

## Veri Ön İşleme Sürecinin Türkiye Trafik Kaza Verilerine Uygulanması

<sup>1</sup>Halenur Kurtoğlu, <sup>2</sup>Özgenur Yıldız, <sup>3</sup>Serkan Kuruoğlu, <sup>4</sup>Adem Kepçe, <sup>\*5</sup>Zeliha Çağla Kuyumcu  
<sup>1,2,3,4</sup>Bilgisayar ve Bilişim Bilimleri Fakültesi, Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü, Türkiye  
<sup>\*5</sup>Mühendislik Fakültesi, Sakarya Üniversitesi İnşaat Mühendisliği Bölümü, Türkiye

### Özet

Türkiye’de meydana gelen ölümlü ve yaralanmalı trafik kazaları verilerinin T.C. Emniyet Genel Müdürlüğünden temin edilmesi çalışmanın başlangıç noktasını oluşturmaktadır. Bu çalışmada veri madenciliğine ilişkin genel bir bilgi verilerek Türkiye sınırları içinde bulunan devlet karayolu kesimine ait 2015-2021 yıllarını kapsayan, 3.532 km Otoyol, 30.965 km Devlet Yolu ve 34.029 km İl Yolu olmak üzere toplam 68.526 km yol ağında yedi yılda (2015-2021) gerçekleşen ölümlü ve yaralanmalı trafik kaza verileri, veri ön işleme teknikleri kullanıldıktan sonra verilere ait tanımlayıcı bilgiler elde edilmiştir. Bu incelenen yedi yılda, motorlu taşıt sayısında her yıl artış olmuş, ölümlü ve yaralanmalı kaza sayısı bu oranda artmamıştır. 2019-2020 yıllarında kaza sayısındaki düşüşün özel bir sebebi olup bu COVID-19 pandemidir. Bu yıllarda sokağa çıkma yasakları, evde çalışma gibi durumlardan dolayı trafikteki taşıt sayısının azalması kaza sayısının da azalmasını sağlamıştır.

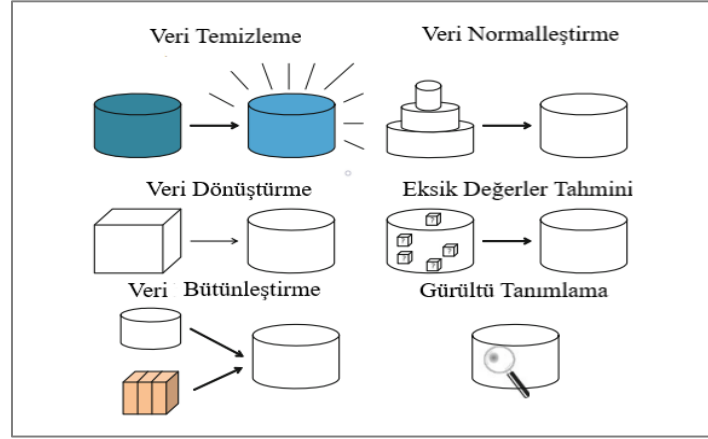
**Anahtar Kelimeler:** Veri ön işleme, veri madenciliği, trafik kazası, kaza

### 1. Giriş

Verilerin geçmişe dönük yapılacak incelemelerde kullanılabilmesi, geleceğe dönük tahminlerde bulunabilmesi amacıyla verileri saklama ihtiyacı duyulur. Bilgi teknolojisinin gelişimi ile verileri saklamak için birçok alanda veri tabanları kullanılmıştır. Veri tabanlarındaki bu verilerin hacmi hızla artmıştır ve bu veriler büyük veriyi oluşturmaktadır. Büyük veriler incelenmediği sürece pek bir anlam ifade etmez. Bazı önemli bilgileri içlerinde saklarlar. Büyük verilerden faydalı bilgiye ulaşma süreci veri madenciliği olarak isimlendirilir. Veri madenciliğinde veri kümesinin büyüklüğünden kaynaklanan en fazla zaman alıcı aşama, verilerin ön işleminden geçirilmesi aşamasıdır [1]. Veri madenciliği uygulamalarında kaynakların %80’ i verilerin ön işleminden geçirilmesi ve temizlenmesi süreçleri için harcanmaktadır [2]. Bu aşamada veri üzerinde bilgi keşfi yapılır.

