Int. Per. of Recent Tech. in App. Eng. 2020; 2(2): 42-50
DOI: 10.292228/porta.1

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Crime Data Analysis with Association Rule Mining

Birliktelik Kural Çıkarımı ile Suç Veri Analizi

Duygu ÇALIŞKAN[1], Kazım YILDIZ[2], Buket DOĞAN[2], Abdulsamet AKTAŞ[2]

[1]Marmara University, Institue of Pure and Applied Sciences, Department of Computer Engineering 34722, Istanbul, Turkey

[2]Marmara University, Technology Faculty, Department of Computer Engineering, 34722, Istanbul, Turkey

Abstract

Along with the positive developments of the globalizing world, new types of crime such as social media fraud, drug trafficking and vehicle robbery, which have disrupted community welfare and order, have also emerged. With developments in information technology, it is possible to record real-time various data related to subject of crimes, location and time information, type of crime. By analyzing these recorded raw data using various data mining methods, it is possible to extract information that can be used to identify the data or for prediction purposes. In this study, an analysis of the association rules on the NIBRS Crime dataset which includes real crime cases from July 2016 to April 2018 in the state of Maryland in USA was carried out using R program with Apriori algorithm and Rapid Miner with FP-Growth algorithm. With these association rules created, the time intervals, the districts, the types of crimes and the frequency of the occurrences are analyzed and the results of the algorithms are presented. With the results of this analysis; for organizations which are responsible for maintaining the peace and social order, such as security forces and law enforcement agencies; it is possible to follow useful information such as which crimes are committed more frequently and in which time period of day the criminals are more active.

Keywords: Crime Analysis, Association Rule, Apriori, FP-Growth

Öz

Küreselleşen Dünya'nın hayatımıza kattığı olumlu gelişmeler ile birlikte, toplum refahını ve düzenini bozan sosyal medya dolandırıcılığı, uyuşturucu ticareti, araç hırsızlığı vb. gibi yeni suç türleri de ortaya çıkmıştır. Bilişim teknolojisindeki gelişmeler sayesinde bu suçların konusu, konum ve zaman bilgileri, suç türü gibi olaya ilişkin çeşitli verilerin gerçek zamanlı kayıt altına alınabilmesi mümkün olmaktadır. Kayıt altına alınan bu ham verilerin çeşitli veri madenciliği yöntemleri kullanılarak analiz edilmesi ile veriyi tanımlayan veya öngörü amaçlı kullanılabilecek bilgilerin ortaya çıkartılması mümkündür. Bu çalışmada, veri madenciliği uygulamalarından R programı ile Apriori algoritması ve Rapid miner programı ile FP-Growth algoritması kullanılarak, ABD'nin Maryland eyaletinde 2016 yılının Temmuz ayından 2018 Nisan ayına kadar meydana gelen suç verilerinden oluşan NIBRS Crime veri seti üzerinde birliktelik kuralları analizi uygulaması gerçekleştirilmiştir. Oluşturulan bu birliktelik kuralları ile hangi saat aralıklarında, hangi semtte, ne tür suçların, ne sıklıkla gerçekleştirildiği analiz edilmiş ve algoritmaların sonuçları sunulmuştur. Bu analiz sonucunda çıkan sonuçlar ile güvenlik güçleri ve kolluk kuvvetleri gibi toplumun huzurunu ve düzenini korumakla görevli olan kuruluşların; hangi semtte, hangi suçların daha sık işlendiği veya suçluların hangi saat aralığında daha aktif olduğu gibi faydalı bilgileri takip etmesi mümkün olmaktadır.

Anahtar Kelimeler: Suç Analizi, Birliktelik Kuralları, Apriori, FP-Büyüme Algoritması

I. INTRODUCTION

With the development of technology and civilization in the world, the amount of data produced increases day by day. In this increasing world of data, the use of information systems is of great importance for the storage and interpretation of data. In this study, crime data, which is an important sociological phenomenon in human history, is examined. Crime is a concept that has existed since the early ages when humans existed. The murder of Adam and Eve's first children, Habil, by his brother, Kabil, is one of the first known criminal cases of premeditated murder [1].

While crimes were committed in simple ways in the early periods, today this situation is quite different. The constant change and development of technology and civilization, as well as changing human life in every sense, has also shown

Sorumlu Yazar/Corresponding Author: Duygu ÇALIŞKAN, Tel: +90 542 449 35 31, e-posta: ehm.duygucaliskan@marmara.edu.tr
Gönderilme/Submitted: 16.08.2020, **Düzenleme/Revised:**08.05.2021, **Kabul/Accepted:** 21.05.2021

its effect in the phenomenon of crime [2]. Existing types of crime have now become more complex and then simple. Crime is neither systematic nor purely coincidental. Therefore, a standard definition of the concept of crime could not be made. Although there have been many definitions to date for the concept of crime, the most accepted of these is "a prohibited act or a whole of actions, or a violation of an obligation that is bound by the laws of society." is the definition [1-3].

