Association Rule Mining on the BIST100 Stock Exchange

Erkut Memiş
Computer Engineering
Ankara Yıldırım Beyazıt University
Ankara, Turkey
erkutmemis@gmail.com

Hilal Kaya
Computer Engineering
Ankara Yıldırım Beyazıt University
Ankara, Turkey
hilalkaya@ybu.edu.tr

***Abstract*— Data mining is the process of finding valuable information in data. Association mining is one of the fields of the data mining. It is used to discover all associations between items in a dataset. It also discovers hidden patterns in the large dataset. Patterns are discovered using predefined constraints (such as minimum support and minimum confidence values).**

Data mining on stock market is an important research area. This research field and its results are valuable information for all of the investors. There are many studies on this field in the world. But to the best of our knowledge, there is not any study on BIST100 stock exchange that finds associations between stocks. In this study, we focused on BIST100 stock exchange, used 87 different BIST100 stocks daily data for dates between 10/21/2013 and 10/19/2018 from "finance.yahoo.com". We figured out association rules on BIST100 stock exchange using historical data, Apriori algorithm and two different condensation methods.

Keywords*— *BIST100, Stock, Apriori, Association Rule, Finance

I. Introduction

Data mining on stock market data is one of the attracted research subject due to its financial benefits. Finding hidden patterns and associations in the stock exchange data can help us to understand market behaviors and movements of stocks. In the past, generally two analysis methods (technical analysis and fundamental analysis) were used to predict movements in stock market, but associations between all companies in the stock exchange were considered in a limited extend. Technical analysis is based on historical stock market prices and time series analysis techniques are used in this method. The second method, fundamental analysis trusts company information (company value, loans, profits, investments etc.) and economic values (interest rate, inflation, unemployment etc.). Recently, there are many studies based on association mining on stock exchange.

Association rule mining exposes hidden patterns on a dataset items by using frequently occurred items with some constraints. Association rules show regularities in a database. Association mining was first introduced by Agrawal, Imielinski and Swami [1] and it is used in different areas to find hidden patterns. Stock markets are one applied area of the association mining. And researches show that most the world's stock markets are integrated and associated each other and also companies in a stock market have relations each other. Wu and Su studied on four major international stock markets and they show that there are significant dynamic relations between of them [2]. They also find that the returns of large markets lead to the returns of small markets. Almohaimeed used an Apriori type algorithm to identify the association rules between S&P 500 and the stock prices [3].

Ariya has studied stock forecasting using association rule mining. Ariya selected eight stock groups which are financial, agro-industry, consumer product, service, properties & construction industry group, resource, technology and industry. Each group has four randomly selected sample stocks [4]. Apriori algorithm is used to discover associations between items in databases. She presented the way to prepare data for running on apriori algorithm and the result showed unknown relationship of stocks.

Kumar and Kalia studied association rule mining on the stocks dataset of thirteen years period i.e. from 1996 to December 2008 of NSE stock exchange that amounts to 3252 days was used [5]. First item sets are mined for a given minimum support and based on these item sets obtained; the association rules are computed for a given minimum confidence. The generated patterns help investors to build their portfolio and using these patterns investment strategies may be planned.

Karpio, Lukaseiewicz, Orlowski and Zabrowski studied mining associations on the Warsaw Stock Exchange [6]. Their study implements the association rules using a data mining approach to explore co-movement between stock items listed on the Warsaw Stock Exchange. To explore associations they used Apriori algorithm.

Şimşek and Özdemir have studied relation between Turkish twitter messages and stock market index. They examined a Turkish twitter data set using data mining techniques. They created emotional corpus and evaluated the happiness and unhappiness level of tweets. They show that stock market and tweet data relation is about approximately 45% [7].

Yıldırım and Yüksel studied investigation of relationship between social media and stock price daily movement direction: sentiment analysis implementation. In their study, they select a telecommunication company and whose shares are processed in Borsa Istanbul. For a selected period, daily opening and closing prices of this company have collected and daily raise and fall of the stock price has been classified accordingly. Sentiment analysis has been conducted for the same period and daily polarity values have been obtained. In their study, public sentiment analysis and stock price movement direction is conducted. They used Spearman's rank correlation test. According the test result, a negative and moderation correlation exists between daily stock price and the public sentiments in tweets [8].

Öztürk and Çiftçi have studied to predict exchange rate movements with twitter content. They used daily exchange rate of USD/TR and twitter sentiments have been collected with the keywords #USD/TR, USD/TR, Dollar, #Dollar. If exchange rate increases, its value selected as 1 for a given date, or its value selected 0. Collected tweets are sorted 3 different categories depending on their sentiments: Buy, Sell, and Neutral. They find that there is a significant relationship between Twitter sentiment and USD/TR movement [9].

