Pamukkale Univ Muh Bilim Derg, 24(2), 266-275, 2018

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: Adana örneği

Analysis of injury traffic accidents with machine learning methods: Adana case

Cevher ÖZDEN[1], *Çiğdem ACI*[2*]

[1]Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Çukurova Üniversitesi, Adana, Türkiye.
efeozden@gmail.com

[2]Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Mersin Üniversitesi, Mersin, Türkiye.
caci@mersin.edu.tr

Geliş Tarihi/Received: 21.07.2016, Kabul Tarihi/Accepted: 13.12.2016 doi: 10.5505/pajes.2016.87847
* Yazışılan yazar/Corresponding author Araştırma Makalesi/Research Article

Öz

Bu çalışmada, Adana ilinde 2005 ile 2014 yılları arasında meydana gelen yaralanmalı trafik kazalarına ait aylık bazdaki sayısal veriler ile aynı yıllara ait aylık bazdaki meteorolojik verilerden oluşturulan bir veri kümesi kullanılarak yaralanmalı kaza sayısı ve yaralı sayısı tahmini yapacak modeller geliştirilmiştir. Tahmin modellerinde, İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı (İBÇK-YSA), Fonksiyon Uydurma Yapay Sinir Ağı (FU-YSA), Genelleştirilmiş Regresyon Yapay Sinir Ağı (GR-YSA), Regresyon Ağacı (RA), Destek Vektör Makinesi (DVM) ve Çoklu Doğrusal Regresyon (ÇDR) Analizi yöntemleri kullanılmıştır. Çalışma sonucunda, DVM yönteminin her iki tahmin senaryosunda da en başarılı sonuçları verdiği görülmüştür. Yaralanmalı Kaza Sayısı tahminlerinin RA yöntemi dışında Yaralı Sayısı tahminlerinden daha başarılı olduğu saptanmıştır. Ayrıca, önceki yıllarda gerçekleşen kazalara ait yol ve hava verilerini kullanarak gelecek yıllar için uygun önlemler almanın mümkün olduğu sonucuna varılmıştır.

Anahtar kelimeler: Trafik kazası, Tahmin modeli, Makine öğrenmesi

Abstract

In this study, a dataset is created using numeric data of injury traffic accidents in monthly base between 2005 and 2014 years in Adana province and meteorological data of the same years in order to develop prediction models which estimate number of traffic accidents involving injury and number of injured people. Feedforward Multilayer Artificial Neural Network, Function Fitting Artificial Neural Network, Generalized Regression Artificial Neural Network, Regression Tree, Support Vector Machine and Multiple Linear Regression Analysis methods were used in the prediction models. As a result of the study, SVM gives the most successful results for both prediction scenarios. Prediction of the number of traffic accidents involving injury is more successful than prediction of number of injured people except Regression Tree method. In addition, it has concluded that it is possible to take precautions using road and weather data of the accidents which occurred in previous years.

Keywords: Traffic accident, Prediction model, Machine learning

1 Giriş

Yarattığı maddi ve manevi kayıplar nedeniyle, modern taşımacılık sisteminin en önemli ve üzerinde en çok durulan kısmını kazalar oluşturmaktadır. Trafik kazalarının oluşmasına birbirinden oldukça farklı çok sayıda değişken sebep olduğu için, trafik güvenliğini sağlamak oldukça karmaşık ve zor bir alan olarak ortaya çıkmaktadır. Özellikle sınırlı sayılabilecek ülke kaynaklarının doğru ve sonuç getirecek önlemlere sarf edilebilmesi için trafik kazalarının arkasında yatan sebeplerin doğru olarak belirlenmesi gerekmektedir. Kazalardan sonra trafik memurlarınca tutulan kaza tutanak ve analizleri yardımıyla ana faktörler orantısal olarak belirlenmektedir. Buna göre trafik kazalarına sebep olan faktörler olarak insan, taşıt, yol ve hava koşulları öne çıkmaktadır [1]. İnsan faktörü %99 ile neredeyse tek başına kazaların tamamını açıklamaktadır. Hava koşulları ile trafik kazaları arasında doğrudan bağlantı kurulmasının zorluğundan ötürü trafik kazalarının ana faktörleri arasında hava koşulları yer almamaktadır. Bu sebeple hava koşullarının insan, yol, taşıt gibi diğer kaza faktörlerine olan etkisi ise incelenmesi gereken bir alandır.

Literatürde farklı yöntemlerle geliştirilmiş birçok kaza tahmin modeli bulunmaktadır. Bu çalışmalar ya istatistiksel ya da makine öğrenmesi yöntemleri kullanılarak gerçekleştirilmiştir. Shon ve Shin [2] Yapay Sinir Ağı (YSA), Lojistik Regresyon, Karar Ağacı yöntemlerini kullanarak Kore'de meydana gelen trafik kazalarının şiddetini etkileyen faktörleri incelemişlerdir ve üç model sonuçlarında belirgin bir farklılık bulamamışlardır. Özgan ve diğ. [3] Sivas il sınırları içinde 1999-2002 yılları arasında meydana gelen kazalara ait raporları kullanarak sonraki yıllarda meydana gelebilecek kazaları istatistiksel yöntemlerle tahmin etmeye çalışmışlardır. Chong ve diğ. [4] trafik kazalarında yaralanma derecesini tahmin etmek amacıyla YSA, Regresyon Ağacı (RA), Destek Vektör Makinesi (DVM) yöntemleri ve YSA ile RA yöntemlerini karma şekilde kullanan hibrit bir model yer almıştır. Geliştirilen bu hibrit modelde kaza verileri önce RA ile düğümlere ayrılmakta daha sonra kaza verileri bulundukları düğüm bilgileriyle birlikte YSA modeline aktarılmakta ve ağın eğitimi bu verilerle gerçekleştirilmektedir. Yapılan analiz ve karşılaştırmalar sonucunda geliştirilen hibrit modelinin diğer tüm bireysel yöntemlerden daha iyi sonuç verdiği görülmüştür. Chang ve Wang [5] ise Karar Ağacı tekniğini kullandıkları çalışmalarında araç tipinin kaza şiddetine etki eden en önemli faktör olduğu sonucuna ulaşmışlardır. Buna göre motosiklet ve bisiklet sürücüleri ile yayalar kazadan daha olumsuz etkilenmektedirler. Murat ve Şekerler [6] Denizli ilindeki trafik kazaları açısından önemli noktaları belirlemek amacıyla k-ortalamalar ve Bulanık Mantık yöntemlerini kullanmıştır.

