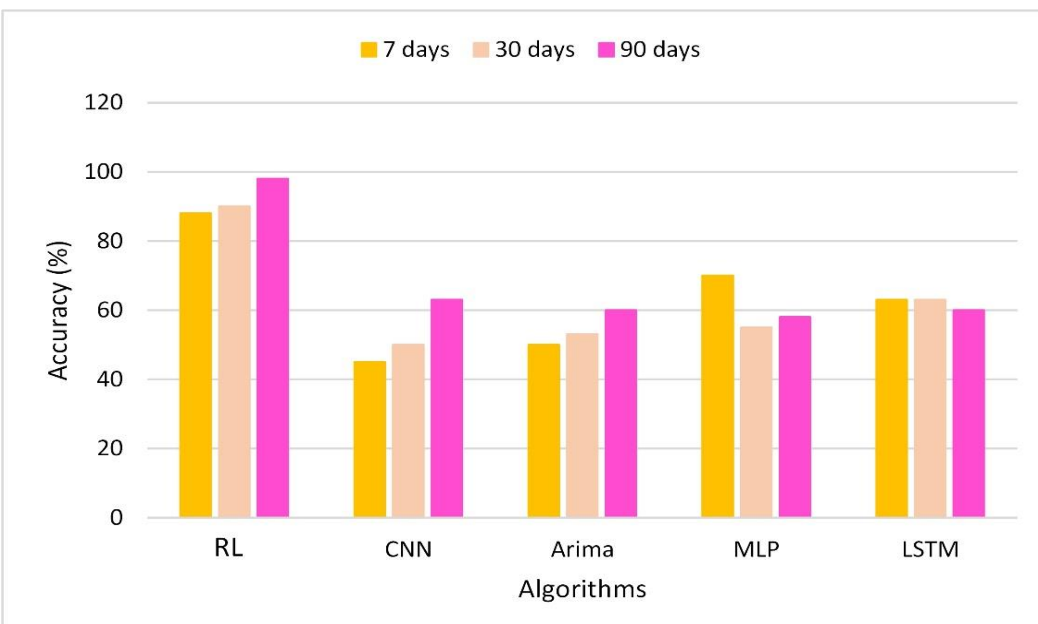


Figure 39. Classification model accuracy

4.2.3.2 Experimental Results of Cryptocurrency Prediction Based on Knowledge Discovery

Feature extraction is an essential part of pre-processing data since it helps improve the performance of the proposed model. Various approaches are used to extract the features. The RL algorithm specifies the importance of components first. Check cross-correlation and multi-collinearity next to decrease the number of features. Pearson correlation and variance inflation factors were used to determine this. Feature subset result includes low correlations and high values that are not multi-collinear. This process continues every seven days, 30 days, and 90 days. The feature selection process creates the best option for the mth day when the prediction for that day arrives. E.g., 7-day prediction systems cannot predict 90th-day exchange rates. The extraction of raw features is shown in Table 13.

Table 13. The raw feature technical indicators

Features	Description
Size of block	Transaction information based on Blockchain cryptographical links
Transaction	Send and receive payment records
Difficulty	The daily mining difficulty records based on the number of blocks
Sent Records	Daily payment records distinct digital coin addresses
Transaction value Avg.	The average value of digital coins transactions
Mining Profitability	Daily tera-hash profit based on USD

Ratio fee reward	Based on user transaction records, the ratio of rewards sent out
The average transaction fee	The median digital coin transaction
Avg. transaction fee	The received transaction fees from minor for verification
Block Time	Block mining requires time
Value of the average transaction	The median transaction value of digital coins
Hashrate	Computer power used by digital coins each day
Active addresses	In the transaction, the participants' addresses

Digital coins' exchange rate features are shown in Figure 40. The RL algorithm was applied to evaluate the importance of factors in this process.

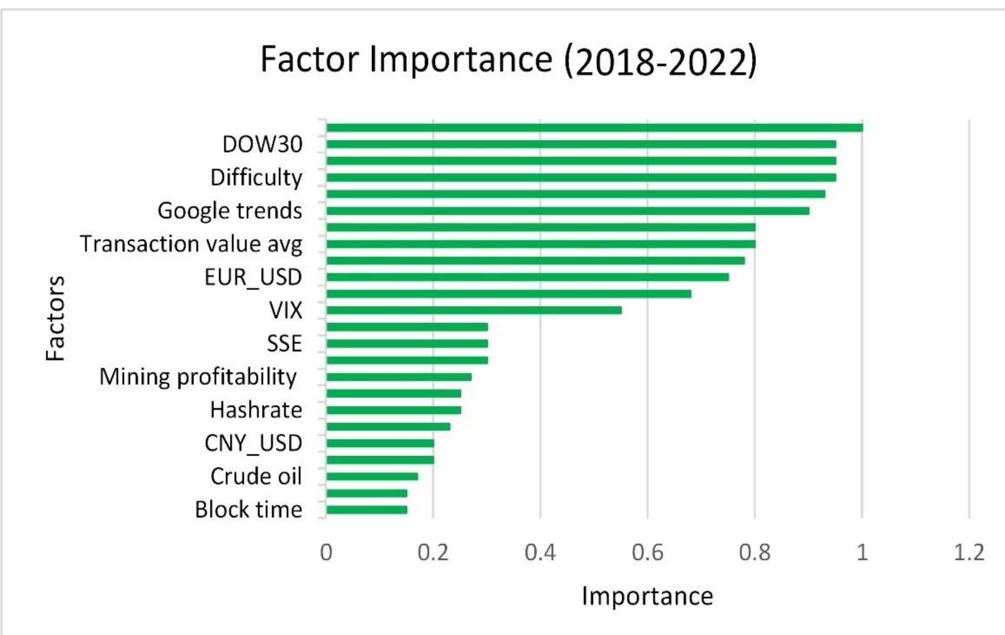


Figure 40. Factor importance differences 2018-2022

A good performance can be achieved for ML algorithms by selecting the correct variables for the input. A model needs to select input variables correctly to determine how their relationship relates to the goal value. Cross-correlations evaluate the similarity between variables and delayed values in the presented exchange rate system. The Equation below gives the cross-reaction coefficient between An and B_{n+m}. A represents the first variable, and B represents the second variable for the cross-correlated. ˆA represents the first variable mean, and $B^{\wedge}$ represents the second variable mean. x_m is the cross-correlation m_{th} order.

$$\boldsymbol{x_m} = \frac{\sum_{i=1}^{y-m}(A_i - A^{\wedge})(B_{i+m} - A^{\wedge})}{\sqrt{\sum_{i=1}^{y}(A_i - A^{\wedge})2\sum_{i=1}^{y}(B_{i+m} - B^{\wedge})}} \tag{15}$$

4.2.3.3 Experimental Results of Cryptocurrency Prediction Based on Blockchain

In blockchain frameworks, complex systems are modeled as networks, which is an essential perspective. Interconnected components and economic systems make up the social, physical, and technical network. Studies of complex networks have been conducted over the past 20 years to analyze the properties and structures of real-world networks, such as small-world phenomena, scale-free properties, and mechanisms of similar network formation. The network analysis improves in the cryptocurrency transaction study when identifying user characteristics and checking the network structure and temporal properties. Much media coverage has been given to Bitcoin as the first cryptocurrency. The digital coin world was also flooded with other cryptocurrencies at that time. Decentralized networks and computer cryptography are the two critical foundations of cryptocurrency. A public ledger called a blockchain is used to store encrypted transnational data. Nodes are distributed across the network through the register, and cross-validation is enabled through computation power. There are no limitations from a centralized authority on cryptocurrency processing. The transaction process function is shown in Figure 41. The blockchain network includes functions for creating, deleting, updating, and modifying assets and participants. A JavaScript smart contract was used to implement the transaction processor functions. Cryptocurrency records are updated based on event information and registry data provided by ShareRecord.

```
/**
* share Cryptocurrency Repositry Record with User
*@param {composers.CryptocurrencyRepositry.shareRepositryRecordsWithUser} newDetail - the update
CryptocurrencyRepositry transaction}
*@transaction
*/
async function shareRepositryRecordWithUser(record) {
        //payBill.user.balanceDue = payBill.bill.amount;
        return getAccetRegistry('composers.CryptocurrencyRepositry.Repositry')
              . then (function (assetRegistry) {
              record.cryptocurrencyRepositry.manager = record.userID;
              console.log(record.cryptocurrencyRepositry.manager);
               let shareRecordEvent = Cryptocurrency.newEvent('composer.CryptocurrencyRepositry',
               'shareRepositryRecordWithUsernotification');
               shareRecordEvent.cryptocurrencyRepositry = record.cryptocurrencyRepositry;
               emit (shareRecordEvent);
               return assetRegistry.update(record.cryptocurrencyRepositry);
         })
         .catch(function (error)  {

         });

}
```

Figure 41. Transaction process function for the cryptocurrency blockchain platform

Domain model elements can be controlled with access control language (ACL). Business network environments may have rules that specify what roles and users are allowed to make changes. Participants in this network are allowed to change the network by utilizing the ACL rules shown in Figure 42.

