Creation of a recommendation system to recommend cryptocurrency portfolio using Association rule mining

MSc Research Project
Data Analytics

Girish Kandan Murugan
Student ID: X19199791

School of Computing
National College of Ireland

Supervisor: Sean Heaney

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Girish Kandan Murugan
Student ID:	X19199791
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Sean Heaney
Submission Due Date:	20/04/2021
Project Title:	Creation of a recommendation system to recommend crypto-currency portfolio using Association rule mining
Word Count:	5589
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	23rd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

Creation of a recommendation system to recommend cryptocurrency portfolio using Association rule mining

Girish Kandan Murugan
X19199791

Abstract

Cryptocurrencies have emerged well in the past decade starting from bitcoin to very latest ones. Although there are certain inconsistencies people have started using cryptocurrency for trading and transaction purposes. There are asset managers, individual traders, stock managers who add cryptocurrencies to their portfolios. Hence, we propose a cryptocurrency portfolio recommender system using association rule mining (ARM) which analyses the cryptocurrency dataset with the suggestion of cryptocurrencies ranked in a basket. The main objective of this research is to help the traders and the managers involved in crypto market in making decisions to invest in group of cryptocurrencies when maximum profit transaction evidence is available for investment.

Our recommender system we propose is not the same as other systems as it determines the correlation between the cryptocurrencies and recommends a portfolio. Existing research in cryptocurrency mostly concentrates on portfolio management, portfolio optimization, Cryptocurrency price prediction, Crypto price trend forecasting etc. We have used the association rule technique which was never used in cryptocurrency portfolio recommender system creation. It's not always feasible to use the traditional ARM technique as the rules created will be exponential and getting the most relevant rule will be difficult to choose.

We have done a thorough research on our cryptocurrency historical dataset that we got from Kaggle. We have compared our portfolio we created with the ranking of the crypto currency available in the actual crypto market data and the CRIX (Crypto Index) values. The result of our research shows the top 10 cryptocurrency portfolios for short selling(Intraday trading) purpose which has Return of interest of more than 99 percent. .

Keywords - Block-chain, Crypto currency, Association rule mining (ARM), Recommendation system, Apriori algorithm, Intraday trading, KDD

1 Introduction

We propose a cryptocurrency portfolio recommender system using association rule mining (ARM) which analyses the cryptocurrency dataset with the suggestion of cryptocurrencies ranked in a basket. The main objective of this research is to help the traders and the managers involved in crypto market in making decisions to invest in group of cryptocurrencies when maximum profit transaction evidence is available for investment.

Our recommender system we propose is not the same as other systems as it determines the correlation between the cryptocurrencies and recommends a portfolio. Existing research in cryptocurrency mostly concentrates on portfolio management, portfolio optimization, Cryptocurrency price prediction, Crypto price trend forecasting etc. We have used the association rule mining (ARM) technique which was never been used in cryptocurrency portfolio recommender system creation. Association rule mining is one of the traditional techniques used for creating recommendation systems which is also called as market basket analysis as it is commonly used in retail or e-commerce business as customers mostly tend to get a product along with another main product. It's not always feasible to use the traditional ARM technique as the rules created will be exponential and getting the most relevant rule will be difficult to choose.

In this study, we first offer a technique for portfolio building based on historical cryptocurrency price data that is based on standard ARM. If we notice that cryptocurrencies are not performing as expected, we present a technique for rebalancing the portfolio at regular intervals utilising rules generated using ARM. This strategy produced outcomes with crypto portfolios that are far superior with more than 99% of return of interest (ROI) in intraday transaction.

We have discussed the methodology and the associated concepts with a running example on the cryptocurrency dataset downloaded from Kaggle throughout the paper. There are 2005 cryptocurrencies involved here in this research across different years. This consists of cryptos with high market cap value to low market cap value with intraday transactions and transactions across different years with attributes like opening price, closing price etc. along with the coin rank. In most cases, our research highlights the use of soft computing techniques of ARM in the creation of an effective recommender system.

The following is a breakdown of the paper's structure. We discuss about the related work associated with our paper, followed by the methodology section which has KDD described and the background of ARM is discussed followed by the implementation and flow including a brief description of how the support confidence framework is used to generate frequent item sets and association rules. The final two sections are the evaluation and result analysis which beings this paper to conclusion.

1.1 Motivation

In today's crypto market, there are over 6700 different cryptocurrencies that are publicly traded all over the world. According to CoinMarketCap, the overall cryptocurrency market capitalization on April 13, 2021 was roughly with 2.2 billion dollars, Bitcoin, the most prominent cryptocurrency, valued at more than 1.2 billion dollars.

Bitcoin was the first cryptocurrency to be released, and it remains the most popular due to the massive amount of data recorded in blockchain and its market capital. Because of the rise in bitcoin investors in recent years, as well as the popularity of other cryptocurrencies and the fact that Bitcoin has been widely utilized as a medium of exchange and in many regions of the world it used is an online money.

Existing research in cryptocurrency mostly concentrates on portfolio management, portfolio optimization, Cryptocurrency price prediction, Crypto price trend forecasting etc. We have used the association rule mining (ARM) technique which was never been used in cryptocurrency portfolio recommender system creation.

These are the driving forces behind our research into the cryptocurrency market and the development of a recommendation system that uses association rule mining to recommend cryptocurrency portfolios, allowing investors to profitably invest in cryptocurrencies. Fayyad et al. (1996).