Qiuping ve Subing [7] Grey tahmin modeli ve YSA modelini kombinasyon halinde kullanarak trafik kazalarını tahmin etmeye çalışmışlardır. Çalışma sonuçları, artma ve rasgele dağılım olmak iki temel karakteristiğe sahip trafik kazalarının daha doğru tahmin edilebildiğini göstermektedir. Hibrit model, Grey Tahmin Modeli ve YSA modelinin tek başlarına uygulanmasıyla elde edilen tahmin değerlerinden daha yüksek tutarlılık göstermiştir. Durduran [8] Konya-Afyon otoyolunda meydana gelen kazalara neden olan faktörleri incelemek amacıyla Coğrafik Bilgi Sistemi tabanlı bir platform ile DVM ve YSA yöntemlerini kullanmıştır. Trafik kazalarını beş faktörlü tabandan en yüksek korelasyonu sağlayan tek boyuta indirgeyerek gelecek trafik kazalarını tahmin etmiş ve tahmin tutarlılığını YSA için %61.79 ve DVM için ise %67.42 olarak bulmuştur. Akgüngör ve Doğan [9], Regresyon Analizi, YSA ve Genetik Algoritma yöntemlerini kullanarak İzmir ili için trafik kaza tahmin modelleri geliştirmişlerdir. Modellerin geliştirilmesinde, nüfus, araç sayısı ve kaza sayısı parametrelerine ait 1986-2005 yılları arasına değerler kullanılmıştır. Analiz sonucunda, Smeed ve Andreassen kaza modellerinin daha iyi performans sergilediği, eğitim periyodunda YSA, test periyodunda ise Genetik Algoritma kaza modelinin en iyi performansa sahip olduğu görülmüştür. Wang ve diğ. [10] Bulanık Mantık tabanlı bir kaza tahmin modeli geliştirmişler ve bu modeli Harbin şehrinde meydana gelen 133 kaza verisine uygulamışlardır. Analiz sonucunda 41 kurallı bir tahmin modeline ulaşmışlar ve üretilen tahmin sonuçlarının yüksek tutarlılık gösterdiğini gözlemlemişlerdir. Çalışma sonucunda, bulanık mantık yönteminin diğer yaygın kullanılan YSA, DVM gibi makine öğrenimi yöntemlerine alternatif olarak kaza tahmin modellerinde başarıyla kullanılabileceği sonucuna ulaşılmıştır. Yu ve Abdel-Aty [11] DVM yöntemini kullanarak Amerika Colorado eyaletinde yer alan dağlık bir otoyol için gerçek zamanlı kaza riski analizi gerçekleştirmiştirler. Veri kümesi olarak 2010-2011 yılları arasında meydana gelen trafik kazalarını ve otoyol üzerinde kurulu 30 trafik gözetleme radarından gelen anlık verileri kullanmışlardır. RA yöntemi ile en önemli açıklayıcı değişkenleri belirledikten sonra bu değişkenleri kullanarak anlık kaza analizi için üç farklı Bayesian lojistik regresyon modeli geliştirmişlerdir. Daha sonra bu regresyon modelleri ile çeşitli kernel fonksiyonları kullanarak geliştirdikleri DVM modellerinin tutarlılığını sınamışlardır. Çalışma sonucunda, Radyal temelli kernel fonksiyon ile geliştirilen DVM modelinin en başarılı tahmine ulaştığı, ayrıca güçlü ve az sayıda değişken ile yapılan DVM analizlerinin daha başarılı olduğu sonucuna varmışlardır. Li ve diğ. [12] trafik kazalarında yaralanma derecesini analiz etmek amacıyla 326 farklı otoyol noktasında meydana gelen kaza verilerini kullanmışlardır. Analiz için Sıralı Probit Modeli ve DVM yöntemini kullanmışlar ve sonuçları karşılaştırmışlardır. Çalışma sonucunda %44.0 doğru tahmin oranına ulaşan Sıralı Probit modeline göre, DVM yönteminin %48.8 ile nispeten daha yüksek doğru tahmin gerçekleştirdiği sonucuna ulaşılmıştır. Çelik ve Oktay [13] Erzurum-Kars otoyolunda 2008-2013 yılları arasında meydana gelen 11.771 trafik kazasını incelemişlerdir. Kazaları yaralanma derecesine bağlı olarak ölümlü, yaralanmalı ve maddi hasarlı olmak üzere sınıflandırdıktan sonra multinominal lojik analiz yöntemiyle yaralanma derecesine etki eden faktörleri belirlemişlerdir. Buna göre, 65 yaş üstü sürücü, ilkokul mezunu sürücü, tek araçlı kazalar, otoban veya şehirlerarası yolda meydana gelen kazaların ve kaza alanında yaya kaldırımı bulunması faktörlerinin ölümlü kaza riski üzerinde etkili faktörler olduğu sonucuna ulaşılmıştır. Kwon ve diğ. [14] Kaliforniya Otoyol Devriyesi ekiplerince 2004-2010 yılları arasında tutulan kaza tutanaklarını incelemişler ve çalışmalarında Naif Bayes ile Karar Ağacı sınıflandırma yöntemlerini kullanarak kaza faktörlerini nispi önemlerine göre belirlemişlerdir. Çalışma sonucunda risk faktörlerinin büyük kısmının birbirine güçlü derecede bağlı olduğu ve kazalara etki eden sadece birkaç önemli faktörün söz konusu olduğunu bulmuşlardır.

İncelenen çalışmalarda trafik kaza analizlerinin son yıllarda üzerinde önemle durulan bir alan olduğu görülmüştür. Bu alandaki çalışmalarda, istatistiksel modeller bağımlı ve bağımsız değişkenler arasında önceden tanımlanmış sabit varsayımlara göre hareket etmektedir ve bu varsayımların çiğnenmesi halinde tutarlı sonuç vermemektedir. Diğer taraftan, makine öğrenmesi yöntemleri bağımlı ve bağımsız değişkenler arasında önceden tanımlı herhangi bir ilişkinin varlığını gerektirmemekte ve tahmin problemlerinde oldukça başarılı sonuçlar vermektedir.

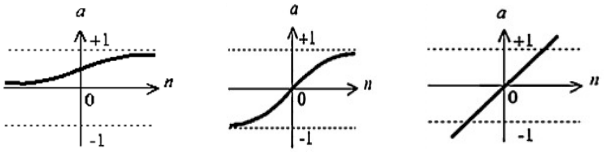
2 Yöntem

Bu bölümde tahmin modellerinde kullanılan makine öğrenmesi yöntemleri hakkında özet bilgiler verilecektir.

2.1 İleri beslemeli çok katmanlı (İBÇK) yapay sinir ağı ve fonksiyon uydurma (FU) yapay sinir ağı

İBÇK, yapay sinir ağları içerisinde yapısı itibariyle en basit olanıdır. Girdi, gizli ve çıktı olmak üzere üç katmandan oluşur. Girdi katmanı sisteme verilerin girdiği ve gizli katmana doğru işlendiği katmandır. Bu ağlarda her işlem döngüsünde hata geriye doğru dağıtılarak öğrenme gerçekleştirilmektedir. Lineer olmayan bir aktivasyon fonksiyonu ağırlıklı girdi sinyallerinin toplamını nakletmektedir. Gerçek gözlem sonuçları ağın sonuçları ile karşılaştırılmakta ve ağın hatası hesaplanmaktadır. Daha sonra hesaplanan ağ hatası sistem tarafından geriye yayılmakta ve katsayıların ağırlıkları güncellenmektedir [16]. İBÇK genel olarak bir veya daha fazla gizli sigmoid nöron katmanı içerir. Bu nöronlardan sonra lineer nöronların çıktı katmanı yer alır. Ağ, transfer fonksiyonlu nöron katmanları vasıtasıyla, girdi ve çıktı vektörleri arasındaki lineer olan ve olmayan ilişkileri öğrenir. Daha sonra ağ, lineer çıktı katmanı ile -1 +1 sınırları dışında değerler üretir [17].

YSA transfer fonksiyonları ile girdi ve çıktı parametreleri kullanılarak olayların tepkileri benzetilir [18]. Log-sigmoid transfer fonksiyonu (LOGSIG) yaygın olarak kullanılan transfer fonksiyonlarından birisidir. Bu fonksiyonla, 0 ile 1 aralığında çıktı üretilir. Bir diğer alternatif olan hiperbolik tanjant transfer fonksiyonunda (TANSIG) ise çıktı, bipolar sigmoid şeklinde ve -1 +1 aralığındadır. Pureline (PURELIN) transfer fonksiyonunda ise lineere yakın değerler temsil edilebilmektedir [19] (Şekil 1).