Savaş and Can have studied Euro□Dollar Parity and Effects of The Real Exchange Rates on The Index of IMKB 100 [10]. They investigate the mutual effect and causality between IMKB100 index, real exchange rate and Euro/Dollar parity.

Selçuk and Kendirli have studied on the causality between the dollar exchange rate and Istanbul Stock Exchange National 30 Index (BIST-30) was investigated with "Granger Causality Test" [11].

In this study, BIST100 stock exchange was focused on, 87 different BIST100 stocks daily data was used for dates between 10/21/2013 and 10/19/2018. We used Apriori algorithm to identify association rules on BIST100 stock exchange. There are some researches like these but to the best of our knowledge, there aren't any studies on relations between BIST100 stocks.

II. MATERIALS AND METHOD

A. BIST100 Stock Market Data

In this study, the stock items listed on the BIST100 Stock Exchange were analyzed. 87 different BIST100 stock items' daily data were collected for dates between 10/21/2013 and 10/19/2018 from "finance.yahoo.com". These data includes Date, Open, High, Low, Close, Adj. Close, Volume columns. Sample stock item data can be shown in Table I.

For the purpose of analysis, "positive", "negative" and "same" tags have been defined using items' close values. Close values of items fluctuate over days. The difference between close values within the range of [-1%; 1%] are called "same".

TABLE I. SAMPLE STOCK ITEM DATA

Date	Open	High	Low	Close	Adj Close	Volume
21.10.2013	29,799999	29,9	29,200001	29,200001	25,479435	303459
22.10.2013	29,4	29,6	29,200001	29,299999	25,566694	126933
23.10.2013	29,4	29,5	29,299999	29,4	25,653952	165327
24.10.2013	29,4	29,4	28,9	29,299999	25,566694	134552
25.10.2013	29,200001	29,200001	28,9	29	25,30492	114069
28.10.2013	29,1	29,200001	28,799999	28,799999	25,130402	100740

If difference is bigger than 1% of item close value is called "positive" otherwise it is called "negative". Using these rules, we created new files that include date and related stocks names columns and value of columns includes "same", "positive" and "negative" tags. The new file includes all stock items with transformed values can be shown in Table II.

TABLE II. ALL STOCK ITEMS TRANSFORMED DATA

Date	Stock1	Stock2	Stock3	…	Stock87
21.10.2013	same	same	same	…	same
22.10.2013	negative	positive	positive	…	positive
23.10.2013	negative	positive	negative	…	negative
24.10.2013	positive	positive	negative	…	positive
25.10.2013	positive	positive	same	…	positive
28.10.2013	negative	positive	positive	…	negative
…	…	….	…	…	…

B. Association Rule Mining

Frequent patterns are patterns that occur frequently in a data set. Frequent patterns have an important role in association mining, data classification, clustering and other data mining techniques. Market basket analysis is a basic form of frequent data mining. This process analyses relations between items in basket to find customer buying habits. [12] So significant amount of terminology used to describe both the data (e.g. transactions) and the output (e.g. itemsets) is borrowed from the supermarket analogy [13].

Frequent item sets are used to generate association rules. Association mining was first introduced by Agrawal, Imielinski and Swami [1]. The research results show the hidden patterns in large datasets. The association rules algorithm reveals the relationships between items or features that occur frequently in databases. For instance, in market basket analysis case, if people buy item A also buy item B, we can say that there is a relationship between item A and item B. The association rule mining commonly defined as follows [12].

Let $I = \{I_1, I_2, \ldots, I_m\}$ be an itemset. Let D be a set of database transactions. Each transaction (T) is a nonempty itemset of I ($T \subseteq I$). Let A be a set of items. A transaction T is said to contain A If $A \subseteq T$. $A \Rightarrow B$ is called an association rule where $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \phi$ $A \subset I$ and $B \subset I$. The rule $A \Rightarrow B$ has **support** s and **confidence** c values. The ratio of transactions that contain AUB to transactions in D gives us s value. The ratio of transactions that contain AUB to transactions that include A give us c value. That is as seen in (1) and (2):

$$\text{Support}(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$\text{Confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \quad (2)$$

Rules that provide minimum support and minimum confidence values are selected as association rules [12].

Finding association rules consists two phases;

1. At first; all the frequent item sets that provide minimum support are calculated.

2. In the second; the association rules are selected that support minimum confidence [13].

There are several additional measures to eliminate uninteresting rules. **Lift** is one of these correlation measures.