1.2 Research Question

How feasible is to develop a recommendation system to recommend cryptocurrency portfolio using association rule mining?

1.3 Justification

We reviewed a number of papers that had useful information regarding developing a bitcoin recommendation engine. The employment of fuzzy techniques and public perception-based systems was successful.

Association rule mining, also known as market basket analysis, is a conventional technique for developing recommendation systems that is typically employed in retail or e-commerce businesses since customers frequently purchase a product in conjunction with another key product.

We create an Association rule mining (ARM)-based cryptocurrency portfolio recommendation system that analyzes crypto data to select a group of cryptocurrencies. In comparison to other existing methods, our recommendation system will detect the correlation between various cryptocurrencies and will recommend a cryptocurrency portfolio. Existing approaches suggest buying or selling bitcoins, but not building a portfolio.In terms of generating Association rules, we actually apply the support confidence framework relationship. The process of association rule mining was always thought to be impractical since the number of association rules generated would be exponential, making it difficult to choose the best rule.

However, we will use this method in this reasearch to see if it is possible to build a cryptocurrency portfolio recommendation system utilizing association rule mining, as this has never been done before.

2 Literature Review

There are a variety of recommendation systems that can be used to re-commend a product or several products to users across various domains. Similarity measurements, collaborative filtering, new Fuzzy approach, as-sociation rule, Machine learning approaches, and other techniques were employed. Because the focus of our research is on applying association rules to develop a recommendation system for cryptocurrency portfolios, we will focus mostly on research papers that have successfully tested their research using ARM and articles that involve the creation of cryptocurrency portfolios. Some of the publications we'll be using as research references are listed below.

2.1 Research works related to Cryptocurrency portfolio

According to Jiang and Liang (2017), managing a portfolio is nothing but the process of decision making to allocate funds to invest in various financial products. A convolutional neural network is used in this paper by the author on a dataset with historic prices of cryptocurrencies and gets the desired portfolio as the output. The training of data is done with 0.7 years' price data in a reinforcement manner and getting the maximum return which is the considered as the reward function. Author also does another experiment called as back test trading experiment by for 30 minutes in the same market, which leads to achievement of returns worth 10 times in 1.8 month's period. The author performs other strategies to perform the back tests and compares it with neural network. He states that this method can be implemented in any financial markets.

In the recent years, cryptocurrency usage has been widely spread along with the financial automated consultancy. This consultancy doesn't exploit the potentiality of this huge market which has great future. Because of this reason, Giudici et al. (2020)proposed an approach to create an efficient allocation strategy for portfolio which involves cryptocurrencies. This method is the extension of Markowitz model combined with Random matrix theory and network measures. With this method the portfolio weights are achieved to enhance their risk-return profiles. The author proved that this method outperforms other existing methods with relatively low level of risk.

Over the past few years, there has been a significant increase in the modern technology of blockchain and cryptocurrencies. In addition, a few of the early investors of this type of distinct digital evolution have made drastic profits. The tools such as Heuristic algorithm and differential evolution have been exponentially used in the portfolio optimization. The author Mba et al. (2018)states that they have formulated two new approaches from this differential evolution method: the GARCH-differential evolution (GARCH-DE) and the GARCH-differential evolution t-copula (GARCH-DE-t-copula). The author has differentiated the two models with DE (benchmark) as a sole and multi-period optimisations on the portfolio which consists of the five cryptoassets. These assets comes under the CvaR constraint, which is considered as risk measures. This study has proved that the GARCH-DE-t-copula outruns the DE and GARCH-DE techniques by all means. For these notably elusive assets, the GARCH-DE-t-copula has demonstrated the risk-control ability, hence it proves the ability of t-copula to apprehend the dependence structure in fat-tailed distributions.

The investor must be able to contemplate the effectiveness of the income ratio of the portfolio asset orderly to notify and filter the risk of the investment. This research paper Čuljak et al. (2020) enables us to identify and elaborate the advantages of the classification of sectoral cryptocurrency portfolio optimization and its execution. It is structured into six optimization targets such as MinVar, MinCVaR, MaxSR, MaxSTARR, MaxUT and MaxMean. Over the same duration, the established portfolio is then differentiated with the standards of the CRIX index. The outcome proves that, out of six, five of the portfolio strategies has better performance if the cryptocurrencies from financial, exchange and business services sectors are being included

For the investors, estimating the cryptocurrency prices is vital. The author Sun et al. (2020)uses Gradient Boosting Decision Tree (GBDT) algorithm, Light Gradient Boosting Machine (LightGBM), to show the price trend in cryptocurrency market. To use the market details, the author combines the daily data of 42 types of primary cryptocurrencies with the lead economic indicators. It confirms that the robustness of the LightGBM model is beneficial than the other methods. In addition, the comprehensive stability of the cryptocurrencies affects the performance of this forecast. With this, the investors can develop a relevant cryptocurrency portfolio and alleviate the risks.

In this study, Uhliarik (n.d.) the author considers the sustainability of portfolios with the cryptocurrencies and predicts whether adding them in a portfolio can improve the performance. For this purpose, the author builds up two cryptocurrency indices, one with unrestricted weights and the other with maximum weight of 30 percent for any single constituent for the duration from January 2017 to December 2019. Eventually, the author adds these indices in a portfolio, which is then enhanced by the usage of three various techniques, and tests their out-of-sample performance. However, when considering the transaction costs and other such factors, the real-life investors are not benefited from investing into cryptocurrencies. This reasoning is aided by the results of the other two optimization methods as parametric and Black-Litterman.