Crime analysis examines the relationship between crime and criminal, and includes data on the regional distribution of crimes. Crime analysis takes an important place in the field of safety and security in terms of providing foresight before a crime is committed [4]. Crime analysis is a concept that involves preventing crime from occurring, identifying existing crimes and crime trends and taking necessary measures against them [5]. The crime analysis has an important role in the security and safety in terms of finding the connection between the crime and the criminal, getting information about the rate of territories of the criminals which they have existed and providing foresight before committal.

An important issue in crime analysis is the representation of criminals according to a profile. This topic can be divided into three main topics: inductive, deductive and geographic profiling. Inductive profiling is carried out with the help of the characteristics of known criminals. Deductive profiling is created by the responses and information provided by the victim. Geographical profiling is shaped by taking advantage of the geographical features of the place where the crime is committed [6]. Crime data mining is one of the most popular techniques used for crime analysis.

Data mining is the process of acquiring meaningful and large-scale information from large datasets to help decision-making for future processes. [7]. In other words, it is process of obtaining the valuable data among the large-scale data. In this way, it is possible to discover the relationships between the data and, if necessary, to make predictions [8]. Data mining techniques such as clustering, classification and association rules (Apriori algorithm, FP-Growth algorithm) are some of the effective methods used in crime analysis. Through these techniques, it is possible to identify criminal patterns inference of relationships between data that appear to be unconnected. In the clustering technique, similarities and relationships in crime records can be found. The clustering technique allows the study of the proximity and distance of crime distributions by regions. Outlier detection stands out as a technique used to establish a pattern of abnormal conditions in records. Fraud detection is a criminal data mining method used in the study of issues such as network attacks. If there is unusual activity in an area, it is usually a sign of an incident [9].

It is essential to take a look at some important studies in the field of crime data mining to date. These studies provide important examples for the use of data mining in the field of crime.

The COPLINK project is significant from work in the field of crime data mining. In the COPLINK project, the police unit and a team from the University of Arizona worked together to extract assets from criminal records. As a result of this study, it was possible to infer those associated with the crime [10].

Ozgul, Atzenbeck, Çelik and Erdem have introduced a prediction model called Crime Prediction Model (CPM) to solve unsolved terrorist incidents [11]. Mason used the naive bayes method of classification to analyze crime data in his study [12]. Mittal, Goyal, Sethi and Hemanth analyzed crime data using machine learning algorithms [13]. By using Thangamuthu, Vadivel and Priyadharshini clustering algorithms; They analyzed crime data with the k-means algorithm and improved the crime-based forms [14]. Ma, Chen, and Huang used a two-step clustering algorithm called AK-Modes for looking similar event subsets from large datasets automatically [15].

In this study, it is aimed to improve the application of association rules analysis on NIBRS Crime data consisting of crime data that occurred between July 2016 and April 2018 in Maryland, USA [16]. Unlike other approaches in the crime prediction literature, it offers the design and implementation of a proactive method to predict crime trends. Crime data crime analysis was conducted on actual crime data provided by the U.S. General Services Administration [16]. Association rules were applied on dataset which contain 98.272 criminal data.

This article is organized as follows. Chapter 2 presents the most important approaches in crime data mining literature and the most representative projects in such a research area. Chapter 3 data describes training the step-by-step model and evaluation of real data. Finally, Chapter 4 is the conclusion of the article and summarizes future work.

II. METHODOLOGY

2.1. Association Rule Mining

Data mining represents the process of discovering significant new relationships and trends through the processing of data stacks; Thanks to this process obtains eduseful meaningful information. [17].

Data mining methods are divided into two as predictive and descriptive. In predictive models, a model is developed from data with known results and it is aimed to estimate the result values for data sets with unknown results. In the descriptive model, it is aimed to find data by combining with clustered relationships. Supervised (Classification); It is represented by Neural Network, K-Nearest Neighbor, Decision Tree, Bayesian , Genetic Algorithms, Decision Support Machines, Fuzzy Set and other methods. And Unsupervised; represented by Clustering, Association Rules, Sequential Pattern Analysis, Extreme Value Analysis and other methods. [18,19].

Association rules are one of the most popular data mining methods. Support and confidence in the model discovered by the association rule technique are two interestingness measures of this rule [19,20]. Definition:

Let *I = {I1, I 2, ..., I m}* be a binary set of attributes that we will call products.

Let *T = {t1, t2, ..., tn}* denote the operations in the database. The value each t_k = will take is 0 or 1. If t_k = 0 l_k is not bought, if t_k = l_k = 1 = means purchased. There is a separate record for each transaction in the database. Now for X ⊆I the value t_k corresponding to each l_k in X is t_k = 1.

