

Suç Veri Madenciliği Yardımıyla Hırsızlık Suçları Hakkında Kural Çıkarımı

Rule Extraction for Theft Crimes by the Help of Crime Data Mining

Hidayet Takçı¹, Şengül Hayta²

¹Bilgisayar Mühendisliği Bölümü, Cumhuriyet Üniversitesi
htakci@cumhuriyet.edu.tr

²Bilgisayar Mühendisliği Bölümü, Haliç Üniversitesi
sengulbayrak@halic.edu.tr

Özet

Suç analizi; suçun aydınlatılması ile ilgili sistematik işlemler bütünüdür. Bugüne kadar sıklıkla reaktif yöntem kullanılmıştır fakat bugünlerde proaktif yöntem kullanılmaya başlanmıştır. Proaktif yöntem için bir teknik veri madenciliğidir. Ayrıca suçu oluşturan nedenler arasında bağ kurma için de veri madenciliği uygun bir tekniktir. Veri madenciliğinin konuya uygunluğundan dolayı suç veri madenciliği kavramı ortaya çıkmıştır. Bu çalışmada hırsızlık suçu suç veri madenciliği yöntemleriyle ele alınarak suçu oluşturan unsurlar analiz edilmiştir. Analiz işlemi birliktelik kuralları madenciliği ile yerine getirilmiştir. Elde edilen kurallar ile önleyici tedbirlerin alınması imkânı vardır. Çalışmamızda geçmiş hırsızlık olayları için gerçek veri yerine gerçek veri ile benzer özelliklere sahip yapay veri kullanılmıştır. Yapay verinin üretimi için bir uygulama geliştirilmiş olup çalışmamızda detayları sunulmuştur.

Abstract

Crime analysis is a systematic set of operations about disclosure of crime. In this field, so far, reactive method has been usually used but nowadays proactive method has begun to use. One of the techniques for proactive method is data mining. Data mining is a useful technique for finding of relationships between criminal elements. Because of the data mining is suitable for crime analysis crime data mining concept has spread. In this study, relationships between criminal elements were explored for theft crime. With this aim, association rule mining was used. Obtained rules may be used for preventive measures. In the experiments artificial data was generated. This generated data has same distribution with the real crime data.

1. Giriş

Suç kavramı için bugüne kadar birçok tanım yapılmasına rağmen bunlardan en çok kabul göreni, "yasaklanan bir eylem ya da eylemler bütünü veya toplum yasaları ile sınırları çizilen bir yükümlülüğün dışına çıkmaktır" tanımıdır[1].

Asayiş suçlarından insan ticaretine, cinsel suçlardan çocuk suçluluğuna hatta bilişim suçlarına kadar suçların çok sayıda çeşidi vardır. Suçla mücadelede suçun tipine göre çeşitli yöntemler geliştirilmiş olmakla birlikte bütün suçlar için ortak bir çalışma suç analizidir.

Suç analizi; suçun oluşmadan önlenmesi, mevcut suçların ve suç eğilimlerinin tespit edilmesi ve bunlara karşı gerekli tedbirlerin alınmasını da içeren bir kavramdır[2]. Suç analizi günümüzde emniyet ve güvenlik alanında oldukça önemli bir yer tutmaktadır. Bunun başlıca sebepleri; suç ve suçlu arasında ilişki kurmak, suçların bölgesel olarak dağılımı konusunda bilgi sahibi olmak ve en önemlisi suç işlenmeden önce öngörülerde bulunabilmeyi suç analizinin sağlayabilmektedir. Suç analizinin birbiri ile bağlantılı beş farklı aşaması vardır. Bunlar; suç verilerinin toplanması, verilerin sınıflandırılması, veri analizi, elde edilen bilgilerin dağıtımı ve geribildirimlerin değerlendirilmesidir. Bu aşamaların yerine getirilmesiyle tam bir suç analizi gerçekleştirilebilir[3]. Suç analizi terimi, genel bir kavramı ve polisiye ortamda uygulanan bir disiplini belirtir. Bu disiplin, suç analizinin amaç, kapsam, veri ve analiz teknikleri olarak farklılaşan alt kümelerine ayrılır[4].

Suç analizinde önemli bir konu suçluların bir profile göre temsil edilmesidir. Bu konu; tümevarımsal, tümdengelsel ve coğrafi profillemeye olarak üç ana başlıkta incelenebilir. Tümevarımsal profillemeye, bilinen (suçu sabit görülmüş) suçluların özellikleri yardımıyla yerine getirilir. Tümdengelsel profillemeye, mağdurun verdiği yanıtlar ve bilgiler ile oluşturulur. Coğrafi profillemeye ise suçun işlendiği yerin coğrafi özelliklerinden yararlanılarak şekillendirilmektedir [5].