There has been creative and developed technologies in the Forth industrial revolution that are now inspiring the traditional economies. The expert investors and individuals are being attracted by the ideas and the technological financial tools, to dig into a broader investment spectrum and to develop their portfolios. For a long term practicality, it is essential to go beyond the conventional portfolios and working on the state-of-the-art strategies that follows the technological advancements. Hence, to consider the needs of all societal segments, there was a requirement to expose technological and digital financial alteration. This Ma et al. (2020) study also affects the variegation with the five of cryptocurrencies from November 2015 to November 2019 on four of the traditional asset portfolios. This denotes that when short sales are being used, the results may have an improvement. However, adding multiple cryptocurrencies in a portfolio contributes to better results for diversification, and Ethereum has increased diversification opportunity when compared to Bitcoin.

Financial portfolio management is the way of intermittently reassigning a fund into various financial investment productions, with the intention to maximise the profits. On one hand, the financial machine learning methods are used to assume the price trends, whereas, the learning based portfolios management are useful to make decisions depending upon the direct changes in the price. Eventhen, the reinforcement learning methods are not adequate as they are restricted in extracting the information of changes in price at a single-scale level. However, the author Shi et al. (2019) has initiated a novel Ensemble of Identical Independent Inception (EI3) convolutional neural network, with the purpose to convey the limitation of existing reinforcement learning based portfolio management methods. Using EI3, even while sharing the similar network parameters, many assets are processed independently. In addition, even the price movement details for each product are collected at multiple scales with wide network and used to make trading decisions. Regarding the portfolio management problem, the author uses EI3 to bring forward a recurrent reinforcement learning framework, to produce a deep machine learning solution. Both in hiking and declining environments, a broad experiments on the cryptocurrency datasets shows the excellence of the author's method over existing competitors.

2.2 Research works related to ARM technique

In this research Amaral et al. (2019) is the application of artificial intelligence algorithms was used to estimate the value of cryptocurrencies based on previous data. After the value has been anticipated, two methods for investment recommendation have been introduced. The first one is utilized to create an investment recommendation based on the anticipated result. The second method employs the Mamdani system, which suggests what to do with the money invested. Three cryptocurrencies are employed here, and the results of both algorithms were compared to the actual quote after historical data was analyzed.Both methodologies were examined based on the results, which show that all three cryptocurrencies have a total assertiveness rate of above 90%. Based on the practical experiments conducted, this strategy appears to be promising and competitive with other alternative ways mentioned in the literature.

Bibi et al. (2019), has developed a method for identifying the top places in the globe where cryptocurrencies are extensively used. Opinion mining is conducted out in various regions to determine public sentiment in bitcoin. Above that, topic modeling was utilized to determine the semantics of users' interests. The findings were primarily utilized to assist business specialists in learning more about public demand and making the best cryptocurrency investment. Data acquisition and pre-processing are the approaches employed in this literature along with Recommendation system (Sentiment analysis, Location retrieval / Cryptocurrency world identification, Users concerns and interests detection, Analysis Visualization)

Tewari and Barman (2016) has suggested a book recommendation system that creates recommendations based on the participation of the target users' trusted friends and association rule mining, which aids in determining the current reading trends of distinct Internet users. The trust system is crucial, and it allows users to designate different levels of confidence to their friends. When it comes to making recommendations, the level of trust that one user has in another is crucial. With the support of his trust network buddies, TBRS is used to deliver book recommendations to the target user. When it comes to e-commerce advice, this is one of the most trustworthy approaches.

Moonen et al. (2016) has looked into a few different aspects that influence change recommendation. The researcher tested the change histories of eight large open-source systems and two large industrial systems in a number of ways. The researcher has succeeded in controlling the size of the change set from which a suggestion will be formed, measuring evolutionary coupling strength, and accounting for massive historical changes while inferring these couplings. The findings of this study were utilized to develop a number of practical guidelines for using ARM to recommend changes.

Jooa et al. (2016) offered a recommendation system that was built and executed to analyze customers' usage patterns and personal tendencies utilizing association rule analysis and collaborative filtering with data acquired from customer companies using NFC (Near Field Communication). This recommendation algorithm, which was preferably employed in the suggested system, used the findings for data analysis and distance information obtained via GPS (Global Positioning System), which will be useful in recommending local companies to those who are likely to visit.

Alsalama (2013) has suggested a project in which association rule mining and a content-based method were coupled to create a hybrid recommendation system framework. In this study, this is done in two-dimensional spaces. Movie Lens datasets were used by the researcher. Additionally, WEKA software was utilized to construct association rules and perform data mining tasks in order to implement and evaluate the framework.

A book suggestion system is provided here in this research Jomsri (2014) that will be useful for individuals based on their needs. The researcher employed the association rule technique to discover the link between the books that customers are interested in and their availability in the system database by category. The researcher's purpose was to assist clients in their book searches and provide better search results. In the end, the expense of preserving the books was decreased as well.