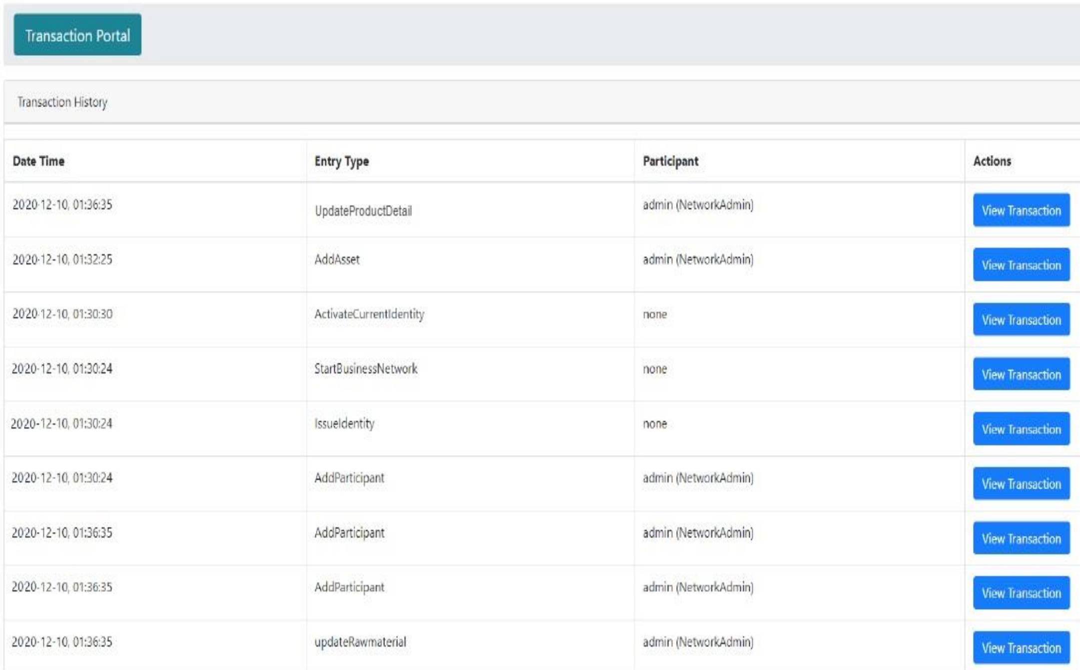
```
//6
rule UserUpdateCryptocurrencyOrderTransaction {
	description: "Only User can update Cryptocurrency Order transaction"
	participant: "composers.participant.User"
	operation : READ, CREATE, UPDATE
	resource: "composers.ordersdetails.UpdateorderStatus"
	action:Allow
}

//7
rule UserUpdateTransactionRecords {
	description: "Only User can update Transaction records"
	participant: "composers.participant.User"
	operation : READ, CREATE, UPDATE
	resource: "composers.prescription.UpdateTransactionRecords"

//8
rule UserViewTransaction {
	description: "Only User can view Transactions"
	participant: "composers.participant.User"
	operation : READ, CREATE, UPDATE
	resource: "composers.prescription.ViewTransactions"
```

Figure 42. Definition of access control in the cryptocurrency system

Using the blockchain network in the cryptocurrency system, an authorized user can add, update, or delete product details. To add new information, the user must fill out both the web form and the entries on the blockchain network. Similar to updating their information, users can update their knowledge in the blockchain network by submitting an update request. As shown in Figure 43, the proposed system presents the blockchain network's transaction history portal. All activities completed about transactions are shown in this portal. This information includes the date, time, entry type, participants, and actions. Also provided is the transaction log file information for the network.



Transaction Portal

Transaction History

Date Time	Entry Type	Participant	Actions
2020-12-10, 01:36:35	UpdateProductDetail	admin (NetworkAdmin)	View Transaction
2020-12-10, 01:32:25	AddAsset	admin (NetworkAdmin)	View Transaction
2020-12-10, 01:30:30	ActivateCurrentIdentity	none	View Transaction
2020-12-10, 01:30:24	StartBusinessNetwork	none	View Transaction
2020-12-10, 01:30:24	IssueIdentity	none	View Transaction
2020-12-10, 01:30:24	AddParticipant	admin (NetworkAdmin)	View Transaction
2020-12-10, 01:36:35	AddParticipant	admin (NetworkAdmin)	View Transaction
2020-12-10, 01:36:35	AddParticipant	admin (NetworkAdmin)	View Transaction
2020-12-10, 01:36:35	updateRawmaterial	admin (NetworkAdmin)	View Transaction

Figure 43. Cryptocurrency transaction portal history

The transaction latency is presented in Figure 44 as the maximum, minimum, and average. A comparison is made between different user groups, each with a different number of members. There was an average latency of 450 milliseconds for the 540 user process, 167 milliseconds for the 270 user process, and 145 milliseconds for the 90 user process. Based on this procedure, the average latency for the 45 users was 52 (ms).

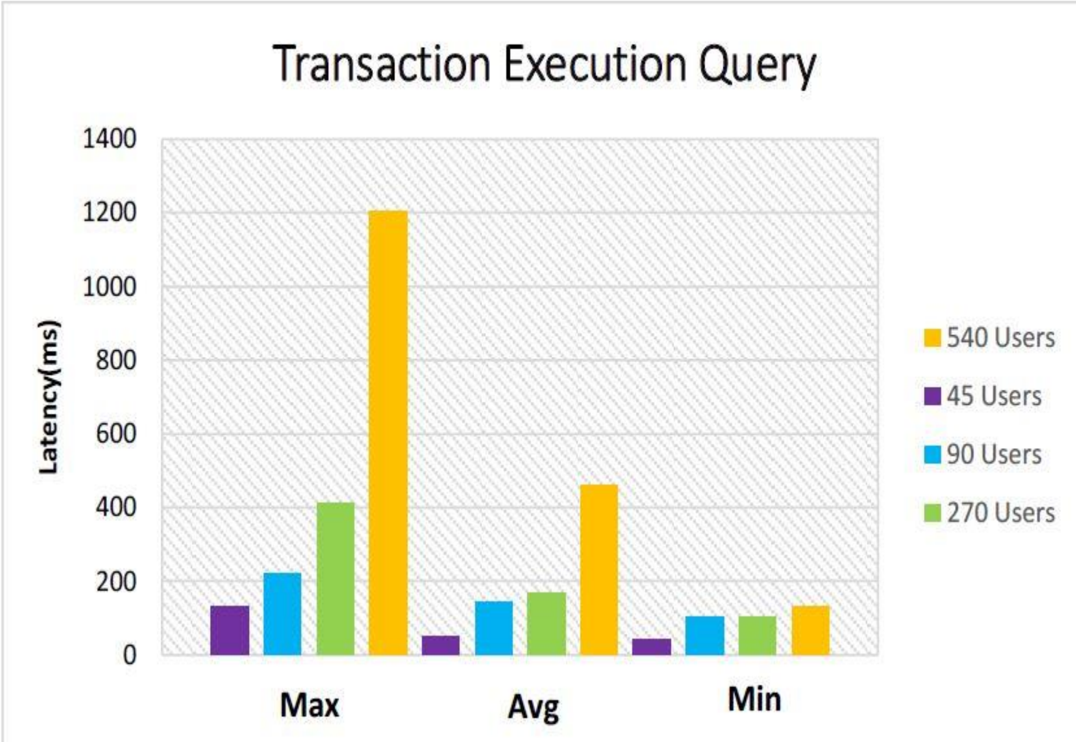


Figure 44. Latency of query transaction

A chart depicting the network's response times based on the number of users can be found in Figure 45. In the first, second, and third testing periods, 90, 270, and 320 users evaluated the system's performance. A slight change was observed in the system's response time in the third test group, while no difference was observed for the first and second tests.

