



Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması

Can Aydın^{1*}

¹ Dokuz Eylül Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İzmir, Türkiye

(İlk Geliş Tarihi 10 Eylül 2018 ve Kabul Tarihi 16 Kasım 2018)

(DOI: 10.31590/ejosat.458613)

Öz

Kalabalık şehirlerde kent içerisinde itfaiye istasyonlarının doğru yer seçimi, yangınlara hızlı müdahale etmek, can ve mal kaybını en aza indirmek açısından çok hayati bir konudur. İtfaiye istasyonu yer seçiminde; kent bütününe belirli bölgelere ayırarak belirlenen her bir bölge için itfaiye istasyonu ihtiyacının sorgulanması gerekmektedir. Bu çalışmada da mevcut itfaiye istasyonlarından yola çıkarak makine öğrenmesi algoritmaları kullanarak bölgelere göre itfaiye istasyonu ihtiyacının sınıflandırılması gerçekleştirilmiştir. Çalışma kapsamında her bir bölgeye ait, itfaiye araçlarının o bölgeye ulaşım süreleri, bölgenin nüfus yoğunluğu, bölgeye giden ortalama ana ve yardımcı araç sayısı verileri ile bölgedeki itfaiye istasyonu bulunma durumu verileri kullanılarak istasyon ihtiyacının tahmini için sınıflandırılma çalışması gerçekleştirilmiştir. Bu çalışmadaki amaç İzmir Büyükşehir Belediyesinin belirlediği 808 bölgeye dair itfaiye istasyonu ihtiyacı sınıflandırılmasında en başarılı sınıflandırma algoritmasının tespit edilmesidir. 2015-2017 tarihleri arasındaki yangın kayıtları analiz edilerek bölgelerin sınıflandırılmasında %93.84 ile en başarılı algoritmanın Random Forest algoritması olduğu tespit edilmiştir. En başarılı algoritma tespit edilirken doğruluk, ortalama mutlak hata (MAE), kök hata kareler ortalaması (RMSE) ve Kappa değerleri göz önüne alınmıştır.

Anahtar kelimeler: Makine Öğrenmesi, Yer Seçim, Coğrafi Bilgi Sistemleri, Yönetim Bilişim Sistemleri.

Classification of fire station requirement using machine learning algorithms

Abstract

In crowded cities, the right location of the fire stations in the city is a very vital issue to intervene quickly in fires and to minimize loss of life and property. In selecting the location of the fire station, it is necessary to question the need for a fire station for each region determined by dividing the whole city into specific zones. In this study, firefighting station needs to be classified according to the regions by using machine learning algorithms. Within the scope of the study, a classification study was carried out for the estimation of the station necessity by using data which transportation times of the fire brigades, the population density of the region, the average number of main and assistant vehicles and the presence of the fire station in the region. The purpose of this study is to determine the most successful classification algorithm in the classification of fire station requirements for the 808 regions determined by İzmir Metropolitan Municipality. By analyzing the fire records between 2015-2017, it was determined that the most successful algorithm is Random Forest algorithm with 93.84% classification of the regions. In determining the most successful algorithm, accuracy, mean absolute error (MAE), root mean square error (RMSE) and Kappa values are taken into consideration.

Key words: Machine Learning, Location Choice, Geographical Information System, Management Information System.

