T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TRAFİK KAZASI OLUŞUMU ETKENLERİNDEN YARARLANARAK MAKİNE ÖĞRENME YÖNTEMLERİ İLE KAZA ŞİDDETİ TAHMİNİ

Mehmet Ayhan ÇETİN

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

ŞANLIURFA
2020

Dr. Öğr. Üyesi Kerim KARADAĞ danışmanlığında, Mehmet Ayhan ÇETİN'in hazırladığı **"Trafik Kazası Oluşumu Etkenlerinden Yararlanarak Makine Öğrenme Yöntemleri İle Kaza Şiddeti Tahmini**" konulu bu çalışma 29/05/2020 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Harran Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

İmza

Danışman : Dr. Öğr. Üyesi Kerim KARADAĞ .....

Üye : Dr. Öğr. Üyesi Mehmet Emin TENEKECİ .....

Üye : Dr. Öğr. Üyesi Faruk KÜRKER .....

Bu Tezin Elektrik Elektronik Mühendisliği Anabilim Dalında Yapıldığını ve Enstitümüz Kurallarına Göre Düzenlendiğini Onaylarım.

Doç. Dr. İsmail HİLALİ
Enstitüsü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

İÇİNDEKİLER

Sayfa No

ÖZET i
ABSTRACT ii
TEŞEKKÜR iii
ŞEKİLLER DİZİNİ iv
ÇİZELGELER DİZİNİ v
SİMGELER ve KISALTMALAR DİZİNİ vi
1. GİRİŞ 1
1.1. Avrupa Ülkelerine Ait Trafik Kaza Bilgileri 2
1.2. Trafik Kazalarını Etkileyen Faktörler 4
1.2.1. Karayolu tipinin ve trafik hacminin trafik kazalarına etkisi 4
1.2.2. Şerit genişliğinin trafik kazalarına etkisi 5
1.2.3. Banket genişliğinin trafik kazalarına etkisi 5
1.2.4. Taşıt hızının trafik kazalarına etkisi 6
1.2.5. Yatay kurbaların trafik kazalarına etkisi 7
1.2.6. Boyuna eğimin trafik kazalarına etkisi 7
1.3. Kaza Tahmin Yöntemleri 8
1.4. Makine Öğrenmesi ve Çeşitleri 9
2. ÖNCEKİ ÇALIŞMALAR 11
3. MATERYAL ve YÖNTEM 16
3.1. Materyal 16
3.2. Yöntem 17
3.2.1. Normalizasyon 18
3.2.1.1. Min-Max normalizasyonu 18
3.2.1.2. İstatiksel veya z-score normalizasyonu 19
3.2.1.3. Medyan normalizasyonu 19
3.2.1.4. Sigmoid normalizasyonu 20
3.2.1.5. D min max normalizasyonu 20
3.2.2. Destek vektör makinaları (DVM) 21
3.2.2.1. Destek vektör makinelerinde regresyon 23
3.2.2.2. Destek vektör makineleri yönteminde çapraz doğrulama 24
3.2.3. K- en yakın komşu algoritması (KNN) 24
3.2.3.1. KNN parametreleri 25
3.2.4. Karar ağaçları (KA) 26
4. ARAŞTIRMA BULGULARI ve TARTIŞMA 29
4.1. Bulgular ve Tartışma 29
4.1.1. ROC eğrisi ve AUC değeri 31
4.1.2. Çapraz doğrulama (cross validation) 31
4.1.3. Algoritma sonuçları 32
5. SONUÇLAR ve ÖNERİLER 43
KAYNAKLAR 44
ÖZGEÇMİŞ 48

ÖZET

Yüksek Lisans Tezi

TRAFİK KAZASI OLUŞUMU ETKENLERİNDEN YARARLANARAK MAKİNE ÖĞRENME YÖNTEMLERİ İLE KAZA ŞİDDETİ TAHMİNİ

Mehmet Ayhan ÇETİN

Harran Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Kerim KARADAĞ
Yıl: 2020, Sayfa: 48

Trafik kazaları genellikle ulaşımın temel unsurları olan insan, taşıt, karayolu ve çevre koşulları sonucunda ortaya çıkan ve maddi hasar, yaralanma, ölüm gibi sonuçlara sebebiyet veren olayların bütünüdür. Trafik kazalarını etkileyen birçok faktör sayılsa da ülkemizde genelde ilk akla gelen etken sürücü hataları olarak göze çarpmaktadır. Bir trafik kazası oluştuğunda öncelikle raporlar tutulur ve sürücülere hatalar paylaştırılır. Çünkü kaza raporunu tutan görevlilerin yolun geometrisi vb. konularda bilgisi yoktur. Oysaki sürücü hatalarının yanında; hava koşulları, yolun geometrik özellikleri gibi birçok faktör trafik kazalarının oluşmasında önemli etkenlerdendir. Yapılan çalışmada Yeni Zelanda Ulaştırma Ajansından alınan; yolların eğimi, şerit sayısı, aydınlığı, kuru veya ıslaklığı, olabilecek hız sınırı, yollarda bulunan yaya, bisikletli ve ağaç gibi bilgilerin bulunduğu veri seti kullanılarak yollarda oluşan kazaların ciddi veya basit hasarlı oluşu tahmin edilmiştir. Tahminleme işlemi makine öğrenme algoritmalarından K en yakın komşu, Destek Vektör Makinaları ve Karar Ağaçları kullanılarak yapılmıştır. Elde edilen performans değerlerinin yüksek oluşu, bu tür yöntemlerin olası trafik kazalarının önlenmesi çalışmalarında yardımcı olabileceğini göstermiştir.

ANAHTAR KELİMELER: Trafik Kazaları; Yol Faktörleri; K En Yakın Komşu; Destek Vektör Makinaları; Karar Ağaçları.

ABSTRACT

MSc Thesis

ESTIMATION OF ACCIDENTAL VIOLENCE BY USING MACHINE LEARNING METHODS USING THE TRAFFIC ACCIDENT

Mehmet Ayhan ÇETİN

Harran University
Graduate School of Natural and Applied Sciences
Department of Electrical and Electronics Engineering

Supervisor: Assist. Prof. Dr. Kerim KARADAĞ
Year: 2020, Page: 48

Traffic accidents are total of the incidents which are generated from the basic elements of traffic that are human, vehicle, environmental aspect and cause financial loss, wounding, death. Although there are many factors which affect the accidents, in our country the first thing coming to mind is driver failure. When an accident happens firstly a report is kept and the fault is allocated to drivers .Because the attendants are not aware of the subjects like geometry of roads. Besides driver failure, there are many different factors such as dimate, geometric properties of road which cause traffic accidents. It is estimated by the information of abliquity, moistness, dryness of the road taken from New Zeland Transport Agency, whether the accidents are serious or ordinary. Estimation process is made by using K near neighbour, support vector machine and decission tree from machine learning algorithms. The high levels of performance valve shows that this kind of methods can help traffic accidents. of probable

KEY WORDS: Traffic Accidents; Road Factors; K Near Neighbour; Support Vector Machine; Decission Tree.

TEŞEKKÜR

Yüksek lisans öğrenimimin tüm aşamalarında benden yardımlarını esirgemeyen danışman hocam Dr. Öğretim Üyesi Kerim KARADAĞ'a ve Dr. Öğretim Üyesi M. Emin TENEKECİ'ye teşekkürlerimi sunmayı bir borç bilirim. Yüksek lisans ders dönemimde bana şevk veren ve destek olan Dr. Öğretim Üyesi Faruk KÜRKER'e çok teşekkür ederim. Ayrıca özellikle hayatımın her safhasında bana destek olan eşime, çocuklarıma ve tüm aileme teşekkür ederim. Tezimin araştırma kısmında bana destek olan Şanlıurfa Trafik Şube Müdürlüğü çalışanlarına ve Şanlıurfa Büyükşehir Belediye Başkanlığı Trafik ve Ulaştırma Daire Başkanı Can HALLAÇ Bey'e teşekkür ederim. Ayrıca Dr. Öğretim Üyesi Nurettin BEŞLİ'ye değerli katkılarından dolayı çok teşekkür ederim.

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. Makine Öğrenmesi Yöntemleri 9
Şekil 3.1. DVM çalışma şeması 22
Şekil 3.2. KNN çalışma şeması 25
Şekil 3.3. KA çalışma şeması 26
Şekil 4.1. Çalışmanın akış diyağramı 30
Şekil 4.2. KA Sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri 32
Şekil 4.3. KA sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri 33
Şekil 4.4. DVM sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri 34
Şekil 4.5. DVM sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri 34
Şekil 4.6. KNN sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri 35
Şekil 4.7. KNN sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri 36
Şekil 4.8. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri 37
Şekil 4.9 Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri 38
Şekil 4.10. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve DVM sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri 39
Şekil 4.11. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri 40
Şekil 4.12. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KNN sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri 40
Şekil 4.13. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KNN sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri 41

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 1.1. Avrupa ülkelerindeki trafik kazası bilgileri ve bin kişiye düşen otomobil sayısı..... 2
Çizelge 1.2. Yeni Zelanda 2018 yılına ait sürücü çeşidine ve yaş gruplarına göre ölümlü kaza sayıları 3
Çizelge 1.3. Bazı AB ülkelerindeki hız limitleri 6
Çizelge 2.1. KNN ve geliştirilen derin öğrenme modelinin doğruluklarının karşılaştırılması 15
Çizelge 3.1. Trafik kazası veri seti değişkenler ve özellikleri 16
Çizelge 4.1. Çalışmada kullanılan performans ölçüm yöntemi 31
Çizelge 4.2. On parçaya bölünmüş veri setine ÇD uygulama örneği 32
Çizelge 4.3. 10 ÇD ile oluşturulan veri setinin doğruluk ölçütü performans değerleri 42
Çizelge 4.4. %60 eğitim %40 test ile oluşturulan veri setinin doğruluk ölçütü performans değerleri..... 42

SİMGELER ve KISALTMALAR DİZİNİ

AUC ROC	Area Under The Curve Receiver Operating Characteristic
BG	Banket Genişliği
CV	Cross Validation
DVM	Destek Vektör Makinaları
DVR	Destek Vektör Regresör
DVS	Destek Vektör Sınıflandırıcı
EDS	Elektronik Denetim Sistemi
EG	Eğim
KA	Karar Ağaçları
KGM	Karayolları Genel Müdürlüğü
km/s	Kilometre/saat
KNN	En Yakın Komşuluk
KS	Kaza Sayısı
MATLAB	Matrix Laborartory
ŞG	Şerit Genişliği
TÜİK	Türkiye İstatistik Kurumu
YKT	Yol Kenarı Tehlike Oranı
YKY	Yatay Kurba Yarıçapı
YOGT	Yıllık Ortalama Günlük Trafik Hacmi
YSA	Yapay Sinir Ağları

1. GİRİŞ

Trafik kazaları genellikle ulaşımın temel unsurları olan insan, taşıt, karayolu ve çevre koşulları sonucunda ortaya çıkan ve maddi hasar, yaralanma, ölüm gibi sonuçlara sebebiyet veren olayların bütünüdür. Trafik kazalarını etkileyen birçok faktör olmasına rağmen ülkemizde ilk akla gelen kaza sebebi sürücü hataları olarak değerlendirilmektedir. Kaza raporlarını tutan görevlilerin yolun geometrisi vb. konularda bilgi sahibi olmadıklarından kaza raporlarında hatalar sürücülere paylaştırılmaktadır. Oysaki sürücü hatalarının yanında; hava koşulları, yolun geometrik özellikleri gibi birçok faktör trafik kazalarının oluşmasında önemli etkenlerdendir.

Trafik kazalarının oluşumunda, sürücü davranışları, karayolunun geometrisi ve çevresel faktörler ve araçların özellikleri olmak üzere üç temel faktör etkili olmaktadır. Bunların içerisinde sürücü davranışları tek başına %57 iken, yol ve araç kusurları ile beraber %93 oranına ulaşmaktadır (Rumar, 1985).

Trafik kazalarının oluşmasına birbirinden farklı birçok değişken sebep olduğu için, trafik güvenliğini sağlamak oldukça karmaşık bir alan olarak ortaya çıkmaktadır. Özellikle sınırlı sayılabilecek ülke kaynaklarının doğru ve sonuç getirecek önlemlere sarf edilebilmesi için trafik kazalarının meydana gelmesinin nedenlerinin doğru bir biçimde belirlenmesi gerekmektedir. Kazalardan sonra trafik memurlarınca tutulan kaza tutanaklarına bakılırsa, ana faktörler belirlenmektedir. Buna göre trafik kazalarının sebepleri insan, taşıt, yol ve hava koşulları olarak ön plana çıkmaktadır (Baş, 2015).

Ülkemizde trafik kazalarının artış göstermesinin en önemli sebebi, çözüm için yapılan araştırma ve çalışmalarda doğru hedef ve stratejinin belirlenememesi, sorunun çözümünde bilimsellik ve mühendislik çalışmalarının gerektiğinin tam anlaşılamamış olmasıdır (Özgan, 2003).

Türkiye'de trafik kaza istatistikleri incelendiğinde kusurların %88,6' sının sürücülere, %9,4'ünün yayalara, %0,6' sının taşıt arızalarına, %0,5' inin yolculara,

%1’ inin ise yol kusurlarına ait olduğu görülmektedir (TÜİK, 2014).

Bu verilerde görüldüğü gibi ülkemizde karayolunun geometrik özelliklerinin trafik kazalarındaki payı oldukça azdır. Türkiye’de 2015 yılında ölümlü yaralanmalı trafik kazasına neden olan toplam 210 bin 498 kusura bakıldığında kusurların %89,3'ünün sürücü, %8,8'inin yaya, %0,9'unun yol, %0,6'sının taşıt ve %0,4'ünün yolcu kaynaklı olduğu görülmüştür (TÜİK, 2015)

1.1. Avrupa Ülkelerine Ait Trafik Kaza Bilgileri

Avrupa ülkelerindeki trafik kazaları incelendiğinde, ölü sayısının en fazla olduğu ülkenin Türkiye olduğu görülmektedir. Çizelge 1.1. de Avrupa ülkelerindeki trafik kazası bilgileri ve bin kişiye düşen otomobil sayısı verilmiştir. Çizelge 1.1. incelendiğinde, 2018 yılında Türkiye’de meydana gelen toplam 186.352 kazada 6.675 kişinin öldüğü görülmektedir (KGM Trafik Kazaları Özeti 2018).

Çizelge 1.1. Avrupa ülkelerindeki trafik kazası bilgileri ve bin kişiye düşen otomobil sayısı

ÜLKE	Ölümlü ve yaralanmalı kaza sayısı	Ölü sayısı	Bin Kişiye Düşen Otomobil Sayısı	Birmilyon Otomobile Düşen Ölü Sayısı	Birmilyon kişiye düşen ölü sayısı
TÜRKİYE	186.532	6.675	151	538	81
ALMANYA	308.145	3.206	555	71	39
AVUSTURYA	38.466	432	550	90	49
BELÇİKA	40.096	637	505	112	56
BULGARİSTAN	7.404	708	443	225	99
ÇEKYA	21.386	611	502	117	58
DANİMARKA	2.882	211	429	87	37
ESTONYA	1.468	71	534	103	54
FİNLANDİYA	4.752	258	608	78	47
FRANSA	57.515	3.477	479	108	52
GÜNEY KIBRIS RUM YÖNETİMİ	650	46	595	92	54
HIRVATİSTAN	10.457	307	374	201	74
HOLLANDA	18.749	533	481	65	31
İNGİLTERE	142.846	1.860	484	59	28
İRLANDA	5.893	186	428	92	39
İSPANYA	102.362	1.810	492	80	39
İSVEÇ	14.086	270	477	57	27
İTALYA	175.791	3.283	625	87	54

Çizelge 1.1. **(devam)**

ÜLKE	Ölümlü ve yaralanmalı kaza sayısı	Ölü sayısı	Bin Kişiye Düşen Otomobil Sayısı	Birmilyon Otomobile Düşen Ölü Sayısı	Birmilyon kişiye düşen ölü sayısı
LETONYA	3.792	158	341	235	81
LİTVANYA	3.201	192	456	151	67
LÜKSEMBURG	941	32	662	83	55
MACARİSTAN	16.627	607	338	186	62
MALTA	1.437	23	615	82	51
POLONYA	33.664	3.026	571	143	80
PORTEKİZ	32.299	563	470	118	55
ROMANYA	30.751	1.915	279	360	97
SLOVAKYA	5.602	275	390	132	51
SLOVENYA	6.495	130	523	120	63
YUNANİSTAN	11.318	824	486	159	76
AVRUPA BİRLİĞİNE ÜYE OLAN ÜLKELER (AB-28)	1.099.075	25.651	507	100	50

Çalışmada Yeni Zelanda ülkesine ait trafik kazaları verileri kullanılmıştır. Çizelge 1.2'de ise Yeni Zelanda'da 2018 yılında meydana gelen ölümlü trafik kazalarının, sürücü tipi ve yaş grubuna göre sınıflandırması görülmektedir (www.transport.govt.nz).

