

Trafik Kazalarında Birliktelik Kuralı Analizi: Ankara İli Örneği

Association Rules on Traffic Accident: Case Of Ankara

İsmet SÖYLEMEZ¹, Ahmet DOĞAN², Uğur ÖZCAN³

Özet

Bu çalışmada Ankara ilinde 2010 yılında meydana gelen trafik kazaları için veri madenciliği tekniklerinden birliktelik kuralı analizi kullanılarak kural çıkarımı çalışması yapılmıştır. Bu kurallar sayesinde farklı hava koşullarında (karlı, yağmurlu vb.), kazanın gerçekleştiği yollar (cadde, sokak, otoyol vb.) ve yol bölünmüşlüğü (bölünmüş veya bölünmemiş yol gibi) gibi durumlarda hangi olasılıklarla kazalarda insanların zarar görmediği (kaza sonucu kimsenin ölümü veya yaralı olmama durumu) belirlenmiş olacaktır. Birliktelik kuralı analizi için farklı algoritmalar kullanılmaktadır. Bu çalışma için Apriori algoritması seçilmiş ve sonuçların elde edilmesi için de SPSS Clementine 12.0 programı kullanılmıştır. Apriori algoritması için öncelikle sık geçen öğe kümelerini bulmak gerekir. Daha sonra sık geçen öğelerin birleştirilmesi işlemi gerçekleştirilir. Bu çalışmada ilk olarak; veri ön işleme yani kaza verilerinin düzenlenmesi ve kayıp verileri tamamlama ya da o verinin çıkarılması işlemi gerçekleştirilmiştir. İkinci aşama ise, aykırı değerlerin belirlenip çıkartılması ve veri tiplerinin 1-0 (binary) tipine dönüştürülmesidir. Üçüncü olarak; Apriori algoritmasının uygulanması kullanılarak kural çıkarımın gerçekleştirilmiş ve elde edilen kurallar değerlendirilmiştir.

Anahtar kelimeler: Apriori algoritması, Birliktelik kuralı, Kaza verileri.

Abstract

In this study, association rules analysis of the data mining techniques are used for data of traffic accidents in 2010 and some rules are obtained. With this rules, what is the possibility of accident which resulted anybody injured for "different weather conditions (snowy, rainy etc.)", "where the accidents occurred (street, road etc.)" and "way situations (separated road or not)". Different algorithms are used to analyze the association rules. Apriori algorithm is selected for this study and SPSS Clementine 12.0 is used for this algorithm. Firstly, frequency of items are found. Then, items are grouped. In this study, data preprocessing is done and missing values are filled or rejected. In the second phase, outliers are rejected and data type is converted type of 1-0 (binary). In the third phase, Apriori algorithm is applied and results are evaluated.

Keywords: Apriori algorithm, Association rules, Accident data.

1. GİRİŞ

Günümüzde veri tabanlarında çeşitli veriler kaydedilmektedir. Bu verilerin akademik çalışmalar dışında faydalı bir şekilde kullanımı çok azdır. Genellikle veriler depolanır ve zamanla veri deposundan silinir. Veri madenciliği teknikleri yardımı ile bu tip veriler değerlendirilerek mevcut sistemlerin hataları veya eksikleri tespit edilip sistemde iyileştirmeler gerçekleştirilebilir. Günümüzde araç kullanımı oldukça fazladır. Özellikle büyük şehirlerde trafik kazaları artan araç kullanımı ile önlenemez hal

almaktadır. Trafik kazaları maddi sonuçların yanında istenmeyen bir durum olan yaralanma veya ölüm ile sonuçlanmaktadır. Bu çalışmada Türkiye'nin başkenti olan Ankara ilinde meydana gelmiş olan trafik kaza verileri kullanılmıştır. Ankara'da artan yaralı ve ölümlerle sonuçlanan kazaların minimizasyonu için veri madenciliği tekniklerinden olan birliktelik kuralı analizi kullanılacaktır. Birliktelik kuralı analizi büyük veri tabanlarında farklı ilişkiler ve korelasyonları bulmak için kullanılır. Bu ilişki kuralları belirli bir güven değerinde tespit edilir.

¹Abdullah Gül Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 38039, Kocasinan, Kayseri, ismet.soylemez@agu.edu.tr

²Osmaniye Korkut Ata Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Karacaoğlan Yerleşkesi, 80000, Osmaniye

³Gazi Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 06570, Maltepe, Ankara

Bu çalışmanın amacı, gerçek hayat verilerinin literatürde mevcut olan yöntemlerden en uygun olanı kullanılarak analiz edilmesidir. Yapılan detaylı araştırma sonucu literatürde yer alan yöntemlerden birliktelik kuralı analizi ve algoritmalar arasında da Apriori algoritması seçilmiştir. Bu algoritma ile Türkiye'nin başkenti olan Ankara ilinde 2010 yılında meydana gelen trafik kazaları verileri kullanılarak analiz yapılmıştır. Bu çalışmanın literatüre katkısı ise bir gerçek hayat uygulaması olması ve bu sonuçlar sayesinde Ankara'da ve hatta tüm Türkiye'de farklı durumlarda ve özelliklerdeki yollarda ne tür önlemlerin alınması gerektiğidir. Çalışmanın bir başka çıktısı olarak, elde edilen sonuçlar ve önerilerle kazaların yol açtığı ölüm, yaralanma gibi travmatik durumların ya da sadece maddi hasarla sonuçlanan maliyet durumlarının neden olabileceği etkileri azaltma çabaları ve bu yolla ülkeye katma değer yaratma olarak görülebilir.

Bu çalışmanın kalan bölümleri şu şekildedir: ikinci bölümde veri madenciliği birliktelik kuralı analizi ve Apriori algoritmasından bahsedilecektir. Üçüncü bölümde SPSS Clementine 12.0 programı sayesinde elde edilmiş olan kurallar ve bu kuralların yorumlanması yer almaktadır. Son bölümde de genel olarak değerlendirme verilmektedir.