A association rule is expressed as follows:

X ⇒ I , X is a subset of I. l_j is any element in I and this element is not in X. In order to say that X ⇒ I rule is appropriate for T, it will be necessary to mention a certain level of confidence. That is, how much of all the X in T provide k to I should be expressed with the value c%. In this case, we can express the association rule with a confidence level of 0 ≤ c ≤ 1 as follows: X I c j ⇒ | . Trust level also expresses the strength of the rule. There are two values used in expressing the mentioned relationships: Support and Trust. These values are numerical values and we need to define some numerical terms to describe them. Let D be the database of transactions, and let N be the number of transactions in D.Each Di process is a set of products. [19,21].

Let support (X) be the ratio of transactions involving the X product set:

Confidence(X) = {I | I ∈ D ∧ I ⊇ X} / N

I is an element set and |. | also shows the number of elements of the set.

The support value of an association rule is the ratio of both the previous and the next transactions to the total number of transactions. Confidence value is the rate at which transactions that include the previous one also have the next one. Support and confidence values for A ⇒ C partnership are as follows [22].

support (A => C) = support (AUC)

confidence(A => C) = support (AUC) / support (A)

Rules that meet both the minimum support threshold (minup) and the minimum confidence threshold (minconf) are called strong. A set of items is called a set of itemset. An itemset that contains K element is a k-itemset [19].

If the confidence value is 100%, the rule is true in all data analyses, and these rules are called "definite".

In the association rule, the relationship between the items is calculated by the criteria of support and confidence. The support criterion tells you how often the link between items is in the data. The confidence criterion tells you what probability B will be with A. In order for the association of the two elements to be important, both the criteria of support and confidence must be as high as possible [21,23].

2.1.1 Apriori Algorithm

In the literature, there are different algorithms that produce association rules. Apriori algorithm is the most widely used association rule mining algorithm known one in association rule inference algorithms [18].

Apriori is a classic and widely used association rule algorithm. The Apriori algorithm works by scanning the database multiple times to find frequent item sets.In the first scan, there is a found set of frequent items with one element, providing the minimum support value. Frequent sets of items found in the previous search in on going scans are new potential frequent items called candidate clusters,Used to produce sets [24].

Steps of the Apriori algorithm:

1. Determining the minimum number of supports and the minimum confidence value.
2. Finding the frequency value (number of repetitions) of each item in the item sets and calculating the support values.
3. Having support lower than the minimum support value disabling items.
4. A new table is created with our products that have a support value equal to or above the minimum support value we previously determined. After creating our new table, we repeat what we did in the first step in our new table. Only this time the frequency value indicates the existence of both objects at the same time.
5. Removing sets of items that are below the minimum support value.

6. Establishing trinity, quaternary ect. partnerships.
7. Those associations who exceed the minimum support value removal of anything else
8. Extracting association rules from triple, quaternary ect. partnerships.

The basic approach in this algorithm, whose pseudo code is defined in Figure 1. If the k-element cluster meets the minimum support criterion, the subsets of this cluster also meet the minimum support criterion. The support value of a set of items is not greater than the support value of its subset. All empty objects of a favorite object set non-subsets are also frequent. [25].

```
L1 = { large 1 – itemsets } ;
for ( k = 2 ; Lk-1 ≠ Ø ; k++ ) do begin
  Ck = apriori_gen(Lk-1); // New candidates
  for all transaction t ∈ D do begin
    Ct = subset(Ck, t);  // Candidates in t
    for all candidate c  ∈ Ct do
      c.count++;
   end
  Lk = {c ∈ Ck | c.count ≥ minsup};
   end
end
 Answer  = ∪k Lk ;
```

Figure 1. Apriori Algorithm Pseudocode

2.1.2 Frequent Pattern (FP) Growth Algorithm

The FP Growth algorithm is the improvement of the Apriori algorithm. It is used to find a frequent set of items in a database without candidate creation. FP-Growth algorithm, one of the association rules, shows higher performance than other algorithms.

The FP-Growth algorithm consists of two steps: Creating the FP Tree and extracting common patterns from the FP Tree.

The database needs to be scanned twice to build the FP tree. The first scan selects frequently used items, it is then ordered in descending order to build the list. The second scan creates the FP-Tree. First, operations are reordered according to the F list by removing non-frequent items. The reorganized transactions are later added to the FP Tree. The FP-Growth entry is the FP-Tree and the minimum number of supports. In the FP-Growth algorithm, the nodes in the FP-Tree are separated from the least found item in the F-List. All items in the path from the node to the root are collected while visiting each node. These items create the conditional pattern basis for that item. These The conditional pattern base that occurs with the element is a small database of patterns. Later FP-Growth is created. FP-Growth is executed on the small FP-Tree and FP-Tree from the conditional pattern base. The process is repeated repetitively without creating a conditional pattern base [25].