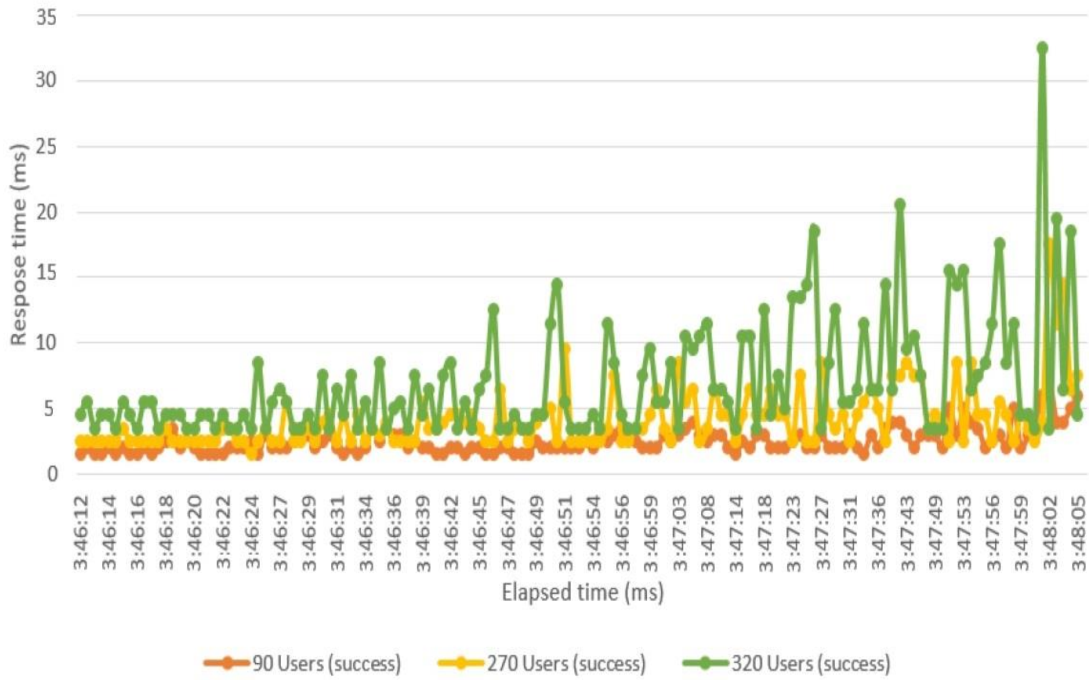


Figure 45. Response time of different requests

4.3 Performance Analysis of Cryptocurrency Prediction

In this section, we mainly focus on the proposed prediction system performance. We cover the whole topic in three subsections: Price management for the prediction, Risk management for the optimization of cryptocurrency framework, and security based on Blockchain.

4.3.1 Price Management for Prediction

Data used in this process is based on the coin market cap website and daily cryptocurrency prices from 2017 to 2022. The missing data information is excluded from the applied algorithm to avoid incompatible processes. Propagating reliable forwarded observations covers the missing parts. Sixty-one cryptocurrencies are listed in the final records of data. In Table 14, you will find samples of information for ten cryptocurrencies. Figures 46, 47, and 48 illustrate how regulation in Asia grew significantly in 2016 and 2017, following a sharp decline in 2018.

Table 14. Cryptocurrencies information

#	Mean	Min	Max
Block	0.0023	-0.5826	1.8873
Dash	0.0038	-0.3159	-.5492
Burst	0.0053	-0.3816	1.5189
GRS	0.1231	-0.4168	1.5154
NAV	0.0228	-0.7797	5.7875
PND	0.0813	-0.8922	6.0000
RDD	0.0225	-0.7891	2.3135
TRC	0.0213	-0.8991	13.0000
VTC	0.0067	-0.4496	1.4153
XRP	0.0039	-0.4600	1.7937

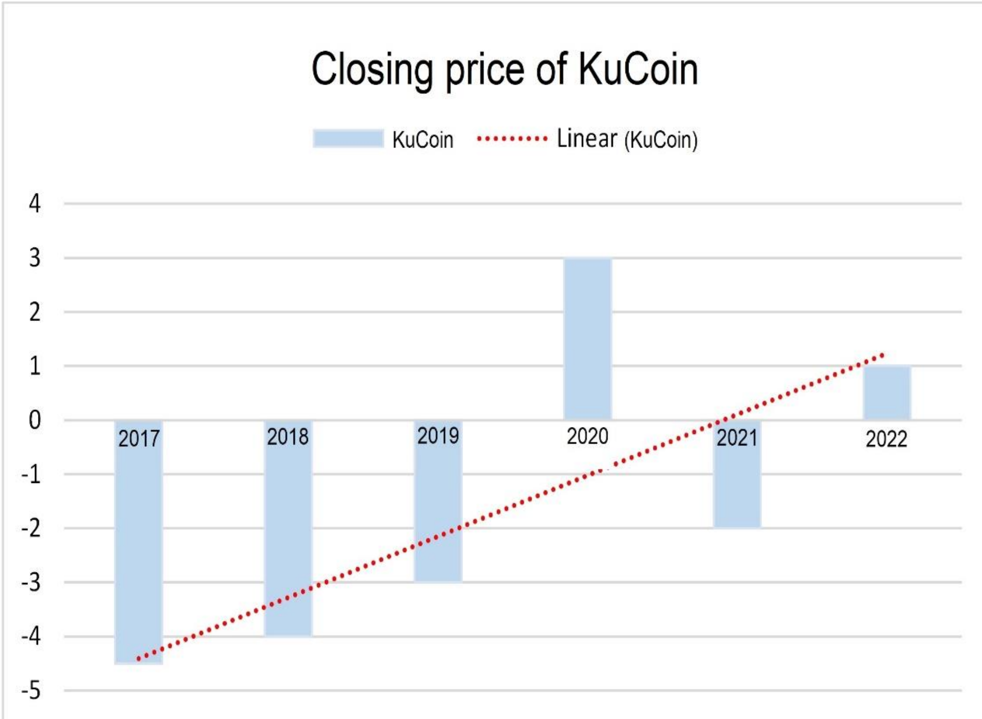


Figure 46. Closing price for KuCoin cryptocurrency

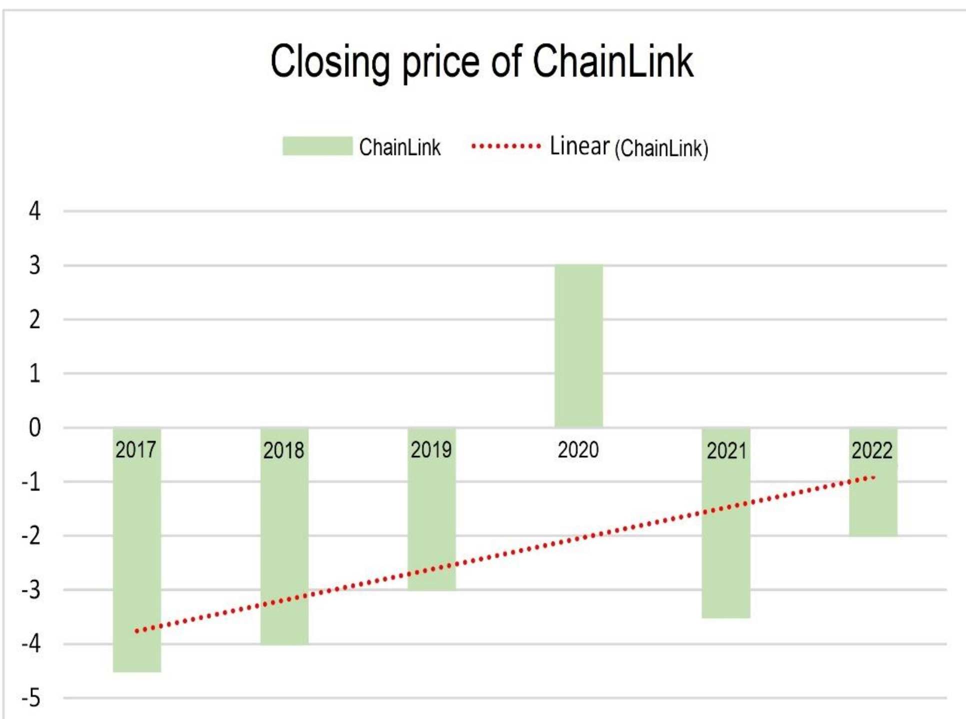


Figure 47. Closing price for ChainLink cryptocurrency

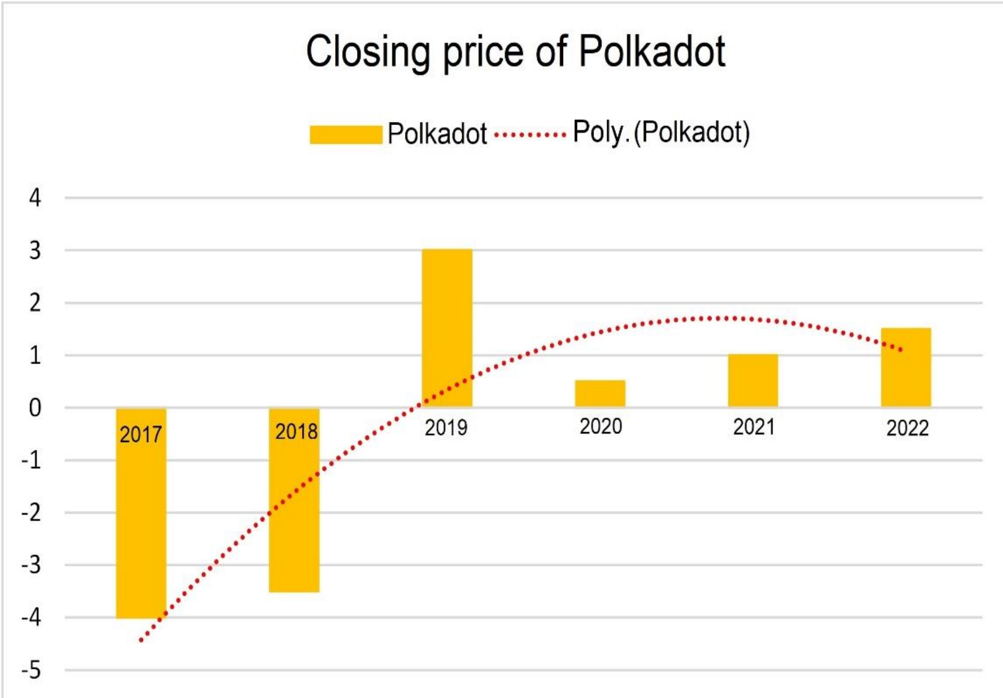


Figure 48. Cryptocurrency closing price for Polkadot

4.3.2 Risk Management for Prediction

As described in Tables 15, 16, and 17, Hierarchical Risk Parity (HRP) was compared using three famous risk-based asset traditional approaches: inverse volatility (IV), minimum variance (MV), and maximum diversification (MD). Table 15 summarizes how the out-of-sample HPR portfolio performed during 350 days of covariance estimation. In rows 0.7718 and 1.7802, you can find HPR annualized volatility and return. In addition, MD generates a greater return of 3.42, and IV reduces volatility to 1.5. When HPR is considered in terms of risk and return, its balance has a high impact, providing the best risk-return result compared to a quick ratio. For the 600-day and 850-day periods, the process is the same.

