



III. Ulaştırma ve Lojistik Ulusal Kongresi

28-29 Kasım, 2019, Gaziantep

<http://www.ulk.ist/>



Trafik Kazalarının İstatistiksel Sınıflandırma Yöntemleri Kullanılarak Değerlendirilmesi

Özet

Arzu Altın Yavuz

Doç. Dr.

Eskişehir Osmangazi
Üniversitesi

aaltin@ogu.edu.tr

Çağdaş Kara

Arş. Gör. Dr.

Eskişehir Osmangazi
Üniversitesi

ckara@ogu.edu.tr

Türkiye’de meydana gelen trafik kazaları, sebep oldukları maddi/manevi kayıplar sebebiyle gündemin ilk sırasında olma durumunu korumaktadır. Trafik kazaları, birçok etkenin bir araya gelmesi sonucu oluşmaktadır. Bunlar, insan, yol, araç, iklim, çevre koşulları olarak sıralanabilmektedir. Trafik kazaları sonucu, telafi edilebilen kazalar olabileceği gibi, telafisinin imkansız olduğu kazalar da olabilmektedir. Trafik kazalarının sayısını ve etkilerini en aza indirebilmek için ülkeler, bu konularda çeşitli stratejiler ve uygulamalar geliştirmektedirler. Genel olarak trafik kazalarının sayısını ve maddi/manevi kayıpları azaltabilmek için kazaya sebep olan etkilerin tespit edilip ortadan kaldırılması gerekmektedir. Trafik kazaları için literatür çalışmaları incelendiğinde, genel olarak diskriminant analizi, lojistik regresyon analizi ve logaritmik doğrusal modellerin kullanıldığı görülmektedir. Bu çalışmada, son dönemde makine öğrenmesi algoritmaları olarak da kullanılmaya başlanılan istatistiksel sınıflandırma ve karar ağaçları teknikleri kullanılarak, trafik kazalarının analizleri gerçekleştirilmiştir. 2012 ile 2016 yılları arasında Antalya ili ve ilçelerinde ölümlü, yaralanmalı ve maddi hasarlı olmak üzere toplam 30232 adet trafik kazası olmuştur. Meydana gelen trafik kazalarının sonucunda, ölümlü ve yaralanmalı olarak gerçekleşen 3181 trafik kazası veri seti olarak kullanılmıştır. Kaza sonucu üzerinde etkili olduğu düşünülerek; kaza yeri, yol tipi, yolun kaplama cinsi, yolun sınıfı, hava durumu, yol yüzeyi, trafik lambasının durumu, aydınlatma, trafik görevlisi durumu, emniyet şeridi durumu ve kaza türü değişkenleri açıklayıcı değişken olarak ele alınmıştır. Sınıflandırma problemi, istatistiksel bir karar verme sürecidir. Bu düzeyde araştırmacı için iki tür karar verme süreci vardır. Grubun ayırt edici özelliklerini inceleyerek ayrımcılığı sağlayan değişkenleri belirlemek ve bu farklı işlevlerin yardımıyla bireyleri doğru gruplara atamaktır. Sınıflandırma problemlerinde varsayımların azlığı ve kullanım kolaylığı nedeniyle Lojistik



regresyon yaygın olarak kullanılmaktadır. Lojistik regresyon analizinin temel amacı, bağımsız değişkenlerle bağımlı değişken arasındaki nedensellik ilişkisini, diğer regresyon yöntemlerinde olduğu gibi en az değişken yardımıyla açıklamaya çalışmaktır. Bu yüzden, trafik kazaları çalışmalarından sıklıkla kullanılmaktadır. Son yıllarda, trafik kazalarının analizinde, karar ağaçları, regresyon ağaçları gibi ağaçlandırma analizleri sıklıkla kullanılır hale gelmiştir. Bu çalışmada, bu analizlerden j48, karar ağaçları (decision tree), regresyon ağaçları ve sınıflandırma (simple cart), rassal ağaçlar (random tree ve random forest) teknikleri kullanılacaktır. j48 tekniği, uyarlanmış bir makine öğrenmesi algoritmasıdır. Bu algoritmanın amacı, en yüksek bilgiyi sağlayan ve en az sayıda dal ile bir karar ağacı oluşturmaktır. Karar ağaçları, çok sayıda gözlem içeren veri setini bir dizi kural kullanarak daha küçük kümelere ve dallara ayırır. Regresyon ağaçları ve sınıflandırma, veri madenciliği ve tahmine dayalı, karmaşık veri setlerinin analiz edilmesine olanak veren parametrik olmayan bir tekniktir. Rassal ağaçlar, karar ağaçları tekniğini kullanarak, tüm alt kümelerdeki dalları araması sebebi ile daha iyi modelleme performansı sunabilmektedir. Tüm bu teknikler, büyük veri ve makine öğrenmesi teknikleri olduğundan, az sayıda gözlem ve çok sayıda bağımsız değişken durumunda bile etkin sonuçlar vermektedir. Ayrıca bu durum lojistik regresyon ve diskriminant analizi gibi gözlem sayısının değişken sayısından fazla olmasını gerektiren klasik tekniklere bir üstünlük yaratmaktadır. Öne sürülen tekniklerin karşılaştırılmasında; doğru sınıflandırma oranı (DSO), ortalama mutlak hata (OMH) ve ortalama hata karesi kökü (OHKK) kullanılmıştır. 2012 ile 2016 yılları arasında Antalya ilinde gerçekleşen ölümlü ve yaralanmalı trafik kazalarının analizi sonucunda, tüm karşılaştırma kriterleri göz önünde bulundurulduğunda en iyi performansı veren tekniğin Random Forest tekniği olduğu belirlenmiştir.

Anahtar Kelimeler:

Kaza Analizi, Makine Öğrenme Yöntemleri, Lojistik Regresyon