(a): LOGSIG. (b): TANSIG. (c): PURELIN.

Şekil 1: Transfer fonksiyonları.

2.2 Genelleştirilmiş regresyon (GR) yapay sinir ağı

GR-YSA radyal tabanlı sinir ağlarının bir çeşididir ve çekirdek regresyon ağlarına dayanmaktadır. Geri beslemeli ağların aksine, GR-YSA'da döngüsel eğitim prosedürüne gerek yoktur.

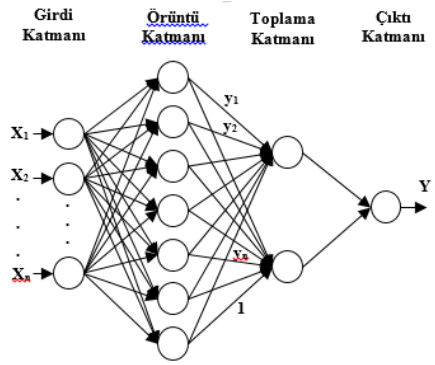
Girdi ve çıktı vektörleri arasında rasgele fonksiyonlarla yakınsaklaştırılır ve fonksiyon tahmini doğrudan eğitim kümesinden alınır. Eğitim kümesinin büyüklüğü arttıkça tahmin hatası sıfıra yakınsar [20].

Girdi katmanında her bir girdi bireysel işlem parametrelerine denk gelir. Girdi katmanı örüntü katmanı tamamen bağlıdır ve her birim bir eğitim kalıbını temsil eder, ayrıca çıktısı girdi değerinin saklanan örüntü değerlerinden uzaklığına olan mesafesidir. Toplama katmanında S- ve D- toplama nöronları olmak üzere iki nöron bulunmaktadır. Her bir örüntü katmanı birimi bu iki nörona bağlıdır. Örüntü katmanının ağırlıklı çıktı toplamı S-toplama nöronu tarafından hesaplanır. Diğer taraftan, örüntü katmanının ağırlıklı olamayan toplamı D-toplama nöronu tarafından hesaplanır. Çıktı katmanında her bir S-toplama nöronunun çıktısı her bir D-toplama nöronuna bölünür, böylece bilinmeyen x girdi vektörü için tahmini Y_i değeri bulunur [20] (Şekil 2).

$$Y_i' = \frac{\sum_{i=1}^{n} y_i \cdot \exp[-D(x,x_i)]}{\sum_{i=1}^{n} \exp[-D(x,x_i)]} \quad (1)$$

$$D(x,x_i) = \sum_{k=1}^{m} \left(\frac{x_i - x_{ik}}{\sigma}\right)^2 \quad (2)$$

Denklem (1) ve Denklem (2)'den anlaşılacağı üzere, Y_i, örüntü katmanındaki *i.* nöron ile S-toplama nöronu arasındaki bağlantının ağırlığını, *n* eğitim kalıbı sayısını temsil *etmektedir*. Gauss fonksiyonu ve girdi vektöründeki öğe sayısı sırasıyla *D* ve *m* ile gösterilmektedir. x_k, x vektörünün *j.* elemanı ve x_{ik} x_i vektörünün *j.* elemanını göstermektedir. Yayılma parametresinin optimum değeri deneysel olarak belirlenir ve σ ile gösterilir. Geleneksel GR-YSA uygulamalarında örüntü katmanındaki tüm birimler aynı tek yayılma parametresine sahiptir [20].



Şekil 2: GR-YSA'nın şematik gösterimi.

2.3 Regresyon ağacı (RA)

Bu yöntem, sınıflandırma, tahmin ve regresyon için sıkça kullanılan bir veri madenciliği yaklaşımıdır (Şekil 4). Kolay yorumu ve anlaşılabilirliği sayesinde karar vericiler için avantaj sağlamaktadır. Entropiye dayalı Sınıflandırma Ağaçları (ID3, C4.5) ve RA olmak üzere iki kategoride birçok algoritma önerilmiştir [21].

Eğitim verisindeki hangi alanların hangi sırada kullanılarak ağacın oluşturulacağı belirlenmelidir. Bu amaçla en yaygın olarak kullanılan ölçüm, Entropi ölçümüdür. Entropi, iletilen bir mesajın taşıdığı enformasyonun beklenen değeridir. Entropi ölçüsü ne kadar fazla ise o alan kullanılarak ortaya konulan sonuçlar da o oranda belirsiz ve kararsızdır. Bu nedenle, karar ağacının kökünde Entropi ölçüsü en az olan alanlar kullanılır. Denklem (3)- Denklem (6) arasında verilen Entropi (H) terimi, tüm durumlara ait P_i olasılıklarına bağlı bir değerdir [21].

$$H(X) = E\big(I(X)\big) \quad (3)$$

$$= \sum_{1\le i\le n} P(X_i).I(X_i) \quad (4)$$

$$= \sum_{i=1}^{n} P(X_i) log_2 \frac{1}{P(X_i)} \quad (5)$$

$$= -\sum_{i=1}^{n} P_i log_2 P_i \quad (6)$$

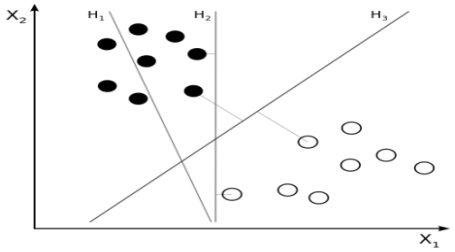
RA'da sınıflar yoktur ve veriler süreklidir. Dallanma işlemi, ayrımlar için tahmin edilen toplam varyansın minimize olması gerekliliği anlamına gelen "artıkların karelerini azaltma algoritmasına" göre gerçekleştirilir. Regresyon ağacında her düğümde minimizasyon problemi Denklem (7)'deki gibi çözülür [22],[23].

$$argmin_{X_j \le X_j^R, j=1\ldots M} [P_l Var(Y_l) + P_r Var(Y_r)] \quad (7)$$

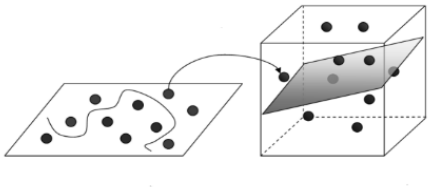
Burada, P_l ve P_r sırasıyla sol ve sağ düğümlerin olasılıklarıdır. *M* eğitim kümesindeki değişkenlerin sayısıdır. X_j *j.* değişkeni, X_j^R ise bu değişkenin en iyi ayrım değerini göstermektedir.

2.4 Destek vektör makinesi (DVM)

DVM, sınıflandırma ve regresyon problemlerinde çalıştırılabilen bir yöntemdir. DVM regresyon metodundaki temel fikir, eldeki eğitim verilerinin karakterini mümkün olduğunca gerçeğe yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uyan doğrusal ayırıcı fonksiyonun bulunmasıdır. Sınıflandırmaya benzer bir şekilde regresyonda da doğrusal olmayan durumların işlenebilmesi için çekirdek fonksiyonları kullanılır. DVM'lerde karşılaşılabilecek iki durum, verilerin lineer olarak ayrılabilecekleri bir yapıda olması veya lineer olarak ayrılamayan yapıda olmasıdır (Şekil 3 ve Şekil 4) [24]. Örneklerin lineer olarak ayrılabilmesi durumunda, iki sınıfı da eş uzaklıkta bölen en uygun ayırıcı düzlemin bulunması amaçlanır [25].



