

VERİ MADENCİĞİ YÖNTEMLERİ KULLANILARAK ŞEHİR BAZINDA İŞLENEN SUÇLARIN ANALİZİ VE SUÇ EĞİLİMLERİ

ANALYSIS OF CRIMES IN URBAN BASIS USING DATA MINING METHODS AND CRIME TRENDS

Emre Ramazan AYDIN – 15542509

Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği, 23110 Elazığ, Türkiye

emramzn@gmail.com

ÖZET

Veri madenciliği, verilerin önceden bilinmeyen anlamlı bilgileri tanımlama ya da tahmin etme tekniklerini içermektedir. Günümüzde birçok ülke hukuksal ceza sistemini iyileştirme yönünde çalışmalar yapmaktadır. Bu çalışmada şehir bazında işlenen suçların analizi veri madenciliği yöntemleri kullanılarak yapılmıştır. Seçilen veri madenciliği yöntemleri doğrultusunda A.B.D 'ye ait şehirlerde 1990-2011 yılları arasında işlenen suçlar yaşlara bağlı olarak saptanmıştır.

Anahtar kelimeler: veri madenciliği, suç analizi, suç eğilimi, ceza hukuku

ABSTRACT

Data mining includes techniques for identifying or estimating meaningful data of previously unknown data. Today, many countries are working to improve the system of legal penalties. In this study, the analysis of city - based crimes was done by using data mining methods. According to selected data mining methods, the crimes committed in the USA between 1990-2011 were determined according to ages.

Keywords: Data Mining, Anlysis of Crime, Crime Trend, Criminal Justice

Giriş

Suç, günümüzde güncelliğini koruyan önemli toplumsal problemlerin başında gelmektedir. Toplumda sürekli işlenen suç olaylarının önemli güvenlik sorunu oluşturduğu bilinmektedir. Bu nedenle suç araştırmalarının günümüzde önemli çalışma konuları içerisinde yer aldığı görülmektedir. Organizasyonlarda belirli bir amaca göre tasarlanan bilgi sistemleri karar almada yardımcı ve yol gösterici kaynak olarak kullanılmaktadır. Cezai yaptırımların caydırıcılığı güncel gelişmelere bakıldığında tartışılması gereken bir konu haline gelmiştir ve büyük bir yoğunlukla uygulanan yaptırımların kimseyi suç işlemekten alıkoyma kabiliyetinin olmadığını düşünmektedir.[1] Uzmanların bu konu üzerindeki çalışmalarının sonucu aile kurumunun işlemeyişini suç oranlarının artmasında ilk sıraya koyuyor. Aile içi problemler varsa ve aile yapısı dağınıksa suç oran daha fazla oluyor.[2]

Suç işlemeye yönelten diğer bir sebep ise ekonomik şartlar olarak görülmekte. Ekonomik şartları refah içinde yaşamaya elverişli olmayan kimi bireylerin de suç işlemeye yöneldiği yadsınamaz bir durum olarak değerlendiriliyor. Bu konuda devletlerin ilgili makamların çalışmaları büyük önem taşıyor. Dünyadaki bütün devletler ülke içi asayişini sağlamakla yükümlüdür. Ülke içi huzur ve düzeni tayin etmek devletin ve hukukun asli görevidir. Bununla birlikte devletler işlenen suçlar karşısında yüksek caydırıcılığı gözeterek ceza kanunları düzenlemekle ülke içi adalet, huzur ve asayişini sağlarlar. Ülkeler bölgesel suç eğilimlerini göz önünde bulundurarak caydırıcı ceza hukuku düzenlemesine gidebilirler. Bu doğrultuda karar destek sistemleri ile çalışılması suç oranının düşmesinde büyük önem taşıyor hale gelmiştir.

1. Materyal ve Yöntem

Veri madenciliği son zamanlarda büyük gelişimler gösteren ve bilimsel çalışmalara ortam sağlayan büyük ölçüde üzerinde çalışılan bir alandır. Veri madenciliğinin temel amacı olarak veri yığınları kullanarak temel model oluşturmaktır. Oluşturulan modeller doğrultusunda çeşitli karar destek sistemleri inşa edilir. Karar destek sistemlerinde veri yığının içinde olmayan veri elde edildiğinde bu verinin yorumlanması sağlanır.