2. LİTERATÜR ARAŞTIRMASI

Bu bölümde literatür üç aşamada incelenecektir. İlk olarak trafik kazaları hakkında ne tür çalışmalar yapıldığı verilecektir. Daha sonra veri madenciliği teknikleri (sınıflandırma, kümeleme vb.) ve diğer bazı teknikler (TOPSIS, AHP vb.) ile uygulama çalışması yapılmış olan çalışmalar sunulacaktır. En son olarak özellikler birliktelik kuralı analizi kullanılarak gerçekleştirilen çalışmalara yer verilecektir.

Trafik kazalarıyla ilgili literatürde yer alan bazı çalışmaları incelediğimizde, Eken vd. (2005), Antalya sınırları içerisinde 1 Ocak 2005 - 31 Ağustos 2005 tarihleri arasında motorlu taşıt kazalarına neden olan faktörleri ve kazaların demografik özelliklerini istatistiksel analiz yaparak incelemiştir. Temel ve Özcebe (2006) çalışmalarında Türkiye'deki trafik kazaları epidemiyolojisini, nedenlerini ve trafik kazalarından korunmak için nelere dikkat edilmesi gerektiği konularını ele almışlardır. Enache vd. (2009) Romanya'da 2005-2007 yılları arasında ki verilerle yapmış oldukları çalışmada trafik kazalarında insan vücudunun en savunmasız bölgelerini tespit etmeye çalışmışlardır. Akgüngör ve Doğan (2010), İzmir ili için 1986-2005 yılları arasındaki verilerle, nüfus, araç sayısı ve kaza sayısını model parametreleri

olarak kullanarak regresyon analizi, yapay sinir ağları ve genetik algoritma ile kaza tahmin modelleri geliştirmeye çalışmışlardır. Aköz ve Karslıgil (2010) çalışmalarında öykünüm ortamında model arabalar ve yollar kullanarak, kavşaklardaki trafik kazalarının insan algısına dayalı incelenmesine ve kazalardan anlamsal çıkarım yapılabilmesine olanak sağlayan görüntü tabanlı trafik kaza analizi yaklaşımını önermişlerdir. Alp ve Engin çalışmalarında (2011) trafik kazaları konusunda uzun süredir birliktelik yapan akademisyenlere anket uygulayarak, TOPSIS ve AHP yöntemleri ile trafik kazalarının nedenleri ile sonuçları arasındaki ilişkiyi matematiksel olarak hesaplanmış ve analiz etmişlerdir. Ali ve Tayfour (2012) çalışmalarında 1991 ve 2009 yılları arasındaki verilerle yol güvenliği ve kaza karakteristiklerini yapay sinir ağları yöntemi ve regresyon teknikleriyle modelleyerek kaza tahminlemesi ve kazalara neden olan faktörler için önleyici tedbirleri ele almışlardır. Şenel B. ve Şenel M. (2013) çalışmalarında, TÜİK (Türkiye İstatistik Kurumu)'den elde ettikleri 2001-2007 yılları arasındaki kaza istatistikleri verileri ile hata ağaç analizi metodolojisini kullanarak trafikteki risk faktörlerini ve bu risk faktörlerinin derecelerini belirlemeye çalışmışlardır. Ona vd. (2013) İspanya'nın Granada ilinde kırsal karayollarında 2005-2008 yılları arasında meydana gelen trafik kazalarını inceleyerek, örtük kümeleme analizi ve bayes ağları tekniklerini birlikte kullanarak trafik kazalarında meydana gelen hasar şiddetlerini belirlemeye çalışmışlardır. Sungur vd. (2014), Türkiye'deki trafik kazalarını analiz ederek Türkiye'de yol güvenliği ve trafik kazaları sorununa dikkat çekmeye çalışmışlardır. Mirzaei vd. (2014), çalışmalarında İran sürücülerinin trafik kurallarına ilişkin bilgi, tutum ve uygulamaları ile bunların trafik kazaları üzerindeki belirleyici etkileri arasındaki ilişkinin değerlendirilmesine yönelik istatistiksel verilere dayalı bir analiz gerçekleştirmişlerdir. Dimitriou ve Vlahogianni (2015) bulanık kural tabanlı bir sistem geliştirerek, bu sistemin trafik ve hava koşulları ile ilgili sınırlı bilgi kullanarak kaza sürelerini tahmin edebileceğini ileri sürmüşlerdir. Hong vd. (2015) hızla yaşanan toplumda yaşlı nüfusta trafik kazalarının görülme sıklığı ve buna bağlı faktörleri incelemişlerdir. Kar vd. (2016) Bhubaneswar/Odisha Hindistan'da trafik kazalarının analizini yapmış kentteki kazaların epidemiyolojik değişkenliklerini belirleyerek trafik polislerine ve kent yönetimine tavsiyelerde bulunmayı hedeflemişlerdir. Waterson vd. (2016) çalışmalarında, Rasmussen tarafından geliştirilen kaza analizleri için kullanılan sistem

tabanlı bir teknik olan Accimap (sistemik kaza gelişimini incelemişlerdir. İncelenen çalışmalar Tablo 1'de özet olarak gösterilmiştir.