### 2. Veri Ön İşleme Teknikleri

Veri ön işleme teknikleri, veri madenciliğinden önce uygulanarak elde edilen sonuçların kalitesi ve/veya veri madenciliği için harcanacak zaman artırılmış olur [3]. Şekil 1’de veri ön işleme teknikleri gösterilmiştir.



Şekil 1. Veri ön işleme teknikleri [3]

## 2.1 Veri Temizleme

Dağınık verilerdeki sistematik sorunları veya hataları düzeltmeyi içerir. Verilerin yanlış değerlere sahip olmasının, yanlış yazılmış, bozulmuş, çoğaltılmış gibi birçok nedeni vardır [3]. Herhangi bir değişkene ilişkin eksik değerlerin doldurulması için farklı yollar vardır. Bunlardan bazıları aşağıda kısaca açıklanmaktadır [4]:

- 1.Eksik değer içeren kayıt veya kayıtlar atılabilir.
- 2.Değişkenin ortalaması eksik değerlerin yerine kullanılabilir.
- 3.Aynı sınıfa ait tüm örneklem için değişkenin ortalaması kullanılabilir.

### 2.1.1 Eksik Değerler Tahmini

Çoğu durumda, makul bir tahminin eklenmesi uygun bir veri değeri boş bırakmaktan iyidir. Bu tür durumlarda veri temizlemenin bir adımı olarak dahil edilir [3]. Tahmin edilen değerler eksik değerlerin yerine koyulur.

### 2.1.2 Gürültü Tanımlama

Veri temizlemenin bir adımı olarak dahil edilen ve aynı zamanda veri dönüşümünde düzeltme olarak da bilinen bu adımın temel amacı, ölçülen bir aralıktaki rastgele hataları veya varyansları tespit etmektir [3].

## 2.2 Veri Bütünleştirme

Farklı sütunlardaki verilerin birlikte değerlendirilebilmesi amacıyla tek bir veri sütununa

dönüştürülmesi işlemidir [5].

### **2.3 Veri İndirgeme**

Veri madenciliği yapıldıktan sonra elde edilecek sonucu değiştirmeyecek olan değişkenler veri setinden çıkarılabilir. Veri indirgeme yöntemlerinden bazıları birleştirme, boyut indirgeme, veri sıkıştırmadır [5].

### **2.4 Veri Dönüştürme**

Veri dönüştürme, anlaşılması daha kolay modeller sağlamak için veri madenciliğinden önce veriler üzerinde gerçekleştirilmesi gereken bir veri ön işleme tekniğidir. Veri dönüştürme; düzeltme, birleştirme, genelleştirme ve normalleştirme gibi değişik işlemlerden biri veya bir kaçını içerebilir. Veri normalleştirme en sık kullanılan veri dönüştürme işlemlerinden birisidir. Veri normalleştirme tekniklerinden bazıları aşağıdaki biçimde sıralanabilir [4]:

1. Min-Max
2. Z Skor
3. Ondalık Ölçekleme

## **3. Veri Ön İşlemede Kullanılan Teknolojiler**

Veri ön işleme yaparken;

- Python
  - Pandas (Python kütüphanesi) ,
  - Pylot (Python kütüphanesi),
- Power BI
- MS Excel Formülleri

kullanılmıştır. Bu teknolojileri nasıl kullandığımız veri seti üzerinde yapılan işlemler bölümünde detaylandırılacaktır.