Çizelge 1.2. Yeni Zelanda 2018 yılına ait sürücü çeşidine ve yaş gruplarına göre ölümlü kaza sayıları

Yaş Grubu	Sürücü	Yolcu	Motosiklet Sürücüsü	Motosikletçi	Yaya	Bisikletçi	Diğer
0-4 yaş		5					
5-9 yaş		3					
10-14 yaş		6			1	1	
15-19 yaş	17	13	4		2	1	
20-24 yaş	28	16	6		1		
25-29 yaş	18	10	2		4		
30-34 yaş	19	10	3		2		
35-39 yaş	8	2	1	1	2	1	
40-44 yaş	7	3	8	1	4	1	
45-49 yaş	11	3	7		2		
50-54 yaş	8		5	1	2		
55-59 yaş	12	1	8	1	2		
60-64 yaş	13	5	1		2		1
65-69 yaş	14	5	1		2		
70-74 yaş	6	3	1		6	1	
75-79 yaş	7	5			3		
80-84 yaş	10	3			5		
85 yaş üstü	6	1	1		1		

1.2. Trafik Kazalarını Etkileyen Faktörler

Trafik kazalarını etkileyen faktörler deyince, ülkemizde genelde ilk akla gelen etken sürücü hatalarıdır. Bir trafik kazası vuku bulduğunda raporlar tutulur ve sürücülere hatalar paylaştırılır. Çünkü kaza raporunu tutan görevlilerin yolun geometrisi vb. konularda bilgisi yoktur. Halbuki sürücü hatalarının yanında; hava koşulları, yolun geometrik özellikleri gibi birçok faktör trafik kazalarının oluşmasında etkendir.

Trafik kazalarının oluşumunda, sürücü davranışları, karayolunun geometrisi ve çevresel faktörler ve araçların özellikleri olmak üzere üç temel faktör etkili olmaktadır. Bunların içerisinde sürücü davranışları tek başına %57 iken, yol ve araç kusurları ile beraber %93 oranına ulaşmaktadır (Rumar, 1985).

Türkiye'de trafik kazası raporlarını düzenleyen görevliler çoğu kez yolun özellikleri ve çevresel faktörlere bakmadan, araç özelliklerini incelemeden sürücüleri suçlamaktadırlar.(Çakır, 1996).

Ülkemizde karayollarının tasarımının gerektiği gibi yapılmaması ve yolların yapım aşamasında yeterli denetimlerin yapılmaması trafik ve yol güvenliğine etkileyen faktörler arasında sayılabilir (Çanakçı, 2001).

Özet olarak trafik kazalarının oluşmasında birçok faktör vardır ve bunlar bir belirsizlik içerisindedir. Bu konuda yapılmış çalışmaların sonuçları yüzdelerle ve yaklaşık rakamlarla açıklanmıştır (Bağırgan, 2006).

1.2.1. Karayolu tipinin ve trafik hacminin trafik kazalarına etkisi

Trafik kazalarını etkileyen en önemli faktörlerden birisi YOGT (yıllık ortalama günlük trafik hacmi) dir. Bu değişken trafik kazaları ile yapılan çalışmalarda çok kullanılmaktadır.

Genellikle trafik yoğunluğu arttıkça trafik kazalarının da arttığı söylenebilir (Satterhwaite, 1981).

Genel olarak bölünmüş karayollarında meydana gelen trafik kaza oranı, bölünmemiş yollardakine göre daha azdır(Bağırgan, 2006).

Şehirlerarası bir karayolunda ortalama günlük trafik 2000 taşıt/gün'den az ise yıllık ortalama günlük trafik yoğunluğu çok etkili değildir (Bağırgan, 2006).

Yıllık ortalama günlük trafik yoğunluğunun 4000 taşıt/gün olduğu şehirlerarası bölünmemiş karayollarındaki trafik kaza tahminlerinde yolun uzunluğunun daha önemli bir etken olduğu gözlemlenmiştir (Cleveland ve ark., 1984)

1.2.2. Şerit genişliğinin trafik kazalarına etkisi

Şerit genişliğinin trafik kazalarının oluşumunda etkili olduğu bilinen bir gerçektir. Fakat emniyet açısından şerit genişliğinin en uygun değerinin ne olduğunu söylemek zordur. Bu konuda birçok çalışma yapılmıştır. Ancak bu çalışmaların sonuçları farklılık göstermektedir. Bunun sebebi her ülkenin koşullarının farklı olmasıdır (Bağırgan, 2006).

Şerit genişliği dar olduğunda bu durum taşıt güvenliği bakımından ve şerit genişliği fazla olduğunda sürücülerin daha hızlı ya da daha rahat taşıt kullanmalarından dolayı trafik kaza oranını arttırmaktadır (Bağırgan, 2006).

Şehirlerarası bölünmemiş karayollarında yapılmış bir çalışmada bir taşıtın minimum sürüş hızı ile geçmesi için şerit genişliğinin 3.3 metre olması gerektiği belirtilmiştir (Yager ve Aerde, 1983).

1.2.3. Banket genişliğinin trafik kazalarına etkisi

Yoldaki trafik hacmi düşükse banket genişliği trafik kazalarını etkilememektedir (Zegeer ve Stewart, 1994). Trafik hacmi fazla ise banket genişliği trafik kazaları üzerinde etkili olmaktadır.

Şehirlerarası bölünmemiş karayolları üzerine yapılan bir çalışmada ;kaplamalı bir banketin her bir kenar için 0.6 metre genişletilmesi durumunda trafik kazalarının %16 azaldığı belirtilmiştir (Zegeer ve ark., 1986).

Yapılan başka bir çalışmada karayollarında banket genişliği ile trafik kazalarının oluşması arasında negatif bir ilişki olduğu belirtilmiştir (Jorgensen, 1978).

YOGT 2600-4500 taşıt/gün aralığında ise banket genişliğinin arttırılması trafik kaza sayısını azaltmaktadır. Banket genişliğinin 2.4 metrenin üzerine çıkması durumunda trafik kazaları artmaktadır (Blensley ve Head, 1960).

Başka bir çalışmada çok şeritli bölünmüş ve bölünmemiş karayollarında taşıtların banketlere uygunsuz bir şekilde park etmesinden dolayı trafik kazalarının arttığı rapor edilmiştir (TRB, 1978).

1.2.4. Taşıt hızının trafik kazalarına etkisi

Dünyada her yıl meydana gelen trafik kazalarının %13'ü aşırı hız ve hız kurallarına uyulmamasından kaynaklanmaktadır(Çubuk ve Hatipoğlu, 2015).

Taşıtların hızı arttıkça durma mesafesi artmakta, direksiyon hakimiyeti azalmakta, çarpışma şiddeti artmaktadır.

Otomotiv sektörü gün geçtikçe daha hızlı otomobiller üretirken, hükümetler de hız limiti belirleme ve hız kontrolü ile ilgili değişik yöntemler geliştirmiştir. Trafikte hız kontrolü noktasında radar ve hız koridorları oluşturarak elektronik denetim sistemi (EDS) ile çeşitli denetimler yapılmaktadır. Çizelge 2.3'de bazı Avrupa Birliği ülkelerinde belirlenen hız limitleri verilmiştir.

Çizelge 1.3. Bazı AB ülkelerindeki hız limitleri

ÜLKE	Şehir içi (km/saat)	Şehir dışı (km/saat)	Otoyol (km/saat)
Almanya	50	100	130
Avusturya	50	100	130
Belçika	50	90-120	120
Danimarka	50	80	110
Finlandiya	50	80-100	120
Fransa	50	90-110	130
Hollanda	50	80-100	120
İngiltere	50	100-110	110
İrlanda	50	100	110
İtalya	50	90-110	130
Lüksemburg	50	90	130
İspanya	50	90-100	120
İsveç	50	70	110
Türkiye	50	90	120

Dünyanın farklı ülkelerinde hız limiti azaltması çalışmaları yapılmaktadır. Fakat buna karşılık ülkemizde hız limitini biraz daha arttırabilir miyiz önerileri gündeme gelmektedir.

Danimarka'da yapılan bir araştırmada hızı azaltmanın ölümlü trafik kazalarını azalttığı belirtilmiştir (Fridstrom ve ark., 1995).

Sürücülerin hız stratejilerini etkileyen birçok faktör vardır. Bunlardan karayolu ile ilgili olarak; yolun genişliği, yolun eğimi, yolun yatay geometrisi, yoldaki çizgiler, yolun üstyapı kalitesi ve yoldaki işaretlemeler söylenebilir (Çubuk ve Hatipoğlu, 2015).

1.2.5. Yatay kurbaların trafik kazalarına etkisi

Bir karayolunun güvenliği üzerinde; yolun yatay geometrik özellikleri, düşey geometrik özelliklerine göre daha etkilidir. Yolun yatay geometrisini belirleyen iki özellik, yatay kurba ve aliynman özellikleridir (İyinam, 1997).

Yatay kurbalarda meydana gelen trafik kazalarında iki önemli değişken vardır ki bunlar; yatay kurbanın yarıçapı ve iki kurba arasındaki mesafedir (Brenac ve Annex, 1994).

Yatay kurbaların yarıçapı arttıkça trafik kazalarında artış olmaktadır. Eğer bir karayolunda art arda yatay kurba varsa bu yol, düz aliynman sonra küçük yarıçaplı yatay kurba gelen bir yoldan daha az tehlikelidir (Srinivasan, 1982; Brenac, 1990).

Yatay kurbalara sahip bir karayolundaki trafik kaza oranı, düz bir karayolundaki kaza oranından yaklaşık üç kat fazladır (Glennon ve ark., 1987).

Yatay kurba yarıçapları kısaldıkça trafik kazaları artış göstermektedir (Perchonok ve ark., 1978; Department of Transport, 1984).

1.2.6. Boyuna eğimin trafik kazalarına etkisi

Karayollarında boyuna eğim tek başına değilse bile, trafik kazalarının meydana gelmesini etkileyen diğer faktörlerle birlikte kazaların oluşumunu etkileyen

etkenlerden biridir.

Boyuna eğimli bir karayolunda meydana gelebilecek trafik kaza oranı, düz eğimli bir karayolundaki kaza oranına göre daha fazladır (Glennon ve ark., 1987).

Boyuna eğimi yüksek olan karayollarında, iniş yönünde meydana gelen trafik kazalarının oranı çıkış yönündekilerine göre 1.5 ile 3 kat daha fazladır (Babkov, 1975).

İnişli bir yolun trafik kazalarına etkisi incelenirken hız da gözden kaçırılmaması gereken bir etken olarak karşımıza çıkmaktadır.

Boyuna eğimi %6 dan az olan karayollarında meydana gelen trafik kazalarında eğimin etkisi olmamakla beraber, bu değerden yüksek eğimli karayollarında trafik kazalarının arttığı tespit edilmiştir (Krebs ve K, 1977; Choveiri ve ark., 1994).

1.3. Kaza Tahmin Yöntemleri

Literatürde değişik yöntemlerle geliştirilmiş birçok kaza tahmin modeli bulunmaktadır. Bu çalışmalarda ya istatistiksel ya da makine öğrenmesi yöntemleri kullanılmıştır. Shon ve Shin (2010) Yapay Sinir Ağı (YSA), Lojistik Regresyon, Karar Ağacı yöntemlerini kullanarak Kore'de meydana gelen trafik kazalarının şiddetini etkileyen sebepleri incelemişlerdir ve üç model sonuçlarında önemli bir farklılık bulamamışlardır

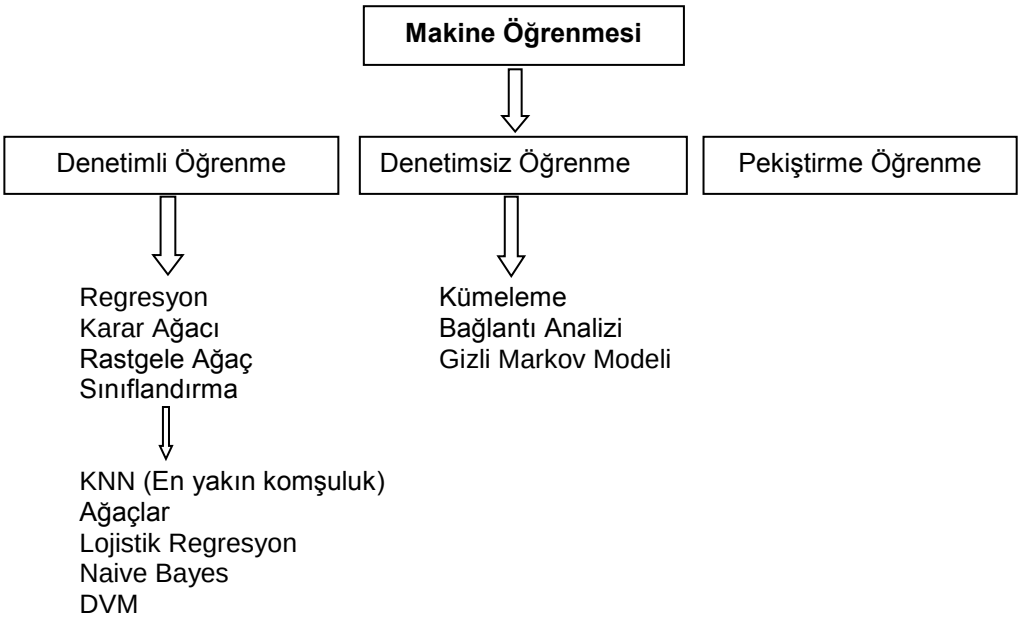
Çelik ve Oktay (2014), Erzurum-Kars otoyolunda 2008-2013 yılları arasında meydana gelen 11.771 trafik kazasını incelemişlerdir. Kazaları yaralanma derecesine göre ölümlü, yaralanmalı ve maddi hasarlı olarak sınıflandırdıktan sonra multinominal lojik analiz yöntemiyle yaralanma derecesine etki eden sebepleri 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 etkenlerinin ölümlü kaza riski üzerinde etkili olduğu sonucuna ulaşılmıştır.

Trafik kaza analizlerinin ve tahminlerinin son yıllarda üzerinde çok çalışılan bir alan olduğu görülmüştür. Yapılan çalışmalarda, istatistiksel modeller bağımlı ve

bağımsız değişkenler arasında önceden tanımlanmış sabit varsayımlara göre hareket etmektedir ve bu varsayımların çiğnenmesi durumunda tutarlı sonuç vermemektedir. Fakat, makine öğrenmesi yöntemleri bağımlı ve bağımsız değişkenler arasında önceden tanımlı herhangi bir ilişkinin varlığını gerektirmemekte ve trafik kaza tahminlerinde oldukça başarılı sonuçlar ortaya çıkarmaktadır.