Tablo 1: Trafik Kazaları Literatür Araştırması

Yazarlar	Yıl	Açıklama	Yöntem
Kar vd.	2016	Bhubaneswar/Odisha'da trafik kazalarının analizi	İstatistiksel analiz
Waterson vd.	2016	Sistemik kaza analizinde Accimap yaklaşımının gelişimi	Accimap Tekniği
Dimitriou ve . Vlahogianni	2015	Yağış ve trafik akışı etkileşimleri ile otoyol kaza süresinin bulanık modellemesi	Bulanık Kural Tabanlı sistem
Hong vd.	2015	Hızla yaşanan toplumda yaşlı nüfusta trafik kazalarının görülme sıklığı ve buna bağlı faktörler dikkate alınmıştır.	İstatistiksel analiz
Sungur vd.	2014	Türkiye'de trafik kazalarının analizi	İstatistiksel analiz
Mirzaei vd.	2014	İran sürücülerinin trafik kurallarına ilişkin bilgi, tutum ve uygulamaları ile bunların trafik kazaları üzerindeki belirleyici etkileri arasındaki ilişkinin değerlendirilmesi.	İstatistiksel analiz
Şenel B. ve Şenel M.	2013	Türkiye'de gerçekleşen trafik kazaları üzerine hata ağacı analizi uygulaması	Hata ağaç analizi metodolojisi
Ona vd.	2013	Kırsal karayolları üzerindeki trafik kazalarının Latent Class Clustering ve Bayes Ağları kullanılarak analizi	Örtük Kümeleme Analizi ve Bayes Ağları
Ali ve Tayfour	2012	Sudan'da Trafik Kazaları kayıplarının İstatistik Modelleme ve Yapay Sinir Ağları Kullanılarak Karakteristiklerinin belirlenmesi ve tahminlenmesi	İstatistiksel modelleme ve yapay sinir ağları
Alp ve Engin	2011	Trafik kazalarının nedenleri ve sonuçları arasındaki ilişkinin çok kriterli karar verme yöntemleri kullanılarak analizi ve değerlendirilmesi	AHP ve TOPSIS
Akgüngör ve Doğan	2010	Farklı Yöntemler Kullanılarak Geliştirilen Trafik Kaza Tahmin Modelleri ve Analizi	Regresyon analizi, yapay sinir ağları ve genetik algoritma
Aköz ve Karslıgil	2010	Kısmi araç yörüngeleri kullanımı ile kavşaklardaki trafik kazalarının görüntü tabanlı çözümlemesi	Görüntü tabanlı trafik kaza analizi
Enache vd.	2009	Ölümcül trafik kazalarının ve risk faktörlerinin analizi	İstatistiksel Analiz
Temel ve Özcebe	2006	Türkiye'de karayollarında trafik kazaları	Türkiye'de trafik kazaları epidemiyolojisi
Eken vd.	2005	Motorlu taşıt kazaları ve risk faktörleri, Antalya ilinde meydana gelen 12185 trafik kazasının analizi.	İstatistiksel analiz

İkinci olarak veri madenciliği teknikleri (sınıflandırma, kümeleme vb.) ve diğer bazı teknikler (TOPSIS, AHP vb.) ile uygulama yapılmış olan çalışmalar Tablo 2 de özetlenmektedir.

Birliktelik kuralı analizi için yapılmış birçok çalışma mevcuttur. Çok fazla verinin bulunduğu müşteri temelli sistemler ile verinin çok fazla depolanıp kaydedildiği sistemlerde veri madenciliğinde birliktelik kuralı analizi çalışmaları yapılabilmektedir (Karaibrahimoğlu, 2014).

Trafik kazaları ile ilgili birçok tez çalışması yapılmıştır. Trafik kazaları için; karar ağaçları (Parıldar,2014), yapay sinir ağları (Yıldırım; Bolakar, 2014), zaman serileri analizi (Bagherinabel, 2014), coğrafi bilgi sistemleri (Ertunç, 2013) teknikleri kullanılarak çalışmalar yapılmıştır.

Sigortacılık sektöründe müşterilerin davranışları ile ilgili Apriori algoritması kullanılarak kural çıkarım çalışması Doğan vd.(2014) tarafından yapılmıştır.

Sakarya ilindeki yangın verileri kullanılarak uygulanan Apriori algoritması ile hangi mevsimlerde hangi yangın türlerinin gerçekleşebileceği konusunda araştırma yapılmıştır (Yurtay vd., 2014).

Atılgan(2011) çalışmasında kaza verileri için karar ağacı ve birliktelik kuralı yöntemini kullanarak trafik kazasında sürücüler ve yayaların kaza sonrasındaki durumuna etki eden faktörler belirlenmiştir.

Tablo 2: Veri Madenciliği ve Trafik Kazaları

Yazarlar	Yıl	Açıklama	Yöntem
Feng vd.	2016	ABD'de 2006-2010 yılları arasında meydana gelen ölümcül trafik kazaları mevsim, haftanın hangi günü, hangi saatlerde, hava durumu, otobüs tipi, yolun hız limiti, şoförün yaşı ve cinsiyeti gibi özellikler dikkate alınarak incelenmiştir.	Kümeleme analizi (K-ortalama)
Ye vd.	2016	Çin'deki otobüs duraklarında güvenliğin sağlanması için ihtiyaçların neler olduğu tespit edilmeye çalışılmıştır.	Temel bileşen analizi ve K-ortalama kümeleme metodu
Xi vd.	2016	Trafik Kazası Nedenlerinin AHP Apriori'ye dayalı bir yöntemle Analizi	AHP Apriori
Dadashova vd.	2016	Yol geometrisi ve trafik göstergeleri kullanılarak kentler arası yol kazalarının sıklığının ve şiddetinin belirlenmesi.	Veri Madenciliği
Xu vd.	2014	Kazaların analizi ve önlenmesi Hillsborough da meydana gelen kazalar için incelenmiştir.	RECDAB(acil bölgeleme metodu)
Lin vd.	2014	Buffalo Nigarai New York da meydana gelen kazalar dikkate alınmıştır.	Birliktelik Kuralı ve Karmaşık ağ algoritmaları
Cirovic ve Pamucar	2013	Sırbistan'da meydana gelen ölümlü 1220 kaza bulanık mantık ile incelenmiştir.	ANFIS
Selmic vd.	2012	Araç kullanıcıların, araç, yaşadıkları yer, çalışma saatleri ve ehliyet sınıfları gibi özellikleri dikkate alarak gruplandırmayı amaçlamaktadır.	K-ortalama algoritması
La Lama vd.	2011	İspanya'da 2008-2000 yılları arasında hayvanların dahil olduğu araç kazaları trafiğin yoğun olduğu saatler, kötü tasarlanmış yol planları gibi özellikler için incelenmiştir.	Kümeleme analizi
Polat ve Durduran	2011	Yeni bir veri önışleme yöntemi geliştirmişlerdir. Sıcaklık, nem, hava şartları gibi özellikler dikkate alınmıştır.	SCAW-Coğrafi bilgi sistemleri
Huang ve Abdel-Aty	2010	Çok aşamalı veri yapıları için model uygunluğu ve tahmin performansı test edilmiştir.	Bayes hiyerarşik modeli
Gundogdu	2010	Konya'da meydana gelen kazaların coğrafi bilgi sistemleri ile analiz edilmesi	Coğrafi bilgi sistemleri(Hot spot analizi)
Johansson vd.	2010	Stockholm'de çocukların korku, üstesinden gelme, ebeveyn/ çocuk düşünceleri gibi 18 adet özellik ve 50 kategori dikkate alınarak kümeleme analizi yapılmıştır.	Kümeleme analizi
Atalay ve Tortum	2010	Türkiye'deki illerin 1997-2006 yılları arası trafik kazalarına göre kümeleme analizi	Kümeleme Analizi
Skyving vd.	2009	2002-2004 yılları arasında İsveç'de meydana gelen 152 adet yaşlı sürücülerin sebep olduğu ölümcül kazalar incelenmiştir.	Kümeleme yöntemi
Guerts vd.	2005	Belçika'da meydana gelen kazaların niçin bazı tip yollarda daha çok gruplandığı sebebi araştırılmıştır.	Birliktelik kuralı
Golob ve Recker	2004	Güney California da meydana gelen 1000 den fazla gündüz/gece ve hava şartları gibi özellikler dikkate alınarak incelenmiştir.	Kümeleme metodu-Nonlineer kanonik korelasyon