Table 15. The return portfolio risk performance = 350

Co-variance matrix sample					Co-variance matrix shrinkage			
#	HPR	IV	MV	MD	HPR	IV	MV	MD
Panel A: Window = 350								
Annualized Return	1.7913	1.6522	1.3528	3.5343	1.4278	1.6522	1.3525	2.4646
Annualized Volatility	0.8829	0.8779	1.4456	1.8673	0.9815	0.8779	1.1612	1.5449
Risk value (10%)	0.0098	0.0005	0.0043	0.0231	0.0235	0.0005	0.0020	0.0046
Conditional risk value (10%)	0.0029	0.0022	0.0005	0.0029	0.0049	0.0022	0.0004	0.0029
Draw down	0.3272	0.3541	0.7169	0.8459	0.3827	0.3541	0.5822	0.7282
Max draw down	0.4435	0.4389	0.7755	0.8834	0.6152	0.4398	0.6922	0.7917
Sharp ratio	0.2716	0.2581	0.1852	0.1828	0.2825	0.1581	0.2161	0.2174
Calmar ratio	5.5185	5.1423	2.1326	4.2325	5.6337	5.1423	2.3983	3.3761
Sortino ratio	0.0072	0.0066	0.0058	0.0220	0.0090	0.0066	0.0063	0.092

Table 16. The return portfolio risk performance = 600

Co-variance matrix sample					Co-variance matrix shrinkage			
#	HPR	IV	MV	MD	HPR	IV	MV	MD
Panel A: Window = 600								
Annualized Return	1.9839	1.7675	1.2262	3.7684	2.4379	1.7675	1.3000	2.6598
Annualized Volatility	0.9311	0.9195	0.9744	1.7148	1.1319	0.9295	0.8484	1.5313
Risk value (10%)	0.1275	0.0282	0.0261	0.0229	0.0039	0.0282	0.0269	0.225
Conditional risk value (10%)	0.0232	0.0282	0.0261	0.1229	0.0039	0.0282	0.0269	0.0225
Draw down	0.3272	0.3539	0.4944	0.8269	0.3827	0.3739	0.2857	0.6883
Max draw down	0.4435	0.4389	0.6134	0.8611	0.5152	0.4389	0.4251	0.7724
Sharp ratio	0.2677	0.2558	0.2295	0.0658	0.2744	0.4389	0.2363	0.0251
Calmar ratio	5.7275	5.3151	2.6293	4.5371	5.6729	5.3151	4.2982	3.7444
Sortino ratio	0.0077	0.0072	0.0066	0.0231	0.0084	0.0072	0.0061	0.0098

Table 17. The return portfolio risk performance = 850

Co-variance matrix sample					Co-variance matrix shrinkage			
#	HPR	IV	MV	MD	HPR	IV	MV	MD
Panel A: Window = 850								
Annualized Return	1.9737	1.8767	1.2161	4.5427	2.4331	1.8767	1.7677	2.9271
Annualized Volatility	0.9832	1.0021	0.9142	1.7928	1.1737	1.0021	1.2177	0.6674
Risk value (10%)	0.0650	0.0596	0.0814	0.0793	0.0546	0.0596	0.0684	0.0716
Conditional risk value (10%)	0.0077	0.0061	0.0048	0.0095	0.0062	0.0061	0.0057	0.0074
Draw down	0.3272	0.3539	0.3213	0.8169	0.3817	0.3539	0.5395	0.6784
Max draw down	0.4159	0.4389	0.4782	0.8525	0.4729	0.3539	0.5385	0.6784
Sharp ratio	0.2583	0.2514	0.2268	0.2139	0.2674	0.2514	0.2291	0.2233
Calmar ratio	6.3246	5.5537	3.6244	5.5124	6.3751	5.5537	3.2626	4.2448
Sortino ratio	0.0079	0.0076	0.0067	0.0253	0.0087	0.0075	0.0076	0.0216

From 2018 to 2022, the cryptocurrency market's risk index was calculated regularly. Uncertain policies produce the most volatility. Indexes move in unison in Figure 49.

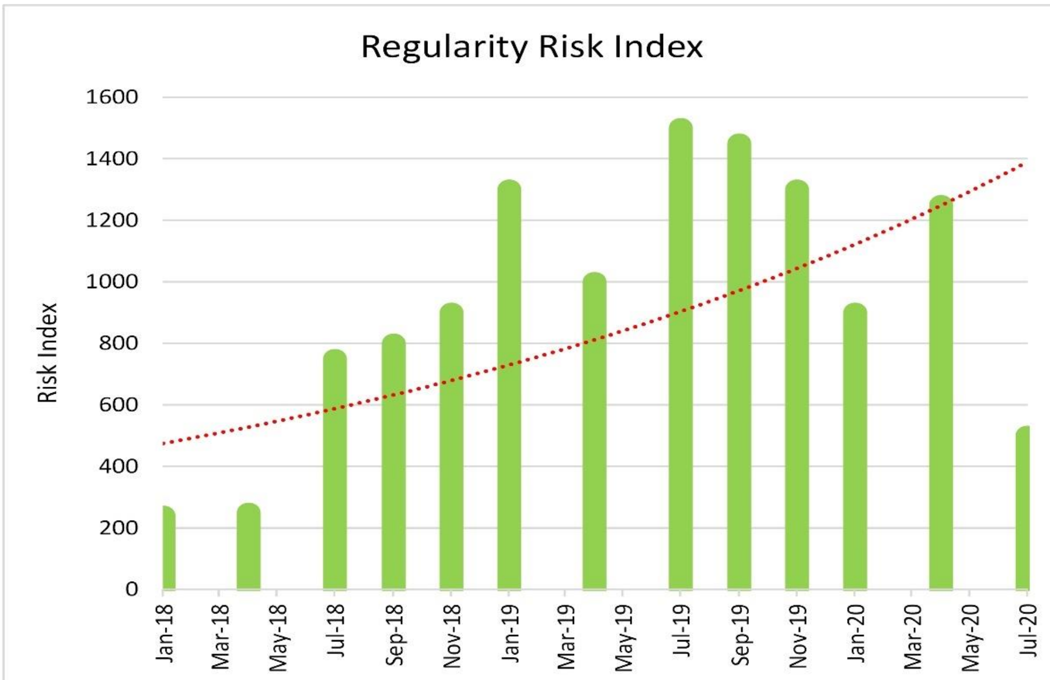


Figure 49. Cryptocurrency market risk index regularity

Information about the results can be found in Table 18. The 30-day records are listed in Table 19. It is evident from both tables that cryptocurrency investors have better results than other investors in terms of portfolios that minimize risk. Ether, too, provides a high-risk reduction when added to the portfolio of risk-minimizing cryptocurrencies. Litecoin offers the highest benefit in decreasing VaR of the four cryptocurrencies mentioned, but Bitcoin's lowest decrease in VaR.