Figure 2 shows the pseudo code of the FP-Growth algorithm. First of all, support of each object in the database values are calculated in the algorithm. The support values correspond to the support threshold given as input to the algorithm; Objects that are greater and equal are put in a list, in descending order. This ordering ensures that uncommon items are not added to the FP-Tree. Thanks to the sorting process, items with a larger support value are closer to the root. If an object in the motion record is not in the created tree, a new node is created for that item and the support value is set to 1. If that item was previously created in the tree only the support value of that node is increased by 1.

```
Algorithm FPGrowth(root ,n,minsupport)
    if node only one path,Y,provided that then
        foreach nodes_of_combination nc in Y do
            pattern p = nc ∪ n
            support = min(support values of nc nodes)
            if p.support > minsupport then
                Output(p);
            end
        end
    else
        foreach ai in nodes do
            pattern p = ai ∪ n
            support = min(support values of ai nodes)
            if p.support > minsupport then
                Output(p);
            end
            create object-conditional patterns;
            create object-conditional FPtree tree;
            if FPtree ≠ Ø then
                Growth(FPtree,p,minsupport);
            end
        end
    end
End
```

Figure 2. General Structure of FP-Growth Algorithm

Then, the growth algorithm is run on the obtained FT-Tree. The Growth that is executed for each item in Figure 3 the general structure of its algorithm is shown. Primarily, the paths that the items passes in the algorithm determines. If there is only one branch, the common set of items is the combination of the items that make up the branch. If there is more than one path, the support value is determined as the minimum support value for that path. These paths then form the basis of the conditional pattern for that item. A conditional pattern tree is created from every conditional pattern basis. Then the algorithm is run on this conditional pattern tree recursively again. When the FP-Growth algorithm ends together the set of frequently visible items is determined.

```
Algorithm FPGrowth(root ,n,minsupport)
        if node only one path,Y,provided that then
                foreach nodes_of_combination nc in Y do
                        pattern p = nc ∪ n
                        support = min(support values of nc nodes)
                        if p.support > minsupport then
                                Output(p);
                        end
                end
        else
                foreach a_i in nodes do
                        pattern p = a_i ∪ n
                        support = min(support values of a_i nodes)
                        if p.support > minsupport then
                                Output(p);
                        end
                        create object-conditional patterns;
                        create object-conditional FPtree tree;
                        if FPtree ≠ Ø then
                                Growth(FPtree,p,minsupport);
                        end
                end
        end
End
```

Figure 3. General Structure of FP-Growth Algorithm

2.2 Dataset

The crimes, which occurred in the state of Maryland, USA between July 2016 and April 2018, were used as a dataset in this study [16]. The preprocessing process includes: correction, completing missing data, removing duplicate data, transforming, integrating, cleaning, normalizing, dimension reduction, etc. are transactions. Parts without analysis value are first deleted from the dataset during the data preprocessing phase. The crime data is then converted to the form appropriate to the data mining algorithms. Some quantitative-qualitative data transformations are needed to make our study conform to the association rules mining algorithm. Some studies on this subject are given below.

- Different criminal characteristics were deduced from date set. Accordingly, the day and month in which the crime was committed is grouped.

- The time zone in which the crime was committed was converted into a 24-hour time zone.

- Attributes that does not have the potential to create association rules such as office code, block address, sector, beat, address number, street prefix, street name have been deleted from the dataset. In Figure 4, the fields of the used data and the content of a sample data can be seen. In the data used, the Crime Name 1, Crime Name 4, and Crime Name 3 fields are text-type fields that store Crime Information. City and Event location is a text-type field that stores information about the city and crime location where the crime was committed. The Event Date and Event Day fields contain the date format that stores the date and day of the crime. The Victims field is an integer field that stores the number of victims. Event Time is a field in the time format that stores