Veri madenciliği modellerini gördükleri işlemlere göre,

1-Sınıflandırma (*Classification*) ve Regresyon (*Regression*)

2-Kümeleme (*Clustering*)

3-Birliktelik kuralları (*Association Rules*)

Olmak üzere üç ana başlık altında incelemek mümkündür [3]. Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları tanımlayıcı modellerdir.[3]

2. Sınıflandırma ve Regresyon

Sınıflandırma ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğitimlerini tahmin eden modelleri kurabilen iki veri analiz yöntemidir.[4] Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır.[4] Örnek olarak sınıflandırma modeli yatırımcıların yatırımlarının riskli olma durumlarını kategorize ederken, regresyon modeli geliri ve mesleği olan potansiyel müşterilerin araba alırken yapacakları harcamaları tahmin etmek için kullanılabilir.

Sınıflandırma ve regresyon modellerinde en yaygın teknikler şunlardır:

- 1 - Yapay Sinir Ağları (Artificial Neural Network)
- 2 - K-En yakın Komşu (K- Nearest Neighbor)
- 3 - Random Forest
- 4 – Naive-Bayes
- 5- C4.5(J48)
- 6 - KStar.

- **J48 Algoritması:** C4.5 Algoritmasının Weka'ya uyarlanmış halidir. Bu algoritmanın amacı karar Ağacını optimize etmektir.
- **Yapay Sinir Ağları:** YSA insan beyninden esinlenerek ortaya çıkmış bir algoritmadır. Esas olarak biyolojik sinir ağlarını taklit eder. Veri seti ile yapılan eğitim sonucunda öğrendiği sistemle, ağa farklı örnek girişler yapıldığında nasıl davranacağına karar verir.[5]
- **K-En yakın Komşu:** Çapraz Doğrulama yöntemini kullanarak bilinmeyen bir örneğin var olduğu sınıfı belirlemek için bu örneğe en yakın olan k tane örnekleme bulan algoritmadır.[6]
- **Random Forest:** Birden fazla karar ağacı oluşturur. Doğru ve istikrarlı bir tahmin elde etmek için onları birleştirir ve karar ormanları elde eder.[7]
- **Naive-Bayes:** Olasılıksal hesaplamalarla verilerin kategorilerini belirleyen algoritmadır. Bu algoritmada esas alınan nokta bilginin sınıflandırılma şeklidir.[8]
- **KStar Algoritması:** Benzerlik fonksiyonlarıyla belirlenen örnek tabanlı bir algoritmadır. Bu algoritma eğitim sınıfındaki örnekler üzerinden test sınıfını oluşturur.

3. Uygulama

Bu çalışmanın temel amacı A.D.B 'ye ait bazı şehirlerde işlenen suçların bulunduğu veri seti üzerinde temel veri madenciliği yöntemleri uygulanmaktır. Şehirlerdeki, cinsiyet ve yaş gruplarına bağlı olarak işlenen suçların analizi yapılmaktadır. Yapılan bu analizler neticesinde şehir bazında işlenen suçlar ve suça yönelimler baz alınarak caydırıcı ceza kanunlarının oluşturulmasına olanak sağlamak ve büyük ölçüde suç işlemeyi önlemektir. Veriler üzerinde analizler, sınıflandırma algoritmalarının deneysel sonuçlarına göre doğruluk yüzdesi en fazla olan seçilmiştir.

3.1 Veri Seti

Bu çalışmada kullanılmış olan veri seti A.B.D 'de bulunan farklı şehirlerdeki yaş aralıklarına bağlı olarak 1990-2011 yılları arasında işlenen suçların kayıtları ile oluşturulmuştur. Bu veri setinde 147 nitelik ve 880 örnek bulunmaktadır.

4. Deneysel Sonuçlar

Veri setinin üzerinde sınıflandırma algoritmaları uygulanmış ve başarı oranı, duyarlılık gibi parametreler değerlendirilmiştir. Sınıflandırma algoritmaları uygulanırken veri setinin içerdiği parametreler önem arz etmektedir. Algoritma uygulanırken referans alınacak sütun seçimi, algoritmanın başarılı sonuç vermesinde büyük rol oynamaktadır. A.B.D 'nin şehirlerinde işlenen suçlar üzerinde sınıflandırma yapıldığında referans olarak alınan parametre şehir sütunu olduğunda uygulanan sınıflandırma yöntemi yüksek başarı yüzdesi vermektedir.

Yöntem	Doğruluk (%)	Duyarlılık
J48 Algoritması	89,38	0,89
Random Forest	97,50	0,97
BayesNet	95,56	0,95
Naive-Bayes	74,09	0,75

Tablo 1. Veri seti üzerinde yapılan sınıflandırma başarı oranları.