Suç veri madenciliği suç analizi için kullanılan en yeni tekniklerden biridir. Her geçen gün gelişen bu teknik gelecek için etkin bir yöntem olarak göze çarpmaktadır. Bununla birlikte, suç tipleri ve güvenlik bağlantıları suç veri madenciliğine önemli bir boyut getirmektedir. Bu çalışmada suç veri madenciliği alanında bir çalışma yapılarak suç

olaylarının birliktelik kuralları yardımıyla anlaşılması hedeflenmiştir.

2. Suç Analizinde Veri Madenciliği

Suç analizi, suç ve suçu oluşturan etkenlerin açıklanması için yapılan çalışmaları ifade eder. Suç analizi, gelişen bilişim teknolojileriyle birlikte çeşitli veritabanları, sayısal haritalar ve coğrafi bilgi sistemleriyle desteklenerek çok daha verimli sonuçlar veren bir yöntem haline gelmiştir. Suç analizinde son dönemde veri madenciliği yoğun olarak kullanılmaktadır [2][6][7][8][9][10][11]. Bu çalışmalardan Brown[2] tarafından yapılan çalışma dikkat çekicidir. Onun çalışmasında[2], suçluları yakalamak için veri füzyonu ve veri madenciliği çalışması etkin olarak kullanılmıştır. Veri füzyonu, özellikle organize suçlarda çoklu kaynaklardan gelen bilgilerin yönetilmesi, birleştirilmesi ve yorumlanmasında kullanılmıştır. Veri madenciliği ise suçlar arasındaki ilişkilerin ortaya çıkarılması ve analizi ile ilgili olarak kullanılmıştır.

Veri madenciliği tekniklerinden kümeleme, sapma tespiti, sınıflandırma ve birliktelik kuralları, suç analizinde kullanılan etkin yöntemlerden bazılarıdır. Bu teknikler sayesinde suç örüntülerinin tanımlanması ve bağlantısız gibi görünen veriler arasındaki ilişkilerin çıkarımı söz konusudur. Kümeleme tekniğinde, suç kayıtlarındaki benzerlik ve yakınlıklar üzerinde ilişkilendirme yapılabilmektedir. Kümeleme tekniği, bölgelere göre suç dağılımlarının yakınlığı ve uzaklığının incelenmesine olanak tanır. Sapma tespiti, kayıtlardaki anormal durumların örüntüsünün oluşturulmasında kullanılan bir teknik olarak göze çarpmaktadır. Dolandırıcılık, ağ saldırıları gibi konuların incelenmesinde kullanılan suç veri madenciliği yöntemidir. Eğer bir bölgede olağan dışı bir hareketlilik varsa bu genellikle bir olaya işaretir.

Sınıflandırma tekniği, örüntüsü bilinen suçların kategorilere ayrılmasında faydalıdır. Spam e-posta ve e-posta göndericilerinin tespiti, bu yöntemin suç analizinde kullanımı için bir örnek olarak gösterilebilir. Sosyal ağ analizi, bir ağdaki ilişkileri ve rolleri tespit etmek için kullanılabilecek teknik olarak açıklanır. Sosyal ağlar üzerinde kısırtma suçlarının tespiti bu alanla ilgilidir. Birliktelik kuralları madenciliği ise büyük veri kümeleri içerisinde nitelikler arası ilişkilerin kurulmasıyla anlamlı senaryolar oluşturmaya yardımcı olmaktadır. Bunun için eldeki veriler, suçun nitelikleri olarak gruplanır ve yapılan analiz sonucu sıklıkla tekrarlayan durumlar için kurallar oluşturulur.

Suç veri madenciliği alanında bugüne kadar yapılmış bazı önemli çalışmalara göz atılması elzemdir. Bu çalışmalar bize veri madenciliğinin suç etki alanında kullanımına iyi örnekler oluşturmaktadır.

Suç veri madenciliği alanındaki çalışmalardan COPLINK projesi önemlidir. Bu projede polis birimi ve Arizona üniversitesinden bir ekip birlikte çalışarak suç kayıtlarından varlık çıkarımı yapmışlardır[7]. Böylece suçla ilintili olanların çıkarımı mümkün olabilmektedir. Bruin ve arkadaşları[12] suçlular ve suç davranışlarını tanımlamak üzere bütün bireylere ait profilleri ikili olarak karşılaştırmışlardır. Profil karşılaştırmada; suçun ciddiyeti, suç sıklığı, suç ortamı ve suçun süresi gibi faktörler kullanılmıştır [12].