Table 18. Risk evaluation of cryptocurrency portfolios

#		Portfolio	Portfolio of risk min	Determined portfolio
KuCoin	Re Red.	0.47258	0.02715	0.29387
	SV Red.	0.22325	0.00073	0.03353
	ES Red.	-1.2454	-0.21266	-1.2789
	VaR Red.	0.04471	0.04592	0.02143
	Risk Red.	0.99792	1.837E-05	0.99183
ChainLink	Re Red.	0.41988	0.02456	0.18388
	SV Red.	0.07759	0.00059	0.02823
	ES Red.	-1.5325	-0.00451	-1.2728
	VaR Red.	0.04378	0.04378	0.02849
	Risk Red.	0.99517	0.02449	0.99231
Polkadot	Re Red.	0.26585	0.02456	0.18471
	SV Red.	0.03753	0.00056	0.02148
	ES Red.	-1.55551	-0.08812	-1.2638
	VaR Red.	0.02681	0.04616	0.02755
	Risk Red.	0.99585	0.004829	0.98924

Table 19. Risk evaluation of cryptocurrency portfolios for 30 days

#		Portfolio	Portfolio of risk min	Determined portfolio
KuCoin	Re Red.	1.39172	1.98181	2.51911
	SV Red.	1.91481	1.21357	5.41574
	ES Red.	-5.1427	-4.2156	-8.3354
	VaR Red.	0.02734	0.02622	0.02846
	Risk Red.	0.61353	0.36192	0.91294
ChainLink	Re Red.	1.31639	1.62473	3.00738
	SV Red.	1.11648	0.51222	3.26936
	ES Red.	-4.2481	-2.6435	-9.021
	VaR Red.	0.04173	0.03794	0.04185
	Risk Red.	0.36258	0.29711	0.82485
Polkadot	Re Red.	3.51581	1.89514	1.58792
	SV Red.	1.25241	0.067644	0.21843
	ES Red.	-30.862	-3.2977	-9.971
	VaR Red.	0.02227	0.04464	0.03571
	Risk Red.	0.95486	0.04822	0.66765

RL and MDD baseline output is shown in Figure 50. MDD reached 13% in this system because MDD in Hold consists of 49%, and MDD in random consists of 64%. In case the baseline loses the funds of investments, the stability of the proposed system reaches almost 1.15.s

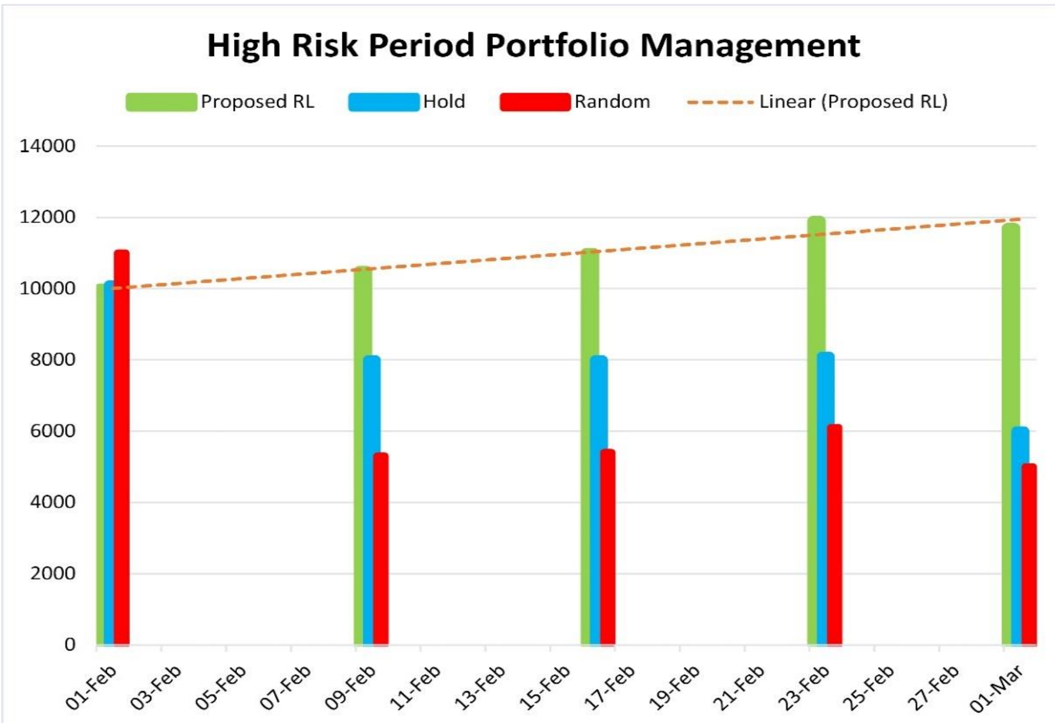


Figure 50. Reinforcement learning-based high-risk management

4.3.3 Evaluation Metrix of the Cryptocurrency Prediction

Data from the KuCoin, ChainLink, and Polkadot electronic wallets and details of the daily transactions are used to conduct the experiments and simulations. Data analysis is then performed to determine the parameters based on the date, high price, low price, open, close, quote volume, and average weighted. Several performance metrics evaluated the trained price prediction model, including mean absolute error, mean absolute percentage error, root means square error, and mean square error. The formulation is divided as below:

- Mean Absolute Error (MAE): The statistic evaluates the difference between the predicted and actual value as defined by the following Equation below:

$$\boldsymbol{MAE = \frac{1}{m}\sum_{n-1}^{m}\frac{|x_n - \hat{x_n}|}{|x|}} \tag{16}$$

- Mean Absolute Percentage Error (MAPE): Analyzing statistics for their predictive accuracy is the purpose of this statistical evaluation.

- $$\boldsymbol{MAPE = \frac{1}{m}\sum_{n-1}^{m}\frac{|x_n - \hat{x_n}|}{|x|} * 100} \tag{17}$$

- Root Mean Square Error (RMSE): As shown in the Equation below, the measurement of error rate and size is based on the target value:

$$\sqrt{\boldsymbol{RMSE = \frac{1}{m}\sum_{n-1}^{m}(\frac{x_n - \hat{x_n}}{x})^2}} \tag{18}$$

- Mean Square Error (MSE): According to the Equation below, the correlation between predicted and actual values is measured by this statistical evaluation:

$$MSE = \frac{1}{m}\sum_{n=1}^{m}(x_n - \hat{x_n})^2 \qquad (19)$$

Figures 51, 52, 53, and 54 show the MSE, RMSE, MAE, and MAPE evaluation metrics. Various lengths in a row are offered, such as three days, seven days, and one month.