## **4. Veri Seti Üzerinde Yapılan İşlemler**

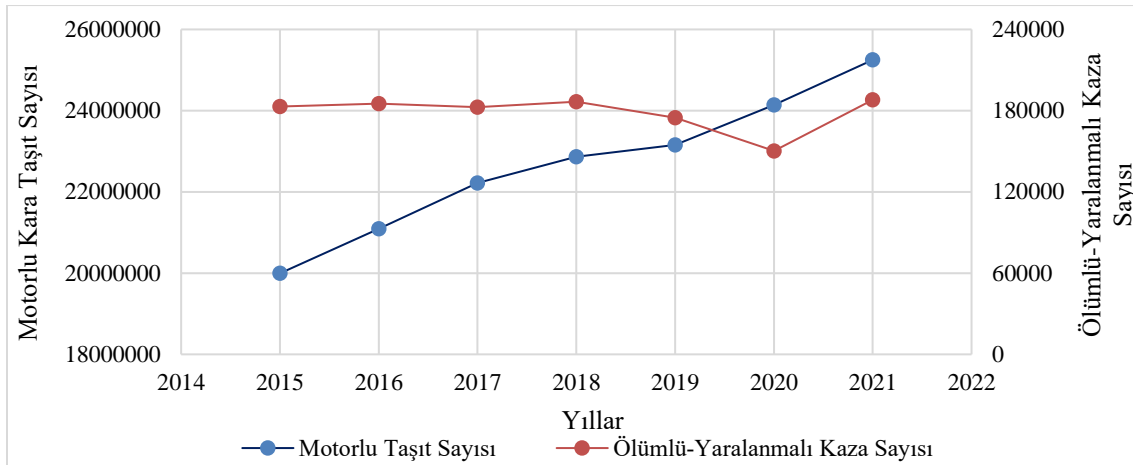
Emniyet birimlerinden üç farklı MS Excel dosyası alınmıştır. Bu dosyalarda bulunan veriler, yıl ayrımı yapılarak ‘Kaza ID’, ‘Taşıt ID’ ve ‘Kaza Yılı’ koşullarına göre tek bir MS Excel dosyasında birleştirilmiştir. MS Excel dosyalarını birleştirmek için Python’ ın ‘Pandas’ kütüphanesi kullanılmıştır.

Öncelikle yazılan Python kodları ile MS Excel dosyaları okutulmuştur. Sonrasında ‘full outer join’ yöntemi ile bu dosyalar birleştirilerek veriler yeni bir dosyaya yazdırılmıştır. İki tablodaki eşleşen kayıtlar ve eşleşmeyen sol ve sağ kayıtlar için ‘full outer join’ kullanılmıştır.

- Hava durumu yağışlı, karlı olan satırlarda yol yüzeyi kuru ise MS Excel’in filtreleme özelliği kullanılarak kuru yazan satırlar ıslak olarak değiştirilmiştir.
- Kaza tarihi verilerinden günler, aylar, mevsimler elde edilmiştir. ‘İç içe if’, ‘metne çevir’, ‘parça al’, ‘soldan’, ‘sağdan’ MS Excel formülleri kullanılmıştır.
- Kaza tarihi verilerinden yıllara göre ramazan bayramları, kurban bayramları ve resmî tatiller elde edilmiştir. Bu şekilde tatil günlerin kaza sayısına etkisi gözlemlenebilir. ‘İç içe if’ MS Excel formülü kullanılmıştır.
- Oluşturulan dosyada yol bölümü adında bir kolon açılmıştır. Bu kolona kazanın kavşakta ya kavşak dışında olması hakkında bilgi veren veriler getirilmiştir.
- Işık durumu adında değişken açılmıştır. Dosyada bulunan gün ışığı durumu ve aydınlatma kolonlarından veri çekilerek gündüz ise gündüz gece ve alacakaranlık ise sokak aydınlatmasının var ya da yok olma durumu yazılmıştır. Bu şekilde kaza anında aydınlatmaya ihtiyaç olup olmadığı bilgisi görülmüştür.
- Kazanın ışıklı-ışıksız kavşakların hangisinde olduğunu görebilmek adına ışık durumu adında kolon açılmıştır.
- Plaka değişkeni şehir ismi olarak düzenlenmiş bunun için, ‘Çok eğer’ formülü kullanılmıştır.
- Yaş grubu adında kolon açılmıştır. Sürücü yaşı kolonundan yaş bilgisi çekilerek 25 ve altındaysa genç, 26-40 aralığındaysa orta yaş 1, 41-64 aralığındaysa orta yaş 2, 65 yaş ve üstüyse yaşlı bilgileri yaş grubu kolonuna yazılmıştır.
- Kaza-sonucu kolonu açılmış ve sonuç-ölümlü ve sonuç-yaralanmalı kolonlarından veri çekilmiştir. Kaza-sonucu kolonuna ölümlü ya da yaralanmalı yazılmıştır.
- Verilerin analiz programlarında kullanılabilmesi için Türkçe karakterler İngilizce karakterler ile değiştirilmiştir. Bunun için ‘Pandas’ kütüphanesinden yararlanılmıştır.