Brown[2], bölgesel suç analizi için bir analiz programı (ReCAP) önermiştir. Brown çalışmasında veri madenciliğini bir suç analizi algoritması olarak ele almıştır. Nath ve arkadaşları[10] suçları çözme prosesini hızlandırmak için suç örüntülerini tespit için K-Means kümelemeyi kullanmıştır. Adderly ve Musgrove[13] seri cinsel saldırganların suçlularını tahmin etmek için Self Organizin Map (SOM) algoritması uygulamıştır. Son dönemde yapılan çalışmalardan birinde Ozgul ve diğerleri[14] çözülmemiş terör olaylarını çözebilmek için Crime Prediction Model (CPM) isiminde bir tahmin modeli ortaya koymuşlardır. L. Ma ve diğerleri[15] ise büyük veri kümelerinden benzer olay alt kümelerini otomatik olarak bulmak için AK-Modes isimli iki adımlı bir kümeleme algoritması kullanılmıştır.

3. Önerdiğimiz Model

Suç veri madenciliğinde bugüne kadar başta kümeleme ve sınıflandırma olmak üzere çeşitli veri madenciliği teknikleri kullanılmıştır. Yapılan çalışmalarda amaç suç olaylarını çözümlenmek, suç faktörleri arasında ilişki kurmak ve suçla ilgili tahminlerde bulunmaktır. Bu çalışmada suç analizi için birliktelik kuralları madenciliği kullanılmıştır. Birliktelik kuralları madenciliği bugüne kadar başta pazar sepeti analizi olmak üzere çeşitli alanlarda başarılı şekilde uygulanmıştır. Birliktelik kuralları madenciliği açıklamalı veri madenciliği kategorisindedir ve suçun açıklanması için de uygun bir tekniktir. Suç verileri üzerinde yapılacak birliktelik kuralları madenciliği ile suçu oluşturan etkenler ve durumlar bu çalışmada ortaya konacaktır. Üzerinde çalışma yapılacak suç türü hırsızlıktır. Bu makalede hırsızlık suçu için etkili faktörler dikkate alınarak suçun meydana gelmesinde bu faktörlerin ne denli etkisi olduğu konusu ele alınacaktır.

3.1. Model Detayları

Suç veri madenciliği çalışması, diğer veri madenciliği süreçleri gibi benzer adımlarda ele alınacaktır. Dolayısıyla çalışmamız, verinin elde edilmesinden suçlarla ilgili çıkarımlarda bulunmaya kadar devam eden bir süreci içermektedir. Burada veri ve ona uygulanacak yöntem model detaylarını verecektir.

3.1.1. Suç Verisi

Suç veri madenciliği için yeteri kadar suç verisine ihtiyaç vardır ve suç verisi elde etmek için karşımızda genellikle iki seçenek bulunur. Bunlardan birincisi doğrudan emniyet teşkilatından veri alınması olup mahremiyet gibi konular nedeniyle gerçek suç verisini elde etmek kolay değildir. İkincisi bir benzetim ortamı sayesinde gerçek dünya koşullarının benzetimi yapılarak yapay veri üretimidir. Bu iki veri kaynağından daha doğru sonucu verecek olan birinci veri kaynağı olmakla birlikte taşıdığı zorluklar nedeniyle yapay veri ile çalışma seçeneği daha uygundur. Yapay veri konusunda ise veri kalitesi benzetim ortamının başarısına bağlıdır. Ne kadar gerçekçi bir sistem tasarlanırsa o kadar doğru veri elde edilecektir. Veri üretimi Crime Data Generator (CDG) adı verilen uygulama ile yapılmıştır. Yapay veri üretiminde gerçek suç istatistiklerinden faydalanarak uygulamanın istatistik tabanlı doğru bir zemine

oturtulmuştur. Üretilen verideki dağılımlar ile (ortalama ve standart sapma) gerçek verinin dağılımları birbirine benzerdir. Veri üretim aşaması modelimizin ilk adımını oluşturmaktadır. Bu adımda üretilen veriler üzerinde suç örüntüsü oluşturma, suç niteliklerinin ilişkilendirilmesi gibi işlemler diğer aşamalarda ele alınacaktır.