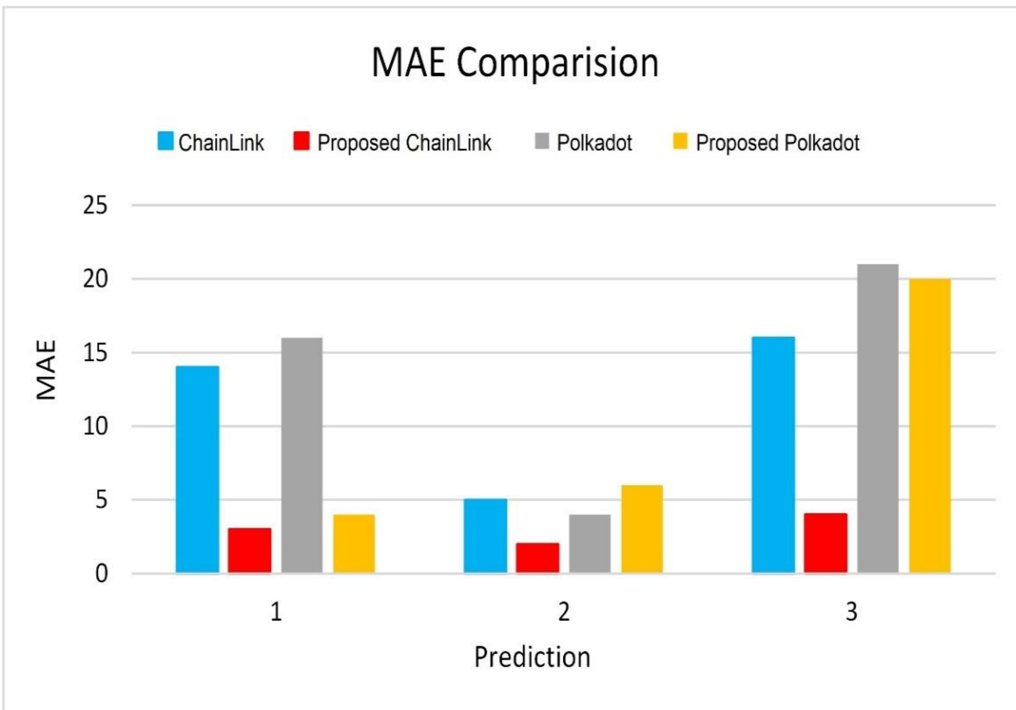


Figure 51. Proposed approach MAE comparison

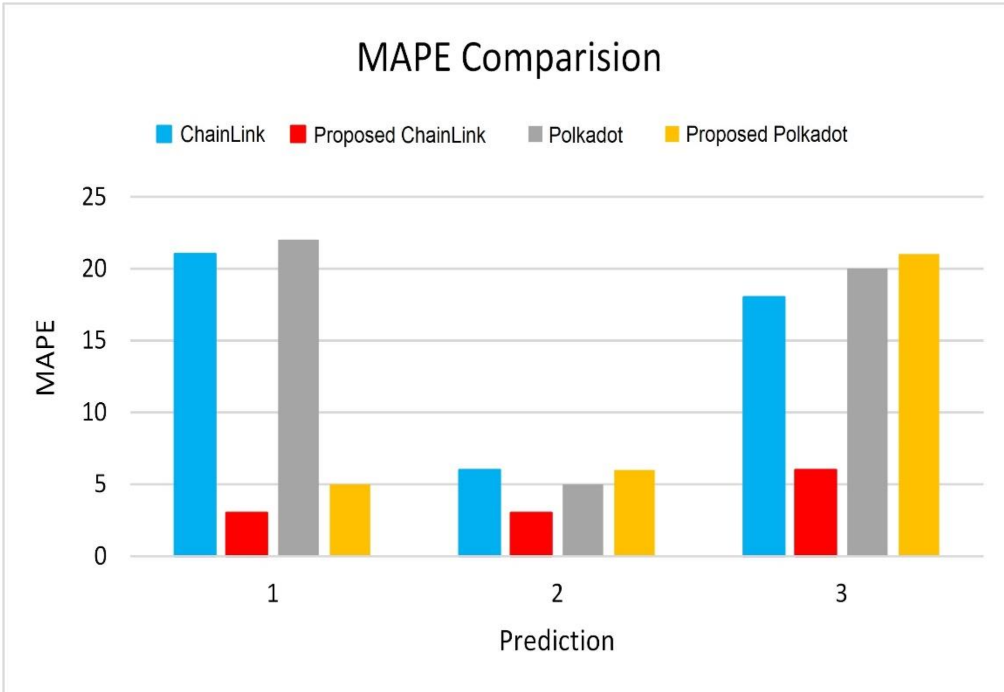


Figure 52. Proposed approach MAPE comparison

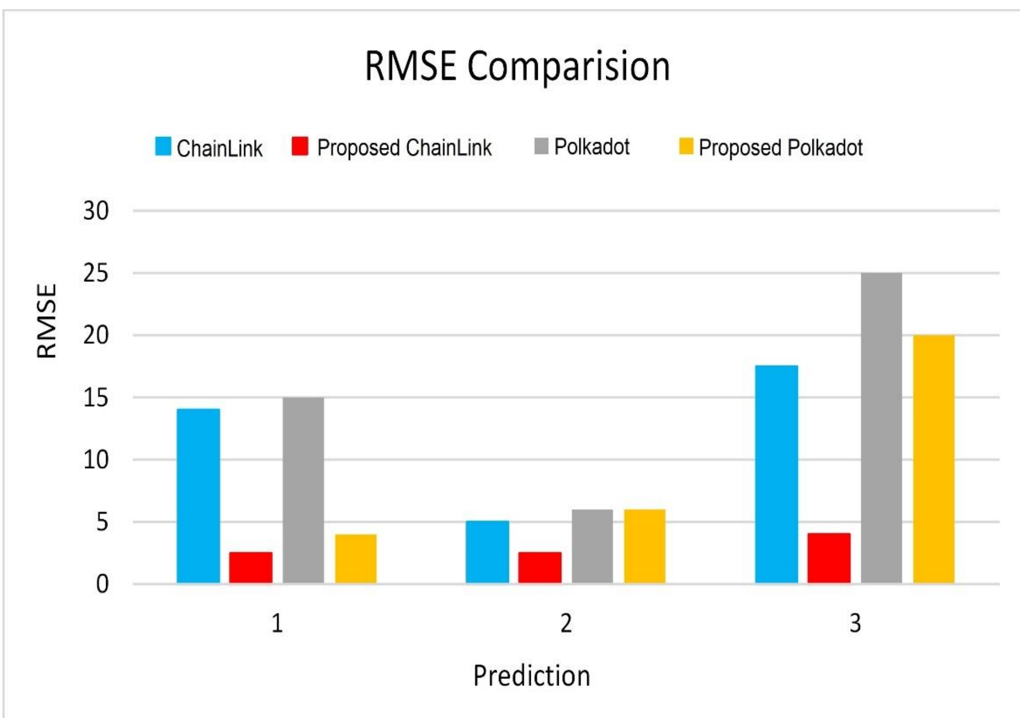


Figure 53. Proposed approach RMSE comparison

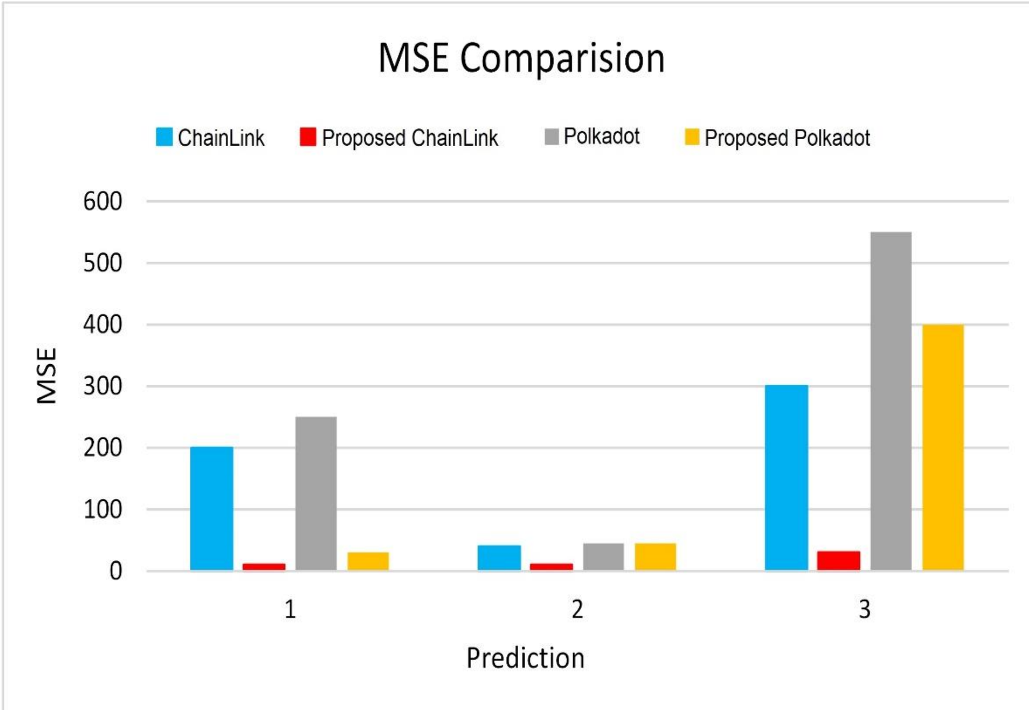


Figure 54. Proposed approach MSE comparison

Chapter 5: Conclusion and Future Directions

Cryptocurrency price prediction is a challenging area in finding a suitable digital coin for future investment, in other words, saving in a secure environment for plans. This thesis proposes a novel architecture based on machine learning and Blockchain for cryptocurrency price prediction. The thesis serves as a prediction system architecture design pattern involving various parameters and goals. The architecture pattern is built-in prediction, optimization, scheduling, and control.

This thesis is segregated into five chapters, presenting each module separately. Chapter 1 provides the introduction detailing the basic conceptual flow of the prediction system. The background knowledge of the proposed study is explained by forming a linkage with the existing prediction systems. The motivation of the thesis is described based on an in-depth critical analysis of the current state-of-the-art in the domain, which also contributed to forming the problem statement and research goal.

Chapter 2 presents the literature review related to the concept of cryptocurrency, knowledge discovery, topic modeling, and Blockchain for the cryptocurrency framework. Chapter 3 presents the proposed cryptocurrency price prediction, which comprises four layers. Layers are independent of one another. Developers can add and remove modules independently without affecting the system because of the decoupled feature. Chapter 4 presents the prediction architecture with detailed design, implementation, and performance analysis mapped on the general cryptocurrency architecture. Finally, chapter 5 concludes the rest of the paper.

This thesis used the Reinforcement Learning (RL) method and the Hierarchical Risk Parity (HRP) asset allocation method to analyze cryptocurrency asset portfolios. Comparisons of reinforcement learning and other methods used in this field show high-performance evaluation results. Due to its learning-based approach, RL is ideal for this process. It allows the system structure to provide the correct information with high accuracy. A high level of diversity and desirable properties are also found in the HRP. Different estimation windows and methodologies were used in analyzing the results. A similar rebalancing of the selected period was also used in the analysis. Applied HRP improves risk management by providing meaningful alternatives to transitional asset allocations. Furthermore, the proposed architecture implements the latest tools and technologies to enhance the presentation of information.

References

[1] Vidal-Tomas, D.; Ibanez, A. Semi-strong efficiency of Bitcoin. Finance Research Letters 2018, 27, 259–265.

[2] Tan, L.; Xue, L. Research on the Development of Digital Currencies under the COVID-19 Epidemic. Procedia Computer Science 2021, 187, 89–96.

[3] Pieters, G. Digital Currencies and Central Banks. In The Palgrave Handbook of Technological Finance; Springer, 2021; pp. 139–160.

[4] Cortez, K.; Rodriguez-Garcia, M.d.P.; Mongrut, S. Exchange Market Liquidity Prediction with the K-Nearest Neighbor Approach: Crypto vs. Fiat Currencies. Mathematics 2021, 9, 56.