#### **4. Türkiye’de 2015-2021 yıllarında gerçekleşen Ölümlü-Yaralanmalı Trafik Kazaları**

Trafik kazaları nedeniyle dünyada her yıl ortalama 1 milyon 350 bin kişi hayatını kaybetmekte, 78 milyon insan yaralanmakta ve bunun yanı sıra önemli ölçüde sosyo-ekonomik kayıplar meydana gelmektedir [6]. Türkiye’de ise COVID-19 pandemisi öncesi 4 yılda (2015-2018) ortalama 7233 kişi hayatını kaybetmiş, 303172 kişi yaralanmıştı. COVID-19 pandemisi ile birlikte sokağa çıkma yasakları nedeni ile trafikte taşıt sayısı ve dolayısıyla kaza sayısı da azalmıştır. 2019-2021 yıllarında 3 yıl için yıllık ortalama 5233 kişi hayatını kaybetmiş, 261372 kişi yaralanmıştır. Bu yıllar için trafiğe kayıtlı motorlu taşıt sayısı ile ölümlü-yaralanmalı trafik kaza sayısının değişimi Şekil 2’de verilmiştir.

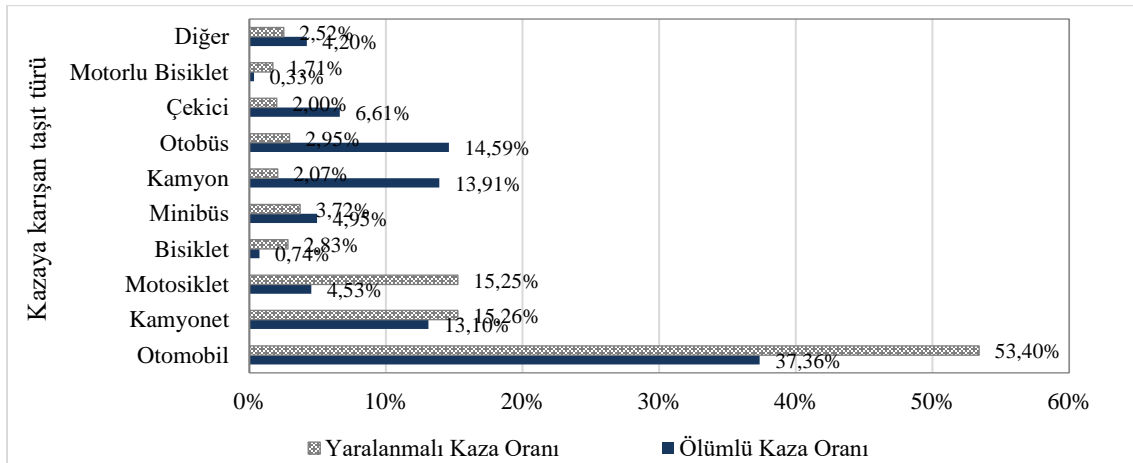


Şekil 2. Yıllara göre motorlu kara taşıtı sayısı ile ölümlü-yaralanmalı kaza sayısı değişimi