Bu çalışmada, yalnızca hırsızlık suçu üzerinde çalışılacaktır ve ikinci yöntem ile suç verileri üretilecek, bu veriler üzerinde analiz çalışmaları yapılacaktır. Bunun için hazırladığımız Crime Data Generator uygulaması ile suç ve suçlu nitelikleri için rastgele değerler atanacaktır. Oluşturulan veri seti üzerinde gerekli düzenlemeler yapıp, analize hazır hale getirilecek ve veri madenciliği aracı ile üzerinde analiz işlemi gerçekleştirilecektir.

3.1.2. Suç Verisi Üretimi

Veri üretimini; suçun işlendiği tarih, gün saat, mekân, mahal, mevsim, suçu işleyen profili (yaş, cinsiyet, eğitim durumu, sabıka kaydı) gibi parametreleri esas olarak yerine getirilecektir. Bu yüzden her bir parametre için olası değerleri tanımlayıp, çalıştırdığımız senaryolar yardımıyla her bir parametrenin olası değerlerinden rastgele seçimler ile örnek olay bilgileri elde edeceğiz. Analiz etmek için yeteri kadar veri toplayana kadar çalıştırma devam edecek ve çalıştırma sonrasında veride anlamsız değerler ortaya çıkarsa bunlar elle düzeltilenecektir.

Hırsızlık suçu için kullanılacak nitelikler Çizelge 1’de görüldüğü gibi suç için; yer (lokasyon), mahal, olay yerine giriş şekli, gün, mevsim olarak çeşitlendirilebilir. Suçlunun nitelikleri ise; yaşı, eğitim durumu, sabıka kaydı, cinsiyeti gibi nitelikler belirleyici olacaktır (Çizelge 2).

Çizelge 1: Suç Nitelikleri

Özellik adı	Açıklama
LOCATION	Yer (lokasyon)
REGION	Mahal
ENTRY	Olay yerine giriş şekli
DAY	Gün
SEASON	Mevsim

Çizelge 2: Suçlunun Nitelikleri

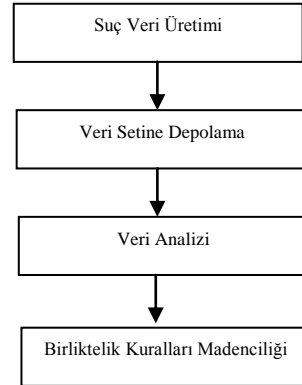
Özellik adı	Açıklama
LOCATION	Yer (lokasyon)
REGION	Mahal
ENTRY	Olay yerine giriş şekli
DAY	Gün
SEASON	Mevsim

Bu nitelikleri içeren suç kayıtlarının oluşturulması için Crime Data Generator uygulamasını hazırlanmıştır. Uygulamada rasgele veriler üretilirken, bu veriler veri setleri olarak depolanacak ve analiz için uygun hale getirilecektir.

3.1.3. CDG Programı

Crime Data Generator, Microsoft Visual Studio’da C++ dilinde tarafımızdan yazılmış bir veri üretici uygulamasıdır. Bu uygulama nesne tabanlı bir yapıya dayalı olup, oluşturduğumuz sınıflar suçun ve suçlunun birer niteliğini temsil etmektedir. Her bir nitelik uygulamanın çalıştırılması ile anlık rastgele veri üretmektedir. Bu üretilen suç verileri metin dosyasına kaydedilir ve analiz için hazır veri seti haline getirilir.

Çizelge 3: Suç Veri Üretiminde İzlenen Adımlar



3.1.4. Birliktelik Kuralları Madenciliği

Birliktelik kuralları madenciliği, bir veri tabanında yer alan nesnelere arasındaki ilişkileri açığa çıkaran bir veri madenciliği tekniğidir. Birliktelik kurallarının kullanıldığı en tipik örnek pazar sepeti uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder.

Bir birliktelik kuralı genellikle $A \Rightarrow B$ [destek = %2, güven = %60] şeklinde ifade edilir. Bu kuralın anlamı; A ve B ürününü alanların oranı bütün işlemler arasında %2’dir. A ürününü alan bir kişinin aynı zamanda B ürününü alma olasılığı %60’dır. Buradan iki bilgi elde edilir; bu işlem ne kadar sık tekrarlanmış ve iki nesne arasında ne kadar ilişki var bilgisidir.