Bu çalışma ile Ankara ili trafik kazaları analizi yapılmış, çıkarılan kurallar sayesinde karayollarında hangi durumlarda, kazaların insanları ne derece etkilediği (trafik kazalarından doğan yaralanma, ölümler ve maddi hasarlı kazalar) ve nasıl sonuçlandığı bulgusuna ulaşılmıştır. Farklı durumlarda ki kazalar dikkate alınarak karayollarında bazı iyileştirmeler yapılmasına ve artan kazaların önlenmesine katkıda bulunulmaya çalışılmıştır. Herhangi bir yaralanma veya ölümlerle sonuçlanmayan kazalar (maddi hasarlı kazalar) için farklı kurallar elde edilmiştir.

3. BİRLİKTELİK KURALI ANALİZİ

Veri madenciliği teknikleri günümüzde hastane, banka, kargo sistemleri gibi çok fazla bilginin depolandığı sistemlerde bazı uygulama çalışmalarında kullanılmıştır. Veri madenciliğinde birçok teknik kullanılmaktadır. Bunlardan en çok kullanılanlardan bir tanesi olan birliktelik kuralı analizi bu çalışmada kullanılmıştır. Birliktelik kuralı analizi için farklı algoritmalar kullanılmaktadır. Bu çalışma için Apriori algoritması seçilmiş ve sonuçların elde edilmesi için de SPSS Clementine 12.0 programı kullanılmıştır. Birliktelik kuralı analizinde apriori algoritması kullanılmıştır. Bu kuralda sık geçen veri kümesi keşfedilmektedir. İlk olarak bir elemanlı sıklık kümelerinden başlayarak tüm sıklık değerleri incelenir. Bu tarama yapılırken güven değerleri de kontrol edilmektedir. İstenen güven değeri %80-90 gibi kabul edilebilir. Yani çıkarılmış olan bir kuralın güven değeri %90 ise bunun anlamı %90 olasılıkla yeni verilerde bu kuralın örneklerini görmemiz mümkün demektir.

Veri madenciliğinde birçok farklı yöntem kullanılmaktadır. Bu yöntemlerden bazıları; sınıflandırma, regresyon analizi, kümeleme, birliktelik analizi, sıralı dizi analizi ve zaman serileri analizi şeklindedir.

Sınıflandırma yöntemi için sonuçları önceden bilinen verilere ihtiyaç vardır. Veriler eğitim ve test verisi olarak ikiye ayrılır. Eğitim verileri ile model oluşturulur ve test verileri ile de modelin ne oranda doğruluğa sahip olduğu hesaplanır. Örneğin, kredi başvurularının risk seviyeleri. Regresyon analizi, sürekli veriler için kullanılan bir tahmin metodudur. Bir ürüne gün, ay veya yıl bazında gelecek tahmini talebin belirlenmesi örnek olarak verilebilir. Kümeleme yöntemi, birbirine benzeyen verileri aynı grupta toplamayı amaçlar. Mümkün olduğunca benzer yapıya sahip verilerin ortak kümelerde yer alması amaçlanır. Ürünlerin gruplandırılması gibi. Birliktelik analizi ise büyük verilerden birbiri ile ilişkili

olanları belirler. Marketten ekmek alan kişilerin genellikle sütte alması örnek olarak verilebilir. Sıralı dizi analizi, verilerin zaman ve nerde gerçekleştiğine göre sıralanmasını gösterir. Zaman serileri analizinde ise, zaman içerisinde verilerin artış ve azalış gibi trendleri incelenir. Borsadaki hisse sentleri bu yöntemin en güzel örneklerinden biridir (Argüden ve Erşahin, 2008)..

Veri madenciliğinde kullanılan bazı algoritmalar ise karar ağaçları, regresyon analizi, bayes ağları, apriori algoritması, K-ortalama vb. şeklindedir. Karar ağaçları en yaygın yöntemlerden birisidir. Ağaç görünümünde bir teknik olup karar düğümlerinden ve yapraklardan oluşur. Değişken ya da değişkenler kullanarak bir olayı etkileyen etmenler ve ne oranda etkilediği dikkate alınır. Verilerin belirlenmiş olan sınıflara ne oranda ait olduğunu gösterir ve Bayes teoremine dayanır. Apriori algoritması sık tekrarlanan verileri bulmak için kullanılır. Kümeleme yöntemlerinde ise K-ortalama, hiyerarşik kümeleme algoritmaları kullanılmaktadır. İlk olarak verinin kaç kümeye ayrılacağı bilgisi girilerek K-ortalama algoritması uygulanmaktadır (Argüden ve Erşahin, 2008). Eğer küme sayısı bilinmiyor ise hiyerarşik kümeleme algoritması kullanılabilir.

Veri madenciliği tekniklerinden birliktelik kuralı analizi dinamik çözümler sunan bir yöntem olmamakla birlikte elde edilen kurallar sayesinde başka şehirlerde ya da ülkelerde meydana gelebilecek kazalarda ne tür durumlarda hangi seviyede bir kaza ile karşılaşılacağı konusunda sonuçlar sunar.