It is described as; The occurrence of itemset A is independent of the occurrence of itemset B if P(AUB) = P(A)P(B); otherwise itemsets A and B are dependent and correlated as events. The **lift** is calculated by (3):

$$\textbf{Lift}(A, B) = P(B|A) = \frac{P(A \cup B)}{P(A)P(B)} = \frac{P(B|A)}{P(B)} = \frac{conf(A \Rightarrow B)}{sup(B)} \quad (3)$$

If lift value equal to 1, then items are independent. If its value less than 1, this says that items are negatively correlated and the other result says that there is a positive correlation between items. [12].

C. Apriori Algorithm

Apriori is an algorithm to find association rules. It was initially introduced by Agrawal and Srikant in 1994. It uses preceding information of frequent item sets. A repetitive process, level-wise search is used in Apriori. In this process, (k+1)-itemsets are calculated from k-itemsets. There are some key concepts used in this algorithm such as **Frequent Itemsets**, **Apriori Property**, **Join** and **Prune** operations. **Apriori property**, which says that each nonempty subsets of frequent item set must also be frequent, improves the algorithm efficiency. Apriori algorithm is two steps processes that are join and prune steps that are defined as follows[12].

The join step: L_{k-1} includes (k-1) itemsets that provide minimum support and confidence values. L_{k-1} is joined with itself to determine candidate k-itemsets, C_k. C_k is used to find L_k [12].

The prune step: L_k is formed from candidate k itemsets, C_k. Each candidate item is controlled to provide minimum support and confidence values to generate L_k. But C_k can be huge and this requires heavy computation. Apriori property is used to reduce the size of C_k. [12].

D. Generating Association Rules from Frequent Item sets

Association rules that satisfy minimum support and minimum confidence are generated from frequent itemsets.

E. Methodology for Association Rule Generation

In this part, we have studied on a question that is finding association rules on BIST100 companies using historical data and Apriori algorithm. This research only focuses on close prices of stocks. 5 years daily data of 87 different BIST100 stocks were collected for date between 10/21/2013 and 10/19/2018 from "finance.yahoo.com". These data includes Date, Open, High, Low, Close, Adj. Close, Volume columns. We transformed this data as we explained in section 2 - BIST100 Stock Market Data. Finally we used Apriori algorithm to find association rules in selected stocks set. We implemented a python program for data transformation and used WEKA to find association rules.

Two different methods were used to find association rules. At first method, we used all 87 stocks to find association rules and we removed best 2 stocks from stocks set. We iterate this process until to find 14 rules. We called this method as "decreasing". At second method we create different stocks sets based on sectors such as, banks, energy, automobile, industry etc. We found association rules on these sets. We called this method as "sectoral".

The research procedure for decreasing method is shown in figure 1. And research procedure for sectoral method is shown in figure 2.

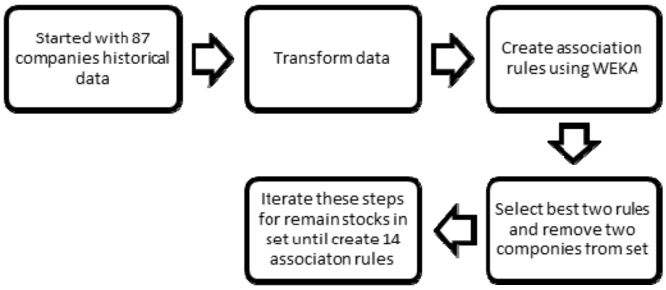

Fig. 1. Proposed decreasing method

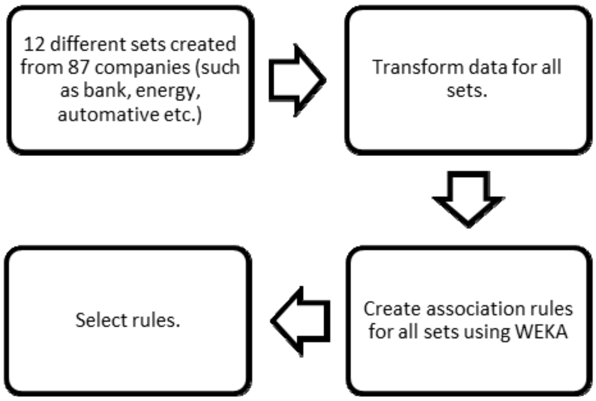

Fig. 2. Proposed sectoral method

III. RESULTS

In decreasing method, after data preparation that includes 87 companies stock data, Apriori algorithm was run in WEKA. Prepared data created from all 87 companies is shown in Table III.

TABLE III. PREPARED STOCK DATA FROM 87 COMPANIES' STOCK DATA CLOSE PRICES.