Veri tabanı taramaları sonrasında buna benzer kurallar çıkar ki buna birliktelik kuralları adı verilir. Bu kurallardan kimisi önemli kimisi önemsiz olabilir. Kuralın önemini anlamak için iki parametre ise destek sayısı (minsup) ve güven ölçütü (minconf) bilgisidir. Yani bir kuralın önemli sayılabilmesi için sağlaması gereken minimum destek eşiği ve minimum güven eşiği nedir? Kural seçimi aşamasında yapılması gereken işlemlerin adımları;

1. Kural destek ölçütü hesaplanır. C ve D durumunu birlikte içeren toplam destek sayısı ifade edilir.

$$\text{destek}(C \rightarrow D) = \text{destek}(C,D) / N \quad (1)$$

(N: tüm veri seti)

2. C ve D durumunun birlikte olma olasılığını veren kural güven ölçütü hesaplanır.

$$\text{güven}(C \rightarrow D) = \text{destek}(C,D) / \text{destek}(C) \quad (2)$$

Birliktelik kuralları belirlenirken bir de eşik değer belirlenmelidir. Hesaplanan destek veya güven ölçütlerinin destek ve güven değerinden büyük olması gerekir.

$$\begin{aligned} \text{Support} &\geq \text{minsup} \\ \text{Confidence} &= \text{minconf} \text{ olmalıdır.} \end{aligned} \quad (3)$$

Bu çalışmada da suç verileri için birliktelik kuralları uygulanacak ve suçu oluşturan unsurlar arasındaki ilişkiler ortaya çıkarılacaktır. Birliktelik kuralları madenciliğinde en sık kullanılan Apriori algoritması uygulamamızda da tercih edilmiştir. Apriori algoritmasında izlenen adımlar ise;

- Birliktelik kurallarının çözümlenmesi için eşik değerlerinin belirlenmiştir.
- Veri tabanındaki veri seti taranarak çözümlenmeye dahil edilecek her kayıttın tekrar sayıları ve destek sayıları hesaplanır. Destek sayıları eşik destek sayısı ile karşılaştırılır. Eşik destek sayısından küçük değerlere sahip kayıtlar çıkarılmıştır.
- Seçilen verilerden ikişerli gruplandırılarak destek sayıları elde edilir. Bu değerler eşik destek sayıları ile karşılaştırılır. Eşik değerden küçük değerler kayıtlardan çıkarılmıştır.
- Üçerli, dörderli, beşerli gibi gruplandırmalar yapılarak bu grupların destek sayılarının eşik değer ile karşılaştırılmıştır.
- Kural destek ölçütüne bakılarak birliktelik kuralları türetilip her kuralın güven ölçütleri hesaplanmıştır.

4. Deneysel Çalışmalar

Deneysel çalışmalar aşamasında kendi imkânlarımızla ürettiğimiz yapay veri kullanılacaktır. Yapay veri Crime Data Generator ismini verdiğimiz uygulama yardımıyla üretilmiştir. Bu veri hırsızlık suçu ile ilgili olup veri üzerinde birliktelik kuralları madenciliği uygulanacaktır. Çalışmamızda suç kayıtlarında tekrarlanan nitelikleri bağlantılı kurallar haline getirip, anlamlı çıkarımlar oluşturulacaktır.

4.1. Veri Seti

Detayları yukarıda verilen yapay veri üretici tarafından üretilen hırsızlık verileri deneylerde kullanılacaktır. Veriler kullanılmadan önce bazı dönüşümlere tabi tutulmuştur. Veri ön işleme aşamasında öncelikle analiz değeri olmayan kısımlar veriden temizlenir. Daha sonra suç verileri veri madenciliği algoritmalarına uygun hale getirilir. Çalışmamızın birliktelik kuralları madenciliği algoritmasına uygun hale getirilebilmesi için bazı nicel-nitel değer dönüşümlerine ihtiyaç vardır. Bu konuda yapılan bazı çalışmalar aşağıda verilmiştir.

Crime Data Generator ile ürettiğimiz verilerde yaptığımız bazı dönüşümler şu şekildedir:

-Suçlu yaşı üretilirken, nicel değerler olan sayısal ifadelerden nitel değerler olan yaş aralıkları için öngördüğümüz tanımlar haline çevrilmiştir. Örneğin 18-25 yaş arası GENÇ, 26-40 yaş arası ORTA, 41-60 yaş arası ORTAUSTU, 61-80 yaş arası YASLI olarak nitelendirilmiştir.