Şekil 3: Lineer örnek uzayı.



Şekil 4: Non-lineer örnek uzayı.

$$(x_i, y_i), x_i \in R^n, y_i \in \{1, -1\}, i = 1, \dots, l \quad (8)$$

Denklem (8) için DVM modeli Denklem (9)'da verilmiştir;

$$\min_{w,b} \frac{1}{2} w^t w + C \sum_{i=1}^{l} \delta(w, b; x_i, y_i) \quad (9)$$

$\delta(w, b; x_i, y_i)$ kayıp fonksiyonu; $C \geq 0$ ise eğitim hatası üzerinde ceza parametresidir. Yaygın olarak kullanılan iki kayıp fonksiyonu Denklem (10) ve Denklem (11)'de verilmiştir;

$$\max(1 - y_i(w^T \sigma(x_i) + b), 0) \quad (10)$$

$$\max(1 - y_i(w^T \sigma(x_i) + b), 0)^2 \quad (11)$$

Burada, σ eğitim verilerini daha yüksek boyutlu bir uzaya taşımak için kullanılan fonksiyonu temsil etmektedir. Her x test verisi için karar fonksiyonu ise Denklem (12)'de verilmiştir;

$$f(x) = sgn(w^T \sigma(x) + b) \quad (12)$$

DVM eğitimi için lineer, polinomsal ve RBF (Radial Basis Function) fonksiyonlarından daha iyi sonuç veren RBF çekirdek (kernel) fonksiyonu kullanılmıştır. Formülü Denklem (13)'te verilmiştir;

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \quad (13)$$

$$\gamma = 1/2\sigma^2 \quad (14)$$

Denklemde $\|x_i - x_j\|^2$ vektörel mesafeyi göstermektedir. γ parametresi serbest değişken olup, fonksiyonun genişliğini belirlemektedir. Değeri arttıkça çan şeklindeki fonksiyonun genişliği daha dar olmakta, azaldıkça genişlik artmaktadır.

2.5 Çoklu doğrusal regresyon analizi (ÇDR)

ÇDR analizi, iki ya da daha çok değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur. Eğer tek bir değişken kullanılarak analiz yapılıyorsa buna tek değişkenli regresyon, birden çok değişken kullanılıyorsa çok değişkenli regresyon analizi olarak isimlendirilir. Regresyon analizi, değişkenler arasındaki ilişkinin varlığı, eğer ilişki var ise bunun gücü hakkında bilgi verir [26].

ÇDR yönteminde iki veya daha çok sayıda bağımsız değişken veya bağımsız değişken fonksiyonu bulunur. Örneğin, p tane bağımsız ve 1 tane bağımlı değişken için çoklu doğrusal regresyon modeli Denklem (14)'te verilmiştir:

$$y_i = \beta_0 + \beta_1 x_i + \dots + \beta_p x_{pi} + \epsilon_i, i = 1, m \quad (14)$$

Buradan hareketle aşağıdaki çok değişkenli regresyon olan Denklem (15) elde edilir;

$$yi = \beta_0 + \sum_{j=i}^{p} X_{ij} \beta_{ij} + \epsilon_i \quad (15)$$

3 Veri Kümesi ve tahmin modelleri

Bu bölümde yaralı sayısı ve yaralanmalı trafik kazalarının analizi için kullanılan veri kümesi ve makine öğrenmesi yöntemlerinin bu verilere nasıl uygulandığı açıklanacaktır.

3.1 Veri kümesi ve ön işleme

Çalışmada, trafik kazaları ve hava parametrelerini içeren ve 2005-2014 yıllarını kapsayan veri kümesi kullanılmıştır. Trafik kazalarına ait bilgiler Adana Trafik Şube Müdürlüğünden, hava parametreleri ise Adana Meteoroloji Bölge Müdürlüğünden alınmıştır. Kullanılan meteorolojik değişkenler: Aylık Sisli Gün Sayısı, Donlu Gün Sayısı, Aylık Ortalama, Sıcaklık (°C), Aylık Minimum Sıcaklıkların Ortalaması (°C), Aylık Maksimum Yağış (mm), Aylık Maksimum Sıcaklıkların Ortalaması (°C), Kuvvetli Rüzgârlı Günler Sayısı, Aylık Toplam Yağış (mm), Aylık Kırağılı Günler Sayısı, Aylık Fırtınalı Günler Sayısı, Aylık Toprak Üstü Minimum Sıcaklığın -0.1 °C ve Küçük Olduğu Günler Sayısı ile Aylık Ortalama Rüzgâr Hızı (msn) seklinde 12 adettir. Kaza raporlarından ise kaza anında havanın ve yolun özelliklerini gösteren; Açık, Sisli, Yağmurlu, Gündüz, Gece, Asfalt Yol, Parke Yol, Bölünmüş Yol, İki Yönlü Yol, Düz Yol ve Viraj olmak üzere 11 adet değişken seçilmiştir. İki veri kümesinden elde edilen toplam bağımsız değişken sayısı 23 olup, her bir değişkene ait veri satırı 118'dir. Bağımlı değişkenler Yaralanmalı Kaza Sayısı ve Yaralı Sayısı olmak üzere 2 adettir. Hazırlanan veri kümesi üzerinde İBÇK-YSA, FU- YSA, GR-YSA, RA, DVM ve ÇDR Analizi yöntemleri kullanılarak yaralanmalı kaza sayısı ve yaralı sayısı tahmini yapacak *modeller* geliştirilmiştir. Geliştirilen tahmin modellerinde kullanılan veri kümesi 10-katlı çapraz doğrulama kullanılarak rastgele eğitim ve test kümelerine ayrılmıştır. Tahmin modellerinin başarısını değerlendirmek için R-kare, Ortalama Mutlak Hata (Mean Absolute Error, MAE), ortalama hata karekök (Root Mean Squared Error, RMSE) ve bağıl karesel hatanın karekökü (Relative Squared Root Error, RSE) kriterleri esas alınmıştır.

İBÇK-YSA, FU-YSA, GR-YSA, RA ve DVM yöntemleri MATLAB ortamında kodlanmıştır. ÇDR Analizi ise MS-Excel ortamında gerçekleştirilmiştir. Analizler öncesi veriler üzerinde çeşitli veri ön işleme teknikleri kullanılmıştır. Bu teknikler için MATLAB yazılımının MapMinMax, MapStd, ProcessPCA fonksiyonlarından faydalanılmıştır [27]. MapMinMax, her veri kümesinin minimum değeri -1 ve maksimum değeri +1 olacak şekilde normalleştirme yapmaktadır. MapStd, her veri kümesinin ortalama değeri 0 ve standart sapması 1 olacak şekilde normalleştirme yapmaktadır [27]. ProcessPCA ise, Temel Bileşenler Analizi ile her veri sırasındaki aralarında yüksek korelasyon bulunan bağımsız değişkenleri temsil edecek faktör tanımlayarak veri kümesinde indirgeme yapmaktadır [15]. Bu yöntemlerden RA tahmin modeli için MapMinMax en iyi sonucu verirken, diğer tüm tahmin modellerinde en iyi sonucu MapStd yöntemi vermiştir. Temel Bileşenler Analizi ise nispeten sınırlı bir iyileşme sağlamış ve bu sebeple çalışmada kullanılmamıştır.