Date	AEFES	AFYON	…	YKBNK	ZOREN
21.10.2013	same	same	…	same	same
22.10.2013	neg	pos	…	pos	pos
23.10.2013	neg	pos	…	neg	pos
24.10.2013	pos	pos	…	same	pos
25.10.2013	pos	pos	…	pos	neg
28.10.2013	neg	pos	…	same	pos
29.10.2013	same	same	…	same	same
30.10.2013	same	pos	…	neg	neg
31.10.2013	neg	pos	…	neg	neg
1.11.2013	neg	pos	…	neg	pos
4.11.2013	neg	neg	…	neg	neg
…	…	…	…	…	…
17.10.2018	same	pos	…	pos	same
18.10.2018	pos	neg	…	neg	neg
19.10.2018	pos	neg	…	neg	neg

When we run Apriori algorithm on this data, we figure out association rules as shown in Table IV.

TABLE IV. GENERATED ASSOCIATION RULES FOR 87 COMPANIES

Weka Apriori -N 10 -T 0 -C 0.98 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1, Insts: 1305, Attrs: 87, Minimum support: 0.2 (261 instances), Minimum confidence: 0.98
1. AKBNK.IS=positive HALKB.IS=positive ISCTR.IS=positive THYAO.IS=positive VAKBN.IS=positive YKBNK.IS=positive 295 ==> GARAN.IS=positive 292 confidence: 0.99
2. AKBNK.IS=positive HALKB.IS=positive ISCTR.IS=positive PGSUS.IS=positive VAKBN.IS=positive YKBNK.IS=positive 282 ==> GARAN.IS=positive 279 confidence: 0.99
3. AKBNK.IS=positive ECILC.IS=positive HALKB.IS=positive ISCTR.IS=positive VAKBN.IS=positive YKBNK.IS=positive 269 ==> GARAN.IS=positive 266 confidence: 0.99
4. AKBNK.IS=negative ISCTR.IS=negative PGSUS.IS=negative SAHOL.IS=negative VAKBN.IS=negative YKBNK.IS=negative 265 ==> GARAN.IS=negative 262 confidence: 0.99
5. GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive TATGD.IS=positive YKBNK.IS=positive 283 ==> VAKBN.IS=positive 279 confidence: 0.99
6. AKBNK.IS=positive ALARK.IS=positive HALKB.IS=positive ISCTR.IS=positive VAKBN.IS=positive YKBNK.IS=positive 280 ==> GARAN.IS=positive 276 confidence: 0.99
7. AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive TSKB.IS=positive 273 ==> VAKBN.IS=positive 269 confidence: 0.99
8. GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive TSKB.IS=positive YKBNK.IS=positive 272 ==> VAKBN.IS=positive 268 confidence: 0.99
9. AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive TRKCM.IS=positive YKBNK.IS=positive 270 ==> VAKBN.IS=positive 266 confidence: 0.99
10. AKBNK.IS=positive ISCTR.IS=positive SAHOL.IS=positive THYAO.IS=positive VAKBN.IS=positive YKBNK.IS=positive 270 ==> GARAN.IS=positive 266 confidence: 0.99

We selected first, fourth and ninth rules form this table. And we removed two result companies that are GARANTI and VAKIFBANK from dataset. We select these rules because of theirs confidence level and providing companies diversity. We iterate this process until to reach 17 rules. We did not set minimum support level. Minimum support took 0.15 and 0.2 values. However, we set minimum confidence level that took 0.85, 0.90, 0.95 and 0.98 values and the created rules are as shown in Table V.