1.4. Makine Öğrenmesi ve Çeşitleri

Makine Öğrenmesi, bilgisayarların belirli bir görevi yerine getirmek için verilere bağlı olarak kararlar vermesini ve yapmasını sağlanmasıdır. Algoritmalara yeni veriler yüklendiğinde zamanla öğrenmek ve geliştirmek için tasarlanan sistemlerdir. Makine Öğrenme algoritması, bir model oluşturabilmek için bir veri seti kullanılarak eğitilir. Bu eğitimden sonra makine öğrenme algoritmasına yeni verileri yüklendiğinde ise, modelin temelinde bir tahmin yapar. Bu tahminin kabul edilebilir olması durumunda algoritma dağıtılır. Tahminin doğruluk derecesi kabul edilebilir değilse veri seti ile tekrar tekrar eğitilir. Bu durum en doğru tahmin yapılıncaya kadar devam ettirilir. Şekil1.2. de görüldüğü gibi Makine Öğrenmesi denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olmak üzere üç sınıfa ayrılmaktadır (Kantardzic, 2011; Çelik U., Akçetin E., ve Gök M., 2017).



Şekil 1.1. Makine Öğrenmesi Yöntemleri

Son yıllarda sınıflandırma problemleri için kullanılan en başarılı algoritmalardan biri Destek Vektör Makineleri'dir. Destek Vektör Makineleri algoritması sınıflandırma problemini kareli optimizasyon problemine çevirip çözdüğü için problemin çözümü ile ilgili öğrenme aşamasında işlem sayısı azalmakta ve dolayısıyla çözüme daha hızlı ulaşılmaktadır (Osowski, Siwekand ve Markiewicz, 2004). Bu özelliğinden dolayı, büyük veri setlerinde oldukça avantajlıdır.

K-en yakın komşuluk (KNN) algoritması, uygulaması kolay denetimli öğrenme algoritmalarındandır. Bu algoritma çoğunlukla sınıflandırma problemlerinin çözümünde kullanılmaktadır.

Karar ağacı (KA), veri madenciliği sınıflandırma algoritmalarından biri olup, önceden tanımlanmış bir hedef değişkene sahiplerdir. Yapı olarak en tepeden en aşağı inen bir stratejiye sahiptirler (Kantardzic, 2011; Çelik ve ark., 2017).

Bir karar ağacı, çok sayıda kayıt içeren bir veri setini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölen bir yapıdır. Karar ağaçları kolay anlaşılabilir ve yorumlanabilir. Karar ağaçları ile hem sayısal hem de kategorik veriler işlenebilir.

Dünya nüfusunun 7.3 milyar olduğu günümüzde Türkiye nüfusu 76 milyon olmuştur. Nüfus artışı bu şekilde devam ederse yaklaşık 50 yıl sonra dünyada 13 milyar insan olacağı tahmin edilmektedir. Bu nüfus artışı karşısında yeryüzü kaynaklarında herhangi bir artış olmamakla birlikte yanlış kullanımlar sebebi ile kaynaklar da azalmaktadır. Süre gelen durum karşısında kaynakları kullanırken iki önemli etkenin özenle uygulanması gerekmektedir. Bunlar sürdürülebilirlik ve kamu yararına kullanma ilkeleridir. Sürdürebilirliği kısaca, kaynakları kullanırken başkalarına ve gelecek nesillerin haklarını düşünerek kullanma ilkesi ile ifade etmek mümkündür (Ayten, 2015).

2. ÖNCEKİ ÇALIŞMALAR

Konu ile ilgili yapılan önceki çalışmalar ve elde edilen temel yaklaşımlar aşağıda özetlenmiştir.

Bolakar (2014), 2012–2013 yılları arasında Erzurum ilinde meydana gelen trafik kazalarının yer analizini yapmıştır. Erzurum'un ilçe haritası Coğrafi Bilgi Sistemleri ortamına sayısallaştırılarak yüklenmiş ve ilçelerdeki trafik kaza sayıları hangi kategoride ise o kategoriye göre harita üzerinde sorgulanmış ve gösterilmiştir. Bu şekilde en fazla kaza, ölüm ve yaralanmanın olduğu ilçeler olarak harita üzerinde belirlenmiştir. Trafik kazalarına etki eden; ilçelerin nüfusu, köy sayısı, yüzölçümü, taşıt sayısı, yolun geometrik özellikleri, kazazedelerin kişisel özellikleri, kazanın zamanı, hava durumu, kaza yapan aracın özellikleri ve benzeri faktörler yapay sinir ağları ve çok değişkenli istatistik metotları ile incelenmiştir. Regresyon ve yapay sinir ağları modellerinden elde edilen sonuçlar karşılaştırıldığında istatistiki test sonuçlarının daha anlamlı olduğu görülmüştür. YSA modelinin duyarlılık analizine göre, en önemli parametrenin taşıt sayısı olduğu, daha sonra da; nüfus, gayri safi milli hasıla, köy sayısı, toplam yol uzunluğu, yüzölçümü, düşey ve yatay kurp sayısı vb. olduğu tespit edilmiştir. Yapay sinir ağları modelinin, trafik kazası ve etkenleri arasındaki ilişkiyi açıklayabildiği görülmüştür.

Çinicioğlu ve ark. (2013), trafik kazaları ve trafik kazalarının oluşmasının nedenlerini Bayes Ağları ile analiz etmişlerdir. Bu çalışmada Silivri Bölge Trafik Şube Müdürlüğü ve İlçe Jandarma Trafik Tim Komutanlığı'ndan alınan maddi hasarlı kaza tespit tutanakları ve içerdikleri bilgiler ışığında elde edilen veri setinden ilgili Bayes Ağı öğrenilmiştir. Bayes Ağı'nın doğru tahminleme oranı ayrılan test datası ile tespit edilmiş ve oluşturulan modelin etkinliği, model için hesaplanan logskorun marjinal modelin logskoru ile kıyaslanmasıyla teyit edilmiştir. Ağdaki değişkenler için duyarlılık analizleri yapılmıştır. Oluşturulan modelin test verisinde % 71,45129 doğru tahminleme oranına ulaşılmıştır.

Özden ve Acı (2016), Adana ilinde 2015-2014 yılları arasında meydana gelen aylık bazda yaralanmalı trafik kazaları verileri ile aynı yıllardaki aylık meteorolojik verilerden oluşan veri kümesi oluşturmuşlardır. Bu veri setinden faydalanarak yaralanmalı trafik kaza sayısını ve yaralı sayısını tahmin edecek bir model geliştirmişlerdir. Model olarak, Destek Vektör Makinesi, Regresyon Ağacı, Genelleştirilmiş Regresyon Yapay Sinir Ağı, Fonksiyon Uydurma Yapay Sinir Ağı, İleri Beslemeli Çok Katmanlı Yapay Sinir Ağı, Çoklu Doğrusal Regresyon yöntemlerini kullanılmıştır. Sonuç olarak, yaralanmalı trafik kaza sayısı ve yaralı sayısı tahminleme işleminde başarı seviyesinin yüksek olduğu modelin Destek Vektör Makineleri olarak tespit edilmiştir. Çoklu Doğrusal Regresyon modelinin başarı seviyesinin en alt sırada olduğu görülmüştür.

Özgan ve ark. (2004), Sivas il sınırları içerisindeki D200/20 devlet karayolu kesiminde t 1999-2002 yılları arasında meydana gelen trafik kazalarının verilerine istatistiksel analizler uygulamışlardır. Bu analiz sonuçları neticesinde aylar bazında meydana gelebilecek kazalara ait tahminleme yapılmıştır. Trafik kaza raporlarından alınan verilerle Regresyon Analizi yapılarak modele ilişkin bir matematiksel denklem oluşturulmuştur. Bu çalışma sonucunda trafik kaza tahminlerinin daha sağlıklı yapılabilmesi için, kaza rapor ve istatistiklerinin kullanılabilir ve sağlıklı bir şekilde tutulmasının önemi vurgulanmıştır.

Kıyıldı (2017), Yapay Sinir Ağları (YSA) yöntemini kullanarak Türkiye'deki trafik kazaları tahmin modellerini araştırmıştır. Bu şekilde daha önce meydana gelmiş olan kazalardan yararlanarak, durumu ortaya koymak ve gerekli önlemlerin alınması hedeflenmiştir. Geliştirilen modellerde kaza sayısı, nüfus sayıları ve taşıt sayısı parametre olarak seçilmiştir. 2001-2005 yılları arasında bu parametrelerin verilerinden yararlanılmıştır. Bu çalışmada çok katmanlı YSA mimarisi kullanılmıştır. YSA mimarisinde ileri beslemeli geri yayılım (feed forward back propagation) öğrenme algoritması olarak kullanılmıştır. Oluşturulan modelle taşıt sayısı, nüfus sayısı ile kaza sayısı arasındaki eğitim setinde ilişki $R^2 = 1$ şeklinde yakalamıştır. Model sonuçlarına göre taşıt sayısının ne nüfus sayısının artışının trafik kaza sayısının artmasına sebeb olduğu tespit edilmiştir. Trafik kazalarını azaltmak için ulaştırma planlarının revize edilmesi, yük taşımacılığının demiryolu ve diğer

ulaştırma sistemlerine dağıtılması gerekmektedir. Ulaştırma için alternatif olarak da hızlı tren çalışmalarına ağırlık verilmelidir.

Kibar (2015), yaptığı çalışmada Ankara-Aksaray-Ereğli bağlantısında meydana gelen kamyon kazaları ile trafik ve karayolunun geometrik özellikleri arasındaki ilişkiyi Yapay Sinir Ağları ve istatistiksel yöntemlerle modellemiştir. Çalışmada ilk olarak bu ilişkinin belirlenmesinde Negatif Binom, Poisson, ve Sıfır Değer Ağırlıklı Negatif Binom Regresyon modelleri kullanılmıştır. Yapılan analizler neticesinde, Negatif Binom Regresyon modelinin kamyon kazalarının analizinde en iyi istatistiksel metot olduğu tespit edilmiştir. Daha sonra, Yapay Sinir Ağları yöntemi ile Negatif Binom Regresyon yöntemi karşılaştırılmıştır. Tahminleme işleminde Yapay Sinir Ağları yönteminin daha iyi performansa sahip olduğu gözlenmiştir. Sonuç olarak, kamyon kazalarını modellemede Yapay Sinir Ağları yöntemi önerilmiştir.

Bayır (2006), tahminleme yöntemleri olarak kullanılan yapay sinir ağları ve çoklu doğrusal regresyon analizi yöntemlerini karşılaştırmıştır. Sonuç olarak yapay sinir ağları yönteminin, veri içerisindeki ilişkileri tanıyarak öğrenmesinden dolayı, model kurulumu ve kullanımı basit ve esnektir. Probleme uygun olarak geliştirilmiş ve eğitilmiş bir yapay sinir ağları modelinin, diğer tahminleme metotlarından daha iyi bir sonuç vereceği anlaşılmıştır.

Çalışkan (2008), yaptığı çalışmada Yapılan çalışmada tcm-knn, k-medoids, k-means, k en yakın komşu olmak üzere dört farklı metot kullanmıştır. Öznitelik seçimleri, öznitelikler arasındaki ilişkiler, veri önişleme ve benzerlik ölçümleri üzerinde çalışma yaparak metotlardan elde edilen sonuçların iyileştirilmesi için, bu metotları birlikte kullanan hibrit üç farklı sistem geliştirmiştir. Kümelemeyi ve sınıflandırmayı beraber kullanan bu hibrit yapılarda kümeleme metotlarıyla veri kümesini alt kümelere ayırmıştır. Daha sonra sınıflandırma metotlarını farklı karakteristiklere sahip her alt kümeler için ayrı ayrı çalıştırmıştır. En son alt kümeler için alınan neticeleri bir araya getirerek tüm veri kümesi için bir netice ortaya çıkarmıştır. Matlab 6.5 proğramı ile geliştirilen tcm-knn, k-medoids, k-means, k en yakın komşu ile geliştirilen metotların neticeleri karşılaştırıldığında, hibrit yapıların daha iyi sonuç verdiği tespit edilmiştir.

Tezer (2018), yaptığı çalışmada Destek Vektör Makinaları, Lojiskik Regresyon ve Yapay Sinir Ağları yöntemlerinin başarı düzeylerini kıyaslamıştır. Üç yöntemin de sınıflama başarısının % 90'nın üzerinde olduğu tespit edilmiştir. Yöntem belirlerken veri setinin iyi tanınması gerektiğini, kullanılan yöntemde hangi fonksiyonun kullanılacağının tespitinde deneyimin önemli olduğunu vurgulamıştır. Lojistik Regresyon yönteminin düşük standart sapma değerine sahip veri setlerinde kullanılmasının daha uygun olacağını belirtmiştir. Büyük veri setlerinde, bağımlı ve bağımsız değişkenler arasında nonlineer bağlantı olan veri setlerinde Yapay Sinir Ağları yönteminin daha iyi sonuçlar vereceğini belirtmiştir. Standart sapmanın yüksek olduğu ve sınıf sayısının farklı olduğu veri setleri için Destek Vektör Makinaları yönteminin uygun olacağını belirtmiştir.

Keskin (2018), yaptığı çalışmada n Karar Ağacı, K- en yakın komşu, Rastgele Orman, Naive Bayes, metotlarını *R* programında yazılan kodlar ile farklı veri çeşitlerine uygulayarak, daha açıklayıcı şekilde sunmuştur. Ayrıca gerçek bir veri seti ile, son zamanlarda makine öğrenmesi yöntemleri için kullanılan KNIME programı ile yukarıda bahsi geçen dört yöntemin performanslarını karşılaştırmıştır. Karar Ağacı, Naive Bayes, K-en yakın komşu, Rastgele Orman metotlarının her birini R programının uygulamaları ile ayrıntılı bir şekilde incelemiştir. Son olarak KNIME programı kullanılarak performans kıyaslaması yapmıştır. Tüm performans ölçütleri sonuçlarını inceleyerek, en iyi sınıflandırma başarısını Rastgele Ormanlar algoritmasının gösterdiğini tespit etmiştir.

Taşdelen (2019), perceptron öğrenme algoritması ve K en yakın komşu (KNN) yöntemlerini kullanarak sağlık verilerinde sınıflandırma çalışması yapmıştır. Ayrıca derin öğrenme üzerine bir model tasarlamış, bu modeli farklı değişkenlerle çalıştırarak sınıflama çalışması yaplmıştır. Bu çalışmada kullanılan veri kümesi ise UCI Makine Öğrenme Deposunda bulunan göğüs kanseri, pima yerlileri diyabet veri tabanı, bupa karaciğer hastalıkları, mamografik kitle verileridir. Geliştirilen derin öğrenme modelinde başarı değerleri en iyi olan neticeler değerlendirilerek, her veri kümesi için en iyi netice alınan değişkenlerin değerlendirilmesi yapılmıştır. Geliştirilen algoritmanın iki ve üç özelliğe sahip versiyonu dört veri setinde de çalıştırılarak doğruluk oranları tespit edilmiştir. KNN algoritmasında k değeri 1,3,5

ve 10 değerleri kullanılarak performans değerlendirmesi yapılmış ve bu sonuçlar çizelge 2.1.'de verilmiştir. Çizelge 2.1.'de görüldüğü gibi, geliştirilen derin öğrenme algoritması ile elde alınan neticeler K en yakın komşu algoritmasına göre daha iyidir.