Row No.	Crime Name1	Crime Name2	Crime Name3	City	Event Locati...	Months	Event Day	Event Time	Police Distri...
1	Crime Agains...	All other Larc...	LARCENY (D...	GERMANTO...	Residence - ...	January	Tuesday	0:22	GERMANTO...
2	Crime Agains...	Shoplifting	LARCENY - S...	SILVER SPRI...	Retail - Other	January	Tuesday	0:04	WHEATON
3	Crime Agains...	Weapon Law ...	WEAPON - C...	SILVER SPRI...	Retail - Drug ...	January	Tuesday	0:04	WHEATON
4	Crime Agains...	Trespass of ...	TRESPASSING	SILVER SPRI...	Retail - Drug ...	January	Tuesday	0:04	WHEATON
5	Crime Agains...	Aggravated A...	ASSAULT - A...	GERMANTO...	Convenience ...	January	Tuesday	0:32	GERMANTO...
6	Crime Agains...	Drug/Narcotic...	DRUGS - MA...	SILVER SPRI...	Parking Lot - ...	January	Tuesday	0:58	SILVER SPRI...
7	Crime Agains...	Disorderly Co...	PUBLIC PEA...	ROCKVILLE	Street - Com...	January	Tuesday	0:57	ROCKVILLE
8	Crime Agains...	Simple Assault	ASSAULT - SI...	GERMANTO...	Residence - ...	January	Tuesday	4:00	GERMANTO...
9	Crime Agains...	Trespass of ...	TRESPASSING	ROCKVILLE	Parking Gara...	January	Tuesday	5:00	ROCKVILLE
10	Crime Agains...	False Preten...	FRAUD - SWI...	ROCKVILLE	Residence - ...	January	Tuesday	8:52	BETHESDA
11	Other	All Other Offe...	POLICE INFO...	ROCKVILLE	Residence - ...	January	Tuesday	8:52	BETHESDA
12	Crime Agains...	Drug Equipm...	DRUGS - NA...	SILVER SPRI...	Street - In veh...	January	Tuesday	8:15	SILVER SPRI...
13	Crime Agains...	Destruction/D...	DAMAGE PR...	TAKOMA PARK	Street - Resid...	January	Tuesday	9:10	SILVER SPRI...
14	Crime Agains...	Drug/Narcotic...	DRUGS - HE...	ROCKVILLE	Residence - ...	January	Tuesday	9:30	ROCKVILLE
15	Other	All Other Offe...	MENTAL ILL...	SILVER SPRI...	Street - Resid...	January	Tuesday	13:40	WHEATON

Figure 4. Dataset example

event time information. And finally, the Location field refers to a float field that stores coordinate information.

2.3. Crime Analysis Using Association Rule Mining

In our study, Apriori and FP-Growth algorithms are used for association rule mining and these algorithms are implemented with Rapid Miner and R programming. Association rules on crime data mining with the factors that constitute the crime, the type of crime, the location of the crime, the time of the crime and the situations will be revealed in this study.

The aim of the studies is to analyze crime events, establish a relationship between crime factors and make predictions about crime. The Crime dataset was obtained from the crime which took place in the US state of Maryland from July 2016 to April 2018.

However, there are no rules as to which types of crime and crime scenes are taken.

2.3.1 R Tool

R is a programming language and interpreter for statistical calculations and graphs. It is a GNU project like the S language that has a wide range in time series analysis, classical statistical tests, clustering, classification and graphical techniques [26]. Our data set is analyzed by the Apriori algorithm in the R program. The flow chart for the proposed methodology is seen in Figure 5. First, the data is pre-processed and then Apriori algorithm produces association rules form dataset.

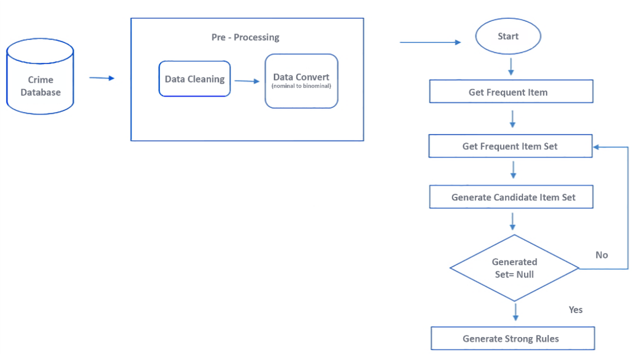

Figure 5. Flow Diagram for Proposed Methodology

2.3.2 Rapid Miner Tool

RapidMiner Studio is a powerful data mining tool for quickly building predictive models [27]. It is developed for the purposes of machine learning, data and text mining, predictive analysis, and business analysis.

The software is often used for commercial applications as well as research and application development. Besides, it supports all process in data mining. So it can be used for preparation of the data, verification, visualization and optimization. RapidMiner developed with open core model [28]. FP-Growth algorithm is implemented in RapidMiner. In Figure 6, the application of association rules process is designed in RapidMiner.

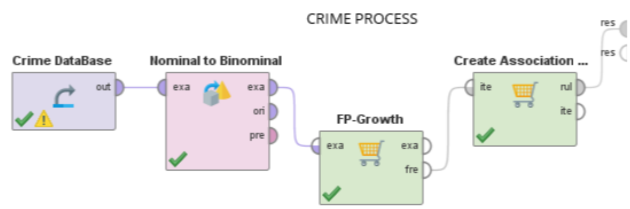

Figure 6. Association Rules Mining Process in RapidMiner

III. ASSOCIATION RULES AND FINDINGS

The rules obtained after the application of association rule mining are listed below. Similar rules were obtained in the R tool and RapidMiner. These rules were established by accepting the support value as 0.01 and the confidence value as 0.8. 325 association rule were obtained from the data set subjected to Apriori and FP-Gwowth Algorithm. In Figure 7, the first 12 rules are listed according to maximum support value. 15 important rules obtained from the interpretation of the rules are given below.