TABLE V. ASSOCIATION RULES USING DECREMENTING METHOD

1	AKBNK.IS=positive HALKB.IS=positive ISCTR.IS=positive THYAO.IS=positive VAKBN.IS=positive YKBNK.IS=positive 295 ==> GARAN.IS=positive 292 confidence: 0.99
2	AKBNK.IS=negative ISCTR.IS=negative PGSUS.IS=negative SAHOL.IS=negative VAKBN.IS=negative YKBNK.IS=negative 265 ==> GARAN.IS=negative 262 confidence: 0.99
3	AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive TRKCM.IS=positive YKBNK.IS=positive 270 ==> VAKBN.IS=positive 266 confidence: 0.99
4	HALKB.IS=positive ISCTR.IS=positive KCHOL.IS=positive PGSUS.IS=positive SAHOL.IS=positive THYAO.IS=positive YKBNK.IS=positive 203 ==> AKBNK.IS=positive 202 confidence: 1
5	AKBNK.IS=positive HALKB.IS=positive KARTN.IS=positive KCHOL.IS=positive THYAO.IS=positive YKBNK.IS=positive 204 ==> ISCTR.IS=positive 202 confidence: 0.99
6	BRSAN.IS=positive DEVA.IS=positive HALKB.IS=positive TATGD.IS=positive 207 ==> YKBNK.IS=positive 200 confidence: 0.97
7	AFYON.IS=negative HALKB.IS=negative IPEKE.IS=negative TATGD.IS=negative 204 ==> KOZAA.IS=negative 197 confidence: 0.97
8	BRSAN.IS=positive GOLTS.IS=positive KCHOL.IS=positive PGSUS.IS=positive 211 ==> THYAO.IS=positive 199 confidence: 0.94
9	ALARK.IS=negative HALKB.IS=negative KCHOL.IS=negative TTKOM.IS=negative 214 ==> SAHOL.IS=negative 201 confidence: 0.94
10	ALARK.IS=positive GOODY.IS=positive KARTN.IS=positive MGROS.IS=positive 211 ==> HALKB.IS=positive 196 confidence: 0.93
11	DEVA.IS=negative GOLTS.IS=negative KARTN.IS=negative PGSUS.IS=negative 220 ==> AFYON.IS=negative 203 confidence: 0.92
12	ALARK.IS=positive ECILC.IS=positive MGROS.IS=positive PETKM.IS=positive 215 ==> GOLTS.IS=positive 196 confidence: 0.91
13	DEVA.IS=negative ECZYT.IS=negative GOLTS.IS=negative GUBRF.IS=negative 222 ==> ECILC.IS=negative 200 confidence: 0.9
14	AKSA.IS=negative PETKM.IS=negative ZOREN.IS=negative 230 ==> TMSN.IS=negative 202 confidence: 0.88
15	BRSAN.IS=negative ISGYO.IS=negative ZOREN.IS=negative 227 ==> VESTL.IS=negative 198 confidence: 0.87
16	AKSEN.IS=negative ANACM.IS=negative GOODY.IS=negative 232 ==> PGSUS.IS=negative 202 confidence: 0.87
17	ANACM.IS=positive ARCLK.IS=positive PGSUS.IS=positive 234 ==> SISE.IS=positive 202 confidence: 0.86

Some graphics for stock prices for companies in decrementing method rules are as presented in figure 3, figure 4 and figure 5.

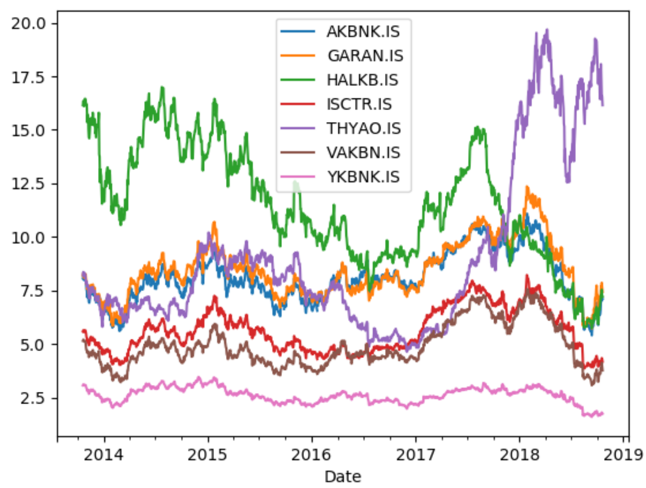

Fig. 3. Historical data graphics for stocks at rule1.