Çizelge 2.1. KNN ve geliştirilen derin öğrenme modelinin doğruluklarının karşılaştırılması

Doğruluk	KNN k=1	KNN k=3	KNN k=5	KNN k=10	Geliştirilen Derin Öğrenme Modeli
Göğüs Kanseri	0.970	0.975	0.970	0.955	0.989
Pima Yerlileri Diyabet	0.647	0.656	0.660	0.673	0.796
Bupa Karaciğer Hastalıkları	0.466	0.533	0.563	0.553	0.728
Mamografik Kitle	0.420	0.375	0.375	0.541	0.809

Tosun (2007), yaptığı çalışmada yapay sinir ağlarının sınıflandırma başarı performansını incelemiştir. Mayıs 2005'de yapılmış olan Defining Critical Factors Affecting Student Success: A Data Mining Approach isimli tez çalışmasının verilerini kullanarak, başarılı ve başarısız öğrencileri tespit etmeye çalışmış ve bu sınıflandırmada etkili olan değişkenleri tespit etmiştir. Çalışmanın amacı Yapay Sinir Ağları ile Karar Ağaçları yöntemlerinin karşılaştırılması ve elde edilen neticelerle üniversite öğrencilerinin başarı kriterlerinin ortaya çıkarılmasıdır. Yapılan uygulama neticesinde yapay sinir ağları başarı oranının %91.77, karar ağaçları yönteminin başarı oranının ise %86 olduğu tespit edilmiştir. Alınan tüm sonuçların analizi sonucunda öğrencilerin derse devam oranları, çalışma saatleri, sınıfları, barındıkları yer, babalarının eğitim seviyesi, babalarının hayatta olup olmaması gibi değişkenlerin öğrencilerin başarısını etkilediği gözlemlenmiştir

3. MATERYAL ve YÖNTEM

Bu tez çalışmasında, Yeni Zelanda Ulaştırma Ajansından alınan trafik kazalarına ait bilgiler kullanılarak Basit ve Ciddi hasarlı kazaların tahmin edilmesi amaçlanmıştır.

Tezin bu bölümü materyal ve yöntem olmak üzere iki ana alt boşluktan oluşmaktadır. Materyal alt başlığında çalışmada kullanılan verilerin nasıl elde edildiğine ilişkin detaylı açıklamalar yer almaktadır. Yöntem alt başlığında ise verilerden hangi özniteliklerin değerlendirildiği ve kullanılan sınıflandırma yöntemleri ile detaylı bilgiler yer almaktadır.

3.1. Materyal

Yapılan çalışmada Yeni Zelanda Ulaştırma Ajansından alınan; yolların eğimi, şerit sayısı, aydınlığı, kuru veya ıslaklığı, olabilecek hız sınırı, yollarda bulunan yaya, bisikletli ve ağaç gibi bilgilerin bulunduğu veri seti kullanılarak yollarda oluşan kazaların ciddi veya basit hasarlı oluşu tahmin edilmiştir. Yaklaşık 565.000 örnek ve 88 öznitelik içeren veri seti incelenmiştir. Veri setine göre trafik kazalarında daha çok etkili olan 17 öznitelik çalışmamızda değerlendirilmiştir. Veri setinden alınan 9780 örneğin 4858 tanesini basit, 4922 tanesini ise ciddi kazalar oluşturmaktadır (colab.research.google.com).

Çalışmada kullanılan veri setindeki değişkenler ve özellikleri Çizelge 3.1'de verilmiştir.

Çizelge 3.1. Trafik kazası veri seti değişkenler ve özellikleri

No	Değişkenler	Özellikler
1	Yaya Sayısı	0-1
2	Bisiklet Sayısı	0-1-2-3
3	Araba İstasyonu Sayısı	0-1-2-3-4-5-6-7-8
4	Hava Durumu	Açık Sağanak yağışlı Hafif yağmurlu Sisli Karlı
5	Şerit Sayısı	0-1-2-3-4-5-6-7-8

Çizelge 3.1. **(devam)**

No	Değişkenler	Özellikler
6	Yolun Yönü	Doğu Batı Kuzey Güney
7	Yolun düz veya tepe oluşu	Düz Tepe
8	Havanın ışığı	Parlak güneşli Karanlık Bulutlu Alaca karanlık
9	Yol aydınlatması	Karanlık Aydınlık
10	Araç Sayısı	Sadece bisiklet Çoklu araç Diğer Tek araç-Araç+bisiklet Diğer birçok araç Araç+yaya
11	Yolun Eğimi	Az eğimli Orta eğimli Çok eğimli Düz yol
12	Yol hattı	1-2
13	Yol İşaretleri	Merkez çizgisi İşaret yok Geçiş çizgisi yok Boyalı ada Alt geçit Yükseltilmiş ada
14	Yolun Islaklığı	Kuru Buzlu Karlı Islak
15	Hız Sınırı	20-30-40-50-60-70-80-90-100
16	Yoldaki Ağaç Sayısı	0-1-2
17	Yolun Kentselliği	Açık yol Şehir yolu
18	Kaza Şiddeti	Basit Ciddi

3.2. Yöntem

Bu çalışmada, trafik kazalarının şiddetini tahmin etme amaçlanmıştır. Tahmin işlemi için verilerin özniteliklerinin çıkarılması ve sınıflandırma işlemine tabi tutulması gerekmektedir. Veri setindeki örnek sayısının ve özelliklerinin çok fazla olması sebebiyle veri analizi bilgisayar aracılığı ile yapılmıştır. İlk önce verilerin boyutu ve özelliklerinin indirgenmesi sağlanmıştır. Daha sonra bu verilerin sınıflandırılması için farklı yöntemler kullanılmış ve elde edilen sonuçlar karşılaştırılmıştır.

3.2.1. Normalizasyon

Normalizasyon işlemi, makine öğrenmesi için veri hazırlamanın bir parçası olarak kullanılan bir tekniktir. Normalleştirmenin amacı, veri kümesindeki sayısal sütunların değerlerini, değerler aralığındaki farklılıkları bozmadan ortak bir ölçeğe uygun bir şekilde değiştirmektir. Makine öğrenmesinde, her veri kümesini normalleştirmeye ihtiyaç olmayabilir. Normalleştirme işlemi, verilerin boyutunu küçültmek veya işlemleri normalleştirilmiş değerlerle uygun aralıklarla yapmak ve daha anlamlı ve kolayca yorumlanabilir sonuçlara ulaşmak için kullanılabilir (Dondurmacı & Çınar, 2014). Literatürde farklı veri normalizasyonu çeşitleri bulunmaktadır. Bunlar ondalık ölçeklendirme, minimum-maksimum (min-max), sigmoid ve z-skor gibi söylenebilir (Jayalakshmi ve Santhakumaran, 2011).

Yeni Zelanda Ulaştırma Ajansından alınan veri setindeki tüm parametreler kullanılmamıştır. Çizelge 3.1. de verilen, trafik kazalarında daha fazla etkisi olan veriler işleme alınarak, normalizasyon işlemine tabi tutulmuşlardır. Normalizasyon işlemi için D-Min-Max modeli kullanılmış ve tüm veriler 0 ile 1 arasında normalize edilmiştir. Veri setimizde 9780 adet veri bulunmaktadır.

Normalizasyonun literatürde geçen bir çok çeşidi bulunmaktadır. Bunlar min-max normalizasyonu, istatistiksel veya z score normalizasyonu, medyan normalizasyonu, sigmoid normalizasyonu, D min-max normalizasyonudur. Bunlar hakkında açıklayıcı bilgiler aşağıda verilmiştir.

3.2.1.1. Min-Max normalizasyonu

Min-Max normalizasyonu, verileri doğrusal biçimde normalize eder. Minimum; bir verinin alabileceği en küçük değer iken, maksimum; verinin alabileceği en büyük değeri ifade eder. Bir veriyi Min-Max yöntemi ile 0 ile 1 aralığına indirgemek için (3.1) nolu eşitlik kullanılır.

$$x' = (x_i - x_{min})/(x_{max} - x_{min}) \quad (3.1.)$$

Bu eşitlikte;

x' = Normalize edilmiş veriyi,

x_i = Girdi değerini,

x_{min} = Girdi seti içerisinde yer alan en küçük sayıyı,

x_{max} = Girdi seti içerisinde yer alan en büyük sayıyı,

göstermektedir.

3.2.1.2. İstatiksel veya z-score normalizasyonu

Veri seti üzerindeki kimi değerlerin 0'dan büyük, kimilerinin ise 0'dan küçük değerlere sahip olması, veriler arasındaki uzaklıkların sonuçlar üzerinde daha etkili olacağını göstermektedir. Verilerin normalizasyonu ile eğitim giriş setindeki her parametrenin, tahminleme işlemine eşit ölçüde desteklemesi sağlanır. Her bir değişkene karşı gelen sütunların aritmetik ortalaması ve standart sapması bulunmuştur. Daha sonra, (3.2) nolu eşitlikte verilen formül ile veriler normalleştirilerek, aralarındaki uzaklıklar ortadan kaldırılır (Yüksek ve ark., 2007: 97-112).

$$x' = (x_i - \mu_i)/\sigma i \quad (3.2)$$

Bu eşitlikte;

x' =Normalize edilmiş veriyi,

x_i = Girdi değerini,

μ_i = Girdi setinin ortalamasını,

σ_i = Girdi setinin standart sapmasını,

göstermektedir.

3.2.1.3. Medyan normalizasyonu

Bu metotta her girdinin medyan değerini alınarak, her örneklem için medyan normalize metodu kullanılır. Medyan aşırı sapmalardan etkilenmez (Jayalakshmi ve Santhakumaran, 2011: 89-93). Bu yöntem için (3.3) nolu eşitlik kullanılır.

$$x' = x_i/\text{Median}(a_i) \quad (3.3)$$

Bu eşitlikte;

x' = Normalize edilmiş veriyi,

x_i = Girdi değerini,

a_i = Girdi setinin medyanını,

göstermektedir.

3.2.1.4. Sigmoid normalizasyonu

Sigmoid normalizasyonu verileri 0 ile 1 veya -1 ile 1 arasında gruplandırır. Bir kaç çeşit doğrusal olmayan sigmoid fonksiyonu bulunmaktadır. Bunlardan Tanjant sigmoid fonksiyonu süreçleri hızlandırma için kullanılabilir (Jayalakshmi ve Santhakumaran, 2011: 89-93). Bu yöntem için (3.4) nolu eşitlik kullanılır.

$$x' = (e^{xi} - e^{-xi}) / (e^{xi} + e^{-xi)} \quad (3.4)$$

Bu eşitlikte;

x' = Normalize edilmiş veriyi,

xi= Girdi değerini,

e = doğal logaritma değerini,

göstermektedir.

3.2.1.5. D min max normalizasyonu

Bu modelin uygulanması için aşağıda eşitlik yardımıyla verilerin tamamı 0,1 ile 0,9 arasında normalize edilmiştir. Veriler bu işlemle boyutsuz hale getirilmiş olur (Doğan ve ark., 2007: 4119-4131). Bu yöntem için (3.5) nolu eşitlik kullanılır.

$$x' = 0.8 * (x_i - x_{min}) / (x_{max} - x_{min}) + 0.1 \quad (3.5)$$

Bu eşitlikte;

x' = Normalize edilmiş veriyi,

x_i = Girdi değerini,

x_{min} = Girdi seti içerisinde yer alan en küçük sayıyı,

x_{max} = Girdi seti içerisinde yer alan en büyük sayıyı,

göstermektedir.

Veri sınıflandırma işlemlerinde bir çok farklı yöntem kullanılmaktadır. Yaptığımız çalışmada verilerin sınıflandırması için destek vektör makinaları, K en

yakın komşuluk, karar ağaçları yöntemleri kullanılmıştır. Bu yöntemler ile ilgili bilgiler aşağıda açıklanmıştır.

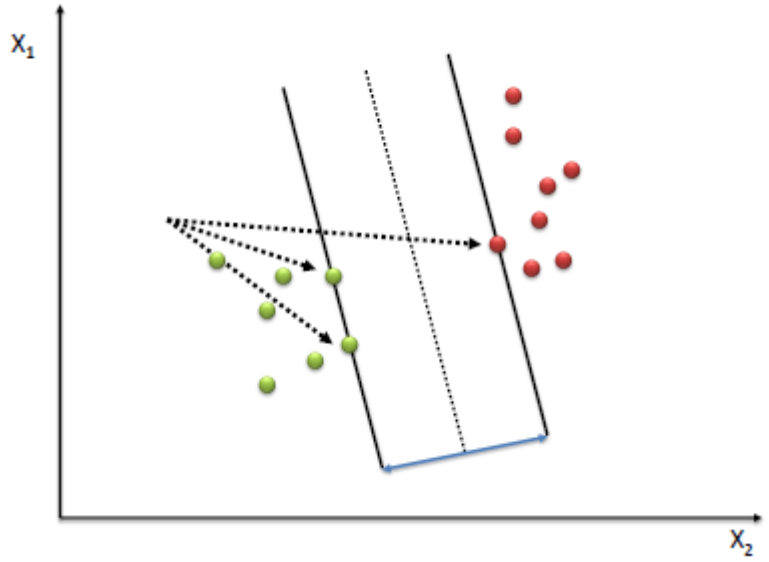
3.2.2. Destek vektör makinaları (DVM)

Destek Vektör Makineleri (DVM), Destek Vektör Regresör (DVR) ve Destek Vektör Sınıflandırıcı (DVS) den oluşur. Veri madenciliği algoritmaları arasında en dayanıklı ve en doğru metotlardandır(Cortes, 1995). DVM'ler Orijinal olarak Vapnik tarafından 1990 larda geliştirilmiştir (Vapnik, V.N, 1995; Vapnik, V.N., 1998), istatistik öğrenme teorisinde köklü sağlam bir teorik temele sahiptir. Geçtiğimiz on yılda, pratik ve teori alanlarında DVM' ler çok hızlı bir biçimde geliştirilmiştir.

İsminden de anlaşıldığı gibi, makine tasarımı destek vektörler olarak hizmet veren eğitim verisinin bir alt kümesinin çıkarımına dayanmaktadır . DVM çok terimli (polinom) makine öğrenmesi, radyal temelli fonksiyon ağ ve iki katmanlı algılayıcı (two-layer perceptron) fonksiyonlarını içermektedir (Haykin, 2001).

DVM' lerin avantajlarının yanında dezavantajları da bulunmaktadır. Mesela büyük veri sayısına sahip kümelerde DVM'nin eğitim işlemi (öğrenmesi) biraz zaman alır. Ancak doğrusal ya da doğrusal olmayan veriler için ayırım yapmaması kullanımını arttırmıştır (Bican, 2014).

DVM' ler ile verilerin karmaşıklığı doğru bir şekilde belirlenebilir (Erastö, 2001). Doğrusal ayrılabilen verilerde, verileri ayırabilecek ve sınıfların sınırları arasını en yüksek değerde tutacak doğru seçilir. Doğrusal ayrılamayan verilerde ise örnek uzayı, yüksek boyutlu bir örnek uzaya aktarılarak, örnekler arasındaki en büyük sınırın bulunması ile sınıflandırma uygulanır. Diğer bir deyişle, sınıfları en iyi ayıran hiper düzlemin tanımlanması yapılır (Bican, 2014). Şekil 3.1. de DVM çalışma şeması verilmiştir.



Şekil 3.1. DVM çalışma şeması

DVM en büyük değerli aralık sınıflandırıcılarındandır. n boyutlu uzayda verileri sınıflandırabilmek için eşitlik (3.6)'da belirtilen bir ayırıcı fonksiyon kullanılır.

$$f(x) = w \cdot xi + b = 0 \tag{3.6}$$

Bu fonksiyonda, w ağırlık vektörünü, b eğilim değerini gösterir. Fonksiyon, ağırlık vektörünün (w) bulunması ile hesaplanır. n tane eğitim verisi olan bir kümede w ve b değerleri

$$|w \cdot xi + b| = 1,\ i = 1$$

olacak şekilde, ayırıcı düzleme en yakın veri noktaları değerlerini içeren bir formda düzenlenebilir. Veriler farklı sınıflardan olduğu için, bazı veriler eşitliği

$$w.xi + b = +1$$

durumunda sağlarken, farklı sınıftaki diğer veriler

$$w.xi + b = -1$$

durumunda sağlar.