2015 yılında motorlu kara taşıt sayısı 19 milyon 994 bin 472 iken her yıl artarak 2021 yılının aralık ayı sonunda 25 milyon 249 bin 119'a ulaşmıştır [7]. Taşıt sayısının artışına rağmen kaza sayısı, pandemi etkisinin dışında yaklaşık olarak aynı seviyelerde kalmıştır. Bu durum trafikte önlemlerin artırıldığı çıkarımını sağlayabilir.

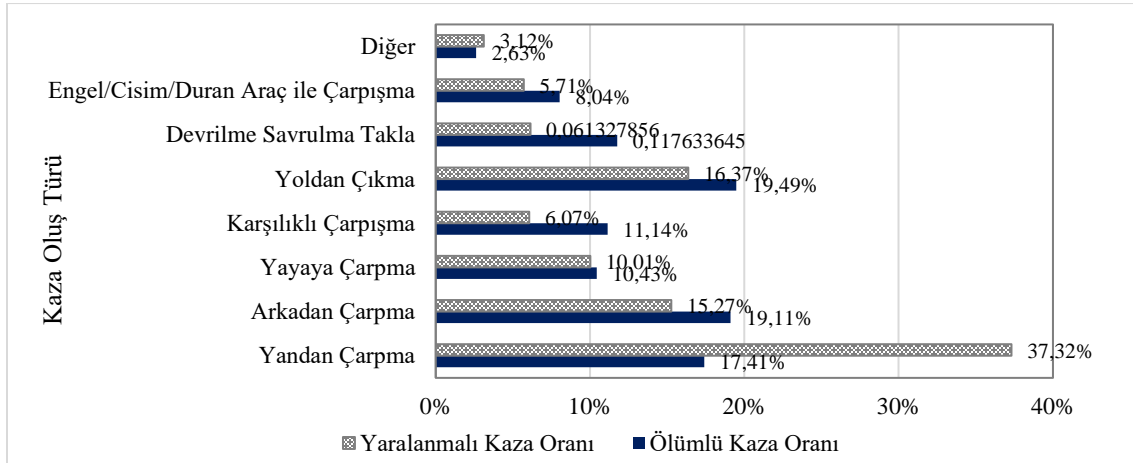
2015-2021 yılları, Türkiye ölümlü-yaralanmalı trafik kaza verileri önışleme sürecinden geçirildikten sonra veriye ait tanımlayıcı grafikler aşağıdaki şekillerde yer almaktadır.

Şekil 3'te bu tip kazalara karışan taşıtların dağılımı görülmektedir. Kazaların çoğunluğuna otomobiller karımış olup, diğer taşıt diye ifade edilen sınıfın içinde; traktör, özel amaçlı kullanılan taşıtlar, arazi taşıtları, ambulans, iş makinesi, tanker, at arabası, tren, tramvay bulunmaktadır. Bu taşıt tiplerinin kaza sayısı diğer tiplere göre oldukça az olduğu için tek bir sınıf altında birleştirilmesi uygun görülmüştür.