A new approach of recommending books to consumers is presented in this study Tewari et al. (2014). To create the most effective and efficient book recommendation system, the researcher combined some of the features of content filtering, collaborative filtering, and association rule mining. Because the recommendation system is one of the most powerful instruments for retaining consumers and increasing profits in a business, the researcher found that this strategy is extremely competitive in providing good outcomes when compared to other methods. Unlike other content-based filtering systems that cannot distinguish between excellent and bad articles, this content-based filtering recommendation system can filter books with high quality material.

Paranjape-Voditel and Deshpande (2013) has presented an association rule-based recommender system for proposing a stock market portfolio to investors (ARM). The researcher's major goal was to assist traders, investors, and stock market managers in making selections about which stock portfolios to invest in based on confirmed transactional data. This system is distinct from other existing systems in that it determines the connection between equities and recommends a portfolio to investors or stock market managers. The existing other approaches suggest buying a single stock based on price volume trends, but they do not suggest building a portfolio.The researcher has undertaken a new attempt to explore if an association rule recommender system can be used to generate a stock portfolio.

Here Cakir and Aras (2012) offered a recommendation engine for the creation of a personalized e-commerce website For customisation, the researcher employed collaborative filtering and the association rule technique. The software was written in c Sharp, and the association rule was created using the apriori algorithm. The testing phase is evaluated using accuracy and coverage, while the deployment phase is evaluated using the basket ratio. As a result of the outcomes, this suggestion method raises the basket ration.

Chen et al. (2014) has contributed to this research article by attempting to use machine learning techniques to cryptocurrency. The purpose was to use machine learning models and methodologies to anticipate and forecast the closing prices of cryptocurrencies using index 30 and nine other members, allowing investors to trade with ease. To acquire the best results, the author applied a variety of machine learning approaches and compared them to each other.

This research Guo et al. (2017b) intends to make e-commerce buying more convenient for customers by avoiding information overload through the use of an interactive personalized recommendation system based on a hybrid algorithm model. This proposed methodology begins by acquiring a list of original suggestion results using various recommendation algorithms. The researcher was able to determine the weights of recommendation algorithms in this way. This interactive assigned weight technique greatly improves consumer demand while also resolving the problem of information overload. Simultaneously, this research has significant ramifications for e-commerce business platforms that provide product recommendations.

The goal is to make e-commerce purchasing more convenient for customers by avoiding information overload with the use of a hybrid algorithm model and an improved apriori algorithm. The researcher Guo et al. (2017a) has increased the effectiveness of data mining in order to make the recommended products more valuable to customers. The modified apriori algorithm is used in the mobile e-commerce recommendation system to tackle problems with the visual interface in mobile terminals.

Bhosale et al. (2016) has proposed a hybrid recommendation system in which the researcher takes into account customer comments while recommending autos on the market. In order to do car categorization, which will offer a rating to a car, technical words relating to cars are considered in the dataset. Mining association rules is also utilized to locate an automobile that is in high demand on the market. This aids in the delivery of faster results for a large dataset. All of the above-mentioned papers were thoroughly reviewed and checked, and they will be used as references in our research.

One of the most important areas in recommendation systems is association rule mining. Because it is frequently utilized in retail and e-commerce, this approach is also known as market basket analysis. The Association rule mining technique is one of the main subjects of this research.

All of the above-mentioned papers were thoroughly reviewed and checked, and they were used as references in our research.

3 Research Methodology

In the preceding sections, we examined the fundamentals of cryptocurrencies and the recommendation system that we will construct. In the interest of the subject, we'll go over how the implementation is carried out in great detail. We'll take a closer look at the issues listed below.

- Approach
- Model Implementation, Design configuration and Research Flow
- Evaluation metrics

3.1 Approach

3.1.1 Data Preparation KDD

Data mining is an unavoidable technique for this study because it is essential for any data-driven study. The amount of data we manage in any business transaction, including voice, video, and photos is expanding on a daily basis. Fayyad and his colleagues (1996) As a result, we'll need a regulated, methodical process that can extract pertinent data and provide automated reports that can be used to improve decision-making in any process. Here, we'll talk briefly about Knowledge Discovery in Databases (KDD), which will be our recommended method for this studyFayyad et al. (1996).

Data mining: This is nothing more than database knowledge discovery, which refers to the extraction of previously insignificant implicit data as well as information that may be beneficial in the future. Data mining entails a number of steps, which we'll go over in detail below.

- Target dataset
- Cleaning process
- Reduction and Integration
- Transformation
- Data Mining
- Evaluation of patterns
- Using Observed Knowledge

3.1.2 Target dataset

During this procedure, we have choosen a dataset which has cryptocurrency historical prices and concentrate on the subset of variables that are relevant to our research.

3.1.3 Cleaning process

We have removed all the undesirable and noisy data from our dataset using this strategy. If any missing values are discovered, they will be carefully cleaned until the essential data is obtained.

3.1.4 Reduction and Integration

Based on the purpose of this research, we have identified all of the useful properties of our data, and we have utilized the neccessary variables from our dataset to a manageable level.

3.1.5 Transformation

In this step, the data is turned into the required format that the data mining algorithm will require.We have created another dataset called as apriori which is uilised for our research where the apriroi algorithm is applied.

3.1.6 Data Mining

In this stage, we have implemented a technique that's carried out based on our study objectives and it's utilized to extract possibly useful patterns. In this stage, we'll use the association rule mining (ARM) approach.In the below figure 1 we can see the architecture of the KDD system.