Ağırlık vektörü değerinin küçük olması, sınır aralığının büyük olmasını sağlar ki bu istenen bir durumdur. DVM, sınır aralıklarını yüksek değerde tutarak YRM prensibini uygulamaya çalışır. DVM üzerinde meydana gelen eksik uyum sıkıntısı, değişik çekirdek fonksiyonları ile giderilebilir (Bican, 2014).

DVM üzerinde sınır aralığının tespiti, veri kümesine göre değişkenlik gösterir. Mesela eğitim verilerinin doğrusal ayrılabildiği durumlarda DVM, VC değerini en küçük değere ulaştırarak, risk ihtimalini azaltır. Bu şekilde doğrusal bir ayırıcı fonksiyon yardımıyla iyi bir genelleme yapılır ve yüksek başarıya ulaşılır.

Doğrusal olmayan veri seti için DVM, sınıflandırma işlemini orjinal uzay üzerinde yapamaz. Bu sebeple veriler, çekirdek fonksiyonları yardımıyla orjinal uzaydan daha büyük boyutlu olan bir uzaya aktarılır. Sonrasında da bu yeni uzay üzerinde sınıflandırma işlemini yapmak için ayırıcı fonksiyon oluşturulur.

DVR yönteminde verilerin özellikleri gerçeğe yakın bir şekilde yansıtılır ve en iyi biçimde sınıflandırma yapan ayırıcı fonksiyonun bulunması amaçlanır. Regresyonda da sınıflandırmada olduğu gibi doğrusal olmayan verilerin işlenmesi için çekirdek fonksiyonları kullanılır.

3.2.2.1. Destek vektör makinelerinde regresyon

Zaman serisi veriler için regresyon, gerçek değerli bir veri setinde öğrenilme işleminin gerçekleşmesi ve gerçek değerli bir tahminleme işlemidir. Regresyon tahminleme işleminde amaç, uygun kriterleri tespit ederek veriler üzerinde öğrenmeyi iyi bir biçimde yapmak ve en iyi değerleri tahmin etmektir. DVM ile yapılan regresyon Destek Vektör Regresyonu (DVR) olarak adlandırılır. Regresyon çözümlerinde, sınıflandırma çözümlerinde ayırıcı olarak kullanılan

$$f(x) = w.\, x + b \tag{3.7}$$

Yukarıda verilen (3.7) fonksiyonu ile beraber hata parametresi olarak (ε) değişkeni kullanılır. Hata parametresi hedefteki sapma miktarını gösterir. Tahmin edilen ve öğrenilen değerler gerçek değerlere sahiptir. Öğrenilen değerler (girdi) xi ve tahmin edilen (çıktı) değerler yi olmak üzere:

$xi \in R\ N, yi \in R\ N\ i = 1$

şeklinde tanımlanır. Hedefteki sapma değerini (ε) her iki yönde de temsil etmek için karar verici fonksiyon (3.8) eşitliğindeki gibi:

$$yi - w\ xi - b \leq \varepsilon\ (2.49) \text{ ve } -yi + w\ xi + b \leq \varepsilon \quad (3.8)$$

olarak kullanılır.

DVR'de hedef değişkenleri etrafında bir aralık olduğu düşünülebilir. Yapılan tahminleme işleminde bu aralığın içerisinde olanların hatasının olmadığını, aralığın dışarısında olanların ise aralığa uzaklıklarına bağlı olarak hatasının olduğu söylenebilir (Bican, 2014).

Hata parametresi (ε) duyarsız kayıp fonksiyonu olarak da adlandırılmıştır (Vapnik, 1998).

(ε) parametresi küçük ve büyük sapma aralığını kontrol eder. Hata parametresi büyük olduğunda destek vektörlerinin sayısında azalma olur ve bu şekilde öğrenme süresi kısalır. Fakat başarıda azalma tespit edilir. Hata parametresi düşük olduğunda ise destek vektörlerinin sayısı artar ve öğrenme zamanı kısalır. Bu durumda da başarıda yükselme tespit edilir.

3.2.2.2. Destek vektör makineleri yönteminde çapraz doğrulama

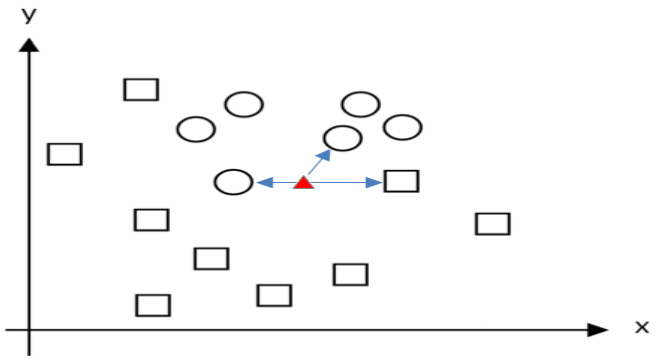
DVM yönteminde başarı oranını arttırmak için çapraz doğrulama yöntemi kullanılmaktadır. Bu yöntemde eğitim kümesi eşit büyüklükte ve birbirinden bağımsız alt kümelere ayrılır. Bu alt kümelerden biri test kümesi olarak seçilir. Diğer kümeler ise eğitilir. Bu işlem bütün alt kümeler için tekrarlanır. Test hatası tüm alt kümelerin hatalarının ortalaması olarak belirlenir.

3.2.3. K- en yakın komşu algoritması (KNN)

K-en yakın komşu algoritması (KNN), uygulamanın kolay olması, eğitme sürecinin kullanışlılığından dolayı sınıflandırma işlemlerinde yaygın olarak kullanılmaktadır. Bu yöntemin büyük miktarda hafıza alanına ihtiyaç duyması, veri

seti ve özelliklerin boyutu fazlalaştıkça işlem yükünün ve maliyetin önemli miktarda artması, performansın k komşu sayısı, uzaklık ölçütü gibi özelliklere bağlı olarak etkilenmesi şeklinde birtakım dezavantajları da vardır (Bhatia, 2010).

KNN algoritması, en temel örnek tabanlı öğrenme algoritmalarından biridir. Eğitim setindeki verilere bağlı olarak öğrenme işlemi gerçekleştirilmektedir. Yeni bir örnek, eğitim setinde bulunan örneklerle benzerliğine göre sınıflandırılmaktadır (Mitchell,1997). Eğitim setindeki örnekler n boyutlu sayısal özellikler ile belirtilmektedir. Her örnek n boyutlu uzayda bir noktayı gösterecek şekilde tüm örnekler n boyutlu bir örnek uzayında tutulur. Yeni bir örnek ile karşılaşıldığında, eğitim setinde bulunan bu örneğe en yakın k tane örnek belirlenerek yeni örneğin sınıf etiketi, k en yakın komşusunun sınıf etiketlerinin çoğunluk oylamasına göre atanır (Han, 2006). Şekil 3.2 de KNN çalışma şeması verilmiştir.



Şekil 3.2. KNN çalışma şeması

KNN algoritmasının performansı için önemli olan diğer bir husus da örnekler arası yakınlığın nasıl ölçümleneceğidir. Örnekler arası yakınlık, Öklid uzaklığı veya başka bir uzaklık ölçütü ile hesaplanabilir.

3.2.3.1. KNN parametreleri

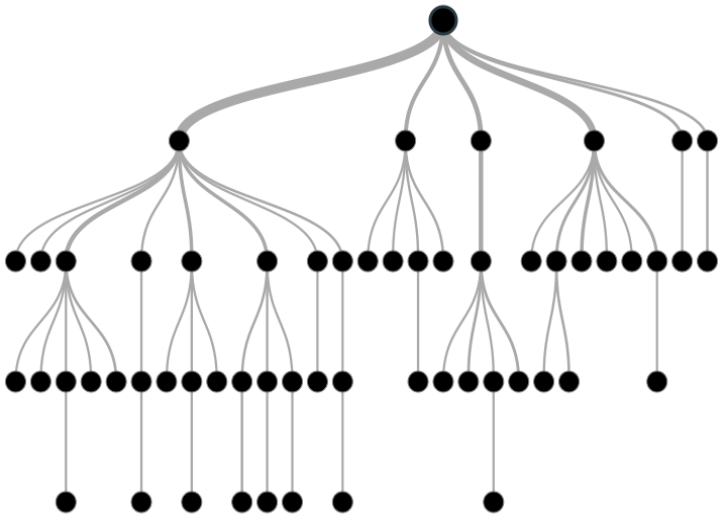
KNN algoritmasının performansını belirleyen önemli parametreler komşu sayısı (*k*), uzaklık ölçütü ve ağırlıklandırma metodudur. Uzaklık ölçütleri olarak, Öklid, Minkowski, Chebyschev, Manhattan ve Dilca uzaklığı kullanılmaktadır.

KNN algoritmasında, komşu sayısı (*k*) parametresinin değerine bağlı olarak sınıflandırma yapılmaktadır. Sınıflandırma sürecinde, *k*=1 için, sadece en yakın komşunun sınıfına atanırken, *k* sayısı örnek sayısına yaklaştıkça veri setindeki tüm veriler dikkate alınıp oylama sonucuna göre seçim yapılmaktadır.

Komşular için ağırlık değerleri atanması ile sınıflandırılmakta olan örneğe daha yakın olan komşu örneklerin, çoğunluk oylamasına daha fazla katkı koyması amaçlanır. Sık olarak kullanılan ağırlık değeri atama yöntemleri, her bir komşunun ağırlığının, *d,* komşular arası uzaklık alınmak üzere, 1/*d veya* 1/*d2* şeklinde alınmasıdır (Doad, 2013).

3.2.4. Karar ağaçları (KA)

Ağaç tabanlı öğrenme algoritmaları, çok sık kullanılan denetimli öğrenme algoritmalarındandır. Genellikle ele alınan bütün problemlerin regresyon ve sınıflandırma çözümlerinde kullanılabilirler. Şekil 3.3. de KA çalışma şeması verilmiştir.



Şekil 3.3. KA çalışma şeması

KA, veri madenciliği sınıflandırmalarında kullanılan algoritmalarından biridir. Önceden tanımlanmış bir hedef değişkene sahiplerdir. (Kantardzic, 2011; Çelik U., Akçetin E., ve Gök M., 2017).

Bir KA, çok sayıda kayıt içeren bir veri kümesini, birtakım kuralları uygulayarak daha küçük kümelere ayırmak amacıyla kullanılan bir yapıdır. Yani basit karar verme adımlarının uygulanmasıyla, büyük veri içeren kayıtları, daha küçük kayıt gruplarına ayırmada kullanılan bir yapıdır.

KA' nın anlaşılması ve yorumlanması kolaydır. KA ile çok çıktısı olan problemler işlenebilir. Az oranda bir veri hazırlığına ihtiyaç duyarken, bu model kayıp değerleri desteklememektedir. Bu modelle sayısal ve sınıflandırmalı veri setleri işlenebilir. (Kantardzic, 2011; Çelik ve ark., 2017).

Bu modelde veriyi güzel açıklayamayan karmaşık ağaçlar meydana gelebilir. Bu şekilde dallar takip edilemeyebilirler. Ayrıca ezbere öğrenme olayı gerçekleşebilir ("over-fitting"). Bu durumda modelin parametrelerine kısıtlama getirilebilir. (Çelik ve ark., 2017).

KA yönteminde bölünmenin nasıl olacağı ağacın doğruluğunu etkileyen faktörlerden biridir. Kategorilendirme ve regresyon problemlerinde bölünme kriterleri genellikle değişkenlik gösterir.

KA, bir düğümü iki veya daha fazla alt düğümde bölmeye karar vermek için birden fazla algoritma kullanmaktadır. Homojenliği arttırmak için alt düğümler oluşturulur. Diğer bir deyişle, düğümün homojenliğinin hedef değişkenlere göre arttığı söylenebilir.

Algoritma seçimi, hedef değişkenin tipine göre değişmektedir. KA' nda en çok kullanılan algoritmalar; sürekli değişkenler için En Küçük Karalere yöntemi, kategorik değişkenler için Entropi, Gini, Sınıflandırma Hatası yöntemi biçimindedir.

Entropi, verilerle alakalı belirsizliğin bir ölçüsüdür. Bir veri kümesinin sadece bir etiketi varsa daha düşük bir entropiye sahip olduğu söylenebilir. Dolayısıyla veriler, entropiyi en aza indirecek bir biçimde bölünmelidir. Bölünmeler iyi yapılması, tahminlerin de iyi olacağı anlamına gelmektedir.

$$H = — \Sigma\, p(x) \log p(x) \tag{3.9}$$

Eşitlik (3.9) da p (x) belirli bir sınıfa ait grup yüzdesini ve H ise entropiyi göstermektedir.

İyi bir bölünmeyi belirlemek amacıyla bilgi kazancını kullanılmaktadır. Bilgi kazancının hesabı eşitlik (3.10) yardımıyla hesaplanır:

$$Gain(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} H(V) \tag{3.10}$$

Eşitlik (3.10) da S orijinal veri kümesini ve D ise kümenin bölünmüş bir parçasını göstermektedir. Her V, S'nin bir alt kümesidir. V'nin tümü ayrıktır ve S'yi oluşturmaktadır. Bilgi kazancı, orjinal veri setinin entropisi ile her özniteliğin entropi değeri arasındaki farktır.

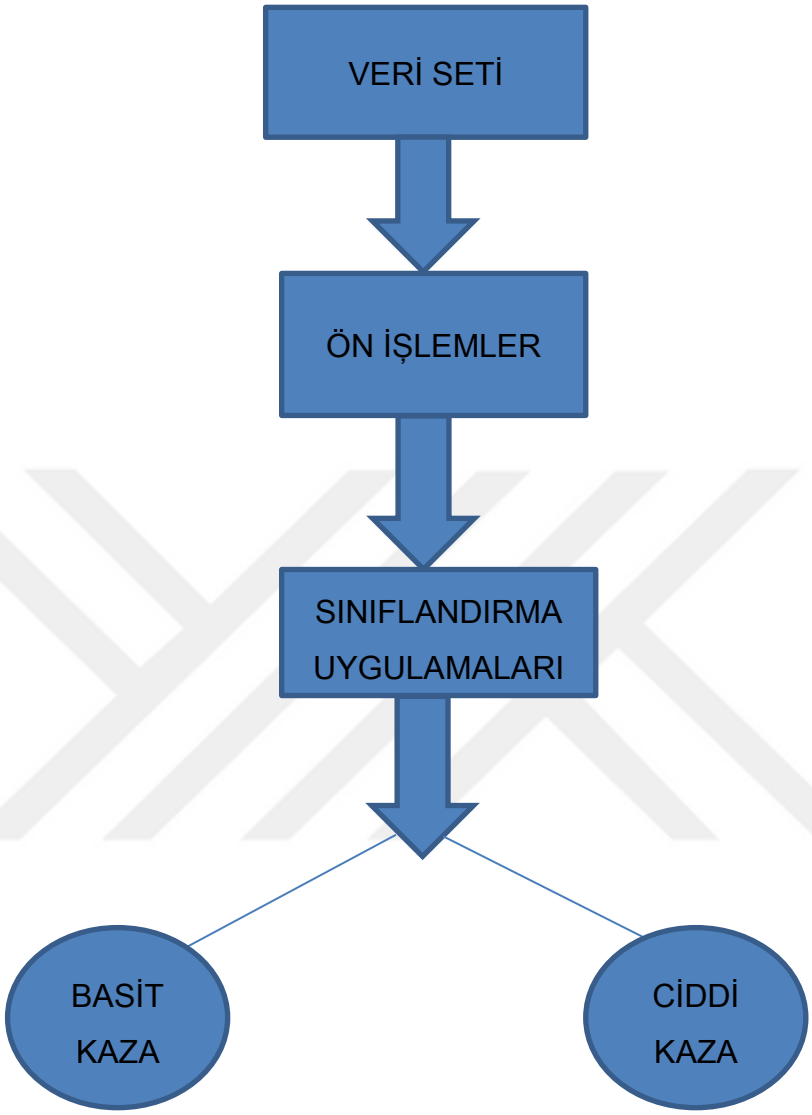
En sık kullanılan KA algoritmaları: CART, ID3, CHAID ve C4.5'dir.