- Motor vehicle theft is the most common type of crime in The city of Silver Spring.
- In the city of Silver Spring, vehicle theft from property crimes is committed at a high rate in the spring.
- The city of Silver Spring has a higher crime rate on Saturday nights.
- In the city of Rockville, crimes against the community are most commonly committed on the street.
- In the city of Gaithersburg, theft against property is the most common crime.
- The police station in Silver Spring often intervenes in simple and second-degree attacks.
- In the city of Gaithersburg, crimes of property crime, damage to private property, destruction, and trespassing are often committed.

- In the state of Maryland, offenses against property and vehicle theft are often committed in the residence's vehicle path.
- Community crimes in Silver Spring ,drug, marijuana, narcotics violations are often committed in vehicles on the street.
- Community crimes in Silver Spring , the use of alcoholic vehicles and the use of alcohol in the vehicle are frequently committed.
- In the city of Gaithersburg, the crime of driving under the influence of drugs is often committed.
- Forced entry into the state of Maryland, burglary offenses are more common in single-family housing.
- In the city of Takoma Park, sudden deaths from other crimes often occur.
- In Germantown, fraud, misbehavior and fraud are committed.
- Child abduction crimes are often committed on Saturday.

Figure 8 shows a network diagram of the rules created for the city of Silver Spring. As shown in the Figure 8 ,for the city of Silver Spring, significant relationships were

No.	Premises	Conclusion	Support ↓	Confidence
27	Police District Name = SILVER SPRING	City = SILVER SPRING	0.018	0.908
22	Crime Name2 = All Other Offenses	Crime Name1 = Other	0.018	0.851
34	Crime Name1 = Other	Crime Name2 = All Other Offenses	0.018	0.991
19	Police District Name = WHEATON	City = SILVER SPRING	0.014	0.781
13	Police District Name = MONTGOMERY VILLAGE	City = GAITHERSBURG	0.012	0.722
24	City = GAITHERSBURG	Police District Name = MONTGOMERY VILLAGE	0.012	0.874
23	Event Location = Street - In vehicle	Crime Name1 = Crime Against Society	0.011	0.870
39	Crime Name2 = Drug/Narcotic Violations	Crime Name1 = Crime Against Society	0.010	1
18	Police District Name = GERMANTOWN	City = GERMANTOWN	0.009	0.765
32	City = GERMANTOWN	Police District Name = GERMANTOWN	0.009	0.986
12	City = ROCKVILLE	Police District Name = ROCKVILLE	0.009	0.697
14	Police District Name = ROCKVILLE	City = ROCKVILLE	0.009	0.722

Figure 7. Top 12 Rules Sorted by Maximum Support Value.

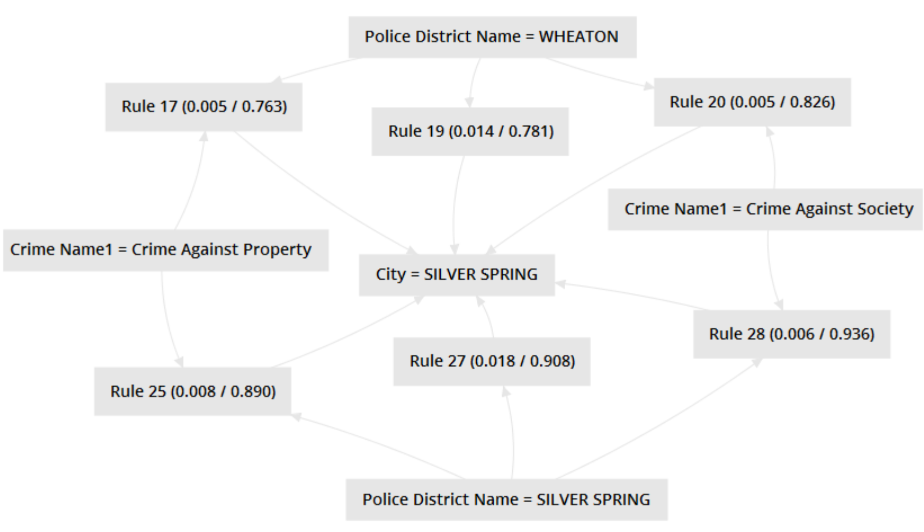

Figure 8. Network Diagram of rules produced for the city of Silver Spring

found between the police district, crime type (crime against property) characteristics. These meaningful relationships are detailed above for the city of Silver String in rule analysis.