-Suçun işlendiği tarih niteliği üzerinden farklı suç nitelikleri çıkarımı yapılmıştır. Buna göre, suçun tarihine göre haftanın hangi günü işlendiği, hangi mevsime denk geldiği gruplandırmaları yapılmıştır.

-Sabıka kaydının olup olmaması boolean değerler olarak atanırken, bu değerlere göre VAR ya da YOK şeklinde sayısal olmayan formlara dönüştürülmüştür(Çizelge 4).

4.2. Birliktelik Kuralları Madenciliğinin Uygulanması

Suçun çeşitli nitelikleri ile birliktelik kuralları ilişkilerinin kurulması(keşfedilmesi) modeli şekillendirecektir. Birliktelikler için önemli bir yöntem olan suç eşleştirme göze çarpmaktadır. Suç eşleştirmede, davranışsal veri ile bölgesel veri arasındaki bağlantılardan çıkacak örüntülerle polisin suçluyu tanımlaması sağlanabilmektedir. Bu noktada eşleştirme için benzerlik tabanlı bir yaklaşımda bulunulur. Birliktelik kuralları madenciliği algoritması Tanagra[16] isimli bir veri madenciliği aracı yardımıyla çalıştırılmıştır. Crime Data Generator ile oluşturduğumuz veri setindeki 10.000 adet veri üzerinde yapacağımız analiz için, birliktelik(association) algoritmalarından Apriori algoritması seçilmiştir.

Algoritmayı kullanırken, destek (support) değerini 0,01 ve güvenilirlik (confidence) derecesini 0,5 olarak set ettik. Bu kabuller doğrultusunda veri setimiz üzerinde algoritma çalıştırılmış ve 260 adet kural bulunmuştur. Analiz sonucunda elde edilen bazı kurallar Çizelge 5’de verilmiştir.

Çizelge 4: Crime Data Generator Tarafından Üretilen Örnek Suç Verisi

DAY	GENDER	AGE	LOCATION	EDUCATION	ENTRY	REGION	RECORD	SEASON
Mon	F	YASLI	Train	Uni	Window	Burg	YOK	Summer
Tue	F	YASLI	Mall	High	Key	Town	YOK	Autumn
Wen	M	YASLI	Home	Coll	Door	Village	YOK	Winter
Thu	M	YASLI	Bus	Uni	Window	Country	YOK	Autumn
Tue	F	YASLI	Street	High	Key	Summery	YOK	Spring
Sat	M	YASLI	School	High	Door	Burg	YOK	Autumn
Fri	M	YASLI	Bus	High	Door	Town	YOK	Winter
Fri	F	YASLI	Mall	Coll	Door	Industry	YOK	Spring
Sun	M	YASLI	Mall	Uni	Door	Summery	YOK	Winter
Sat	M	YASLI	Bus	Prim	Door	Town	YOK	Summer
Fri	F	YASLI	Train	Prim	Balcony	Village	YOK	Spring
Fri	F	YASLI	Street	Coll	Window	Burg	YOK	Winter
Thu	F	YASLI	Hospt	Prim	Key	Summery	YOK	Summer
Sat	F	YASLI	Bus	High	Window	Country	YOK	Autumn
Fri	F	YASLI	Cinema	Prim	Window	Country	YOK	Autumn
Mon	F	YASLI	Street	Coll	Key	Industry	YOK	Winter
...
Mon	M	YASLI	Home	Coll	Window	Summery	YOK	Autumn

Çizelge 5: Crime Data Generator ile Üretilen Veriler Üzerinde Tanagra ile Yapılan Analizi Sonucu Bulunan Bazı Birliklik Kuralları