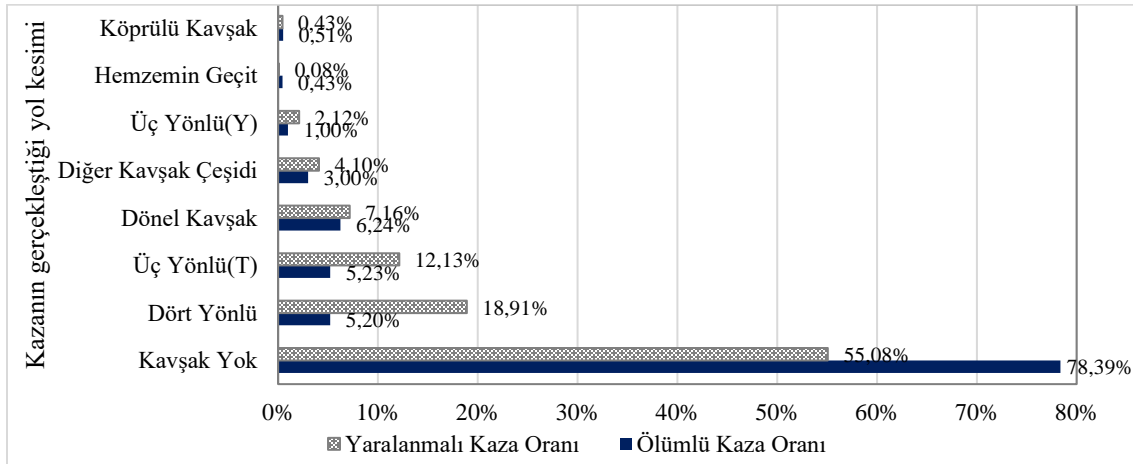
Şekil 3. Ölümlü ve yaralanmalı kazalarda kazaya karışan taşıtların dağılımı

Şekil 4'te kaza oluş türleri dağılımı yer almaktadır. Buna göre, yaralanmalı kazalarda en fazla yandan çarpma, ölümlü kazalarda ise yoldan çıkma türünde kazalar olmaktadır.



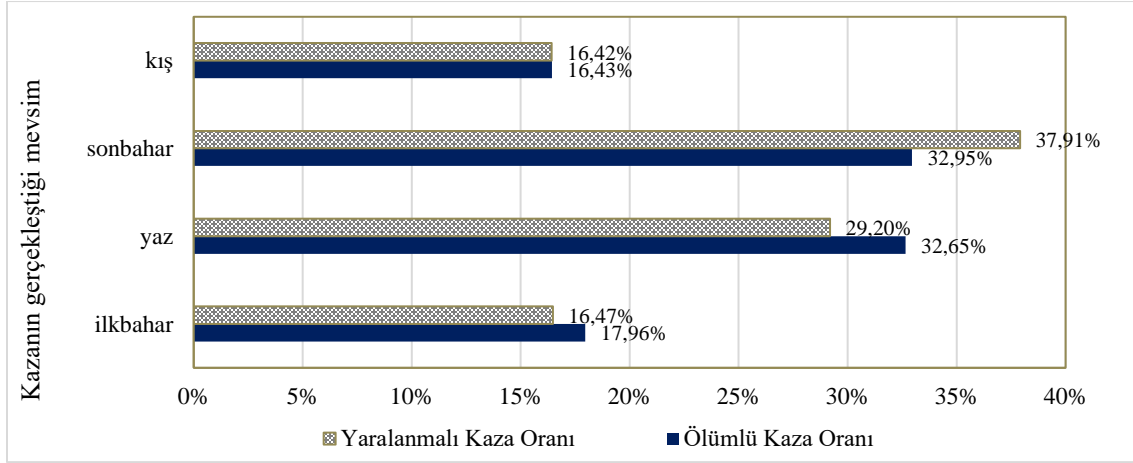
Şekil 4. Ölümlü ve yaralanmalı kazalarda kaza oluş türü dağılımı

Şekil 5'te kazaların gerçekleştiği yol bölümü incelenmiş, ölümlü kazaların %75'inden, yaralanmalı kazaların yarısından fazlasının kavşak olmayan yol kesimlerinde gerçekleştiği görülmüştür.