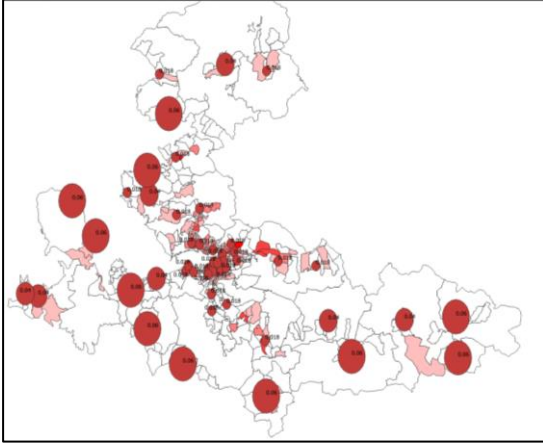
1. Giriş

Son yıllarda kentsel nüfus; yüksek doğum oranları ve artan dış göçler sebebiyle hızla artmaktadır. Artan nüfus, kentlerde yeterli altyapı olmadığından ötürü belediye hizmetlerinin istenilen seviyede verilmemesine sebep olmaktadır. İtfaiye hizmeti de belediye hizmetleri arasında önemli bir yer

tutmaktadır. Nüfus yoğunluğu fazla olan ve plansız yapılaşmış kentlerde itfaiye hizmetlerinde yaşanan aksaklıklar, kaynak yönetimi eksikliği sonucu itfaiye araçlarının olaya zamanında yetişememesiyle sonuçlanmaktadır. Kaynak yönetimi itfaiye araçlarının olay yerine en hızlı şekilde ulaşması için istasyonların uygun şekilde dağılımı yapılmasını kapsamaktadır. Bu dağılım yapılırken itfaiye hizmet alanı sınırı

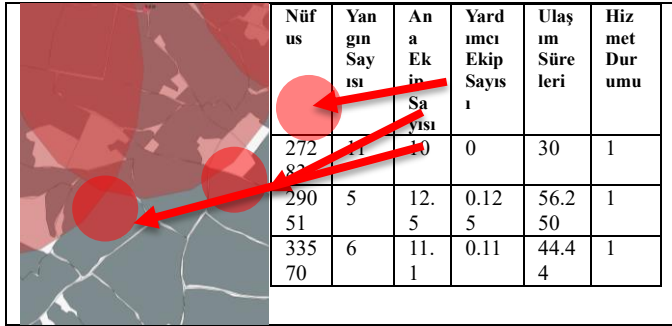
¹ Sorumlu Yazar: Dokuz Eylül Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, İzmir, Türkiye, canaydinn@gmail.com

6030'nolu büyükşehir belediyesi yasası kapsamında il sınırı olarak belirlenen büyükşehir belediyesi hizmet sınırı kabul edilmiştir. Kabul edilen bu hizmet sınırları toplamda 808 adet bölgeye ayrılmıştır. Bu bölgelerin itfaiye istasyonları hizmet alıp almadığını incelemek için ilk olarak mevcut durum tespiti gerçekleştirilmiştir. Mevcut durum incelendiğinde İzmir büyükşehir belediyesi itfaiye daire başkanlığına bağlı bulunan mevcut 50 adet itfaiye istasyonu bulunmaktadır. Dünya standartlarına göre itfaiye istasyonlarının yakın çevrelerinde çıkan olaylara 5 dakika içerisinde müdahale etmesi gerekliliği bulunmaktadır. Bu bağlayıcılık kapsamında İzmir ili içerisinde itfaiye istasyonlarının mevcut trafik yoğunluğu göz önüne alındığında yapılan tampon bölge (buffer) analizi doğrultusunda istasyonların 808 bölgenin 279 tanesine yeterli hizmet veremediği görülmektedir (Şekil 1).



Şekil 1. İtfaiye istasyonları hizmet kapsamları

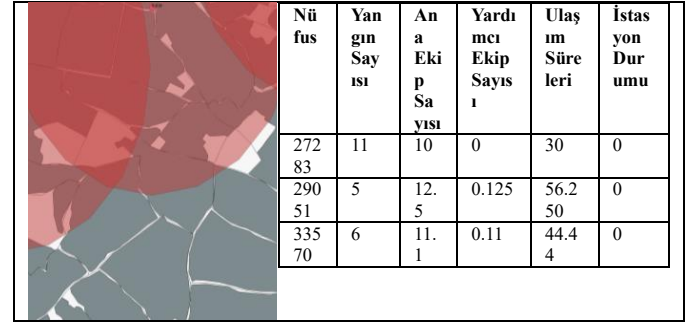
Bu kapsamda itfaiye istasyonlarına atılan tampon bölgeye degen tüm bölgelerin hizmet aldığı kabul edilerek hizmet alma durumuna 1 değeri atılmış hizmet almayan diğer tüm bölgelere 0 değeri atılmıştır (Şekil 2).