Apriori algoritması için öncelikle sık geçen öge kümelerini bulmak gerekir. Daha sonra sık geçen öğelerin birleştirilmesi işlemi gerçekleştirilir. Çıkarılan kuralların güvenilirlik seviyesi ayrıca dikkat edilmesi gereken bir durumdur. Düşük seviyedeki güvenilirlik çıkarılan kuralların doğruluğunu etkiler. Eğer %100 güven değeri için kural elde edilebilirse kuralın tüm veriler için geçerli olduğu sonucuna ulaşılır. Bu çalışmada ilk olarak; veri ön işleme yani kaza verilerinin düzenlenmesi ve kayıp verileri tamamlama ya da o verinin çıkarılması işlemi gerçekleştirilmiştir. İkinci aşama ise, aykırı değerlerin belirlenip çıkartılması ve veri tiplerinin 1-0 (binary) tipine dönüştürülmesidir. Üçüncü olarak; Apriori algoritmasının uygulanması ve SPSS Clementine12.0 programı kullanılarak kural çıkarımının gerçekleştirilmesidir.

Bu çalışmada Ankara ilinde 2010 yılında meydana gelen trafik kazaları için veri madenciliği tekniklerinden birliktelik kuralı analizi kullanılarak kural çıkarımı çalışması yapılmıştır. Bu kurallar

sayesinde farklı hava koşullarında (karlı, yağmurlu, sisli vb.), kazanın gerçekleştiği yollar (cadde, sokak, otoyol vb.) ve yol durumu (bölünmüş veya bölünmemiş yol gibi) gibi durumlarda hangi olasılıklarla kazaların yaralanma, ölüm veya maddi hasar ile sonuçlanacağı belirlenmiştir

Bu çalışmada birliktelik kuralı için aşamalar aşağıdaki gibidir:

- **Adım-1:** Veri ön işleme. Bu aşamada kaza verileri Apriori algoritmasına uygun hale getirilmiştir. Veri ön işleme yani kaza verilerinin düzenlenmesi ve kayıp verileri tamamlama ya da o verinin çıkarılması işlemi gerçekleştirilmiştir.
- **Adım-2:** Aykırı değerlerin belirlenip çıkartılması ve veri tiplerinin 1-0 (binary) tipine dönüştürülmesidir.
- **Adım-3:** Apriori algoritmasının uygulanması ve SPSS Clementine 12.0 programı kullanılarak kural çıkarımının gerçekleştirilmesidir.

- **Adım-4:** Elde edilen kuralların hangi güven seviyeleri için kabul edilebileceğine karar verilir. Çıkarılan kurallar yorumlanır.

Apriori algoritması "ağaç yapısında, k elemanlı bir aday listesinden k-1 elemana baktıktan sonra, alt frekans örüntüsü yetersiz olan elemanları budamakta ve kalan elemanların üzerinden arama yapmaya devam etmektedir" şeklinde tanımlanmaktadır (Sönmez, 2012).

Bu çalışmada kullanılan özellikler hava durumu, yol durumu, kazanın gerçekleştiği yollar vb. literatürde en fazla görülen özelliklerdir (Feng vd. 2016). Bu bağlamda, Ankara ili için kullanılan verilerde eksik bilgi olmayan ve en yaygın kullanılan 11 özellik dikkate alınmıştır.

Bu çalışmada kullanılan değişkenler Tablo 3'de verilmiştir. Buna göre; Yedi farklı hava durumu, 11 farklı kaza yeri, 11 farklı kaza türü, iki tip yol durumu, Yolların dört farklı yatay ve düşey geometrik durumları, Yolların geometrik kavşak durumu için yedi farklı değişken, Yolların geometrik geçiş durumu için beş farklı değişken verileri, kaza bilgisi verileri olarak kullanılmıştır.

Tablo 3: Kaza verilerinde kullanılan değişkenler

Durumlar	Alt durumlar											
Hava durumu	Açık	Bulutlu	Sisli	Yağmurlu	Karlı	Fırtınalı	Tipili					
Kaza yeri	Cadde	Sokak	Otoyol	Devlet yolu	II yolu	Köy yolu	Orman yolu	Servis yolu	Bağlantı yolu	Park alanı	Tesis önü-içi	
Kaza oluşum türü	Karşılıklı çarpışma	Arkadan çarpma	Yandan çarpma/ çarpışma	Duran araca çarpma	Sabit cisme çarpma	Yayaya çarpma	Hayvana çarpma	Devrilme	Yoldan çıkma	Araçtan düşen insan	Araçtan düşen cisim	
Yol geometrik kavşak	Üç yönlü(t)	Üç yönlü(y)	Dört yönlü	Beş veya daha fazla yönlü	Döner	Diğer kavşak	Kavşak yok					
Yol geometrik geçit	Kontrollü demiryolu geçidi	Kontrollsüz demiryolu geçidi	Okul geçidi	Yay geçidi	Geçit yok							
Yol yatay geometrik durumu	Düz yol	Hafif viraj	Korkuluklu sert viraj	Korkuluksuz sert viraj								
Yol düşey geometrik durumu	Eğimsiz	Hafif eğimli	Dik eğimli	Tepe üstü								
Yol bölünmüşlüğü	Bölünmüş	Bölünmemiş										

Ankara trafik kazaları için kullanılacak olan verilerin örnek bir kısmı Tablo 4'de verilmektedir. Örneğin; 1 numaralı kaza verisi "Açık hava durumunda, Caddede, Bölünmüş yolda yaralı ile sonuçlanan bir kazayı göstermektedir. Aynı şekilde 15 numaralı kazada "Yağmurlu havada, Devlet yolunda, Bölünmemiş yolda" hem "Yaralı" hem de "Ölümlü" ile sonuçlanan bir kaza meydana gelmiştir. Bu çalışmada farklı

değişkenler göz önünde bulundurularak insanların zarar gördüğü durumlar için kural çıkarım çalışması yapılmıştır. Bu çalışmada 2010 yılına ait 8897 adet kaza verisi kullanılmıştır.