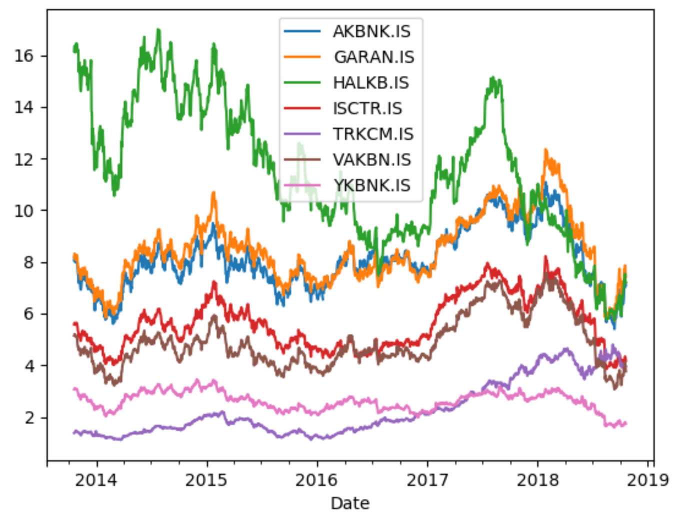

Fig. 4. Historical data graphics for stocks at rule3

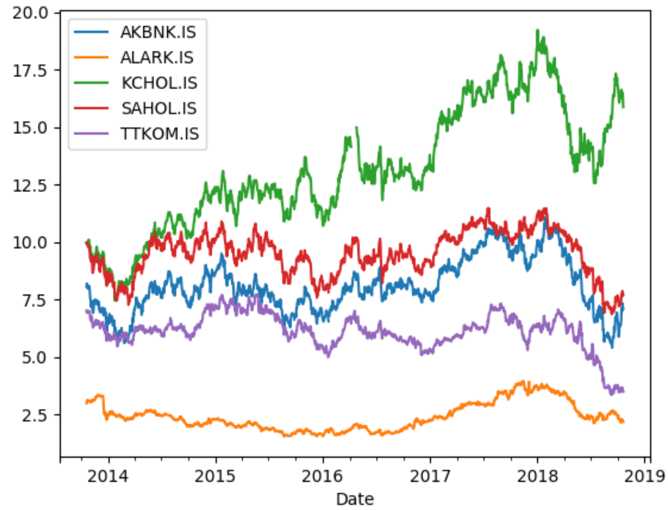

Fig. 5. Historical data graphics for stocks at rule10

In second method, we called it sectoral, we grouped stocks based on sectors (banks, energy companies, holdings, construction, automotive etc.) and also we created some groups that include dominant companies on BIST100.

Our first group is based on banks and prepared data from these companies is shown in Table VI.

TABLE VI. PREPARED STOCK DATA FROM SEVEN BANKS HISTORICAL STOCK DATA CLOSE PRICES.

Date	AKBNK.IS	ALBRK.IS	GARAN.IS	HALKB.IS	ISCTR.IS	VAKBN.IS	YKBNK.IS
21.10.2013	same	same	same	same	same	same	same
22.10.2013	pos	pos	pos	pos	pos	pos	pos
23.10.2013	neg	neg	neg	neg	neg	neg	neg
24.10.2013	neg	pos	neg	neg	pos	same	same
25.10.2013	same	pos	pos	pos	pos	neg	pos
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
15.10.2018	pos	pos	pos	pos	pos	pos	pos
16.10.2018	neg	neg	neg	same	pos	pos	pos
17.10.2018	pos	pos	pos	pos	pos	pos	pos
18.10.2018	neg	neg	neg	neg	neg	neg	neg
19.10.2018	neg	neg	neg	same	neg	neg	neg

When we run Apriori algorithm on this data, we figure out association rules as shown in Table VII.

TABLE VII. GENERATED ASSOCIATION RULES FOR BANKS.

Weka Apriori -N 10 -T 0 -C 0.95 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Insts: 1305, Attrs: 7, Minimum support: 0.25 (326 instances), Minimum confidence: 0.95
1. AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive YKBNK.IS=positive 347 ==> VAKBN.IS=positive 336 confidence: 0.97
2. AKBNK.IS=negative GARAN.IS=negative HALKB.IS=negative VAKBN.IS=negative 345 ==> YKBNK.IS=negative 334 confidence: 0.97
3. AKBNK.IS=positive HALKB.IS=positive ISCTR.IS=positive VAKBN.IS=positive YKBNK.IS=positive 348 ==> GARAN.IS=positive 336 confidence: 0.97
4. GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive YKBNK.IS=positive 370 ==> VAKBN.IS=positive 357 confidence: 0.96
5. AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive YKBNK.IS=positive 369 ==> VAKBN.IS=positive 355 confidence: 0.96
6. AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive 373 ==> VAKBN.IS=positive 358 confidence: 0.96
7. AKBNK.IS=negative GARAN.IS=negative ISCTR.IS=negative VAKBN.IS=negative 368 ==> YKBNK.IS=negative 353 confidence: 0.96
8. AKBNK.IS=negative HALKB.IS=negative ISCTR.IS=negative VAKBN.IS=negative 341 ==> YKBNK.IS=negative 327 confidence: 0.96
9. AKBNK.IS=negative GARAN.IS=negative HALKB.IS=negative ISCTR.IS=negative 343 ==> YKBNK.IS=negative 328 confidence: 0.96
10. AKBNK.IS=positive HALKB.IS=positive ISCTR.IS=positive YKBNK.IS=positive 364 ==> VAKBN.IS=positive 348 confidence: 0.96

We selected first rule from this table. We made same process on other groups and we select some rules from these groups. We did not set minimum support level. Minimum support took 0.1, 0.15 and 0.2 values. However we set minimum confidence level that took 0.70, 0.80, 0.88, 0.90, 0.93, 0.95 and 0.97 values. Finally we created 15 rules that shown in Table VIII.