Tablo 1'de elde edilen doğruluk ve duyarlılık değerleri, veri seti üzerinde uygulanması planlanan yöntemin belirlenmesi noktasında önemli rol oynamaktadır. Bu doğrultuda veri seti üzerinde sınıflandırma yöntemi uygularken doğru sınıflandırmayı yapabilmek için doğruluk yüzdesi en fazla olan yöntemin seçilmesi söz konusudur.

4.1 Veri setinde bulunan veriler üzerinden örnek değerler.

Şehir	Yaş aralığı	Cinsiyet (Ağırlıklı)	Suç Tipi
Columbia	18 - 39	Erkek	Hırsızlık
Wahkiakum	18 - 39	Erkek	Hırsızlık
Ferry	> = 39	Erkek	Hırsızlık
Garfield	12 - 39	Kadın	Hırsızlık
San Juan	18 - 39	Erkek	Uyuşturucu Madde
Pend Oreille	18 - 39	Erkek	Uyuşturucu Madde

Tablo 2. Hırsızlık ve Uyuşturucu tacirliği suçunun şehir, yaş ve cinsiyete bağlı olarak sınıflandırılması.

Şehir	Yaş Aralığı	Cinsiyet (Ağırlıklı)	Suç Tipi
Walla Walla	18 - 39	Erkek	Cinayet
Grant	18 - 39	Erkek	Cinayet
Clallam	40 yaş üstü	Erkek	Ev Soygunu
Mason	40 yaş üstü	Erkek	Ev Soygunu

Tablo 3. Cinayet ve Ev soygunu suçunun şehir, yaş ve cinsiyete bağlı olarak sınıflandırılması.

5. Sonuç

Yaş arttıkça suça eğilimin arttığı sonucu veri madenciliği yöntemlerinin uygulanması neticesinde gözlemlenmiştir. Devletlerin ceza hukukunu düzenlerken yardımcı referans olarak ülkenin belirli bölgelerinde işlenen suçların analizini esas almaktadır. Bu doğrultuda ülkelerin caydırıcı ceza oluşturma konusunda veri madenciliği yöntemleri uygulayarak bir bölgede hangi suça ne derecede eğilim gösterildiği tespit edilerek karar verilebilir. Uygulanan veri madenciliği yöntemleri neticesinde elde edilen sonuçlar ile hukuksal kararlar alınabilir. Bu doğrultuda 1990-2011 yılları arasında A.B.D'nin şehirlerinde yaş ve cinsiyet parametrelerine bağlı olarak işlenen suçların analizi sınıflandırma yöntemleriyle yapılmıştır.

En yüksek başarıma sahip olan ve Tablo 1 'de verilen J48, BayesNet ve Random Forest yöntemleri üzerinden sonuçlar değerlendirilmiştir. Sınıflandırma yöntemlerinin uygulanması neticesinde başarı oranı yüksek algoritmalar şehirlerdeki ağırlıklı suç eğilimi yansıtmaktadır.

Kaynaklar

- [1] Zahir KIZMAZ - Öğrenim düzeyi ve Suç: Suç - okul ilişkisi üzerine sosyolojik bir araştırma. (Erişim tarihi: 05.05.2019)
- [2] Zahir KIZMAZ - Öğrenim düzeyi ve Suç: Suç - okul ilişkisi üzerine sosyolojik bir araştırma. (Erişim tarihi: 05.05.2019)
- [3] İstanbul Ticaret Üniversitesi Dergisi. Serhat ÖZEKES - Veri madenciliği modelleri ve uygulama alanları. (Erişim tarihi : 06.05.2019)
- [4] Data Mining Concepts and Techniques, Han, J.-Kamber, M., Morgan Kaufmann Publishers, 1st Ed., San Francisco, USA, 2000. (Erişim tarihi : 06.05.2019)
- [5]http://www.derinogrenme.com/2017/03/04/yapay-sinir-aglari_
(Erişim Tarihi:14.05.2019)
- [6]<https://veribilimcisi.com/2017/07/20/k-en-yakin-komsu-k-nearest-neighborsknn/>
(Erişim Tarihi : 14.05.2019)
- [7]<https://devhunteryz.wordpress.com/2018/09/20/rastgele-ormanrandom-forest-algoritmasi/> (Erişim Tarihi : 14.05.2019)
- [8]<http://bilgisayarkavramlari.sadievrenseker.com/2008/12/21/bayes-aglari-bayesian-network> (Erişim Tarihi :14.05.2019)