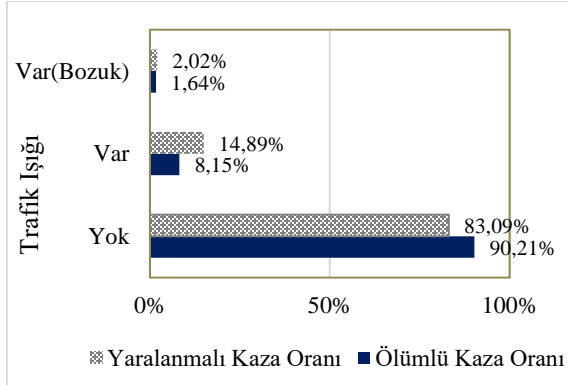
Şekil 5. Ölümlü ve yaralanmalı kazalarda kazanın gerçekleştiği yol kesimi dağılımı

Şekil 6'da kazalara etki eden çevresel koşullardan olan mevsimlere göre kaza dağılımı görülmektedir. Ölümlü ve yaralanmalı kazaların sonbahar ve yaz mevsimlerinde daha çok gerçekleştiği belirlenmiştir.

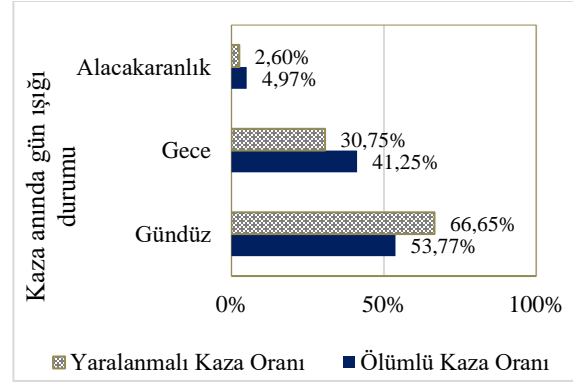


Şekil 6. Ölümlü ve yaralanmalı kazalarda kazanın gerçekleştiği mevsim dağılımı

Şekil 7’de kazaların gerçekleştiği yerlerde ışık durumu incelenmiş buna göre ölümlü ve yaralanmalı kazaların %90 ve %83’ünün trafik ışığı olmayan yerlerde gerçekleştiği görülmektedir. Şekil 8’de ise kazaların gerçekleştiği anda havanın aydınlık durumuna göre kazaların dağılımı yer almaktadır. Buna göre ölümlü ve yaralanmalı kazaların yarısından fazlası gündüz gerçekleşmiştir.

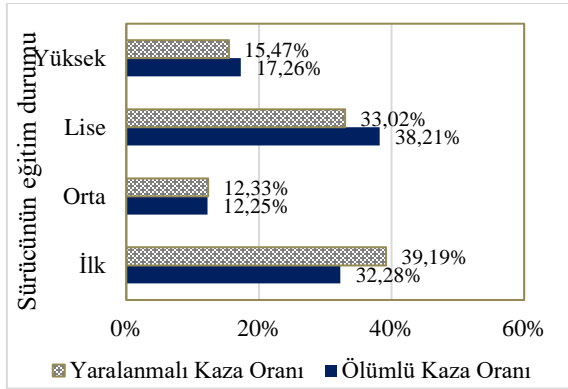


Şekil 7. Ölümlü ve yaralanmalı kazalarda trafik ışığı durumu

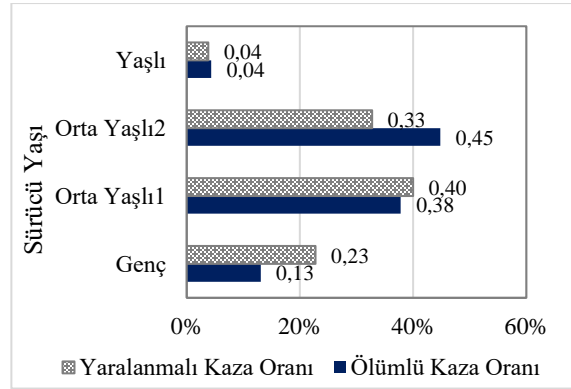


Şekil 8. Ölümlü ve yaralanmalı kazalarda gün ışığı durumu

Şekil 9’da ölümlü ve yaralanmalı kazalardaki sürücülerin eğitim durumuna göre dağılımları verilmiştir. Ölümlü ve yaralanmalı kazalarda en fazla ilkokul ve lise mezunu sürücülerin yer aldığı görülmektedir. Sürücünün eğitim durumu arttıkça kazalara karışma durumunun azaldığı söylenebilir. Şekil 10’da ise sürücü yaşına göre kazaya karışma oranları incelenmiştir.