[5] Atsalakis, G.S.; Atsalaki, I.G.; Pasiouras, F.; Zopounidis, C. Bitcoin price forecasting with neuro-fuzzy techniques. European Journal of Operational Research 2019, 276, 770–780.

[6] Vo, A.; Yost-Bremm, C. A high-frequency algorithmic trading strategy for cryptocurrency. Journal of Computer Information Systems 2020, 60, 555–568.

[7] Engel, C. Lessons for Cryptocurrencies from Foreign Exchange Markets. In Digital Currency Economics And Policy;World Scientific, 2021; pp. 121–180.

[8] van der Merwe, A. Cryptocurrencies and Other Digital Asset Investments. In The Palgrave Handbook of FinTech and Blockchain; Springer, 2021; pp. 445–471.

[9] Makinen, Y.; Kanniainen, J.; Gabbouj, M.; Iosifidis, A. Forecasting jump arrivals in stock prices: new attention-based network architecture using limit order book data. Quantitative Finance 2019, 19, 2033–2050.

[10] Sirignano, J.; Cont, R. Universal features of price formation in financial markets: perspectives from deep learning. Quantitative Finance 2019, 19, 1449–1459.

[11] Zhang, Z.; Zohren, S.; Roberts, S. Deeplob: Deep convolutional neural networks for limit order books. IEEE Transactions on Signal Processing 2019, 67, 3001–3012.

[12] Lahmiri, S.; Bekiros, S. Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals 2019, 118, 35−40.

[13] Tan, X.; Kashef, R. Predicting the closing price of cryptocurrencies: a comparative study. Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, 2019, pp. 1−5.

[14] Poongodi, M., et al. "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system." Computers & Electrical Engineering 81 (2020): 106527.

[15] Aditya Pai, B.; Devareddy, L.; Hegde, S.; Ramya, B. A Time Series Cryptocurrency Price Prediction Using LSTM. In Emerging Research in Computing, Information, Communication and Applications; Springer, 2022; pp. 653−662.

[16] Wu, R.; Ishfaq, K.; Hussain, S.; Asmi, F.; Siddiquei, A.N.; Anwar, M.A. Investigating e-Retailers' Intentions to Adopt Cryptocurrency Considering the Mediation of Technostress and Technology Involvement. Sustainability 2022, 14, 641.

[17] Mendis, G.J.;Wu, Y.;Wei, J.; Sabounchi, M.; Roche, R. A blockchain-powered decentralized and secure computing paradigm. IEEE Transactions on Emerging Topics in Computing 2020.

[18] Liang, J.; Li, L.; Chen, W.; Zeng, D. Targeted addresses identification for bitcoin with network representation learning. 2019 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2019, pp. 158−160.

[19] Zola, F.; Eguimendia, M.; Bruse, J.L.; Urrutia, R.O. Cascading machine learning to attack bitcoin anonymity. 2019 IEEE International Conference on Blockchain (Blockchain). IEEE, 2019, pp. 10−17.

[20] Michalski, R.; Dziubałtowska, D.; Macek, P. Revealing the character of nodes in a blockchain with supervised learning. Ieee Access 2020, 8, 109639−109647.

[21] Toyoda, K.; Ohtsuki, T.; Mathiopoulos, P.T. Multi-class bitcoin-enabled service identification based on transaction history summarization. 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2018, pp. 1153−1160.

[22] Xueshuo, X.; Jiming, W.; Junyi, Y.; Yaozheng, F.; Ye, L.; Tao, L.; Guiling, W. AWAP: Adaptive weighted attribute propagation enhanced community detection model for bitcoin de-anonymization. Applied Soft Computing 2021, 109, 107507.

[23] Jourdan, M.; Blandin, S.; Wynter, L.; Deshpande, P. Characterizing entities in the bitcoin blockchain. 2018 IEEE international conference on data mining workshops (ICDMW). IEEE, 2018, pp. 55−62.

[24] Sun Yin, H.H.; Langenheldt, K.; Harlev, M.; Mukkamala, R.R.; Vatrapu, R. Regulating cryptocurrencies: a supervised machine learning approach to de-anonymizing the bitcoin blockchain. Journal of Management Information Systems 2019, 36, 37−73.

[25] Linoy, S.; Stakhanova, N.; Ray, S. De-anonymizing Ethereum blockchain smart contracts through code attribution. International Journal of Network Management 2021, 31, e2130.

[26] Dey, A.K.; Akcora, C.G.; Gel, Y.R.; Kantarcioglu, M. On the role of local blockchain network features in cryptocurrency price formation. Canadian Journal of Statistics 2020, 48, 561−581.

[27] Saad, M.; Choi, J.; Nyang, D.; Kim, J.; Mohaisen, A. Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions. IEEE Systems Journal 2019, 14, 321−332.

[28] Jay, P.; Kalariya, V.; Parmar, P.; Tanwar, S.; Kumar, N.; Alazab, M. Stochastic neural networks for cryptocurrency price prediction. IEEE Access 2020, 8, 82804−82818.

[29] Sin, E.; Wang, L. Bitcoin price prediction using ensembles of neural networks. 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, 2017, pp. 666−671.

[30] Chen, Z.; Li, C.; Sun, W. Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics 2020, 365, 112395.

[31] Mudassir, M.; Bennbaia, S.; Unal, D.; Hammoudeh, M. Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. Neural Computing and Applications 2020, pp. 1−15.

[32] Wang, Y.;Wang, H. Using networks and partial differential equations to forecast bitcoin price movement. Chaos: An Interdisciplinary Journal of Nonlinear Science 2020, 30, 073127.

[33] Muzafar Rasool Bhat, Majid A Kundroo, Tanveer A Tarray, and Basant Agarwal. Deep lda: A new way to topic model. Journal of Information and Optimization Sciences, pages 1–12, 2019.

[34] Tobias Hecking and Loet Leydesdorff. Topic modelling of empirical text corpora: Validity, reliability, and reproducibility in comparison to semantic maps. arXiv preprint arXiv:1806.01045, 2018.

[35] Jay M Ponte and W Bruce Croft. Text segmentation by topic. In International Conference on Theory and Practice of Digital Libraries, pages 113–125. Springer, 1997. [28] Gerard Salton, Amit Singhal, Mandar Mitra, and Chris Buckley. Automatic text structuring and summarization. Information processing & management, 33(2):193–207, 1997.

[36] Nicola Stokes, Joe Carthy, and Alan F Smeaton. Select: a lexical cohesion based news story segmentation system. AI communications, 17(1):3–12, 2004.

[37] Heidi Christensen, BalaKrishna Kolluru, Yoshihiko Gotoh, and Steve Renals. Maximum entropy segmentation of broadcast news. In Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., volume 1, pages I–1029. IEEE, 2005.

[38] Pei-Yun Hsueh, Johanna D Moore, and Steve Renals. Automatic segmentation of multiparty dialogue. In 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006.

[39] Xiang Ji and Hongyuan Zha. Domain-independent text segmentation using anisotropic diffusion and dynamic programming. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pages 322–329. ACM, 2003.

[40] Sanda Harabagiu and Steven Maiorano. Finding answers in large collections of texts: Paragraph indexing+ abductive inference. In Proceedings of the AAAI Fall Symposium on Question Answering Systems, pages 63–71, 1999.

[41] Charles LA Clarke, Gordon V Cormack, Thomas R Lynam, CM Li, and Greg L

McLearn. Web reinforced question answering (multitext experiments for trec 2001). 2001. [35] Hyeon-Jin Kim, Hyo-Jung Oh, Ji-Hyun Wang, Chung-Hee Lee, and Myung-Gil Jang. The 3-step answer processing method for encyclopedia question-answering system: Anyquestion1. 0. In Annual Conference on Human and Language Technology. Human and Language Technology, 2004.

[42] Yan Xiao, Jacky Keung, Kwabena E Bennin, and Qing Mi. Improving bug localization with word embedding and enhanced convolutional neural networks. Information and Software Technology, 105:17–29, 2019.

[43] Liqun Shao. Text Automation: Title Generation, Summarization and Classification. PhD thesis, University of Massachusetts Lowell, 2018.

[44] Stefan Bunk and Ralf Krestel. Welda: Enhancing topic models by incorporating local word context. In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, pages 293–302. ACM, 2018.

[45] Te-Ming Chang and Wen-Feng Hsiao. Lda-based personalized document recommendation. In PACIS, page 13, 2013.

[46] Cedric De Boom, Steven Van Canneyt, Thomas Demeester, and Bart Dhoedt. Representation learning for very short texts using weighted word embedding aggregation. Pattern Recognition Letters, 80:150–156, 2016.

[47] Jannach, D.; Zanker, M.; Felfernig, A.; Friedrich, G. Recommender Systems: An Introduction; Cambridge University Press: Cambridge, UK, 2010.

[48] Zhang, Z.P.; Kudo, Y.; Murai, T.; Ren, Y.G. Enhancing recommendation accuracy of item-based collaborative filtering via item-variance weighting. Appl. Sci. 2019, 9, 1928.