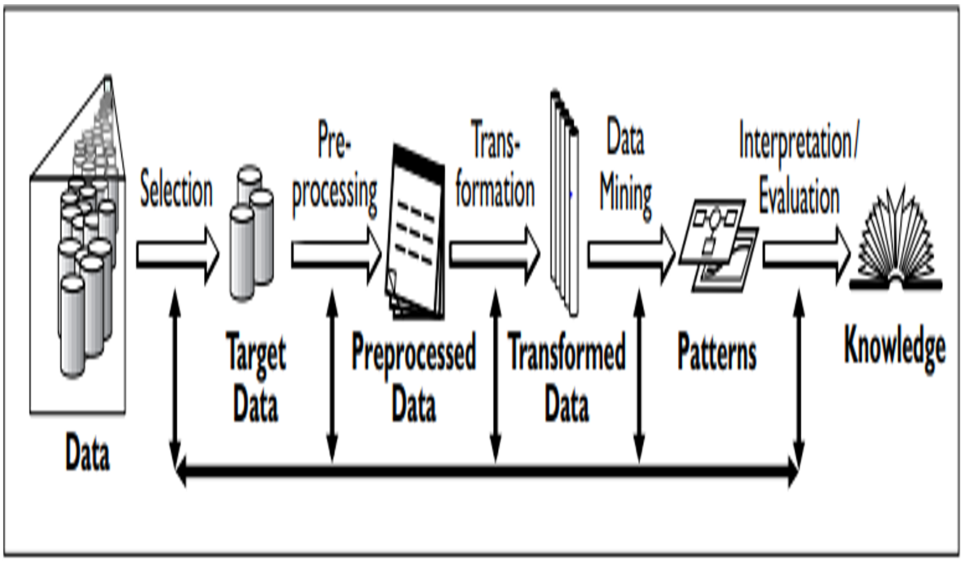

Figure 1: KDD process steps

3.2 Model Implementation, Design configuration and Research flow

In this section, we'll go over the association rule and the Apriori algorithm, as well as how they'll be used in our research. For years, ARM has been employed with tremendous success in fields such as e-commerce, retail, and so on. Using typical ARM to create a recommendation system to promote cryptocurrency portfolio is not easy as more rules will be created. In this next sections, we'll go over the procedure in detail.

3.2.1 Apriori Algorithm

ARM analysis is a popular technique for determining the relationship between two or more objects. In our research, we are using the Apriori algorithm to find the most common sectors. In most cases, the Apriori algorithm will employ three measures.
-Support
-Confidence
-Lift

- **SUPPORT**

This indicator of an item's popularity is based on the proportion of transactions in which the item occurs. If you discover that item sales have exceeded a certain threshold, it will have a substantial influence on your profitability. You might use that percentage as your support threshold. The itemset will then be classified as unique if the support values exceed this threshold. For example, item a receives 4 out of 8 votes, or 50% support.

Itemsets can also contain numerous items. The support for a,b,c, for example, is 2 out of 8, or 25%.

$$Supporta = 4/8$$

- **CONFIDENCE**

This metric indicates how probable it is that the item A will be purchased when the primary item B is. This is written as (a-b). This is determined by the transaction proportion of an item in which item B also appears. One significant disadvantage of this strategy is that it has the potential to misrepresent the significance of an association. This occurs because it only considers the importance of item a and ignores item b. The third strategy, known as lift, takes into account the base popularity of both goods.

$$Confidencea- > b = supporta, b/supporta$$

- **LIFT**

By regulating the popularity of b, this metric reveals the likelihood of how item b will be purchased when item a is purchased. If the lift value is more than 1, it signifies that b will be purchased if item a is purchased, however if it is zero, it means that b will not be purchased if item a is purchased. Since our method is based on support confidence framework we aren't using this in our research.

$$Lifta- > b = Supporta, b/supporta * supportb$$

Below mentioned flow diagram figure 2. shows the work of apriori algorithm.

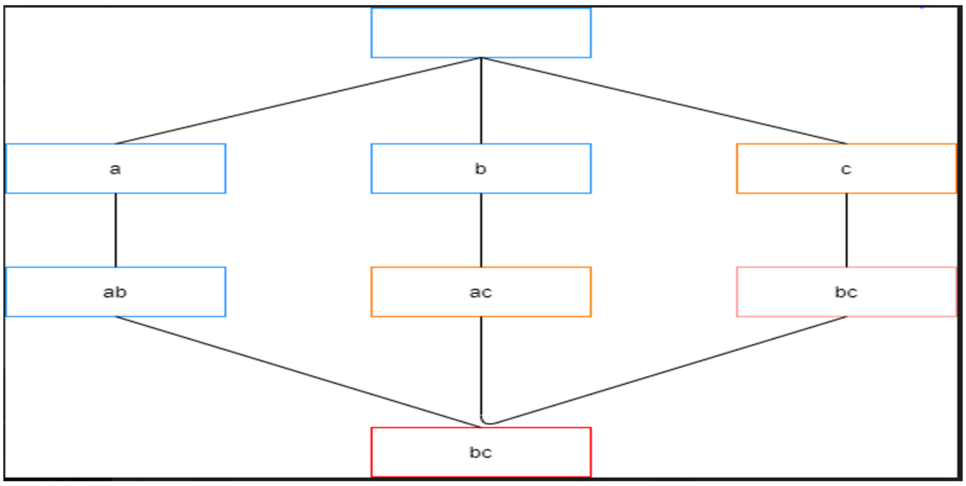

Figure 2: Apriori process flow

3.2.2 ARM Background

Let T = [t1,t2,...,tn] and F = [f1,f1,...,fm] represent the transaction set and the database, respectively, in this problem. Each transaction in f will have a unique transaction ID, as well as a subset of the items in t. A rule is created as $\mathtt{A} \Rightarrow \mathtt{B}$ where A, $\mathtt{B} \subset \mathtt{T}$ and $\mathtt{A}$