Şekil 9. Kazalara karışan sürücülerin eğitim durumu



Şekil 10. Kazalara karışan sürücülerin yaş grubu

Genç sürücülerin (25 ve altı) trafikte sayı olarak az olmasına karşın ölümlü ve yaralanmalı kazalara karışma oranlarının oldukça yüksek olduğu söylenebilir. Orta yaşlı 1 (26-40) sürücülerin yaralanmalı kazalara karışma oranı diğer sürücü gruplarına göre %40 oranla en fazla olanıdır. Orta yaşlı 2 (41-65) sürücülerin ise ölümlü kazalara karışma oranı %45'le en fazla olanıdır. Yaşlı sürücülerin (65 üstü) ölümlü ve yaralanmalı kazalara karışma oranları oldukça düşük olup, %4'tür.

## 5. Sonuç ve Değerlendirmeler

Yapılan çalışmada, Türkiye ölümlü yaralanmalı trafik kaza verileri, önışleme yapılarak düzenlenmiş, hangi etmenlerin can kayıplarına, yaralanmalara neden olduğu incelenmiştir. Trafik kazalarında Türkiye gibi tüm dünyada da azaltmaya yönelik çalışmalar büyük önem arz etmektedir. Çalışmada öne çıkan durumlar;

- Otomobil sürücülerinin trafikte çoğunluğu oluşturduğu bilgisi de göz önünde tutularak, kazalar konusunda önde geldiği,
- Yandan ve arkadan çarpma ile yoldan çıkma türünde oluşan kazaların fazla olduğu,
- Bu kazaların çoğunluğunun kavşak olmayan yol kesimlerinde gerçekleştiği,
- Kazaların sonbahar ve yaz mevsimlerinde daha çok meydana geldiği,
- Çoğu kazanın gerçekleştiği yerde trafik ışığının olmadığı,
- Kazaların çoğunluğunun gündüz olduğu,
- Sürücü eğitim seviyesi arttıkça kazalara karışma durumunun azaldığı,
- Orta yaş gruplarında bulunan sürücülerin daha fazla kazaya karıştığıdır.

İlerleyen çalışmalarda, kazalara neden olan faktörler arasındaki ilişkiler incelenerek kaza oluşumuna birlikte neden olan durumlar tespit edilecektir.

## Kaynaklar

[1] Oğuzlar, Y. D. D. A. (2003). Veri Ön İşleme. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 0 (21), Retrieved from <https://dergipark.org.tr/tr/pub/erciyesiibd/issue/5878/77794>



- [2] Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European Journal of Operational Research*, 156(2), 483–494. [https://doi.org/10.1016/S0377-2217\(02\)00911-6](https://doi.org/10.1016/S0377-2217(02)00911-6)
- [3] García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Vol. 72). Springer International Publishing. <https://doi.org/10.1007/978-3-319-10247-4>
- [4] Roiger, R. J., & Geatz, M. W. (2003). *Data Mining: A Tutorial-based Primer*, Pearson Education. Inc: USA.
- [5] Eda Coşlu. (2013, 25 Ocak). Veri Madenciliği. Akademik Bilişim 2013, Antalya.
- [6] WHO. (2022, June 20). WHO. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>
- [7] TÜİK. (2021). Retrieved 30 June 2022, from <https://data.tuik.gov.tr/Bulten/Index?p=Motorlu-Kara-Tasitlari-Aralik-2021-45703>