N°	Kural öncesi	Kural sonrası	Lift	Support	Confidence
1	"GENDER=F" - "SEASON=Winter" - "REGION=Country"	"ENTRY=Door"	1.349	0.011	0.547
2	"GENDER=F" - "AGE=ORTAUSTU" - "LOCATION=Hospt"	"RECORD=Yok"	1.276	0.012	0.636
3	"AGE=ORTA" - "EDUCATION=Uni" - "SEASON=Summer"	"GENDER=F"	1.224	0.014	0.615
4	"RECORD=Var" - "SEASON=Autumn" - "DAY=Thu"	"GENDER=M"	1.221	0.012	0.607
5	"GENDER=M" - "AGE=ORTA" - "LOCATION=Cinema"	"RECORD=Yok"	1.213	0.010	0.605
6	"ENTRY=Door" - "SEASON=Spring" - "REGION=Industry"	"GENDER=M"	1.203	0.010	0.599
7	"ENTRY=Door" - "SEASON=Spring" - "EDUCATION=Coll"	"GENDER=M"	1.202	0.013	0.598
8	"AGE=ORTAUSTU" - "SEASON=Spring" - "EDUCATION=Coll"	"GENDER=M"	1.202	0.013	0.598
9	"REGION=Village" - "LOCATION=Home"	"GENDER=F"	1.201	0.010	0.604
10	"ENTRY=Door" - "SEASON=Winter" - "REGION=Country"	"GENDER=F"	1.201	0.011	0.603
11	"GENDER=M" - "SEASON=Winter" - "REGION=Town"	"RECORD=Var"	1.197	0.012	0.600
12	"RECORD=Var" - "AGE=ORTAUSTU" - "LOCATION=Train"	"GENDER=M"	1.196	0.010	0.595
13	"AGE=ORTAUSTU" - "EDUCATION=Uni" - "ENTRY=Window"	"GENDER=F"	1.192	0.010	0.599
14	"AGE=ORTAUSTU" - "SEASON=Summer" - "EDUCATION=Prim"	"GENDER=M"	1.162	0.012	0.578
15	"RECORD=Yok" - "EDUCATION=Uni" - "REGION=Country"	"GENDER=F"	1.181	0.013	0.593
16

4.2.1. Yorumlama

Birliktelik kuralları madenciliği uygulandıktan sonra elde edilen kurallar yukarıda yer almaktadır. Bu kuralları anlaşılır hale getirmek istediğimizde karşımıza aşağıdaki bilgiler çıkacaktır. Yorumlaması yapılan kurallar elde edilen 260 kuraldan sadece 15 tanesidir. Örnek amaçlı kullanılmıştır.

1. Kasabada yaşayan bayanlar kış mevsiminde hırsızlık yaparken kapıyı kullanırlar.
2. Hastanede hırsızlık yapan orta yaş üstü bayanların daha önceden sabıka kaydı bulunmamaktadır.
3. Üniversite mezunu, orta yaşlı bayanlar yaz aylarında hırsızlık yaparlar.
4. Sabıka kaydı olan erkekler sonbaharda Perşembe günleri hırsızlık yaparlar.
5. Sinemada hırsızlık yapan orta yaşlı erkeklerin sabıka kaydı vardır.
6. Endüstri bölgesinde bahar aylarında hırsızlık yapan erkekler giriş için kapıyı kullanırlar.
7. Lise mezunu erkekler Bahar aylarında hırsızlık yaparken giriş için kapıyı kullanırlar.
8. Orta yaşlı, lise mezunu erkekler Bahar aylarında hırsızlık yaparlar.
9. Köylerde eve girerek hırsızlık yapanlar bayandır.
10. Bayanlar kışın ilçede hırsızlık yaparken giriş yöntemi olarak kapıyı kullanırlar.
11. Sabıka kaydı olan erkekler kış aylarında şehir merkezinde hırsızlık yaparlar.
12. Sabıka kaydı olan orta yaş üstü erkekler trende hırsızlık yaparlar.
13. Üniversite mezunu, orta yaş üstü bayanlar hırsızlık yapanlar pencereyi kullanırlar.
14. Üniversite mezunu, orta yaş üstü bayanlar hırsızlık yapanlar pencereyi kullanırlar.
15. İlkokul mezunu orta yaş üstü hırsızlar yazın hırsızlık yaparlar.
16. İlçede hırsızlık yapan üniversite mezunu bayanların önceden sabıka kaydı yoktur.

Örnek olarak verdiğimiz bu kurallar gibi daha pek çok kural üretmek mümkündür. Böylelikle çıkarılan analiz sonuçları herkeşe anlaşılır hale getirilebilecektir. Analiz adımlarından sonuncusu olan geribildirimlerin toplanmasında, analiz sonuçlarının anlaşılır olması önemli bir ayrıntıdır.