$\cap$ B =, with A as the antecedent and B as the consequent. In general, association rule mining mines transactions from various transactional data-bases. Different measurements are used by ARM, but we employ the support-confidence framework measure. First, we mine the databases, which necessitates the use of an efficient algorithm because producing frequent itemsets is time-consuming.

In general, association rule mining mines transactions from various transactional databases. Different measurements are used by ARM, but we employ the support-confidence framework measure. In this measure, we first mine the databases, which necessitates the use of an efficient method because generating frequent itemsets is time-consuming in the mining process. Once the frequent itemsets are detected from transactions in the database F using Apriori method algorithm, it will be feasible to create strong association rules.

$$confidence(X \to Y) = P(X|Y) = supportcount(X \cup Y)$$

The conditional probability is the itemset support count, and the support (XUB) is the transaction numbers that have the itemset XUB. whereas the support count X is determined by the trans-action numbers for the itemset X. The following association rules are formed based on this equation: 1. For each frequent itemset n present, n will be generated for nonempty subsets; 2. For every nonempty subsets of d, the output rule is d $\to$ (n { d)

```
if support count(n) / support count(d)  min con
```

where min con is the minimum confidence threshold. As the frequent itemset rules are generated, the minimal support is immediately satisfied. In this scenario, if we assume that a rule's confidence level is 50%, then transactions containing 50% of X and Y are correct. This confidence metric, which serves as a parameter, is used to rank the rules in distinct datasets. As previously said, there are a variety of alternative measurements available, but this is the most appropriate for our system.

3.2.3 Implementation and flow

Our transaction datasets, which will be used for Associate rule mining, will be prepared in a specific way. We'll take into account the closing prices of each cryptocurrency, and each transaction will include all of them. Their closing values are taken into account, and the increase or decrease in the cryptocurrency price on that particular day is expressed as a percentage of the previous day's closing price. The x percent threshold will be determined, and it will be depending on the type of dataset we receive as well as the volatility in the crypto market. This value will be set to .001% for our cryptocurrency dataset because each transaction indicates the number of cryptocurrencies that have changed by x percent or more on that day as we are only considering intra day trading. The number of days that the cryptocurrencies are watched is compared to the number of transactions that have taken place. Because the data will be correlated with other crypto coins, a new dataset will be craeted for applying the apriori rule. On these datasets that were created, ARM wll be applied to obtain the frequent item set rules.

The portfolio is then created for a lock-in term, which is the shortest duration possible. During this time, investors are unable to sell or liquidate their crypto portfolios. For our crypto dataset, the minimal lock-in period is one day. The rules are then developed using the support confidence framework to mine this dataset. These rules are ranked, and the order is determined by confidence. The following procedure is used to create the portfolio::

a) To locate the cryptos that are frequent on our dataset, we use the Apriori algorithm. In our scenario, the dataset is a historical dataset so we use the timeseries function to fetch the cryptos from different years to get the intra day transaction for each crypto and generate a separate dataset. On this dataset, apriori algorithm is used. For these frequently used cryptos, rules are generated. Only the cryptos that are positively connected were considered. Then we take the antecedent and consequent of these rules, which are known as top-k rules. We use the value of k as 2,4 for our dataset, which is dependent on the number of antecedent and consequent cryptcurrencies in the rule. Negatively correlated cryptos are also highlighted, as they will be utilized to re-balance the portfolio if needed later.

b) We'll run the apriori algorithm on the top-k cryptos on every single cryptocurrency once again. We'll be able to generate the most common cryptos in each industry this way. These common cryptos will then be used to calculate the association rules. Following that, the top-k rules will be applied to each . For our dataset, the value of k is taken as 2,4. We'll look at the cryptos with positive correlations again, and we'll keep track of the cryptos with negative correlations so we can utilize them for rebalancing in the future when needed.

c) All of the items in the datasets are brought into play. This stage determines the relative order of the cryptos, which are formed after the rules have been applied to the frequently occurring itemsets. This is critical since the investment will be done in the sequence in which cryptos should be purchased. Assume the rules are constructed in the following order: c1, c2, c3, . We use the apriori technique on these three to discover the most common cryptos. As a result, a rule with similar certainty will emerge. We'd have to go back to the produced rules in the entire database if we needed to figure out the relative ordering for the cryptos in this dataset. If the investment is done with a limited amount, not all cryptos will be purchased. Once the top cryptocurrencies have been invested in, and the cash has been depleted, the buying will be halted. This is an important feature in this case.

d) The amount that will be invested in the portfolio will be determined once the top-k cryptos are generated. If we have a sufficient investment amount, a minimum number of cryptos will be purchased. Otherwise, only the bare minimum of cryptos will be purchased until the supply is depleted. If we have more money, on the other hand, we will increase our investment. Step 3 is critical since the investment is made dependent on the sequence in which the cryptos are present. If the investment is unrestricted, this order won't matter.