In Figure 8 shows a graf drawing of rules created for the city of Silver Spring and Crime against property. As shown in the Figure 9, significant relationships were found between the characteristics of the city of Silver Spring according to crime types (crime against property, drug violations ,vehicle theft, auto theft). In this analysis of these meaningful relationships association rules, the identified motor vehicle theft as

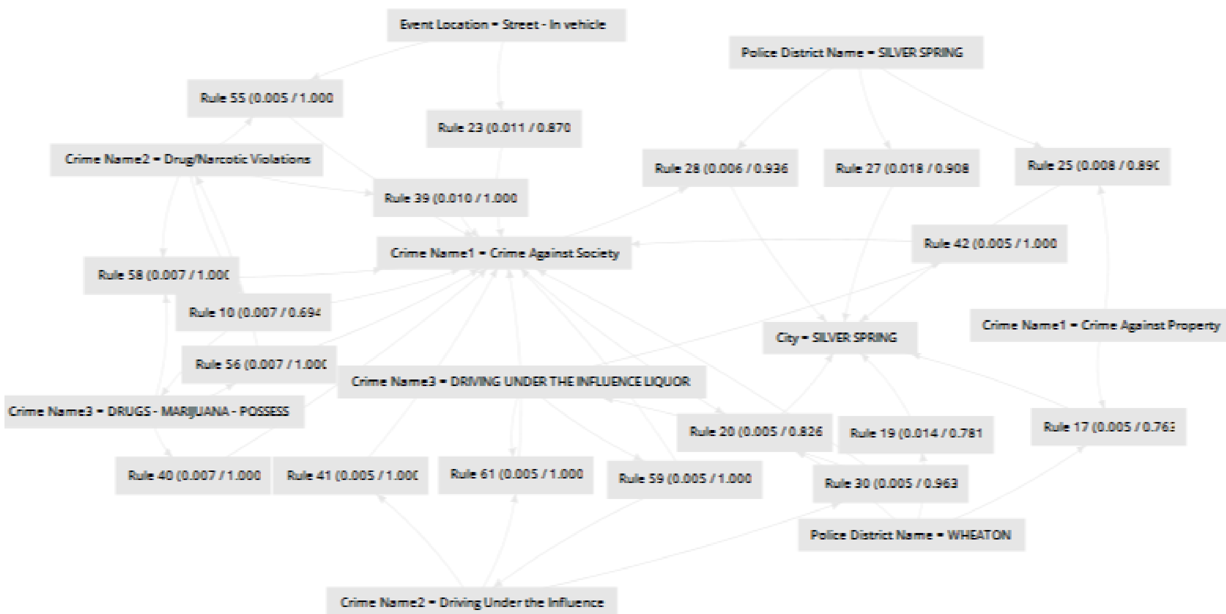

Figure 9. Graf Drawing of Rules Created For The City of Silver Spring and Crime Against Property Crime

the most common type of crime for the city of Silver Spring. Another rule was that the crime of vehicle theft from crimes against property was committed at a high rate in the spring.

4. CONCLUSION

Today's perception of security tends towards preventing a crime that has not yet happened, rather than detecting a crime that has happened. This method, called the proactive method, makes it possible to predict possible crimes in advance. Crime data mining, which is used to prevent crime from occurring, has been used successfully in many studies. Our study also revealed a new practice in this area. In this study, criminal incidents were analyzed on actual data provided by the U.S. General Services Administration. Association rules were applied on 98272 criminal data. Rapid Miner and R Tool were applied to Apriori and FP-Growth Algorithms dataset and 325 rules were created. With the obtained rules, it has become possible to observe the relationships between the attributes that constitute the crimes. In this way, especially according to the results of the analysis, regions that need to be assigned more officers, time intervals can be determined and the use of human resources in the security field can be optimized. In this way, it becomes possible to create safer habitats with the same resources. It may be possible to prevent these crimes before they occur by increasing the controls in the locations at which days and which times are determined to be more dangerous. Thus, the use of technology in a way that directly affects human life and improves the quality of life can be achieved. For example, according to the first rule we obtained from our data set, the city of Silver Spring has a high crime rate on Saturday nights. If more security checks are carried out in this city on Saturday night, meaning resources can be diverted at the right time, the projected crime rate could be reduced. In addition, rules have been obtained about what types of crimes and the scene of the crime. In the city of Silver Spring, significant associations were found between the type of crime (crime against property) characteristics in the Wheaton Police District. If more resources are allocated to the Wheaton Police District and measures are increased, the projected crime rate could be reduced.

With this application, an environment where new data can also be analyzed has been prepared and a pioneering method has been put forward for future studies.