[49] Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. Knowl.-Based Syst.2013, 46, 109–132.

[50] Adomavicius, G.; Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. 2005, 17, 734–749.

[51] Lu, J.;Wu, D.; Mao, M.;Wang,W.; Zhang, G. Recommender system application developments: A survey. Decis. Support Syst. 2015, 74, 12–32.

[52] Albanese, M.; d'Acierno, A.; Moscato, V.; Persia, F.; Picariello, A. A multimedia recommender system. ACM Trans. Internet Technol. (TOIT) 2013, 13, 1–32.

[53] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Analysis of recommendation algorithms for e-commerce. In Proceedings of the 2nd ACM Conference on Electronic Commerce, Minneapolis, MN, USA, 17–20 October 2000; pp. 158–167.

[54] Cho, Y.H.; Kim, J.K. Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce. Expert Syst. Appl. 2004, 26, 233–246.

[55] Cho, Y.H.; Kim, J.K.; Kim, S.H. A personalized recommender system based on web usage mining and decision tree induction. Expert Syst. Appl. 2002, 23, 329–342.

[56] Lawrence, R.D.; Almasi, G.S.; Kotlyar, V.; Viveros, M.; Duri, S.S. Personalization of supermarket product recommendations. In Applications of Data Mining to Electronic Commerce; Springer: Berlin, Germany, 2001; pp. 11–32.

[57] Albanese, M.; Chianese, A.; d'Acierno, A.; Moscato, V.; Picariello, A. A multimedia recommender integrating object features and user behavior. Multimed. Tools Appl. 2010, 50, 563–585.

[58] Albanese, M.; d'Acierno, A.; Moscato, V.; Persia, F.; Picariello, A. Modeling recommendation as a social choice problem. In Proceedings of the Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 329–332.

[59] Amato, F.; Gargiulo, F.; Moscato, V.; Persia, F.; Picariello, A. Recommendation of Multimedia Objects for Social Network Applications. In Proceedings of the EDBT/ICDT Workshops, Athens, Greece, 28 March 2014; pp. 288–293.

[60] Albanese, M.; d'Acierno, A.; Moscato, V.; Persia, F.; Picariello, A. A ranking method for multimedia recommenders. In Proceedings of the ACM International Conference on Image and Video Retrieval, Xi'an, China, 5–7 July 2010; pp. 311–318.

[61] Burke, R.Hybridweb recommender systems. In The AdaptiveWeb; Springer: Berlin, Germnay, 2007; pp. 377–408.

[62] Goldberg, D.; Nichols, D.; Oki, B.M.; Terry, D. Using collaborative filtering to weave an information tapestry. Commun. ACM 1992, 35, 61–70.

[63] Shahbazi, Z.; Byun, Y.C. Product Recommendation Based on Content-based Filtering Using XGBoost Classifier. Int. J. Adv. Sci. Technol. 2019, 29, 6979–6988.

[64] Papagelis, M.; Plexousakis, D.; Kutsuras, T. Alleviating the sparsity problem of collaborative filtering using trust inferences. In International Conference on Trust Management; Springer: Berlin, Germany, 2005; pp. 224–239.

[65] Jamil, F.; Hang, L.; Kim, K.; Kim, D. A novel medical blockchain model for drug supply chain integrity management in a smart hospital. Electronics 2019, 8, 505.

[66] Jamil, F.; Iqbal, M.A.; Amin, R.; Kim, D. Adaptive thermal-aware routing protocol for wireless body area network. Electronics 2019, 8, 47.

[67] Jamil, F.; Ahmad, S.; Iqbal, N.; Kim, D.H. Towards a Remote Monitoring of Patient Vital Signs Based on IoT-Based Blockchain Integrity Management Platforms in Smart Hospitals. Sensors 2020, 20, 2195.

[68] Jamil, F.; Kim, D.H. Improving Accuracy of the Alpha–Beta Filter Algorithm Using an ANN-Based Learning Mechanism in Indoor Navigation System. Sensors 2019, 19, 3946.

[69] Jamil, F.; Iqbal, N.; Ahmad, S.; Kim, D.H. Toward Accurate Position Estimation Using Learning to Prediction Algorithm in Indoor Navigation. Sensors 2020, 20, 4410.

[70] Ahmad, S.; Jamil, F.; Khudoyberdiev, A.; Kim, D. Accident risk prediction and avoidance in intelligent

[71] semi-autonomous vehicles based on road safety data and driver biological behaviors. J. Intell. Fuzzy Syst. 2020, 38, 4591–4601.

[72] Jamil, F.; Kim, D. Payment Mechanism for Electronic Charging using Blockchain in Smart Vehicle. Korea 2019, 30, 31.

[73] Shahbazi, Z.; Byun, Y.C. Towards a Secure Thermal-Energy Aware Routing Protocol in Wireless Body Area Network Based on Blockchain Technology. Sensors 2020, 20, 3604.

[74] Song, H.S.; kyeong Kim, J.; Kim, S.H. Mining the change of customer behavior in an internet shopping mall. Expert Syst. Appl. 2001, 21, 157–168.

[75] Liu, N.N.; Zhao, M.; Xiang, E.; Yang, Q. Online evolutionary collaborative filtering. In Proceedings of the Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 95–102.

[76] Sun, J.Z.; Parthasarathy, D.; Varshney, K.R. Collaborative kalman filtering for dynamic matrix factorization. IEEE Trans. Signal Process. 2014, 62, 3499–3509.

[77] Vinagre, J.; Jorge, A.M. Forgetting mechanisms for scalable collaborative filtering. J. Braz. Comput. Soc. 2012, 18, 271–282.

[78] Zhang, Z.; Kudo, Y.; Murai, T. Neighbor selection for user-based collaborative filtering using covering-based rough sets. Ann. Oper. Res. 2017, 256, 359–374.

[79] Rosaci, D. Finding semantic associations in hierarchically structured groups ofWeb data. Formal Asp. Comput. 2015, 27, 867–884.

[80] De Meo, P.; Fotia, L.; Messina, F.; Rosaci, D.; Sarné, G.M. Providing recommendations in social networks by integrating local and global reputation. Inf. Syst. 2018, 78, 58–67.

[81] Zhang, Z.P.; Kudo, Y.; Murai, T.; Ren, Y.G. Addressing complete new item cold-start recommendation: A niche item-based collaborative filtering via interrelationship mining. Appl. Sci. 2019, 9, 1894.

[82] Mahdi H Miraz and Maaruf Ali. Applications of blockchain technology beyond cryptocurrency. arXiv preprint arXiv:1801.03528, 2018.

[83] TM Navamani. A review on cryptocurrencies security. Journal of Applied Security Research, pages 1–21, 2021.

[84] Douglas J Cumming, Sofia Johan, and Anshum Pant. Regulation of the crypto-economy: Managing risks, challenges, and regulatory uncertainty. Journal of Risk and Financial Management, 12(3):126, 2019.

[85] Irina Cvetkova. Cryptocurrencies legal regulation. BRICS LJ, 5:128, 2018.

[86] Antoon Spithoven. Theory and reality of cryptocurrency governance. Journal of Economic Issues, 53(2):385–393, 2019.

[87] Rian Matanky-Becker and Ella Cockbain. Behind the criminal economy: using uk tax fraud investigations to understand money laundering myths and models. Crime, Law and Social Change, pages 1–25, 2021.

[88] Antoon Spithoven. Theory and reality of cryptocurrency governance. Journal of Economic Issues, 53(2):385–393, 2019.

[89] Rian Matanky-Becker and Ella Cockbain. Behind the criminal economy: using uk tax fraud investigations to understand money laundering myths and models. Crime, Law and Social Change, pages 1–25, 2021.

[90] S Malwa. $1.2 billion in cryptocurrency laundered through bitcoin tumblers, privacy coins. Yahoo! Finance, 2018.

[91] Szde Yu. Does fear of victimization deter online shopping? Journal of Financial Crime, 2018.

[92] Muhammed Miah. Blockchain technology in peer-to-peer elearning: Opportunities and challenges. In Proceedings of the EDSIG Conference ISSN, volume 2473, page 4901, 2020.

[93] Mario Arias-Oliva, Jorge Pelegrin-Borondo, and Gustavo Matias-Clavero. Variables influencing cryptocurrency use: a technology acceptance model in spain. Frontiers in Psychology, 10:475, 2019.

[94] Christian Rueckert. Cryptocurrencies and fundamental rights. Journal of Cybersecurity, 5(1):tyz004, 2019.

[95] Sin, E.; Wang, L. Bitcoin price prediction using ensembles of neural networks. 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, 2017, pp. 666–671.

[96] Azari, A. Bitcoin price prediction: An ARIMA approach. arXiv preprint arXiv:1904.05315 2019.

[97] Weytjens, H.; Lohmann, E.; Kleinsteuber, M. Cash flow prediction: MLP and LSTM compared to ARIMA and Prophet. Electronic Commerce Research 2021, 21, 371–391.