This process's general flow diagram is shown below in the figure 3. We did not include the KDD process in the figure below because it was previously discussed before.

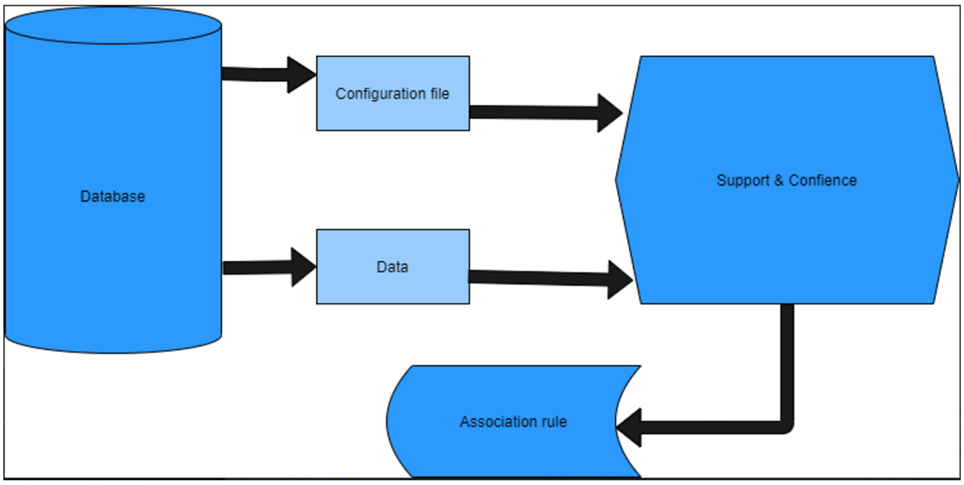

Figure 3: Design and flow diagram of Associate rule mining process

4 Evaluation and Result analysis of Portfolio formation

In any study, the evaluation is crucial since it demonstrates how successful our research model is. So, once the portfolio has been formed, we can begin assessing the timeframe in which an investment can be made. As a variable, we use the investment time. We in this ersearch do a short term investment for one day. If rebalancing is truly necessary, it can be done but in our case it's not necessary as we have creating a portfolio for intra day trading. The following methods will be used to evaluate and rebalance this approach if needed:

a) Following the investment, we must monitor the portfolio to see if it has increased or decreased.
b) We must keep track of the growth and fall of all cryptos.
c) We must also examine the coins that are underperforming..
d) We need to come up with new rules if the prices drop and the dropping cryptocurrencies that are adversely connected across other cryptocurrencies
e) Returns must be determined once the investing period has come to a conclusion. This is accomplished by determining the Return on Investment (ROI).
f) The formula is:

$$ROI = (investmentvalueafterlockin - periodinvestmentcost)/investmentofcost$$

Two new criteria are included here to assess performance in terms of the quality of crypto portfolio suggestion on a profit basis. Precision and adjusting precision are what they are.

$$Precision = correctlyrecommendedcryptos/totalrecommendedcryptos$$

$$Rebalancingprecision = correctlyrecommendedcryptosafterrebalancing/totalcryptos$$

These formulas will be utilised only if the investments are made for longrterm. For our research this is not required as this deals with intra day trading.

Correctly advised cryptos are those that have increased in value without rebalancing during the investment time. Here the percentage value of the connection between support and confidence metrics plays an significant role in evaluating the rules generated by the ARM process. This reccommends higher quality of cryptocurrencies in portfolios. You can check the live ranking of the cryptocurrencies and the intraday market cap, highs and lows in prices of every cryptocurrency from coinmarketcap website [1] to cross validate the cryptocurrencies portfolios. As we have done this research on Intraday trading, buying or selling top ranked cryptocurrencies like Bitcoin or Ehtereum or ETC doesn't guarantee much of ROI. Below are the top 10 cryptocurrency portfolio generated by us from our dataset. The order of investment starts from right to left which is important in the portfolios we have generated.

RANK	CRYPTOCURRENCY PORTFOLIO	Support-Confidence % value(ROI)
1	[{'TTC', 'XRA'}, {'XAUR'}]	0.999180999
2	[{'XRA', 'CSC'}, {'XAUR'}]	0.999154691
3	[{'ARB'}, {'XAUR'}]	0.999113475
4	[{'TTC', 'ARB'}, {'XAUR'}]	0.999101527
5	[{'XRA', 'ARB'}, {'XAUR'}]	0.99908341
6	[{'USDT', 'XRA'}, {'XAUR'}	0.99908341
7	[{'PIGGY', 'XRA'}, {'XAUR'}]	0.999080882
8	[{'ARB', 'CSC'}, {'XAUR'}]	0.999074931
9	[{'COVAL'}, {'TTC'}]	0.990081154
10	[[{'CSC', 'COVAL'}, {'TTC'}]	0.989729225

Figure 4: Top 10 crypto portfolios

In the above figure 4 the symbols represent each cryptocurrency and we can see the top 10 cryptocurrency portfolios generated with antecedent and consequent. We have set minimum support as 50% and minimum confidence as 60% in our case. The support of a rule actually shows the frequency of the items occurring together in the rules. For example in the first row of the table TTC,XRA(Antecedent) and XAUR(Consequent) may appear together in 50% of the transactions. Similarly confidence of a rule actually shows the antecedent and consequent probability that may appear in the same transaction. For example in the first row of the table TTC,XRA might appear in 60% transactions but 50% of them may also have XAUR. So the rule confidence would be more than 80%. So in this way, with the help of the rules set by the support confidence framework approach our evaluation result shows that strong rules have been formed with the performance of support-confidence interval value whose percentage is more than 99% for the top 10 crypto portfolios which acts as the return of interest(ROI) for intraday trading in our research, also shows the performance of the model. So our evaluation is done in this way and it shows that the model performance is meeting the expectations.