Şekil 2. Bölgelere atılan hizmet durumu kapsama problemi

Çalışmanın problem tanımı gerçekleştirilirken yapılan bu analiz çalışmasında bazı hataların yapıldığı tespit edilmiştir. Hatanın detayları incelendiğinde itfaiye istasyonlarına atılan tampon bölgeye sadece degen bölgelere doğrudan 1 değeri atıldığında kapsama alanı hatası gerçekleştirilmiştir (Şekil). Bunun sebebi bazı bölgelerde yerleşim lekeleri bölgenin tampon bölgeye değmeyen tarafında bulunabilmektedir. Bu sebeple başka bir analiz çalışması daha gerçekleştirilerek bölgelerin yerleşim lekeleri çıkarılmıştır. Bu sefer itfaiye istasyonlarına atılan tampon bölge analizindeki koşulumuz tampon bölgenin sadece bölgeye değmesi değil bununla beraber yerleşim lekesine de değmesi olarak değiştirilmiştir. Böylelikle bölgelere atılan hizmet alma durumu değeri değiştirilmiştir. Atılan tampon bölgenin yerleşim lekesine değmesi durumu 1 değeri ise 0 değeri olarak atanmıştır. Sonuç

olarak yapılan itfaiye istasyonlarına atılan tampon bölge (buffer) analizi doğrultusunda istasyonların 808 bölgenin 482 tanesine yeterli hizmet veremediği görülmektedir. Elde edilen bu sonuç problemin tanımlanması ve makine öğrenmesi algoritması için kullanılan veriler için daha iyi sonuçların vermesini sağlamaktadır.



Şekil 3. Bölgelere atılan hizmet durumu

Bu kapsamda yukarıda belirlenen standartlar doğrultusunda yeterli hizmet alamayan bölgelere hizmet verebilmek için yeni itfaiye istasyonu önerilmesi gerekmektedir. Yeni itfaiye istasyonlarının hangi bölgelere kurulması gerektiğini belirlemek için makine öğrenmesi yöntemleri kullanılarak İzmir ili içerisinde bölgelerin itfaiye istasyonu ihtiyacına göre sınıflandırma çalışması gerçekleştirilmiştir. Makine öğrenmesi algoritmaları mimarisi ve gözlemlenen verilerin istatistiksel dağılımları, genellikle bir kayıp (hata) işlevini en aza indirerek gerçekleştirilen sınıflandırma modellerinin eğitimini yönlendirir (Kuncheva, 2004; Marsland, 2009). Eğitilmiş sınıflandırma modelleri daha sonra eğitim verileri içinde mevcut olan sınıfları tahmin etmek için benzer girdi değişkenlerine uygulanır (Hastie et al., 2009; Witten and Frank, 2005).

Çalışmada makine öğrenmesi algoritması ile sınıflandırma sürecini veri ön işleme, sınıflandırma modeli eğitimi ve tahmin değerlendirmesi olmak üzere üç aşama da gerçekleştirilmiştir (Cracknell ve Reading 2014). Veri ön işleme, mevcut verileri temsili bir dizi girdiyi derlemeyi, düzeltmeyi, dönüştürmeyi veya alt kümeyi oluşturmayı amaçlar ve istenen uygulama ile ilgili bilgileri içerecek şekilde veri hazırlamayı sağlar (Guyon, 2008; Hastie vd.,2009). Makine öğrenme algoritmaları mevcut veriler ve istenen uygulama göz önüne alındığında performanslarını optimize etmek için ayarlanmış bir veya daha fazla algoritmadan bir tanesinin seçilmesi gerekmektedir (Guyon, 2009). Eğitim ve performans verilerinin tahmini için k-fold ve çapraz doğrulama (cross-validation) gibi teknikler kullanılması gerekmektedir. Eğitilmiş sınıflandırma modelin performans genellikle k-fold sonuçlarının toplanması veya ortalamasının alınması ile bulunmaktadır. Test edilen verilerin sonuçları da kappa ve doğruluk analizi ile değerlendirilmiştir.