3.2 Tahmin modelleri

İBÇK-YSA, FU-YSA, GR-YSA ve RA yöntemleri MATLAB yazılımı kullanılarak kodlandıktan sonra modellerin en iyi performansı gösterdiği parametrelerin bulunması amacıyla farklı parametreler ile çalıştırılıp sonuçlar kıyaslanmıştır.

İBÇK-YSA ve FU-YSA için 10 adet gizli katman kullanılmış olup Levenberg-Marquardt (trainlm) öğrenme fonksiyonu ile eğitilmiştir. GR-YSA için kullanılan genişlik (spread) parametresi 2.4 olarak belirlenmiştir.

ÇDR Analizi, makine öğrenmesi yöntemlerini istatistiksel bir yöntemle kıyaslama yapılabilmesi için eklenmiş olup 10-katlı çapraz doğrulama bu yöntem için gerçekleştirilmemiştir. Analiz Excel ortamında gerçekleştirilmiş ve takip eden parametreler

kullanılmıştır: Aylık Kırağılı Günler Sayısı (X1), Aylık Fırtınalı Günler Sayısı (X2), Aylık Toplam Yağış (mm) (X3), Aylık Toprak Üstü Minimum Sıcaklığın -0.1 °C ve Küçük Olduğu Gün Sayısı (X4), Aylık Ortalama Rüzgâr Hızı (msn) (X5) ile kaza anında havanın ve yolun özelliklerini gösteren; Açık (X6), Sisli (X7), Yağmurlu (X8), Gündüz (X9), Gece (X10), Asfalt Yol (X11), Parke Yol (X12), Bölünmüş Yol (X13), İki Yönlü Yol (X14), Düz Yol (X15) ve Viraj (X16). Yaralı sayısı ve yaralanmalı kaza sayısı bağımlı değişkenler olarak tanımlanmıştır. Analiz sonucunda elde edilen regresyon denklemindeki katsayılar ve test verileri kullanılarak denklemin tutarlılığı sınanmıştır.

Yaralanmalı kaza sayısı ile gerçekleştirilen ÇDR analizinin Anova Testi sonuçlar Tablo 1'de, analiz sonuçları ise Tablo 2'de görülmektedir.

Tablo 1: Yaralanmalı kaza sayısı-anova testi.

	Reg.	Fark	Top.
df	16	89	105
HT	240838.3	451.5	241289.8
OKH	15052.4	5.07	
F	2967.3		
Önem.	4.90E-114		
R Kare	0.99		

df: Serbestlik derecesi. HT: Hata toplamı. OKH: Ortalama karesel hata. F: Sıfır hipotezi için F testi. Önem: Önem seviyesi. R Kare: Belirlilik katsayısı. Fark: Gözlemlenen değer-tahmin edilen değer. Reg. : Regresyon. Top. : Toplam.

Tablo 2: Yaralanmalı kaza sayısı regresyon analizi sonuçları.

	Katsayılar	Standart Hata	t Stat	P-değeri
Kesişim	-5.763	2.142	-2.690	0.009
X1	0.315	0.250	1.259	0.211
X2	0.000	0.097	0.000	1.000
X3	-0.013	0.007	-1.910	0.059
X4	-0.118	0.153	-0.770	0.443
X5	1.200	1.523	0.788	0.433
X6	0.057	0.026	2.189	0.031
X7	0.182	0.080	2.277	0.025
X8	0.148	0.049	2.997	0.004
X9	0.520	0.085	6.134	0.000
X10	0.571	0.094	6.054	0.000
X11	0.224	0.052	4.306	0.000
X12	0.233	0.087	2.659	0.009
X13	-0.037	0.024	-1.563	0.122
X14	0.007	0.019	0.388	0.699
X15	0.238	0.085	2.790	0.006
X16	0.238	0.096	2.483	0.015

Katsayılar: Regresyon denkleminde bağımsız değişkenlere ait katsayılar. Standart hata: Regresyon katsayılarına ait hata. t-Stat: t-istatistik değeri: P-değeri: İstatistiksel anlamlılık derecesi.

Benzer şekilde yaralı sayısı bağımlı değişkeni için gerçekleştirilen analiz sonuçları Tablo 3 ve Tablo 4'te verilmiştir.

Tablo 3. Yaralı sayısı-anova testi.

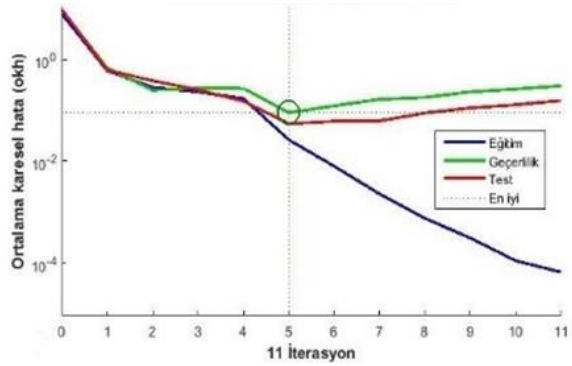
	Reg.	Fark	Top.
df	16	89	105
HT	240838.3	451.5	241289.8
OKH	15052.4	5.07	
F	2967.3		
Önem.	4.90E-114		
R Kare	0.99		

Tablo 4: Yaralı sayısı regresyon analizi sonuçları.

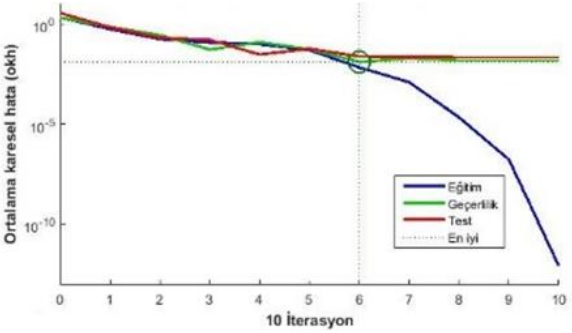
	Katsayılar	Standart Hata	t Stat	P-değeri
Kesişim	-11.963	13.523	-0.885	0.379
X1	-1.767	1.581	-1.118	0.267
X2	0.476	0.613	0.776	0.440
X3	0.090	0.045	2.019	0.047
X4	1.082	0.966	1.120	0.266
X5	-5.839	9.612	-0.608	0.545
X6	0.504	0.164	3.066	0.003
X7	0.370	0.503	0.735	0.464
X8	0.293	0.311	0.943	0.348
X9	0.088	0.536	0.164	0.870
X10	0.209	0.596	0.351	0.726
X11	0.822	0.328	2.506	0.014
X12	0.513	0.552	0.930	0.355
X13	0.052	0.149	0.353	0.725
X14	-0.078	0.121	-0.648	0.519
X15	0.048	0.538	0.090	0.929
X16	1.059	0.604	1.753	0.083

Regresyon analizleri sonucunda elde edilen denklem katsayıları ve test verileri kullanılarak modelin tutarlılığı sınanmış ve sonuçlar çalışmanın bulgular kısmında Tablo 15 ve Tablo 16'da verilmiştir.

Şekil 5'te İBÇK–YSA modelinin hem Yaralı Sayısı hem de Yaralanmalı Kaza Sayısı tahmini için uygulanan eğitimine ait performans gösterilmiştir. Buna göre model Yaralı Sayısı tahmini en iyi eğitim performansı olan 0.0896 hata değerine beşinci iterasyon sonunda, Yaralanmalı Kaza Sayısı tahmini için 0.0134 olan hata değerine ise altıncı iterasyon sonunda ulaşmıştır.