4. ARAŞTIRMA BULGULARI ve TARTIŞMA

4.1. Bulgular ve Tartışma

Bu tez çalışmamızda Yeni Zelanda Ulaştırma Ajansından alınan; yolların eğimi, şerit sayısı, aydınlığı, kuru veya ıslaklığı, olabilecek hız sınırı, yollarda bulunan yaya, bisikletli ve ağaç gibi bilgilerin bulunduğu veri seti kullanılmıştır. Bu veri seti kullanılarak yollarda oluşan kazaların ciddi veya basit hasarlı oluşu tahmin edilmiştir. Veri setinde yaklaşık 565.000 örnek ve 88 öznitelik bulunmaktadır. Veri setine göre trafik kazalarında daha çok etkili olan 17 öznitelik çalışmamızda değerlendirilmiştir. Veri setinden alınan 9780 örneğin 4858 tanesini basit, 4922 tanesini ise ciddi kazalar oluşturmaktadır. Bu çalışmada, trafik kazalarının şiddetini tahmin etme amaçlanmıştır. Tahmin işlemi için verilerin öznitelikleri çıkarılmış ve sınıflandırma işlemine tabi tutulmuştur. Veri setindeki örnek sayısının ve özelliklerinin çok fazla olması sebebiyle veri analizi bilgisayar aracılığı ile yapılmıştır. Veri setine normalizasyon işlemi yapılmıştır. Daha sonra bu verilerin sınıflandırılması için DVM, KNN, KA yöntemleri kullanılmış ve elde edilen sonuçlar karşılaştırılmıştır.

Veri setinden alınan ve Çizelge 3.1.‘de gösterilen parametreler işleme alınarak, normalizasyon işlemine tabi tutulmuşlardır. Normalizasyon işlemi için D-Min-Max modeli kullanılmış ve tüm veriler 0 ile 1 arasında normalize edilmiştir. Veri setimizde 9780 adet veri bulunmaktadır. Veri setine normalizasyon işlemi uygulandıktan sonra sınıflandırma yöntemleri uygulanarak tahminleme yapılmıştır. Yapılan çalışmanın akış diyağramı Şekil 4.1. de verilmiştir.



Şekil 4.1. Çalışmanın akış diyağramı

Kazaların basit ve ciddi olarak tahminleme işleminde makine öğrenme algoritmalarından KNN, DVM ve KA yöntemleri kullanılmıştır.

Makine öğrenmesi yöntemlerinin performans değerlendirmelerinde R2, MSE, RMSE, MAE gibi temel performans göstergeleri kullanılabilinmektedir. Ayrıca modelleri karşılaştırma ölçütleri olarak; doğruluk, hız, sağlamlık, ölçeklenebilirlik, yorumlanabilirlik gibi teknikler kullanılabilir. Yapılan çalışmada doğru tahmin edilen değerlerin tüm örneklere oranını veren doğruluk ölçütü kullanılmıştır. Çizelge 4.1. de

doğru ve yanlış tahmin edilen Confusion Matrix değerleri verilmiştir. Çizelgede doğruyu doğru tahmin etme (DD), doğruyu yanlış tahmin etme (DY), yanlışı doğru tahmin etme (YD) ve yanlışı yanlış tahmin etme (YY) değerleri gösterilmiştir.

Çizelge 4.1. Çalışmada kullanılan performans ölçüm yöntemi

Doğruyu Doğru Tahmin Etme(DD)	Doğruyu Yanlış Tahmin Etme(DY)
Yanlışı Doğru Tahmin Etme(YD)	Yanlışı Yanlış Tahmin Etme(YY)

4.1.1. ROC eğrisi ve AUC değeri

Makine öğrenmesinde önemli konulardan biri de performans ölçümüdür. Bu sebeple bir sınıflandırma probleminin performansını değerlendirmek için AUC değeri ve ROC eğrisinden yararlanılmaktadır. AUC, ROC eğrisi altındaki alandır ve bir değerdir. ROC ise alıcı çalışma karakteristiğini gösteren eğridir. ROC eğrisinin kapsadığı alan ne kadar büyükse ya da AUC değeri ne kadar büyükse sınıflandırma işlemi o kadar başarılı demektir. AUC için en ideal değer 1' dir. ROC eğrisinde y ekseni duyarlılığı, x ekseni ise seçiciliği göstermektedir. Tahminleme işleminde doğruyu doğru tahmin etme (DD), doğruyu yanlış tahmin etme (DY), yanlışı yanlış tahmin etme (YY), yanlışı doğru tahmin etme (YD) ihtimalleri söz konusudur. Duyarlılık (DD/DD+YD) ile ifade edilmektedir. Seçicilik ise (YY/DY+YY) ile ifade edilmektedir. Duyarlılığa doğru pozitiflik oranı, seçiciliğe de yanlış pozitiflik oranı da denilebilir (Kaya, 2017).

Veri seti değerlendirilirken ilk başta onlu çapraz doğrulama kullanılmış, daha sonra veri setinin %60'ı eğitim, %40'ı test olarak işleme alınmıştır.

4.1.2. Çapraz doğrulama (cross validation)

Çapraz doğrulama (ÇD), sınıflandırma modellerinin veri kümesi üzerinde yapılan sınıflandırma sonuçlarının tutarlı olması amacıyla kullanılan bir yöntemdir. Bu metot uygulanırken önce veri setinin kaç parçaya ayrılacağı tespit edilir. Daha sonra bu parçalardan biri test için ayrılarak diğer parçalar sınıflandırıcının eğitimine

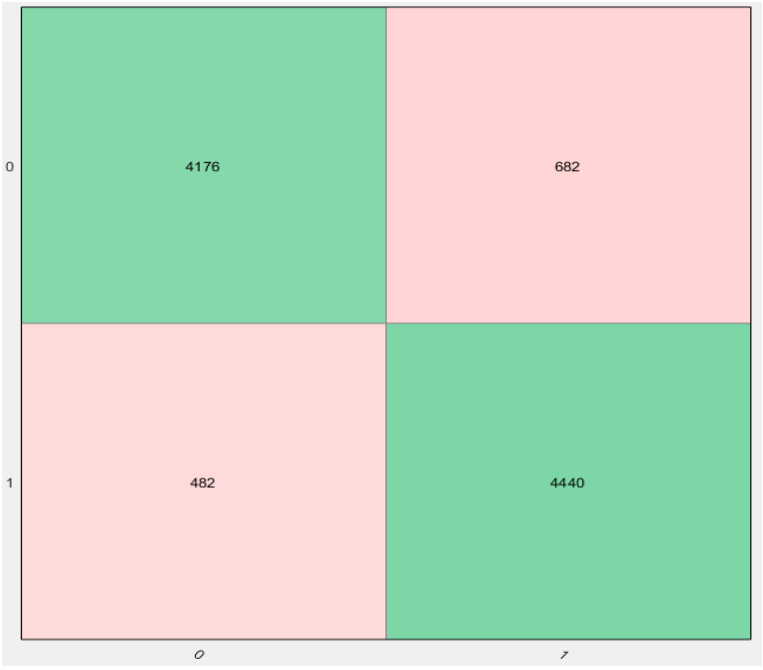
tabi tutulur. Bu işlem tüm parçaların eğitilip test edilmesine kadar devam ettirilir. Sonuçta sınıflandırma sonuçlarının ortalaması alınarak genel sınıflandırma sonucu elde edilir. Çizelge 4.2. de 10 parçaya bölünmüş bir veri setine ÇD uygulaması görülmektedir.

Çizelge 4.2. On parçaya bölünmüş veri setine ÇD uygulama örneği

Adım	1.Parça	2.Parça	3.Parça	4.Parça	5.Parça	6.Parça	7.Parça	8.Parça	9.Parça	10.Parça
1	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	**Test**
2	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	**Test**	Eğitim
3	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	**Test**	Eğitim	Eğitim
4	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	**Test**	Eğitim	Eğitim	Eğitim
5	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	**Test**	Eğitim	Eğitim	Eğitim	Eğitim
6	Eğitim	Eğitim	Eğitim	Eğitim	**Test**	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
7	Eğitim	Eğitim	Eğitim	**Test**	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
8	Eğitim	Eğitim	**Test**	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
9	Eğitim	**Test**	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
10	**Test**	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim

4.1.3. Algoritma sonuçları

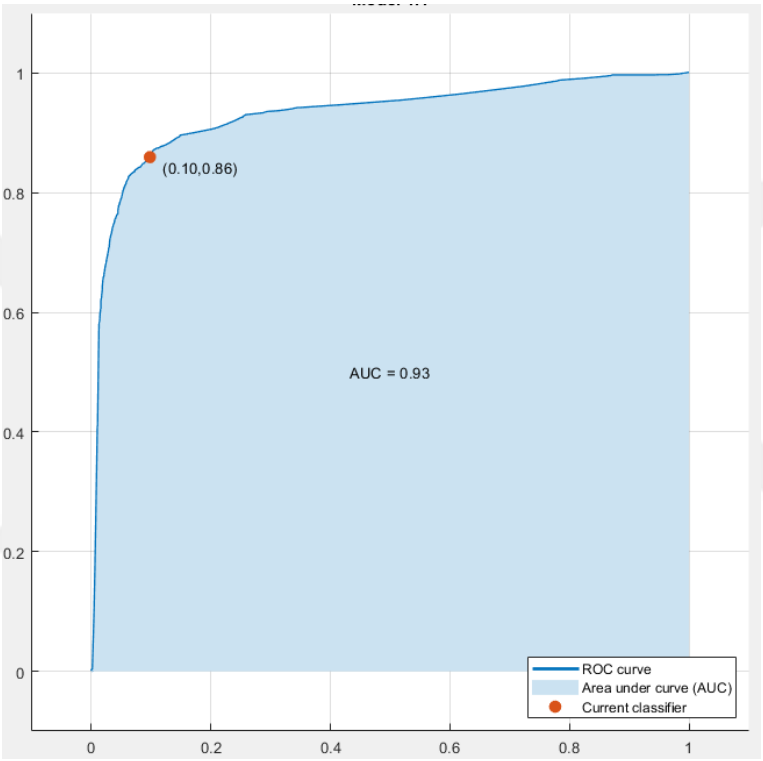
Veri seti normalizasyon işleminden sonra onlu ÇD uygulaması yapılarak KA yöntemi ile sınıflandırma işlemine tabi tutulmuştur. Şekil 4.2. de KA yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.2. KA Sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri

Sonuçlar incelendiğinde KA yöntemi ile elde edilen sayısal değerler şu şekilde olmuştur; 4176 doğruyu doğru, 682 doğruyu yanlış, 4440 yanlışı yanlış, 482 yanlışı doğru tahmin işlemi gerçekleşmiştir.

Aynı şekilde KA sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.3. de verilmiş ve AUC değeri 0.93 olmuştur.



Şekil 4.3. KA sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri

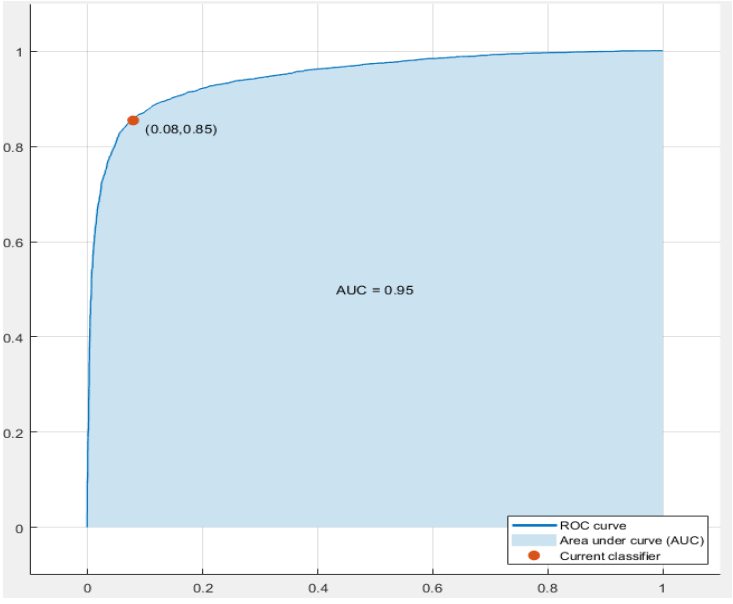
Trafik kaza şiddet tahmini için kullanılan bir diğer sınıflandırma yöntemi DVM olmuştur. Aynı şekilde sınıflandırma işlem süreci 10 ÇD ile uygulanmıştır. Şekil 4.4. de DVM yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.4. DVM sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri

DVM sınıflandırma yöntemi sonuçları incelendiğinde elde edilen sayısal değerler şu şekilde olmuştur; 4153 doğruyu doğru, 705 doğruyu yanlış, 4534 yanlışı yanlış, 388 yanlışı doğru tahmin işlemi gerçekleşmiştir.

DVM sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.5. de verilmiş ve AUC değeri 0.95 olmuştur.



Şekil 4.5. DVM sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri

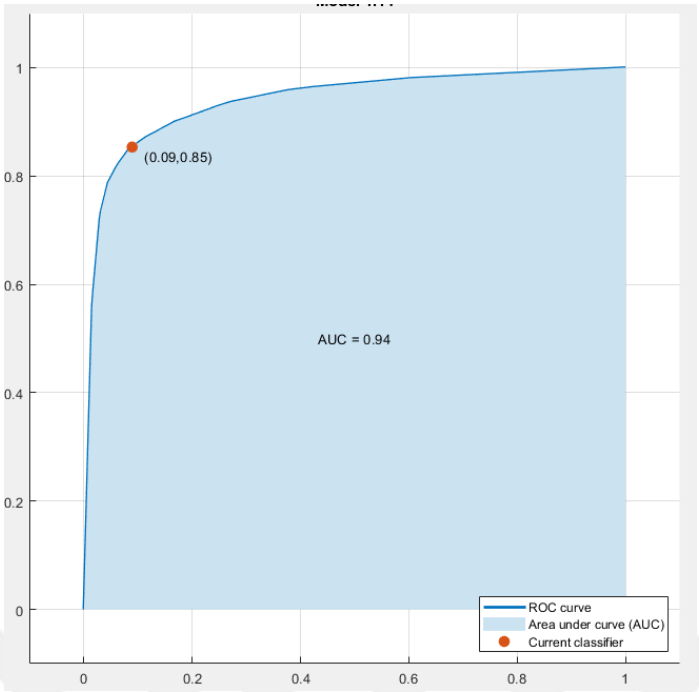
Aynı veri seti tekrar 10 ÇD ile bir başka sınıflandırma yöntemi olan KNN ile değerlendirilip trafik kaza şiddeti tahmini yapılmıştır. Şekil 4.6. da KNN sınıflandırma yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.6. KNN sınıflandırma yöntemi ve 10 ÇD kullanılarak elde edilen Confusion Matrix değerleri

KNN sınıflandırma yöntemi sonuçları incelendiğinde elde edilen sayısal değerler şu şekilde olmuştur; 4146 doğruyu doğru, 712 doğruyu yanlış, 4484 yanlışı yanlış, 438 yanlışı doğru tahmin işlemi gerçekleşmiştir.

KNN sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.7. de verilmiş ve AUC değeri 0.94 olmuştur.