REFERENCES

[1] Benton, W. " Encyclopedia Britannica". Encyclopedia Britannica Inc, Vol. 1,1971.

[2] Yu, C. H., Ward, M. W., Morabito, M., & Ding, W. (2011, December). Crime forecasting using data mining techniques. In *2011 IEEE 11th international conference on data mining workshops* (pp. 779-786). IEEE.

[3] Merriam-Webster-Dictionary, https://www.merriam-webster.com/dictionary ,(09,2019).

[4] Takçı, H., & Hayta, Ş. (2014). Suç Veri Madenciliği Yardımıyla Hırsızlık Suçları Hakkında Kural Çıkarımı. *Eleco Elektrik–Elektronik–Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu*, 694-699.

[5] Brown, D. E. (1998, October). The Regional Crime Analysis Program (ReCAP): a framework for mining data to catch criminals. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)* (Vol. 3, pp. 2848-2853). IEEE.

[6] Turvey, B. E. (Ed.). (2011). *Criminal profiling: An introduction to behavioral evidence analysis*. Academic press.

[7] Ünsal, Ö. (2011). *Mesleki alan seçimlerinin makine öğrenmesi algoritması kullanılarak belirlenmesi* (Doctoral dissertation, Yüksek Lisans Tezi, Gazi Üniversitesi, Bilişim Enstitüsü, Ankara).

[8] Özkan, Y. (2008). Veri madenciliği yöntemleri: Papatya Yayıncılık.

[9] Ergün, K. (n.d.). *Veri Madenciliğine Giriş*. Balıkesir Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği Bölümü, http://kergun.baun.edu.tr/veri_madenciligi_ceng_hafta1.pdf,(June 2019)

[10] Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., ... & Atabakhsh, H. (2003, May). Crime data mining: an overview and case studies. In *Proceedings of the 2003 annual national conference on Digital government research* (pp. 1-5).

[11] Ozgul, F., Atzenbeck, C., Celik, A., & Erdem, Z. (2011, July). Incorporating data sources and methodologies for crime data mining. In *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics* (pp. 176-180). IEEE.

[12] Mason, S. J. (2019). *Analysis of Virginia Crime Data of the Year 2016 Using Data Mining Techniques* (Doctoral dissertation, North Carolina Agricultural and Technical State University).

[13] Mittal, M., Goyal, L. M., Sethi, J. K., & Hemanth, D. J. (2019). Monitoring the impact of economic crisis on crime in India using machine learning. *Computational Economics*, *53*(4), 1467-1485.

[14] Thangamuthu, M. A., Vadivel, M. G., & Priyadharshini, M. A. (2019). Detecting Criminal Method using Data Mining.

[15] Ma, L., Chen, Y., & Huang, H. (2010, November). AK-Modes: A weighted clustering algorithm for finding similar case subsets. In *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering* (pp. 218-223). IEEE.

[16] Data.Gov: Montgomery County-America ,Crime, data.montgomerycountymd.gov,(March 2018).

[17] Agrawal, R., & Shafer, J. C. (1996). Parallel mining of association rules. *IEEE Transactions on knowledge and Data Engineering*, *8*(6), 962-969.

[18] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (Eds.). (1996, February). Advances in knowledge discovery and data mining. American Association for Artificial Intelligence.

[19] Srikant, R., & Agrawal, R. (1997). Mining generalized association rules. *Future generation computer systems*, *13*(2-3), 161-180.

[20] Han, J., & Fu, Y. (1995, September). Discovery of multiple-level association rules from large databases. In *VLDB* (Vol. 95, pp. 420-431).

[21] Zhao, Q., & Bhowmick, S. S. (2003). Association rule mining: A survey. *Nanyang Technological University, Singapore*, 135.

[22] J. Han, M.Kamber. (2006)."Data Mining Concepts And Techniques Second Edition" , San Francisco:Morgan Kaufman.

[23] Şeker, S. E. (n.d.). *Apriori Algoritması*. Bilgisayar Kavramlari, http://bilgisayarkavramlari.com/2011/09/07/apriori-algoritmasi/ , (Nowember 2019)

[24] Borgelt, C., & Kruse, R. (2002). Induction of association rules: Apriori implementation. In *Compstat* (pp. 395-400). Physica, Heidelberg.

[25] Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *ACM sigmod record*, *29*(2), 1-12.

[26] Foundation,What is R?, https://www.r-project.org/about.html,(Nowember 2019)

[27] RapidMiner,Data Mining Tools, https://RapidMiner.com/glossary/data-mining-tools,(Nowember 2019).

[28] *Hofmann, M., & Klinkenberg, R. (Eds.). (2016)*. RapidMiner: Data mining use cases and business analytics applications. *CRC Press*.