Tablo 4: Ankara Trafik Kaza Verileri

	Kaza Numarası	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Hava Durumu	Açık	1	1	1	0	0	1	0	1	0	1	0	1	1	1	0
	Bulutlu	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0
	Sisli	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0

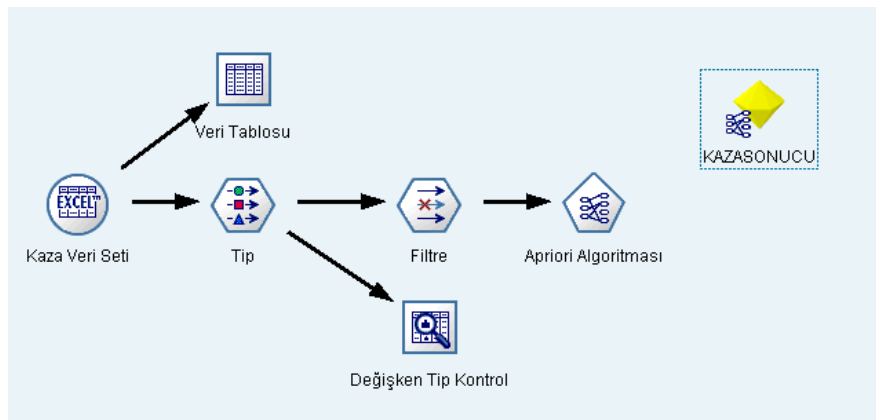
Kaza Yeri	Cadde	1	1	0	1	0	0	0	1	1	1	1	1	0	0	0
	Otoyol	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Yol Bölünmüşlüğü	Bölünmüş Yol	1	1	1	1	0	0	0	1	1	0	0	0	1	1	0
	Bölünmemiş Yol	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1
.
.
.
Kaza Sonucu	Ölü	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1
	Yaralı	1	0	0	0	1	0	1	0	0	1	1	0	0	0	1
	Sağlam	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0
	Park	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

4. TRAFİK KAZALARI İÇİN ELDE EDİLEN KURALLAR

SPSS Clementine 12.0 programı kullanılarak tasarlanan sistemin ekran görüntüsü Şekil 1'de verilmektedir. Burada trafik kaza verileri "Kaza veri seti" adı verilen modül kullanılarak Excel dosyasından okutulmuş ve "veri tablosu" adı verilen modül yardımı ile kontrolü sağlanmıştır. Verilerin tiplerinin uygun hale getirilmesi işlemi gerçekleştirildikten sonra bu değişkenlerin tiplerinin kontrolü yapılır. Bu çalışmada veriler 1-0(binary) olarak kullanılmaktadır. Okunan verilerde istenmeyen değişkenler söz konusu ise "filtre"

modülü ile bu değişkenler ilgili değişkenlerden çıkartılır. Tüm veriler için düzenleme yapıldıktan sonra "apriori algoritması" kullanılarak sonuçlar elde edilir. Bu çalışmada elde edilen kurallar "kaza sonucu" adı verilen sonuç modülünde yer almaktadır. Kullanılan değişkenlerden "kaza sonucu" hariç diğerleri girdi olarak kullanılmış ve kaza sonucu çıktısına etkileri incelenmiştir. Elde edilen kaza sonucu (sağlam, ölü, yaralı ve park durumları için) için hangi değişkenlerin birlikte etki sağladığı ve yüzde kaç oranda daha sonraki kaza durumlarında bu sonuçla karşılaşılacağı sonucuna ulaşılmıştır.

**Şekil 1:** SPSS Clementine 12.0 ekran görüntüsü

Tablo 5’de elde edilmiş olan bazı kurallar verilmektedir. Farklı kuralları görebilmek için en az % 60 güven düzeyinde, % 20 destek seviyesi dikkate alınmıştır. Bu çalışma için 139 adet kural çıkarılmıştır. Kurallar incelendiği zaman bu kuralların akla uygun geldiği görülmektedir. Örneğin, caddede eğimsiz bir bölünmüş yolda açık havada meydana gelen kazada herhangi bir ölü ya da yaralı olmama(yani sağlam olma) olasılığı % 61.462 olarak elde edilmiş olur. Atılğan(2011) yapmış olduğu tez çalışmasında da benzer sonuçları elde etmiştir. Çıkarılan bu kuralların daha sonra gerçekleşecek olan kazalar için gerçekçi birer sonuç olduğu Atılğan(2011)’in çalışmasında elde etmiş olduğu sonuçlar desteklemektedir.

Tablo 5’de verilen birinci kurala baktığımız zaman gerçekleşecek bir kaza için kaza oluş türünde “yandan çarpma durumu”, hava durumu türünde “açık” ve yol yatay geometrik durumu içinde “düz yol” değişkenleri için ölü ya da yaralı olmadan sonuçlanacak bir kaza olma olasılığı % 62.368’dir. Aynı şekilde elde edilen ikinci kurala bakarsak, yolun geometrik kavşak durumu için “kavşak yok”, yol bölünmüşlüğü türü için “bölünmüş yol”, kaza yeri için “cadde” ve yolun geometrik geçit durumu “geçit yok” için insanların sağlam olarak kazadan çıkma olasılığı % 62.241 elde edilmiş olur. Tablo 3’de elde edilmiş olan kurallardan sadece dokuz tanesi yer almaktadır.

Tablo 5: Elde edilen bazı kurallar

Kaza Sonucu	Öncül Değişkenler	Destek Değeri	Güven Değeri
Sağlam	KazaOlusTuru = Yandan carpma/carpisma ve HavaDurumu = Acik ve YolGeoYatay = Duz Yol	20,220	62,368
Sağlam	YolGeoKavsak = Kavsak Yok ve YolBolunmuslugu = Bolunmus ve KazaYeri = Cadde ve YolGeoGecit = Gecit Yok	26,582	62,241
Sağlam	YolBolunmuslugu = Bolunmus ve YolGeoDusey = Egimsiz ve KazaYeri = Cadde ve HavaDurumu = Acik ve YolGeoGecit = Gecit Yok	23,884	61,929
Sağlam	KazaOlusTuru = Yandan carpma/carpisma ve HavaDurumu = Acik	21,142	61,882
Sağlam	YolGeoKavsak = Kavsak Yok ve YolBolunmuslugu = Bolunmus ve KazaYeri =Cadde	28,987	61,768
Sağlam	YolBolunmuslugu = Bolunmus ve KazaYeri = Cadde ve HavaDurumu = Acik ve YolGeoGecit = Gecit Yok	33,056	61,570
Sağlam	YolBolunmuslugu = Bolunmus ve YolGeoDusey = Egimsiz ve HavaDurumu = Acik ve YolGeoGecit = Gecit Yok	33,000	61,000
Sağlam	YolBolunmuslugu = Bolunmemis ve YolGeoYatay = Duz Yol	30,000	61,000
Sağlam	KazaOlusTuru = Yayaya carpma ve KazaYeri = Cadde ve YolGeoYatay = Duz Yol	23,000	60,000