5. Sonuç ve Değerlendirme

Günümüzün güvenlik algısı gerçekleşmiş bir suçu açıklamaktan çıkıp henüz gerçekleşmemiş bir suçu önlemeye doğru meyil halindedir. Proaktif yöntem adı verilen bu yöntem suçların tahmini ve olası suçların önceden görülebilmesini mümkün kılmaktadır. Suçu oluşmadan engellemede kullanılan suç veri madenciliği birçok çalışmada başarıyla kullanılmıştır. Çalışmamızda da bu alanda yeni bir uygulama ortaya konmuştur. Hırsızlık suçunun ele alındığı bu çalışmada; önce suç verileri yapay yöntemle üretilmiş ardından da üretilen bu veriler üzerinde suç analizi yapılmıştır. Yapay veri üretimi Crime Data Generator isimli bir uygulama ile yapılmış ve bu uygulama ile 10.000 adet suç verisi üretilmiştir. Kullanılan suç analiz tekniği birliktelik kuralları madenciliğidir. Birliktelik kuralları madenciliği ile suç örüntülerinin eşleştirilmesi yerine getirilmiştir. Algoritma Tanagra[16] adı verilen bir makine öğrenmesi aracıyla uygulanmış ve algoritma sonrasında 260 adet birliktelik kuralı elde edilmiştir. Elde edilen kurallar sayesinde hırsızlık suçunu oluşturan nitelikler arasındaki ilişkileri gözlemlemek mümkün hale gelmiştir. Ortaya konan uygulama ile gerçek verilerin de analiz edilebileceği bir ortam hazırlanmış ve gelecek çalışmalar için öncü bir yöntem ortaya

konmuştur. Ayrıca farklı suç türleri için de örnek bir çalışma disiplini ortaya konmuştur.

6. Kaynaklar

- [1] Webster's Online Dictionary, <http://www.websters-online-dictionary.org>.
- [2] Brown, D.E., "The Regional Crime Analysis Program (RECAP): A framework for Mining Data to Catch Criminals", *IEEE*, pp. 2848-2853, 1998.
- [3] Demirci, S., Çoban, E., "Adli Birimlerin Uygulamaları İçinde Suç Analizi Kavramı", *Polis Dergisi*, Sayı 30, Ankara, 2002.
- [4] Boba, R., "Crime Analysis and Crime Mapping", *Sage Publications*, ISBN: 0761930922, pp. 5–6, 2005.
- [5] Brent, E., Turvey, M. S., "Deductive Criminal Profiling: Comparing Applied Methodologies Between Inductive and Deductive Criminal Profiling Techniques", *Knowledge Solutions Library, Electronic Publication*, 2002.
- [6] Estivill-Castro, V., Lee I., "Data Mining Techniques for Autonomous Exploration of Large Volumes of Geo-Referenced Crime Data", *Proc. Sixth Int'l Conf. Geocomputation*, 2001.
- [7] Chen, H., Chung, W., Qin, Y., Chau, M., Xu, J. J., Wang, G., Zheng R., Atabakhsh, H., "Crime data mining: an overview and case studies", *ACM International Conference Proceeding Series. Vol. 130*, 2003.
- [8] Dahbur, K. ve Muscarello, T., "Classification system for serial criminal patterns", *Artificial Intelligence and Law Archive* Volume 11, Issue 4 table of contents pp. 251 - 269 ISSN:0924-8463, 2003.
- [9] Oatley, G. C., Zeleznikow, J. ve Ewart, B. W., "Matching and predicting crimes", In A Macintosh, R Ellis, and T Allen (eds.), *Applications and innovations in intelligent systems XII. Proceedings of AI2004 The twenty-fourth SGAI international conference on knowledge based systems and applications of artificial intelligence*, Springer: 19-32, 2004.
- [10] Nath, S.V., "Crime Pattern Detection Using Data Mining", *IEEE/WIC/ACM, pp.41-44, International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, 2006.
- [11] Thongtae, P, Srisuk, S., "An Analysis of Data Mining Applications in Crime Domain, Computer and Information Technology Workshops", Volume , Issue , 8-11 July 2008 ,pp.122 – 126, 2008.
- [12] Bruin, J.S., Cox, T.K., Kosters, W.A., Laros, J., and Kok, J.N., "Data mining approaches to criminal career analysis", in *Proceedings of the Sixth International Conference on Data Mining (ICDM'06) (ICDM'06)*, pp.171-177, 2006.
- [13] Adderly, R., Musgrove, B.P., "Data Mining case study: modeling the behavior of offenders who commit serious sexual assaults", *KDD-'01, Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, 2001, pp. 215-220.
- [14] Ozgul, F., Atzenbeck, C., Celik A., Erdem, Z., "Incorporating data Sources and Methodologies for Crime Data Mining", *IEEE proceedings*, 2011.
- [15] Ma, L., Chen, Y., Huang, H., "AK-Modes: A weighted Clustering Algorithm for Finding Similar Case Subsets", *IEEE*, pp.218-223, 2010.
- [16] <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>.