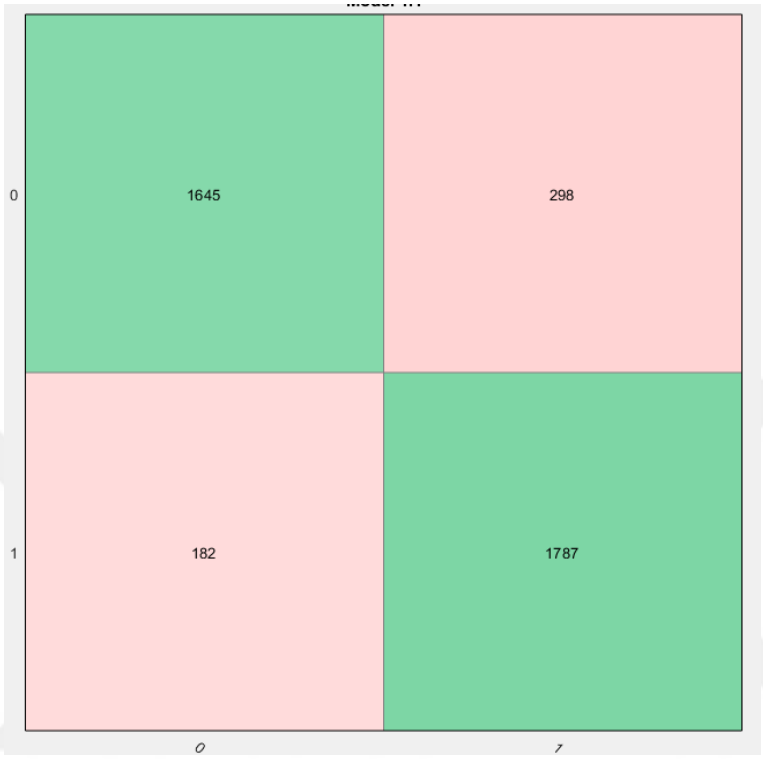
Şekil 4.7. KNN sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri

Yeni Zelanda trafik kazaları verilerinden oluşan veri seti kullanılarak kaza şiddeti tahmini KA, DVM ve KNN sınıflandırma yöntemleri kullanılarak yapılmış ve performans ölçütü olarak doğruluk kriteri kullanılmıştır. Kullanılan üç yöntemin başarı performansı karşılaştırıldığında en başarılı sonuç % 88.82 ile DVM yöntemi kullanılarak elde edilmiştir. Geriye kalan diğer iki yöntemin başarı performansları şu şekilde olmuştur; KA sınıflandırma yöntemi doğruluk başarısı % 88.09 ve KNN yönteminin doğruluk başarısı ise % 88.24 olmuştur.

ROC eğrisinden elde edilen AUC değerleri karşılaştırıldığında ise yine aynı şekilde en yüksek değer DVM sınıflandırma yöntemi kullanılarak elde edilmiş ve AUC değeri 0.95 olmuştur. Bir sonraki başarılı AUC değeri ise 0.94 olmuş ve KNN sınıflandırma yöntemi ile elde edilmiştir. KA yöntemi ile elde edilen AUC değeri ise 0.93 olmuştur.

Bir önceki uygulamada kullanılan aynı veri seti eğitim ve test verileri diye ayrıştırılıp sınıflandırma işlemine alınmıştır. Veri setinin % 60 ı eğitim, % 40 ı ise test verisi olmuştur. Aynı şekilde bir önceki uygulamada olduğu gibi tahminleme işleminde, KA, DVM ve KNN sınıflandırma yöntemleri kullanılmıştır.

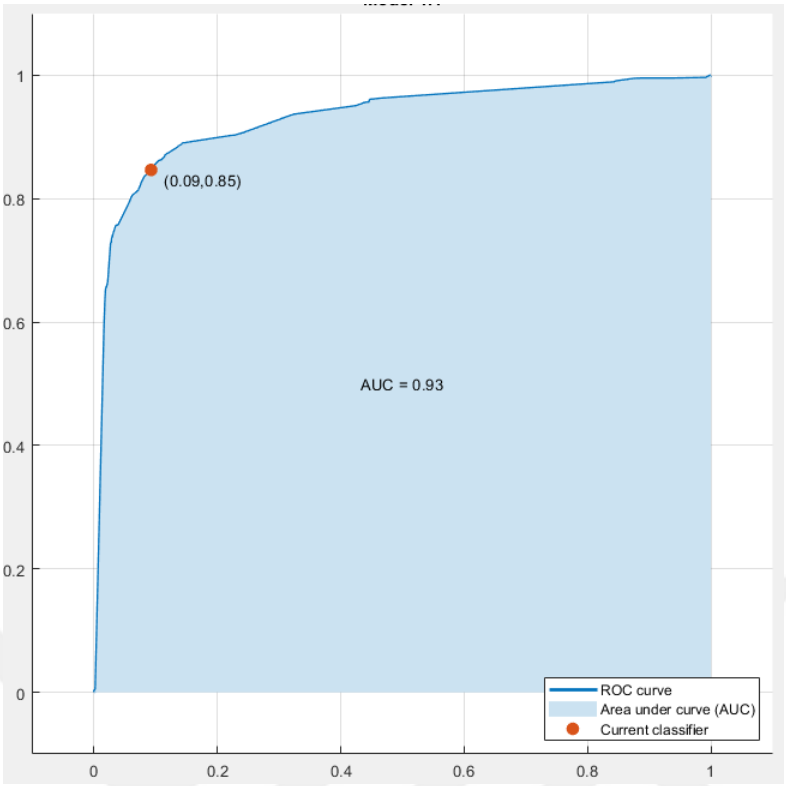
Eğitim ve test verileri ayrı olacak şekilde düzenlenmiş veri seti ilk önce makine öğrenme yöntemlerinden KA uygulanarak yapılmıştır. Şekil 4.8. de KA yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.8. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri

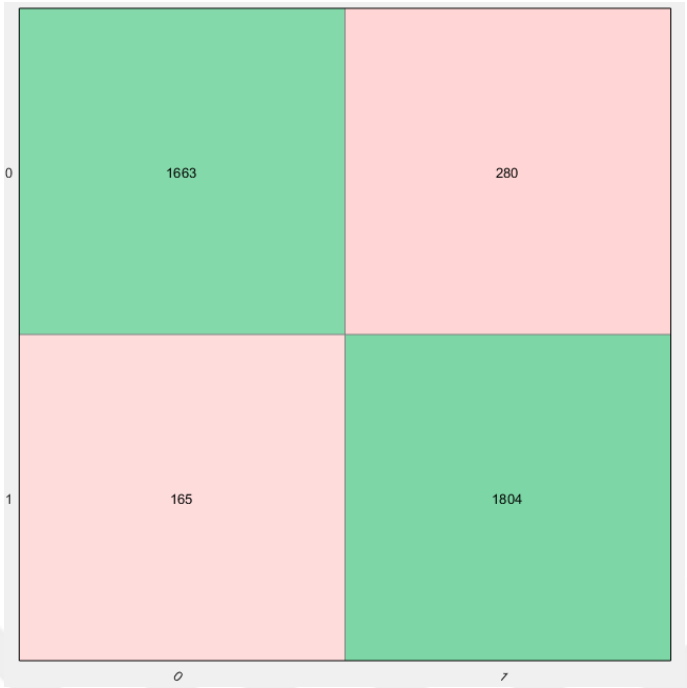
Sonuçlar incelendiğinde KA yöntemi ile elde edilen sayısal değerler şu şekilde olmuştur; 1645 doğruyu doğru, 298 doğruyu yanlış, 1787 yanlışı yanlış, 182 yanlışı doğru tahmin etme işlemi gerçekleşmiştir.

Aynı şekilde KA sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.9. da verilmiş ve AUC değeri 0.93 olmuştur.



Şekil 4.9 Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri

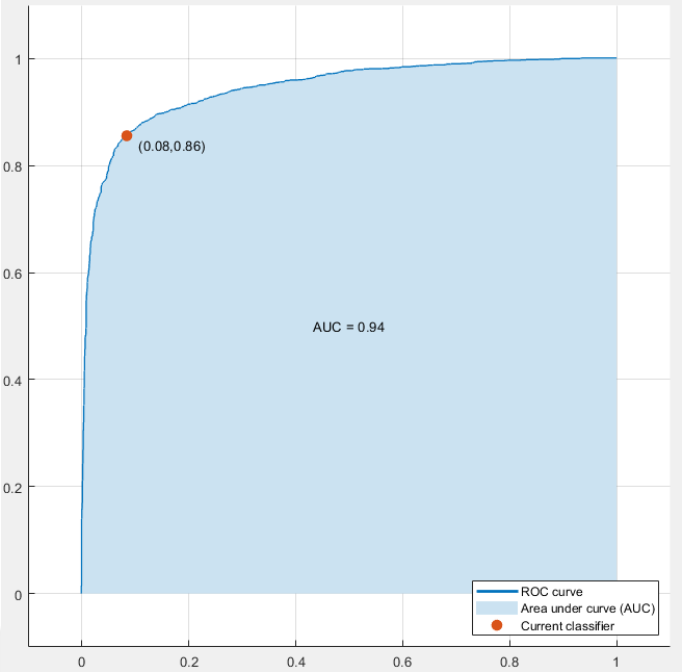
Yine aynı şekilde eğitim ve test verilerinin ayrı ayrı düzenlenerek oluşturulan veri seti DVM sınıflandırma yöntemi uygulanarak incelenmiştir. Şekil 4.10. da DVM yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.10. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve DVM sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri

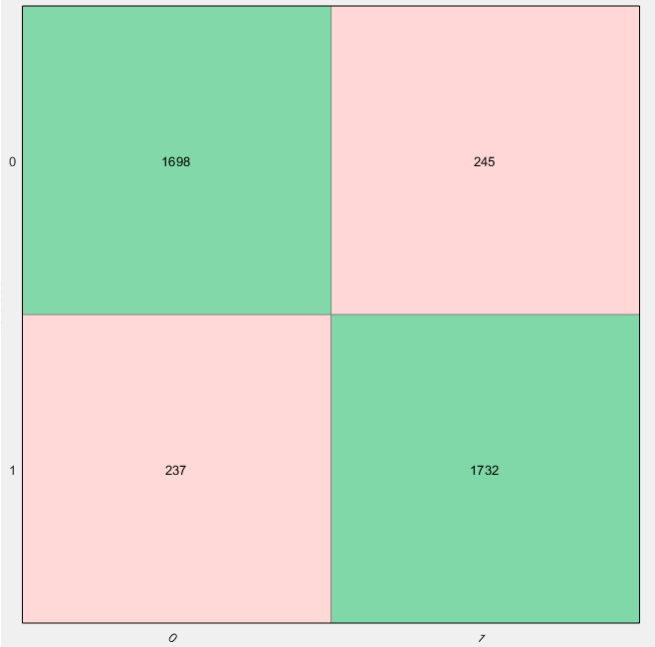
DVM sınıflandırma yöntemi sonuçları incelendiğinde elde edilen sayısal değerler şu şekilde olmuştur; 1663 doğruyu doğru, 280 doğruyu yanlış, 1804 yanlışı yanlış, 165 yanlışı doğru tahmin işlemi gerçekleşmiştir.

DVM sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.11. de verilmiş ve AUC değeri 0.94 olmuştur.



Şekil 4.11. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KA sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri

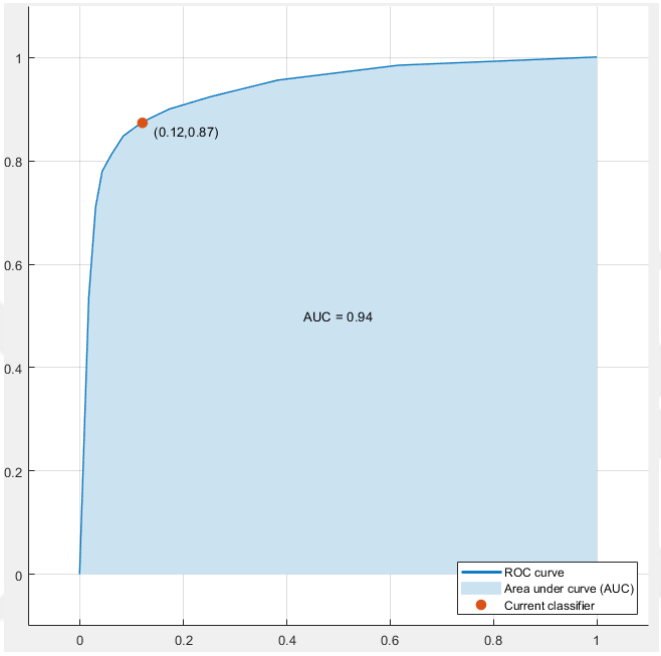
Bir önceki sınıflandırma yöntemi uygulamasında kullanılan veri seti bu kez KNN sınıflandırma yöntemi uygulanarak işleme alınmıştır. Şekil 4.12. de KNN yöntemi kullanılarak elde edilen Confusion Matrix değerleri verilmiştir.



Şekil 4.12. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KNN sınıflandırma yöntemi kullanılarak elde edilmiş Confusion Matrix değerleri

KNN sınıflandırma yöntemi sonuçları incelendiğinde elde edilen sayısal değerler şu şekilde olmuştur; 1698 doğruyu doğru, 245 doğruyu yanlış, 1732 yanlışı yanlış, 237 yanlışı doğru tahmin işlemi gerçekleşmiştir.

KNN sınıflandırma yöntemi kullanılarak elde edilen ROC eğrisi ve AUC değeri Şekil 4.13. de verilmiş ve AUC değeri 0.94 olmuştur.



Şekil 4.13. Veri seti %60 eğitim-%40 test şeklinde ayrılmış ve KNN sınıflandırma yöntemi kullanılarak elde edilmiş ROC eğrisi ve AUC değeri

Eğitim ve test verileri ayrıştırılarak oluşturulan veri setine KA, DVM ve KNN sınıflandırma yöntemleri uygulanarak kaza şiddeti tahmini yapılmaya çalışılmıştır. Başarı performans ölçütü olarak doğruluk kriteri kullanılmıştır. Kullanılan üç yöntemin başarı performansı karşılaştırıldığında en başarılı sonuç % 92.84 ile KA yöntemi kullanılarak elde edilmiştir. Geriye kalan diğer iki yöntemin başarı performansları şu şekilde olmuştur; DVM sınıflandırma yöntemi doğruluk başarısı % 88.62 ve KNN yönteminin doğruluk başarısı ise % 87.67 olmuştur.

ROC eğrisinden elde edilen AUC değerleri karşılaştırıldığında ise yine aynı şekilde en yüksek değer DVM ve KNN sınıflandırma yöntemleri kullanılarak elde edilmiş ve AUC değeri 0.94 olmuştur. diğer sınıflandırma yöntemi olan KA ile elde edilen AUC değeri ise 0.93 olmuştur.

Veri setinin 10 ÇD ayrışma türüne göre uygulanmış sınıflandırma yöntemlerinin doğruluk kriterine göre başarı performans değerleri Çizelge 4.3. de verilmiştir.

Çizelge 4.3. 10 ÇD ile oluşturulan veri setinin doğruluk ölçütü performans değerleri

Kullanılan Yöntemler	Başarı (%)	DD YD	DY YY
DVM	88.82	4153	705
		388	4534
KNN	88.24	4146	712
		438	4484
KA	88.09	4176	682
		482	4440

Çizelge 4.3. de görüldüğü gibi en yüksek doğruyu doğru tahmin sayısı KA sınıflandırma yöntemi ile 4176 olmuştur. En yüksek yanlışı yanlış tahmin sayısı ise DVM sınıflandırma yöntemi ile 4534 olarak gerçekleşmiştir.