5. SONUÇ

Dünya’da ekonomik ve sosyal yönden hızlı bir şekilde gelişmekte olan ülkelerinde birisi olan Türkiye’nin başkenti Ankara iline ait 2010 yılı verileri kullanılarak birliktelik kuralları elde edilmiştir. Bu çalışmada veri madenciliği tekniklerinden olan birliktelik kuralı analizi için Apriori algoritması kullanılmıştır. Ankara ili için kazaların analizi yapılmış ve çıkarılan kurallar sayesinde karayollarında alınabilecek bazı önlemleri yer almaktadır. Bu çalışma hangi durumlarda hangi olasılıklarla kazaların

sonucunda insanların zarar görüp görmeyeceği sonucuna ulaşmamızı sağlamıştır.

Elde edilen tüm kurallara bakıldığı zaman caddelerde, açık havalarda, eğimsiz, kavşağın olmadığı ve bölünmüş yol gibi durumlarda kaza sonucu insanların sağlam çıktığı sonucuna ulaşılabilir. Fakat maddi hasarla sonuçlanan bu tip kazalar ülkemizde büyük maliyetlere sebep olmaktadır. Bu sebeple kazalarda önlemler öncelikle ölüm ya da yaralanma ile sonuçlanan değişkenler için sonra da sadece maddi hasara sebep olan (insanların zarar görmediği)

değişkenler için alınmalıdır. Bu çalışma ile gerçek hayatta farklı durumlarda ve özelliklerdeki yollarda ne tür önlemlerin alınması gerektiği tespit edilmiştir.

Sadece maddi hasarla sonuçlanan trafik kazalarında alınacak önlemler ile bile gereksiz maliyetlerden ortadan kaldırılabileceği öngörülmektedir.

KAYNAKLAR

1. Golob, T. F. and W. W. Recker (2004) "A Method For Relating Type Of Crash To Traffic Flow Characteristics On Urban Freeways" *Transportation Research Part a-Policy and Practice*, 38(1), 53-80.
2. Geurts, K., Thomas, I, Wets, G. (2005) "Understanding Spatial Concentrations Of Road Accidents Using Frequent İtem Sets." *Accident Analysis and Prevention*, 37(4), 787-799.
3. Eken, C., Görmez, H., Başhekim, M., Akyürek, S., Kartal, M. (2005) "Motorlu Taşıt Kazaları Ve Risk Faktörleri: Antalya İlinde Meydana Gelen 12185 Trafik Kazasının Analizi" *Türkiye Acil Tıp Dergisi*, 5(4), 175-180.
4. Temel, F., & Özcebe, H. (2006) "Türkiye'de Karayollarında Trafik Kazaları" *Sted*, 15(11), 192-200.
5. Argüden Y. Ve Erşahin B. (2008) "Veri Madenciligi: Veriden Bilgiye, Masraftan Değere", *ARGE danışmanlık yayınları*, No:10.
6. Enache, A., Chatzinikolaou, F., Enache, F., & Enache, B. (2009) "The Analysis Of Lethal Traffic Accidents And Risk Factors" *Legal Medicine*, 11, 327-330.
7. Skyving, M., Berg, H.Y., Laflamme, L. (2009) "A Pattern Analysis Of Traffic Crashes Fatal To Older Drivers" *Accident Analysis and Prevention*, 41(2), 253-258.
8. Huang, H. and M. Abdel-Aty (2010) "Multilevel Data And Bayesian Analysis In Traffic Safety" *Accident Analysis and Prevention*, 42(6), 1556-1565.
9. Johansson, K., Hasselberg, M., Laflamme, L. (2010) "Young Adolescents' İndependent Mobility, Related Factors And Association With Transport To School. A Cross-Sectional Study." *Bmc Public Health*, 10.
10. Gundogdu, I. B. (2010) "Applying Linear Analysis Methods To GIS-Supported Procedures For Preventing Traffic Accidents: Case Study Of Konya." *Safety Science*, 48(6), 763-769.
11. Akgüngör, A. P., & Doğan, E. (2010) "Farklı Yöntemler Kullanılarak Geliştirilen Trafik Kaza Tahmin Modelleri ve Analizi" *Int.J.Eng.Research & Development*, 2(1), 16-22.
12. Atalay, A., & Tortum, A. (2010) "Türkiye'deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi" *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 16(3), 335-343.
13. Aköz, Ö., & Karslıgil, M. E. (2010) "Video-Based Traffic Accident Analysis At İntersections Using Partial Vehicle Trajectories" *In 2010 IEEE 18th Signal Processing and Communications Applications Conference*, (pp. 499-502).
14. la Lama, G. C. M. D., Sepuveda, W.S. Villarroel, M. Marria, G.A. (2011) "Livestock Vehicle Accidents in Spain: Causes, Consequences, and Effects on Animal Welfare" *Journal of Applied Animal Welfare Science*, 14(2), 109-123.
15. Polat, K. and S. S. Durduran (2011) "Subtractive Clustering Attribute Weighting (SCAW) To Discriminate The Traffic Accidents On Konya-Afyonkarahisar Highway İn Turkey With The Help Of GIS: A Case Study" *Advances in Engineering Software*, 42(7): 491-500.
16. Atılgan, E. S. (2011) "Karayollarında Meydana Gelen Trafik Kazalarının Karar Ağaçları Ve Birliktelik Analizi İle İncelenmesi" Hacettepe Üniversitesi, İstatistik Anabilim Dalı, Yüksek Lisans Tezi.
17. Alp, S., & Engin, T. (2011) "Trafik Kazalarının Nedenleri Ve Sonuçları Arasındaki İlişkinin TOPSIS Ve AHP Yöntemleri Kullanılarak Analizi Ve Değerlendirilmesi" İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 19,65-87.
18. Ali, G., & Tayfour, A. (2012) "Characteristics and Prediction of Traffic Accident Casualties In Sudan Using Statistical Modeling and Artificial Neural Networks" *International Journal of Transportation Science and Technology*, 1(4), 305-318.
19. Selmic, M., MAcura, D. Teodorovic, D. (2012) "Ride Matching Using K-means Method: Case Study of Gazela Bridge in Belgrade, Serbia." *Journal of Transportation Engineering* 138(1), 132-140.
20. Ertunç, E. (2013). "Coğrafi bilgi sistemleri yardımıyla trafik kazalarının analizi: Antalya örneği", Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Harita Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi.