İtfaiye istasyonu yer seçimi ve yangınların tahminlenmesi kapsamında yapılan çalışmalar incelendiğinde birçok farklı çalışmaya rastlanmıştır. 2017 yılında Connor tarafından itfaiye istasyonlarının yangın yönetimi çerçevesinde konumlarının belirlenmesinde boosted regression tree (BRT) makine öğrenmesi yöntemi kullanılmıştır (Connor vd. 2017). BRT yöntemi, değişken etkileşimleri otomatik olarak tanımlayan ve modelleyen ve aday değişkenlerin geliştirilmiş doğrusal veya geliştirilmiş katkı modellerinden daha az (yangın çapı gibi) yaygın keskin süreksizliklerin üstesinden gelebilen esnek bir parametrik olmayan makine öğrenme yaklaşımıdır (De'ath G. 2007; Elith ve diğerleri, 2008). Bir diğer çalışma da

Betanzos tarafından yapay sinir ağları yöntemi ile yangın tahmin modeli gerçekleştirilerek belirlenen bölgeler 0.789 doğrulukla dört kategoride sınıflandırılmıştır (Betanzos vd. 2003). Orman yangınlarının hangi bölgelerde çıkacağını uzun vadeli tahmin etmek için bulanık makine öğrenmesi yöntemleri kullanarak yüksek ve düşük yangın riskli alanların sınıflandırılması gerçekleştirilmiştir (Iliadis 2005). Orman yangınlarının tahmini için bir diğer çalışmada meteorolojik veriler ile veri madenciliği yöntemleri kullanılarak tahminleme gerçekleştirilmiştir. Çalışma da destek vektör makineleri, random forest algoritmaları kullanılmıştır. Sınıflandırma algoritmasında sıcaklık, nem oranı, yağmur ve rüzgar verileri kullanılmıştır (Cortez 2007). 2006 yılında Amatulli vd. tarafından karar ağaçları yöntemi ile uzun vadeli yangın riski değerlendirilmesi gerçekleştirilmiştir (Amatulli 2006). Yangın oluşumları modeli için random forest algoritması ile mekânsal ekonometrik modellerin karşılaştırılması gerçekleştirilmiştir. Çalışma kapsamında itfaiye istasyonlarının noktalarının konumlarının belirlenmesinde makine öğrenmesi yöntemlerinin daha iyi sonuçları verdiği tespit edilmiştir (Song 2017). 2007 yılında genetik algoritma yaklaşımı kullanılarak havaalanı yangın istasyonlarının optimum lokasyonlarının tahminlenmesi sağlanmıştır. Çalışma ilk olarak ihtiyaç duyulan yangın istasyonu ve ikinci olarak optimum yerlerin belirlenmesi olmak üzere iki kısımda gerçekleştirilmiştir (Tzeng ve Chen 2007). Makine öğrenmesi algoritmaları birçok farklı alanda kullanıldığı gibi uydu görüntülerinin sınıflandırılması için de kullanılmaktadır. Örneğin petrol sızıntılarının tespiti için makine öğrenme algoritmaları kullanılmıştır (Kubat vd 1998). 2014 yılında da uydu görüntüleri kullanarak jeolojik haritaların yapımında yine makine öğrenmesi algoritmaları kullanılmıştır. Çalışma da Naive Bayes, en yakın komşu analizi, random forest, destekçi vektör makineler ve yapay sinir ağları algoritmaları sonuçları karşılaştırılmıştır. Çalışmanın sonuçları incelendiğinde en iyi sonuç veren algoritmanın random forest algoritması olduğu tespit edilmiştir. İtfaiye istasyonları yer seçim problemlerinde amaç tüm itfaiye istasyonlarının optimum seviyede hizmet vermesini sağlamaktır. Bu konuda yapılan çalışmalar incelendiğinde hizmet alanı, nüfus yoğunluğu, ulaşılabilirlik, trafik yoğunluğu, olay sayısı gibi kriterler kullanılarak yangın çıkma olasılığı hesaplanmıştır (Valinski 1955). İtfaiye istasyonları yer seçiminde olay sayılarına göre en çok talep gelen bölgelere hizmet götürebilmek, istasyondan olay yerlerine gitmek için kat edilen en uzun mesafeyi en aza indirmek, istasyondan ortalama ulaşım sürelerini en aza indirmek gibi amaçlar bulunmaktadır. Amaçlar doğrultusunda bu zamana kadar bu problemin çözümünde optimizasyon yöntemlerinden olan doğrusal programlama veya kapsama modelleri gibi yöntemler kullanılmıştır.