Veri seti ayrıştırma işlemi %60 eğitim %40 test olacak şekilde uygulanmış çalışmada sınıflandırma yöntemlerinin doğruluk kriterine göre başarı performans değerleri Çizelge 4.4. de verilmiştir. Çizelge 4.4. de görüldüğü gibi en yüksek doğruyu doğru tahmin sayısı KNN sınıflandırma yöntemi ile 1698 olmuştur. En yüksek yanlışı yanlış tahmin sayısı ise DVM sınıflandırma yöntemi ile 1804 olarak gerçekleşmiştir.

Çizelge 4.4. %60 eğitim %40 test ile oluşturulan veri setinin doğruluk ölçütü performans değerleri

Kullanılan Yöntemler	Başarı (%)	DD YD	DY YY
DVM	88.62	1663	280
		165	1804
KNN	87.67	1698	245
		237	1732
	92.84	1645	298
KA		182	1787

Veri setine onlu çapraz doğrulama uygulandıktan sonra yapılan sınıflandırma işleminde en başarılı olan DVM yönteminin performansı % 88.82 iken, % 60 eğitim % 40 test uygulandıktan sonra yapılan sınıflandırma sonucunda en başarılı olan KA yönteminin performans değeri % 92.84 olmuştur.

5. SONUÇLAR ve ÖNERİLER

Yapılan çalışmada Yeni Zelanda Ulaştırma Ajansından alınan; yolların eğimi, şerit sayısı, aydınlığı, kuru veya ıslaklığı, olabilecek hız sınırı, yollarda bulunan yaya, bisikletli ve ağaç gibi bilgilerin bulunduğu veri seti kullanılarak yollarda oluşan kazaların ciddi veya basit hasarlı oluşu tahmin edilmiştir. Tahminleme işlemi makine öğrenme algoritmalarından KNN, DVM ve KA yöntemleri kullanılarak yapılmıştır. Genel olarak kullanılan üç yöntemden alınan performans değerlerinin yüksek oluşu, bu tür yöntemlerin olası trafik kazalarının önlenmesi çalışmalarında yardımcı olabileceğini göstermiştir. KA %92.84 oranıyla en başarılı yöntem olarak tespit edilmiştir.

Türkiye'de trafik kaza istatistikleri daha detaylı ve düzenli tutulduğu taktirde bu yöntemlerle trafik kaza tahminleri çalışmaları daha verimli olacaktır. Bu tahminler ışığında gerek karayollarının alt yapısında düzenlemeler yapılarak, gerekse trafik kazalarının diğer etmenleriyle ilgili tedbirler alınarak kazalar minimum seviyeye indirilebilir.

Trafik kazalarının oluşumu çok sayıda etmene bağlı olduğundan, yapılan bu çalışmadaki parametreler dışında bir çok parametre daha dikkate alınarak, daha geniş çalışmalar yapılabilir.

KAYNAKLAR

ATİKTÜRK TAŞDELEN, D., 2019. Veri Madenciliğinin Tıp ve Sağlık Hizmetlerinde Uygulamaları. Yüksek Lisans Tezi, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 55s.

BABKOV, V.F., 1975. Road Conditions and Traffic Safety, MIR Publishers, Moscow.

BAĞIRGAN, N., 2006. Şehirlerarası Karayollarında Trafik Güvenliği Tahmini. Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Isparta, 156s.

BAŞ, F., 2015. Geçici Fiziksel Özelliklerden Yorgunluk ve Uykusuzluğun Sürücü Davranışı Üzerine Etkisi. Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Erzurum, 72s.

BAYIR, F., 2006. Yapay Sinir Ağları ve Tahmin Modellemesi Üzerine Bir Uygulama. Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 122s.

BHATİA, N. and VANDANA, 2010. Survey of nearest neighbor techniques. International Journal of Computer Science and Information Security, 8(2):302-305.

BİCAN, B.. 2014. Zaman Serilerinin Öngörüsü İçin GKA Tabanlı DVR Metodları. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 79s.

BLENSLEY, R.C., HEAD, J.A., 1960. Statistical Determination of Effect of Paved Shoulder Width on Traffic Acciden Frequency, HRB, Bull. 240, pp.1-23.

BOLAKAR, H., 2014. Yapay Sinir Ağları ile Trafik Kazalarının Modellemesi: Erzurum ili Örneği, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Erzurum, 109s.

BRENAC, T., 1990. Speed, Safety and Highway Design. Recherche Transports Securite, English Issue, No:5.

BRENAC, T., ANNEX TX (E), 1994. Curves on Two-lane Roads. Safety Effects of Road Design Standards, SWOV, Leidschendam The Netherlands.

CHOVEİRİ, E.M., LAMM, R., KLOECKNER, J.H., MAİLOENDER, T., 1994. Safety Aspects of Individual Design Elements and Their Interactions on Two-Lane Highways: International Perspective, Transportation Research Record 1445, pp.34-46.

CLEVELAND, D.E., KOSTYNİUK, L.P., TİNG, K., 1984. Geometric Design Element Groups and High-Volume Two-Lane Rural Highway Safety, In Transportation Research Record 960, TRB, National Research Council, Washington, D.C.

CORTES, C., VAPNİK V., 1995. Support Vector Networks. Machine Learning, vol.20, , pp.273-297.

ÇAKIR, Ş., 1996. Trafik Kazalarında Delillerin Saptanması. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara, 116s.

ÇALIŞKAN, S. K., ve SOĞUKPINAR, İ., 2008. KxKNN: K-Means ve K En Yakın Komşu Yöntemleri İle Ağlarda Nüfuz Tespiti. EMO Yayınları, 120-24.

ÇANAKÇI, H., 2001. Yol Altyapısı Çalışmalarındaki Uygulama Hatalarının Yol Ve Trafik Güvenliğine Etkisi. Trafik ve Yol Güvenliği Kongresi, Ankara.

ÇELİK AK, OKTAY E., 2014. 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.

ÇELİK, U., AKÇETİN, E. ve GÖK M., 2017. Rapidminer ile Uygulamalı Veri Madenciliği kitabında "Karar Ağaçları" Bölümü (Şebnem Özdemir), Pusula Yayıncılık.

ÇİNİCİOĞLU, E. N., ATALAY, M., ve YORULMAZ, H., 2013. Trafik Kazaları Analizi İçin Bayes Ağları Modeli. Bilişim Teknolojileri Dergisi, 6(2), 41.

ÇUBUK, K VE HATİPOĞLU, S., 2015. Trafik Planlaması ve Uygulaması A.B.D., Fen Bilimleri Enstitüsü, Gazi Üniversitesi.

DOAD, P.K. and BARTERE, M.M., 2013. A Review : Study of Various Clustering Techniques. International Journal of Engineering Research & Technology, 2(11):3141-3145.

DOĞAN, E.; SEBAHATTİN I. ve MEHMET S., 2007. Günlük Buharlaşmanın Yapay Sinir Ağları Kullanarak Tahmin Edilmesi. İMO Teknik Dergi, ss. 4119 -4131.

DONDURMACI, G.A. ve ÇINAR, A., 2014. Finans Sektöründe Veir Madenciliği Uygulaması, Akademik Sosyal Araştırmalar Dergisi, 2(1): 258-271.

DUYGU, K., ve TÜRK M., 2017. Biyoelektriksel İşaretlerde Rahatsızlık Teşhisinin Yorumlanması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 29(1), 261-267.

ERASTÖ, P., 2001. Support Vector Machines-Backgrounds and Practice, Academic Dissertation for The Degree of Licentiate of Philosophy. Rolf Nevanlinna Institute, Helsinki.

FRİDSTROM, L., et al., 1995. Measuring the Contribution of Randomness, Exposure,Weather, and Daylight to the Variation in the Road Accident Counts, Accident Analysis and Prevention, 27(1): 1-20.

GLENNON, J.C., 1987. Relationship Between Saffety and Key Highway Features. A Synthsis of Prior Research, State of the Art Report 6. Transport Research Board, National Research Council, Washington, D.C.

HAN, J. and KAMBER, M., 2006. Data mining: concepts and techniques. Morgan Kaufmann Publishers, Burlington.

HAYKİN, S., 2001. Neural Networks: A Comprehensive Foundation, Tsinghua University Press,.

https://colab.research.google.com/gist/shakasom/382a5cc0c59e24fdae29679de6d7e88a/crash-analysis.ipynb#scrollTo=m218tlnzKeFH.

https://www.transport.govt.nz/mot-resources/road-safety-resources/road-deaths/.

İYİNAM, A.F., 1997. Karayolu Güvenliği ile Yol Geometrik Standartları Arasındaki İlişkilerin Analizi, İTÜ Fen Bilimleri Enstitüsü, Doktora Tezi, İstanbul.

JAYALAKSHMİ, T. and SANTHAKUMARAN, A., 2011. Statistical Normalization and Back Propagation for Classification. International Journal of Computer Theory and Engineering, 3(1), pp. 1793-8201.

JORGENSEN, R., ASSOCİATES, Inc., 1978. Cost and Safety Effectiveness of Highway Design Elements, National Cooperative Highway Research Program, Report 197, Transportation Research Board, National Research Council, Washington, D.C.

KANTARDZİC, M., 2011. Data Mining: Concepts, Models, Methods, and Algorithms.

KGM Trafik Kazaları Özeti 2018

KİBAR, F. T., 2015. Türkiye'de kamyon kazaları ile trafik ve karayolu geometrik özellikleri arasındaki ilişkinin istatistiksel ve yapay sinir ağları yöntemleri ile modellenmesi. Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Trabzon, 154s.

KİYİLDİ, R. K., 2017. Türkiye için Yapay Sinir Ağları Yöntemi ile Trafik Kazası Tahmini Araştırması. In 5th International Symposium on Innovative Technologies in Engineering and Science 29-30 September 2017 (ISITES2017 Baku-Azerbaijan).

KREBS, H.G., KLOECKNER, J.H., 1977. Investigation of the Effects of Highway and Traffic Conditions Outside Built-Up Areas on Accident Rates, Techinal Journal: Forschung Strassenbau und Strassenverkehrstechnik, Vol. 223.

KÜRŞAT KESKİN, A., 2018. Makine Öğrenmesi Sınıflandırma Algoritmalarının İncelenmesi. Sinop Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Sinop, 42s.

MİTCHELL, T., 1997. Machine Learning. McGraw Hill, New York.

OSOWSKİ, S., SİWEKAND, K., and MARKİEWİCZ, T., 2004. MLP and SVM Networks – a Comparative Study. Proceedings of the 6th Nordic Signal Processing Symposium – NORSIG.

ÖZDEN, C., ve ACI, Ç., 2018. Makine öğrenmesi yöntemleri ile yaralanmalı trafik kazalarının analizi: Adana örneği. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 24(2), 266-275.

ÖZGAN, E., 2003. Sivas İli Çevre Devlet Karayollarında Meydana Gelen Trafik Kazalarının Çok Yönlü Klinik Araştırması ve Kritiği. Gazi Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Ankara, 176s.

ÖZGAN, E., ULUSU, H. Ve YILDIZ, K., 2004. Trafik kaza verilerinin analizi ve kaza tahmin modeli. Sakarya Üniversitesi, Fen Bilimleri Enstitüsü Dergisi, 8(1), 160-166.

PERCHONOK, K., RANNEY, T.A., BAUM, A. S., MORRİS, D. F., EPPİCH, J. D., 1978. Hazardous Effects of Highway Features and Roadside Objects, Vol. 1 and 2, Reports FHWA-RD-78-201 and FHWA-RD-78-202, FHWA, U.S. Department of Transportation.

RUMAR, K. 1985. The Role of Perceptual and Cognitive Filters in Observed Behavior: Human Behavior in Traffic Safety, New York: Plenum Press.

SATTERTHWAİTE, S.P., 1981. A Survey of Research into the Relationships Between Traffic Accidents and Traffic Volumes, Transport and Road Research Laboratory Supplementary Report 692, United Kingdom.

SOHN SY, SHİN H., 2010. Pattern recognition for road traffic accident severity in Korea. Ergonomics, 44(1), 107-117.

SRİNİVASAN, S., 1982. Effect of Roadway Elements and Environment on Road Safety, Institution of Engineers, Vol. 63.

TEZER, D., 2018. Yapay Sinir Ağları, Lojistik Regresyon ve Destek Vektör Makinesi İstatistik Yöntemlerinin Sınıflandırmadaki Karşılaştırılması. Biruni Üniversitesi Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 91s.

TOSUN, S., 2007. Sınıflandırmada Yapay Sinir Ağları ve Karar Ağaçları Karşılaştırması: Öğrenci Başarıları Üzerine Bir Uygulama. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 115s.

TRB, 1978. Cost and Safety Effectiveness of Highway Design Elements, NCH RPT 197.

Türkiye İstatistik Kurumu Karayolu Trafik Kaza İstatistikleri, 2014. http://www.tasimadunyasi.com/gundem/karayolu-trafik-kaza-istatistikleri-2014-h2071.html

Türkiye İstatistik Kurumu Karayolu Trafik Kaza İstatistikleri, 2015. http://www.tuik.gov.tr/PreHaberBultenleri.do;jsessionid=x3gTYfLVG20dQYhnJyy7GDnS9G1CTrgpL0tP2Bf7YJsk6x92Bsz1!704814975?id=21611.

VAPNİK, V.N. 1995. The Nature of Statistical Learning Theory, New York, Springer–Verlag,.

VAPNİK, V.N., 1998. Statistical Learning Theory, New York, NY, Wiley.

VAPNİK, Vladimir N., 1998. Statistical Learning Theory. New York, Wiley Interscience.

YAGER, M., VAN AERDE, R., 1983. Geometric and Environmental Effects on Speeds of 2-Lane Highways, Transportation Research, Vol. 17A, No 4.

YÜKSEK, A. G.; BİRCAN, H., ZONTUL, M. ve KAYNAR, O., 2007. Sivas İlinde Yapay Sinir Ağları İle Hava Kalitesi Modelinin Oluşturulması Üzerine Bir Uygulama. Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, 8(1), ss. 97-112.

ZEGEER, C.V., REİNFURT, D.W., HUMMER, J., HERF, L., HUNTER, W., 1986. Safety Effects of Cross-Section Design for Two-Lane Roads, FHWA-RD-87-008, Federal Highway Administration, Washington, D.C.

ZEGEER, C.V., STEWART, R., COUNCİL F., and NEUMAN, T.R., 1994. Accident Relationship of Roadway Width on Low-Volume Roads, Paper 940864 presented at the 73rd Annual Meeting, National Research Council, Washington, D. C.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Mehmet Ayhan ÇETİN

Uyruğu : TC

Doğum Yeri ve Tarihi : Şanlıurfa 17.03.1975

Telefon : 05055810316

Faks :

e-mail : cetinayhan2@hotmail.com

EĞİTİM

Derece	Adı, İlçe, İl	Bitirme Yılı
Lise	: Şanlıurfa Teknik ve End. Meslek Lisesi, Merkez, Şanlıurfa	1992
Üniversite	: Gazi Üniversitesi Teknik Eğitim Fakültesi, Merkez, Ankara Elektrik Öğretmenliği	1997
Yüksek Lisans	: Harran Üniversitesi Elektrik-Elektronik Müh., Haliliye, Şanlıurfa	Devam Ediyor

İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
1997-1998	Nizip Mesleki Eğitim Merkezi	Elektrik Öğretmenliği
1998-1999	Tercan Endüstri Meslek Lisesi	Elektrik Öğretmenliği
1999-2002	Nizip Mesleki Eğitim Merkezi	Elektrik Öğretmenliği
2002-	Urfa Mesleki ve Teknik Anadolu Lisesi	Elektrik Öğretmenliği

UZMANLIK ALANI

YABANCI DİLLER

BELİRTMEK İSTEĞİNİZ DİĞER ÖZELLİKLER

YAYINLAR