21. Şenel, B., & Şenel, M. (2013) "Risk Analizi: Türkiye'de Gerçekleşen Trafik Kazaları Üzerine Hata Ağacı Analizi Uygulaması" *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 13(3), 65-83.
22. Cirovic, G. and D. Pamucar (2013) "Decision support model for prioritizing railway level crossings for safety improvements: Application of the adaptive neuro-fuzzy system." *Expert Systems with Applications*, 40(6): 2208-2223.
23. De Oña, J., López, G., Mujalli, R., & Calvo, F. J. (2013) "Analysis Of Traffic Accidents On Rural Highways Using Latent Class Clustering And Bayesian Networks" *Accident Analysis & Prevention*, 51, 1-10.
24. Sungur, İ., Akdur, R., & Piyal, B. (2014) "Türkiye'deki Trafik Kazalarının Analizi" *Ankara Medical Journal*, 14(3), 114-124.
25. Lin, L., Wang, Q. And Sadek A.W. (2014) "Data Mining and Complex Network Algorithms for Traffic Accident Analysis." *Transportation Research Record*, 2460, 128-136.
26. Xu, P. P., Huang, H:L. Dong, N. Abdel-Aty, M. (2014) "Sensitivity Analysis In The Context Of Regional Safety Modeling: Identifying And Assessing The Modifiable Areal Unit Problem" *Accident Analysis and Prevention*, 70, 110-120.
27. Karaibrahimoğlu, A. (2014) "Veri Madenciliğinden Birliktelik Kuralı İle Onkoloji Verilerinin Analiz Edilmesi: Meram Tıp Fakültesi Onkoloji Örneği", Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi.
28. Parıldar, O. (2014) "Trafik Kazalarının Sınıflandırılmasında Karar Ağacı Kullanımı: Bodrum İlçesi Örneği", Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi.
29. Yıldırım, U. (2014) "Ankara'da Şehir İçinde Meydana Gelen Trafik Kazalarının Analizi, Kritik Noktaların Belirlenmesi Ve Bir Yapay Sinir Ağı İle Modellenmesi", Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Kazaların Çevresel ve Teknik Araştırması Anabilim Dalı, Doktora Tezi.
30. Yurtay, Y., Yurtay, N., Çelebi, N., Bacınoğlu, N. Z. Ve Ak, G. (2014) "Sakarya İline Ait Yangın Verilerinin Veri Madenciliği Yöntemleriyle Değerlendirilmesi", *ISITES2014 Karabük*, Türkiye.
31. Doğan, B., Erol, B., ve Buldu, A. (2014) "Sigortacılık Sektöründe Müşteri İlişkileri Yönetimi İçin Birliktelik Kuralı Kullanılması", *Marmara Fen Bilimleri Dergisi*, DOI: 10.7240/mufbed.56489, 3, 105-114.
32. Bagherinabel, E. (2014) "Trafik Kazalarının Zaman Serisi Analizi İle Değerlendirilmesi", Bülent Ecevit Üniversitesi, Sağlık Bilimleri Enstitüsü, Halk Sağlığı Anabilim Dalı.
33. Bolakar, H. (2014) "Yapay Sinir Ağları İle Trafik Kazalarının Modellemesi: Erzurum İli Örneği", Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, İnşaat Mühendisliği Anabilim Dalı, Ulaştırma Mühendisliği Bilim Dalı, Yüksek Lisans Tezi.
34. Mirzaei, R., Hafezi-Nejad, N., Sabagh, M. S., Moghaddam, A. A., Eslami, V., Rakhshani, F., & Rahimi-Movaghar, V. (2014) "Dominant Role Of Drivers' Attitude In Prevention Of Road Traffic Crashes: A Study On Knowledge, Attitude, And Practice Of Drivers In Iran" *Accident Analysis & Prevention*, 66, 36-42.
35. Dimitriou, L., & Vlahogianni, E. I. (2015) "Fuzzy Modeling Of Freeway Accident Duration With Rainfall And Traffic Flow Interactions" *Analytic Methods in Accident Research*, 5, 59-71.
36. Hong, K., Lee, K. M., & Jang, S. N. (2015) "Incidence And Related Factors Of Traffic Accidents Among The Older Population In A Rapidly Aging Society" *Archives of gerontology and geriatrics*, 60(3), 471-477.
37. Xi, J., Zhao, Z., Li, W., & Wang, Q. (2016) "A Traffic Accident Causation Analysis Method Based On AHP-Apriori" *Procedia engineering*, 137, 680-687.
38. Kar, S., Das, S. C., Tiwari, A., & Pharveen, I. (2015) "Pattern Of Road Traffic Accidents In Bhubaneswar, Odisha" *Clinical Epidemiology and Global Health*, 4(3), 115-119.
39. Dadashova, B., Ramirez, B. A., McWilliams, J. M., & Izquierdo, F. A. (2016) "The Identification Of Patterns Of Interurban Road Accident Frequency And Severity Using Road Geometry And Traffic Indicators" *Transportation Research Procedia*, 14, 4122-4129.
40. Waterson, P., Jenkins, D. P., Salmon, P. M., & Underwood, P. (2016) "Remixing Rasmussen: The Evolution Of Accimaps Within Systemic Accident Analysis" *Applied Ergonomics*, 59(B), 483-503.
41. Feng, S. M., Li, Z.N., Ci, Y.S., Zhang, G.h. (2016) "Risk Factors Affecting Fatal Bus Accident Severity: Their Impact On Different Types Of Bus Drivers" *Accident Analysis and Prevention*, 86, 29-39.
42. Ye, Z. R., Wang, C. Yu, Y.B. Shi, X.M. Wang W. (2016). "Modeling Level-Of-Safety For Bus Stops In China." *Traffic Injury Prevention*, 17(6), 656-661.