5 Conclusion and Future work

In this study, we described how to design a cryptocurrency portfolio recommender system using association rule mining and a support confidence framework. Our research

[1] Reference URL : `https://coinmarketcap.com/`

demonstrates how soft computing techniques such as ARM can be used to construct an effective recommender system. The outcomes of this strategy on our historical cryptocurrency dataset acquired from the Kaggle website[2] which is a public resource have been exceedingly good, producing good returns achieving support confidence interval of 99% which acts as ROI for intra day trading. This shows that the proposed method is more than enough to be used in any dataset with similar charactericstics.

Using the same technique or any other method, we can extend this research of cryptocurrency portfolio recommender system we presented for long term trading for an year or more. It should be investigated whether various approach factors such as support, confidence, the threshold to include cryptos in dataset can be determined at runtime using the features available in the dataset.

References

Alsalama, A. (2013). A hybrid recommendation system based on association rules.

Amaral, V. L., Affonso, E. T. F., Silva, A. M., Moita, G. F. and Almeida, P. E. M. (2019). New fuzzy approaches to cryptocurrencies investment recommendation systems, *International Fuzzy Systems Association World Congress*, Springer, pp. 135–147.

Bhosale, M., Ghorpade, T. and Shedge, R. (2016). On demand recommendation using association rule mining approach, *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, IEEE, pp. 1302–1306.

Bibi, S., Hussain, S. and Faisal, M. I. (2019). Public perception based recommendation system for cryptocurrency, *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, IEEE, pp. 661–665.

Cakir, O. and Aras, M. E. (2012). A recommendation engine by using association rules, *Procedia-Social and Behavioral Sciences* **62**: 452–456.

Chen, J., Miller, C. and Dagher, G. G. (2014). Product recommendation system for small online retailers using association rules mining, *Proceedings of the 2014 International Conference on Innovative Design and Manufacturing (ICIDM)*, IEEE, pp. 71–77.

Čuljak, M., Tomić, B. and Žikovic, S. (2020). Benefits of sectoral cryptocurrency portfolio optimization, *Available at SSRN 3661600* .

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). The kdd process for extracting useful knowledge from volumes of data, *Communications of the ACM* **39**(11): 27–34.

Giudici, P., Pagnottoni, P. and Polinesi, G. (2020). Network models to enhance automated cryptocurrency portfolio management, *Frontiers in artificial intelligence* **3**: 22.

Guo, Y., Wang, M. and Li, X. (2017a). Application of an improved apriori algorithm in a mobile e-commerce recommendation system, *Industrial Management & Data Systems* .

Guo, Y., Wang, M. and Li, X. (2017b). An interactive personalized recommendation system using the hybrid algorithm model, *Symmetry* **9**(10): 216.

[2]Reference URL : `https://www.kaggle.com/jessevent/all-crypto-currencies`

Jiang, Z. and Liang, J. (2017). Cryptocurrency portfolio management with deep reinforcement learning, *2017 Intelligent Systems Conference (IntelliSys)*, IEEE, pp. 905–913.

Jomsri, P. (2014). Book recommendation system for digital library based on user profiles by using association rule, *Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014)*, IEEE, pp. 130–134.

Jooa, J., Bangb, S. and Parka, G. (2016). Implementation of a recommendation system using association rules and collaborative filtering, *Procedia Computer Science* **91**: 944–952.

Ma, Y., Ahmad, F., Liu, M. and Wang, Z. (2020). Portfolio optimization in the era of digital financialization using cryptocurrencies, *Technological forecasting and social change* **161**: 120265.

Mba, J. C., Pindza, E. and Koumba, U. (2018). A differential evolution copula-based approach for a multi-period cryptocurrency portfolio optimization, *Financial Markets and Portfolio Management* **32**(4): 399–418.

Moonen, L., Di Alesio, S., Binkley, D. and Rolfsnes, T. (2016). Practical guidelines for change recommendation using association rule mining, *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering*, pp. 732–743.

Paranjape-Voditel, P. and Deshpande, U. (2013). A stock market portfolio recommender system based on association rule mining, *Applied Soft Computing* **13**(2): 1055–1063.

Shi, S., Li, J., Li, G. and Pan, P. (2019). A multi-scale temporal feature aggregation convolutional neural network for portfolio management, *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1613–1622.

Sun, X., Liu, M. and Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on lightgbm, *Finance Research Letters* **32**: 101084.

Tewari, A. S. and Barman, A. G. (2016). Collaborative book recommendation system using trust based social network and association rule mining, *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, pp. 85–88.

Tewari, A. S., Kumar, A. and Barman, A. G. (2014). Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining, *2014 IEEE International Advance Computing Conference (IACC)*, IEEE, pp. 500–503.

Uhliarik, A. (n.d.). Evaluation of cryptocurrency portfolios.