TABLE VIII. ASSOCIATION RULES USING SECTORAL METHOD

1	AKBNK.IS=positive GARAN.IS=positive HALKB.IS=positive ISCTR.IS=positive YKBNK.IS=positive 347 ==> VAKBN.IS=positive 336 confidence: 0.97
2	AKSEN.IS=negative ANELE.IS=negative GEREL.IS=negative ODAS.IS=negative 157 ==> ZOREN.IS=negative 134 confidence: 0.85
3	CCOLA.IS=positive TATGD.IS=positive ULKER.IS=positive 183 ==> MGROS.IS=positive 140 confidence: 0.77
4	ALARK.IS=positive KCHOL.IS=positive TAVHL.IS=positive TKFEN.IS=positive 157 ==> SAHOL.IS=positive 144 confidence: 0.92
5	AFYON.IS=negative KRDMD.IS=negative TKFEN.IS=negative 234 ==> GOLTS.IS=negative 197 confidence: 0.84
6	DOAS.IS=negative FROTO.IS=negative GOODY.IS=negative TTRAK.IS=negative 158 ==> TOASO.IS=negative 133 confidence: 0.84
7	ARCLK.IS=positive PGSUS.IS=positive 325 ==> THYAO.IS=positive 272 confidence: 0.84
8	EREGL.IS=positive GOLTS.IS=positive GUBRF.IS=positive MGROS.IS=positive SAHOL.IS=positive SISE.IS=positive THYAO.IS=positive 135 ==> VAKBN.IS=positive 133 confidence: 0.99
9	ECILC.IS=positive EREGL.IS=positive GOLTS.IS=positive MGROS.IS=positive SAHOL.IS=positive SISE.IS=positive THYAO.IS=positive 134 ==> VAKBN.IS=positive 132 confidence: 0.99 lift:(2.12) lev:(0.05) [69] conv:(23.89)
10	ECILC.IS=positive KARTN.IS=positive MGROS.IS=positive OTKAR.IS=positive SISE.IS=positive 145 ==> GOLTS.IS=positive 141 confidence: 0.97
11	EREGL.IS=positive GUBRF.IS=positive SISE.IS=positive THYAO.IS=positive ZOREN.IS=positive 143 ==> SAHOL.IS=positive 137 confidence: 0.96
12	EREGL.IS=positive GUBRF.IS=positive KARTN.IS=positive SAHOL.IS=positive TOASO.IS=positive 141 ==> THYAO.IS=positive 135 confidence: 0.96
13	EREGL.IS=positive KARTN.IS=positive TOASO.IS=positive ZOREN.IS=positive 145 ==> THYAO.IS=positive 137 confidence: 0.94
14	DGKLB.IS=negative ECILC.IS=negative GUBRF.IS=negative KARTN.IS=negative MGROS.IS=negative 152 ==> ZOREN.IS=negative 136 confidence: 0.89
15	DGKLB.IS=negative GUBRF.IS=negative KARTN.IS=negative MGROS.IS=negative ZOREN.IS=negative 154 ==> ECILC.IS=negative 136 confidence: 0.88

Some graphics for stock prices for companies in sectoral method rules are as seen in figure 6, figure 7 and figure 8.