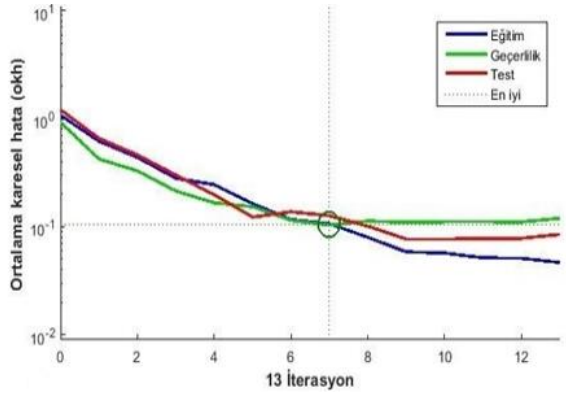
(a): Yaralı sayısı.



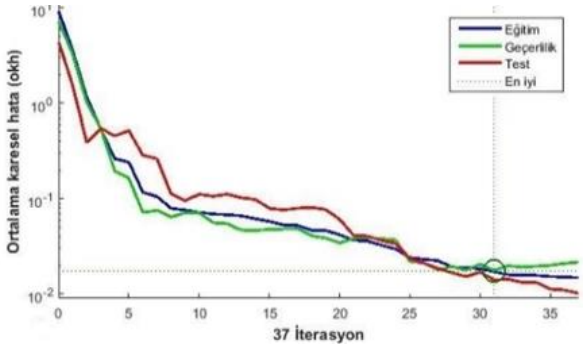
(b): Yaralanmalı kaza sayısı.

Şekil 5: İBÇK-YSA eğitim eğrisi.

Benzer olarak, Şekil 6'da FU-YSA modelinin her iki tahmin parametresi için uygulanan eğitimine ait performans gösterilmiştir. Buna göre Yaralı Sayısı tahmini için en iyi eğitim performansı olan 0.1039 hata değerine yedinci iterasyon sonunda, Yaralanmalı Kaza Sayısı tahmini için en iyi performansı olan 0.0177 değerine ise otuz birinci iterasyon sonunda ulaşmıştır.



(a): Yaralı sayısı.



(b): Yaralanmalı kaza sayısı.

Şekil 6: FU-YSA eğitim eğrisi.

4 Bulgular ve tartışma

Modellerin performansının değerlendirilmesinde R-kare, MAE, RMSE ve RSE esas alınmıştır [28]. Denklem (16)'dan Denklem (20)'ye kadar verilen formüllerde N, test için kullanılan veri sayısı; 0_i , tahmin edilen değer; t_i, gözlenen gerçek değer ve $\bar{t}$ ise gözlenen değerlerin ortalamasını temsil etmektedir.

$$R^2 = \sqrt{1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - O_m)^2}} \quad (16)$$

MAE, ölçüm değerleri ile model tahminleri arasındaki mutlak hatayı sorgular ve bu değer sıfıra yaklaştıkça, modelin tahmin yeteneğinin o kadar iyi olduğunu gösterir [15].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|o_i - t_i| \quad (17)$$

RMSE, ölçüm değerleri ile model tahminleri arasındaki hata oranını belirlemek amacıyla kullanılmaktadır. Sıfıra yakın değerler, tahmin gücünün yüksek olduğunu göstermektedir [15].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(0_i - t_i)^2}{N}} \quad (18)$$

RAE, ölçüm değerleri ile model tahminleri arasındaki göreli mutlak hatayı yüzdesel olarak sorgular. Yüzde değer (%) sıfıra yaklaştıkça, modelin tahmin gücü yüksek olur [15].

$$RAE\ (\%) = \frac{\sum_{i=1}^{N}\left|\frac{(0_i - t_i)}{o_i}\right|}{N}\ x\ 100 \quad (19)$$

RSE, ölçüm değerleri ile model tahminleri arasındaki göreli hata oranını belirlemek amacıyla kullanılmaktadır. Bu değerin sıfıra yaklaşması, modelin tahmin gücünün yüksek olduğunu gösterir [15].

$$RSE\ (\%) = \sqrt{\frac{\sum_{i=1}^{N}(0_i - t_i)^2}{\sum_{i=1}^{N}(ti - \bar{t})^2}}\ x\ 100 \quad (20)$$

Yaralanmalı Kaza Sayısı ve Yaralı Sayısı tahminleri için 10 katlı çapraz doğrulama kullanılarak elde edilen sonuçlar ve bu sonuçların ortalamaları alınarak modellerin performansı değerlendirilmiştir. 10 katlı çapraz doğrulama sonuçları çalışmanın sonunda Tablo 5'den Tablo 16'ya kadar verilmiştir. Modellere ait ortalama değerler ise Tablo 15 ve Tablo 16'da gösterilmiştir.

Yaralı Sayısı tahmini için yapılan analiz ortalamalarına göre (Tablo 15); DVM en yüksek R-kare, en düşük RAE ve RSE oranlarına sahiptir. FU-YSA ise MAE ve RMS kriterlerine göre en başarılı yöntem olmuştur. Diğer taraftan, GR-YSA ve İBÇK-YSA'ya ait sonuçlar bu modellerin yüksek tahmin başarısını göstermektedir. Ancak, RA ve ÇDR metotları tüm mukayese ölçütlerinde oldukça kötü performans sergilemiştir. DVM yöntemine ait en yüksek ve en düşük R-kare değerleri sırasıyla 0.984 ve 0.883 olmuştur. GR-YSA yöntemi için en yüksek ve en düşük R-kare değerleri 0.969 ve 0.907 olurken bu değerler İBÇK-YSA için 0.970 ve 0.765; FU-YSA için 0.963 ve 0.826; RA yöntemi için 0.836 ve 0.377 olmuştur.

Tablo 5: GR-YSA modeline ait yaralı sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.940	0.256	0.317	63.392	33.118
2	0.955	0.233	0.288	43.769	30.030
3	0.936	0.299	0.376	38.394	39.263
4	0.935	0.214	0.294	40.657	30.656
5	0.969	0.221	0.258	25.553	26.910
6	0.957	0.205	0.275	65.610	28.671
7	0.907	0.293	0.378	131.422	39.465
8	0.946	0.243	0.284	86.096	29.650
9	0.954	0.261	0.300	53.981	31.338
10	0.913	0.296	0.366	144.922	38.187
Ort.	0.941	0.252	0.313	69.379	32.729

Tablo 6: İBÇK-YSA modeline ait yaralı sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.902	0.259	0.342	42.395	35.693
2	0.765	0.397	0.602	100.550	62.888
3	0.970	0.224	0.273	37.992	28.493
4	0.901	0.288	0.333	67.061	34.754
5	0.892	0.325	0.391	60.368	40.831
6	0.904	0.271	0.361	32.839	37.654
7	0.911	0.315	0.358	84.913	37.342
8	0.855	0.279	0.440	37.224	45.981
9	0.969	0.153	0.186	19.914	19.387
10	0.915	0.266	0.319	66.232	33.297
Ort.	0.898	0.278	0.360	54.949	37.632

Tablo 7: FU-YSA modeline ait yaralı sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.885	0.280	0.370	46.966	38.680
2	0.937	0.205	0.243	144.494	25.351
3	0.911	0.301	0.320	265.031	33.391
4	0.947	0.204	0.245	26.246	25.585
5	0.920	0.201	0.285	33.280	29.809
6	0.954	0.195	0.232	35.094	24.263
7	0.963	0.245	0.296	32.879	30.935
8	0.910	0.207	0.293	29.116	30.636
9	0.903	0.264	0.323	267.246	33.691
10	0.826	0.308	0.406	1573.348	42.410
Ort.	0.916	0.241	0.301	245.370	31.475

Tablo 8: RA modeline ait yaralı sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.669	33.956	42.635	10.966	72.887
2	0.817	25.970	32.985	9.155	43.833
3	0.377	47.263	56.724	19.754	96.259
4	0.774	36.474	48.695	11.183	52.657
5	0.727	35.275	42.495	12.653	59.902
6	0.555	48.622	63.120	17.158	70.156
7	0.818	28.766	33.710	9.941	46.865
8	0.627	36.486	44.439	14.817	69.194
9	0.836	29.421	34.217	12.331	41.038
10	0.734	28.786	35.279	11.560	51.818
Ort.	0.693	35.102	43.430	12.952	60.461

Tablo 9: DVM modeline ait yaralı sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.984	5.194	8.276	1.983	12.541
2	0.977	8.496	11.387	3.019	15.942
3	0.955	13.642	17.691	5.061	22.284
4	0.883	16.979	21.571	7.374	34.863
5	0.919	18.065	25.580	5.556	31.469
6	0.953	16.622	18.388	6.678	23.080
7	0.971	9.492	11.747	4.130	17.001
8	0.925	14.600	21.255	5.968	30.721
9	0.978	10.748	13.358	3.950	17.471
10	0.888	13.629	21.456	5.978	34.643
Ort.	0.943	12.747	17.071	4.970	24.002

Tablo 10: GR-YSA modeline ait yaralanmalı kaza sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.963	0.262	0.315	71.720	32.859
2	0.979	0.234	0.275	87.568	28.739
3	0.980	0.245	0.301	59.605	31.479
4	0.972	0.218	0.254	32.402	26.561
5	0.945	0.283	0.341	72.703	35.614
6	0.986	0.191	0.222	32.708	23.212
7	0.990	0.197	0.234	201.047	24.454
8	0.977	0.250	0.299	40.176	31.216
9	0.963	0.225	0.270	47.418	28.221
10	0.962	0.271	0.341	29.386	35.594
Ort.	0.972	0.238	0.285	67.473	29.795

Tablo 11: İBÇK-YSA modeline ait yaralanmalı kaza sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.992	0.075	0.092	7.693	9.609
2	0.991	0.107	0.124	17.136	12.941
3	0.971	0.148	0.178	55.536	18.604
4	0.972	0.119	0.184	34.489	19.217
5	0.982	0.105	0.130	24.217	13.571
6	0.993	0.064	0.090	19.216	9.377
7	0.995	0.065	0.077	10.769	7.998
8	0.985	0.117	0.133	16.907	13.907
9	0.989	0.079	0.105	11.880	10.950
10	0.986	0.087	0.116	14.539	12.078
Ort.	0.986	0.097	0.123	21.238	12.825

Tablo 12: FU-YSA modeline ait yaralanmalı kaza sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.789	0.321	0.442	61.183	46.211
2	0.969	0.154	0.199	27.919	20.830
3	0.976	0.136	0.179	24.024	18.732
4	0.983	0.103	0.126	13.670	13.136
5	0.984	0.118	0.147	31.014	15.310
6	0.986	0.086	0.118	14.912	12.282
7	0.962	0.180	0.224	91.619	23.360
8	0.889	0.239	0.330	45.302	34.470
9	0.982	0.103	0.139	35.632	14.523
10	0.975	0.128	0.153	23.453	16.022
Ort.	0.949	0.157	0.206	36.873	21.488

Tablo 13: RA modeline ait yaralanmalı kaza sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.779	19.738	30.468	10.174	49.071
2	0.708	18.449	31.490	8.064	59.023
3	0.724	23.098	32.527	11.503	56.863
4	0.724	21.932	31.222	13.135	53.432
5	0.575	21.483	32.737	10.599	73.328
6	0.650	29.774	36.458	13.544	66.053
7	0.804	18.360	24.548	10.406	55.596
8	0.497	31.334	34.450	18.858	95.230
9	0.564	28.943	35.061	17.089	68.170
10	0.828	19.047	23.504	10.061	44.079
Ort.	0.685	23.216	31.246	12.343	62.084

Tablo 14: DVM modeline ait yaralanmalı kaza sayısı tahmin sonuçları.

Kat	R^2	MAE	RMSE	RAE	RSE
1	0.996	2.698	3.492	1.523	7.126
2	0.998	1.427	2.569	0.700	5.378
3	0.996	2.870	3.773	1.969	9.168
4	0.997	2.303	3.313	1.187	5.762
5	0.996	2.548	3.165	1.256	6.458
6	0.996	2.462	3.553	1.506	7.325
7	0.996	3.810	4.156	2.616	8.819
8	0.994	4.043	4.656	2.141	7.976
9	0.993	2.891	3.855	1.765	8.502
10	0.986	2.893	3.669	1.925	12.451
Ort.	0.995	2.795	3.620	1.659	7.897

Tablo 15: Yaralı sayısı ortalama analiz sonuçlarının karşılaştırılması.

Yöntem	R^2	MAE	RMSE	RAE	RSE
GR-YSA	0.941	0.252	0.313	69.380	32.729
İBÇK-YSA	0.898	0.278	0.360	54.949	37.632
FU-YSA	0.916	0.241	0.301	245.370	31.475
RA	0.693	35.102	43.430	12.952	60.461
DVM	0.943	12.747	17.071	4.970	24.002
ÇDR	0.482	52.998	71.719	14.067	184.599

Yaralanmalı Kaza Sayısı tahmini için yapılan analiz ortalamalarına göre (Tablo 16); DVM yine en yüksek R-kare. en düşük RAE ve RSE oranlarına sahiptir. En düşük MAE değerini FU-YSA, en düşük RMSE değerini ise İBÇK-YSA sergilemiştir. GR-YSA her hangi bir kriterde en iyi değere sahip olmasa da yüksek başarı göstermiştir. RA tüm mukayese ölçütlerinde oldukça kötü performans sergilerken. ÇDR yüksek R-kare ve düşük RAE değeriyle Yaralanmalı Kaza Sayısı tahmininde başarılı olmuştur. DVM yöntemine ait en yüksek ve en düşük R-kare değerleri sırasıyla 0.998 ve 0.986 olmuştur. GR-YSA

yöntemi için en yüksek ve en düşük R-kare değerleri 0.990 ve 0.945 olurken bu değerler İBÇK-YSA için 0.995 ve 0.971; FU-YSA için 0.986 ve 0.789; RA yöntemi için 0.828 ve 0.497 olmuştur.

Tablo 16: Yaralanmalı kaza sayısı ortalama analiz sonuçlarının karşılaştırılması.