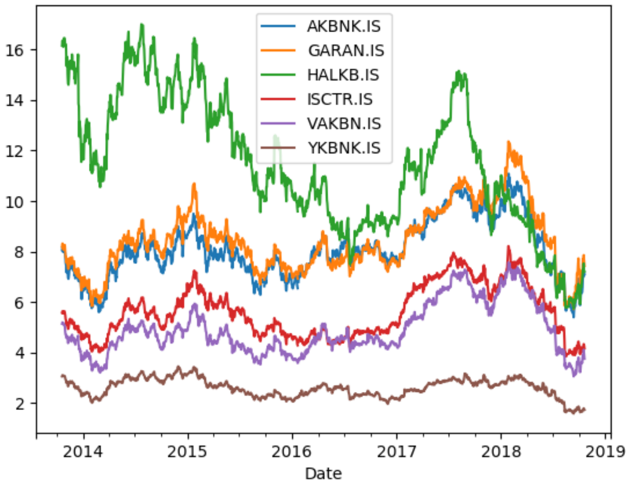

Fig. 6. Historical data graphics for stocks at rule1

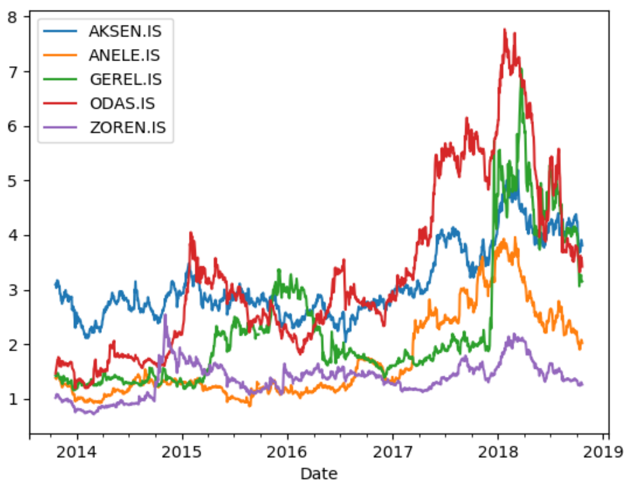

Fig. 7. Historical data graphics for stocks at rule2

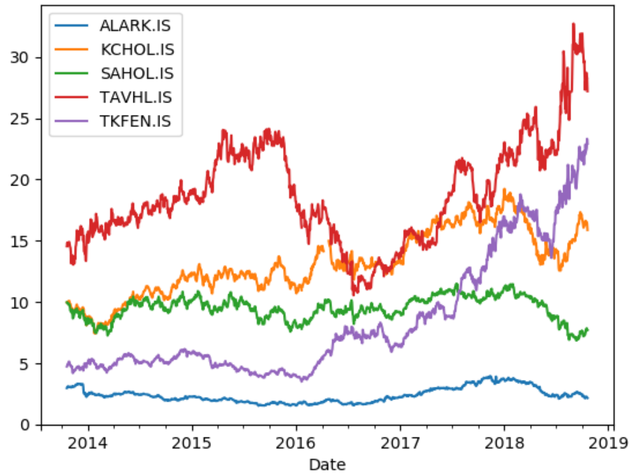

Fig. 8. Historical data graphics for stocks at rule4

IV. DISCUSSION

When the results are examined, it is shown that there are strong relationships between stocks in the dataset. It is found that sectoral based stocks generally move together or they have same envelope for their historical price data. And also stocks at rules that have higher confidence than 0.90 move together or they have same envelope for their historical price data.

We consider that association rule mining is beneficial to find hidden patterns in large financial datasets.

V. CONCLUSION

The main objective of this study is to figure out association rules between BIST100 stocks using Apriori algorithm. We used two methods for stocks selection. In first method we started with all 87 stocks, found association rules between them and removed two stocks at the best two rules and go on to find new association rules. In second method we divided stocks on sectoral base sets, and determined association rules on these sets.

When the results are examined, we showed that there are strong relationships between stocks in the dataset. We found that sectoral based stocks generally move together or they have same envelope for their historical price data. And also stocks at rules that have higher confidence than 0.90 move together or they have same envelope for their historical price data.

We consider that association rule mining is beneficial to find hidden patterns in large financial dataset.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "Database mining: a performance perspective," IEEE Trans. Knowl. Data Eng., vol. 5, no. 6, pp. 914–925, 1993.

[2] C. Wu and Y.-C. Su, "Dynamic Relations Among International Stock Markets," 1998.

[3] A. S. Almohaimeed, "Using Tweets Sentiment Analysis to Predict Stock Market Movement," 2017.

[4] A. Ariya, "Stock Forecasting by Association Rule Mining," in Proceedings of the 21st Asia-Pacific Conference on Global Business, Economics, Finance & Social Sciences (AP18Taiwan Conference), 2018.

[5] M. Kumar and A. Kalia, "Association Rules Mining in the Stock Data."

[6] K. Karpio, P. Šukasiewicz, A. Oraowski, and T. Z¡bkowski, "Mining Associations on the Warsaw Stock Exchange," vol. 123, 2013.

[7] M. U. Şimşek and S. Özdemir, "Analysis of the relation between Turkish twitter messages and stock market index," 2012 6th Int. Conf. Appl. Inf. Commun. Technol. AICT 2012 - Proc., no. May, 2012.

[8] M. Yildirim, C. Arsun YÜKSEL, A. Kelimeler, and D. Analizi, "Sosyal Medya İle Hisse Senedi Fiyatinin Günlük Hareket Yönü Arasindaki İlişkinin İncelenmesi: Duygu Analizi Uygulamasi Investigation of Relationship Between Social Media and Stock Price Daily Movement Direction: Sentiment Analysis Implementation," Yayına Kabul Tarihi (Acceptance Date, pp. 14–11.

[9] S. S. Ozturk and K. Ciftci, "A Sentiment Analysis of Twitter Content as a Predictor of Exchange Rate Movements," Rev. Econ. Anal., vol. 6, no. 2, pp. 132–140, 2014.

[10] İ. Savaş, İ. Can, "Euro-Dolar Paritesi ve Reel Döviz Kuru'nun İMKB 100 Endeksi'ne Etkisi", Eskişehir Osmangazi Üniversitesi İibf Dergisi, Nisan 2011, 6(1), 323□339.

[11] S. Kendirli, M. Çankaya, "Dolar Kuru'nun Borsa Istanbul-30 Endeksi Üzerindeki Etkisi Ve Aralarindaki Nedensellik Ilişkisinin Incelenmesi", Celal Bayar Üniversitesi Sosyal Bilimler Dergisi. 14. 10.18026/cbusos.48113.

[12] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.

[13] C. C. Aggarwal, Data Mining: The Textbook. 2015.