Yöntem	R^2	MAE	RMSE	RAE	RSE
GR-YSA	0.972	0.238	0.285	67.473	29.795
İBÇK-YSA	0.986	0.097	0.123	21.238	12.825
FU-YSA	0.950	0.157	0.206	36.873	21.488
RA	0.685	23.216	31.247	12.343	62.084
DVM	0.995	2.795	3.620	1.659	7.897
ÇDR	0.935	15.577	18.813	5.909	75.822

5 Sonuç

Bu çalışmada çeşitli meteorolojik parametreler ve yol durumu ile kazaların şiddeti ve meydana geliş sıklığı arasındaki ilişki incelenmiştir. Bu amaçla, altı farklı model kullanılarak yaralı sayısı ve yaralanmalı kaza sayısı tahmin edilmiş ve model sonuçları karşılaştırılmıştır. Buna göre her iki tahmin senaryosu için de DVM beş başarı kriterinin üçünde (R-kare, RAE, RSE) en yüksek performansı göstererek ön plana çıkmıştır. Yaralı sayısı tahmininde ise iki kriterde en iyi sonucu veren FU-YSA ile yüksek R-kare değerlerine sahip GR-YSA ve İBÇK-YSA yöntemleri de başarılı sayılabilir. Yaralı sayısı tahmin senaryosunda tüm kriterlerde en kötü tahmin sonucunu ise ÇDR analizi vermiştir. Yaralanmalı Kaza Sayısı tahmininde ise DVM'in haricinde, birer kriterde en iyi değeri veren İBÇK-YSA ve FU-YSA ile yüksek R-kare değerine sahip GR-YSA ve ÇDR başarılı sonuçlar vermiştir. Bu tahmin senaryosunda RA en kötü tahmini gerçekleştirmiştir.

Alınan sonuçlardan anlaşılmaktadır ki ayrıntılı meteorolojik veriler ile kaza verileri birleştirildiğinde yaralanmalı kaza sayısı ve yaralı sayısı tahmini yapacak modeller yüksek doğrulukla geliştirilebilmektedir. Sıradan kaza tutanaklarında hava durumu parametreleri çok genel tutulmaktadır (sisli, yağmurlu vs.). Daha ayrıntılı veriler kullanıldığında ise herhangi bir ilde o günkü hava koşullarında oluşabilecek yaralanmalı kaza sayısı ve kazalardaki yaralı sayısı tahmin edilebilmektedir. Geliştirilen tahmin modelleri sadece Adana için değil diğer iller için de genelleştirilebilecek yapıdadır.

Çalışmada kullanılan trafik veri kümesinde sürücü ile ilgili bilgilerin bulunmaması sebebiyle kazalardaki sürücü faktörü incelenememiştir. İleriki çalışmalarda bu faktörün de bulunduğu daha detaylı ve daha fazla örneklem içeren veri kümeleri kullanılarak daha ayrıntılı analizlerin yapılması planlanmaktadır.

6 Kaynaklar

[1] Baş Fİ. Geçici Fiziksel Özelliklerden Yorgunluk ve Uykusuzluğun Sürücü Davranışı Üzerine Etkisi. Yüksek Lisans Tezi, Atatürk Üniversitesi, Erzurum, Türkiye, 2015.

[2] Sohn SY, Shin H. "Pattern recognition for road traffic accident severity in Korea". *Ergonomics*, 44(1), 107-117, 2010.

[3] Özgan E, Ulusu H, Yıldız K. "Trafik kaza verilerinin analizi ve kaza tahmin modeli". *SAU Fen Bilimleri Enstitüsü Dergisi*, 8(1), 160-166, 2004.

[4] Chong M, Abraham A, Paprzycki M. "Traffic accident analysis using machine learning paradigms". *Informatica*, 29, 89-98, 2005.

[5] Chang L, Wang H. "Analysis of traffic i njury severity: an application of non-parametric classification tree techniques", *Accident Analysis and Prevention,* 38(5), 1019-1027, 2006.

[6] Murat YŞ, Şekerler A. "Trafik kaza verilerinin kümelenme analizi yöntemi ile modellenmesi". *İMO Teknik Dergi*, Yazı 311, 4759-4777, 2009.

[7] Qiuping W, Subing L. "An information renewal GNN model for road traffic accident forecasting", *International Conference on Transportation Engineering*, Chengdu, China, 25-27 July 2009.

[8] Durduran SS. "A decision making system to automatic recognize of traffic accidents on the basis of GIS platform", *Expert Systems with Applications,* 37(12), 7729–7736, 2010.

[9] Akgüngör AP, Doğan E. "Farklı yöntemler kullanılarak geliştirilen trafik kaza tahmin modelleri ve analizi", *International Journal of Engineering, Research and Development*, 2(1), 16-22, 2010.

[10] Wang H, Zheng L, Meng X. "Traffic accidents prediction model based on fuzzy logic", *Advances in Information Technology and Education,* 201, 101-108, 2011.

[11] Yu R, Abdel-Aty M. "Utilizing support vector machine in real-time crash risk evaluation", *Accident Analysis and Prevention*, 51, 252-259, 2013.

[12] Li Z, Liu P, Wang W, Xu C. "Using support vector machine models for crash ınjury severity analysis", *Accident Analysis and Prevention*, 45, 478-486, 2012.

[13] Çelik AK, Oktay E. "A multinomial logit analysis of risk factors in fluencing road traafic injury severities in the Erzurum and Kars Provinces of Turkey", *Accident Analysis and Prevention*, 72, 66-77, 2014.

[14] Kwon OH, Rhee W, Yoon Y, "Application of classification algorithms for analysis of road safety risk factor dependencies", *Journal of Accident Analysis and Prevention*, 75, 1-15, 2015.

[15] Yapraklı TŞ, Erdal H. "Firma başarısızlığı tahminlemesi: makine öğrenmesine dayalı bir uygulama", *Bilişim Teknolojileri Dergisi*, 9(1), 21-31, 2016.

[16] Garret GJ, Wu C, "Knowledge-based modeling of material behavior with neural networks". *Journal of Engineering Mechanics*, 117(1), 132-153, 1991.

[17] Sen Z. *Yapay Sinir Ağları İlkeleri*, Su Vakfı Yayınları, İstanbul, Türkiye, 2004.

[18] MATLAB, "Neural Network Toolbox". http://cda.psych.uiuc.edu/matlab_pdf/nnet.pdf, (29.11.2016).

[19] Patil SR. Regionalization of an Event Based Nash Cascade Model for Flood Predictions in Ungauged Basins. PhD Thesis, Institute für Wasserbau, Aachen, Deutchland, 2008.

[20] Okkan U, Dalkılıç HY, "Radyal tabanlı yapay sinir ağları ile kemer barajı aylık akımlarının modellenmesi". *İMO Teknik Dergi*, 379, 5957-5966, 2012.

[21] Orhan U, Hekim M, Özer M. "EEG signals classification using the k-means clustering and a multilayer perceptron neural network model", *Expert Systems with Applications*, 38(10), 13475-13481, 2011.

[22] Özkan K. "Sınıflandırma ve regresyon ağacı tekniği (srat) ile ekolojik verinin modellenmesi". *Süleyman Demirel Üniversitesi Orman Fakültesi Dergisi*, 13, 1-4, 2012.

[23] Şengür D, Tekin A. "Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini", *Bilişim Teknolojileri Dergisi,* 6(3), 7-16, 2013.
[24] Stack Overflow. "What is the relation between the number of Support Vectors and training data and classifiers performance?". http://stackoverflow.com/questions/9480605/what-is-the-relation-between-the-number-of-support-vectors-and-training-data-and (29.11.2016).
[25] Kaya H, Gündüz-Öğüdücü Ş. "A distance based time series classification framework". *Information Systems*, 51, 27-42, 2015.
[26] Sykes AO. "An ıntroduction to regression analysis", *Chicago Working Paper in Law & Economics,* 20, 1993.
[27] Karacabey AA, Gökgöz F. *Emeklilik Fonlarının Portföy Analizi*. 1. Baskı, Ankara, Türkiye, Siyasal Kitabevi, 2005.
[28] Yakut E, Elmas B, Yavuz S. "Yapay sinir ağları ve destek vektör makineleriyle borsa endeksi tahmini", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1), 139-157, 2014.