Bu çalışmada ise makine öğrenme yöntemleri ile sınıflandırma yöntemi kullanılarak itfaiye istasyonlarının ihtiyacı tahminlenmiştir. Sınıflandırma için itfaiye araçlarının o bölgeye ulaşım süresi, nüfus yoğunluğu, bölgeye giden ortalama ana ve yardımcı araç sayısı ve bölgedeki itfaiye istasyonu bulunma durumu verileri kullanılmıştır. Sınıflandırma için kullanılan veri seti; coğrafi bilgi sistemleri araçları kullanılarak belirli bir model doğrultusunda elde edilmiştir. Örneğin nüfus yoğunluğu tematik haritalama yöntemi kullanılarak elde edilmiştir. Bölgelerin toplam yangın sayıları da yine SUM fonksiyonu ile elde edilerek mekânsal olarak ortaya konmuştur. Çalışmanın amacı tüm bölgelerin beş

dakikada içerisinde hizmet alabilmesi için bölgelerin itfaiye istasyonu ihtiyacını belirlemektir.

2. Yöntem

2.1. Veri

Çalışma kapsamında kullanılan veri İzmir Büyükşehir Belediyesi itfaiye daire başkanlığı bünyesinde tutulmuş yangın raporlarının dijitalleştirilmesi ile elde edilmiştir. 2015-2017 yılları arasında 5400 adet yangın raporları dijitalleştirilmiş ve mekânsal olarak coğrafi bilgi sistemleri ortamında harita üzerine işaretlenmiştir. Yangın raporundan bölgelere göre itfaiye araçlarının yangınlara ortalama ulaşım süreleri hesaplanmıştır. Bunun yanında yangına giden ana ekip araç sayısı ve yardımcı araç ekip sayısı hesaplanmıştır. Ana ekip araç sayısı ve yardımcı ekip araç sayısı çıkan yangının büyüklüğü ve kapsamıyla doğrudan ilgilidir. Bu kapsamda bu sayının yüksek olması çıkan yangının büyük bir yangın olduğu anlamına gelmektedir. Bu bilgilerin yanında bölgelerin nüfus yoğunluğu ve bölge de itfaiye istasyonu olup olmadığı bilgisi de elde edilmiştir. Nüfus yoğunluğu hesaplanırken bölgelere göre İzmir büyükşehir belediyesi 2030 yılı projeksiyon nüfus verisi, öğrenci nüfusu ve istihdam nüfusları toplanarak yeni bir nüfus verisi elde edilmiştir. Nüfus verisi hesaplanırken gelecek nüfus verisine öğrenci ve istihdam verisinin eklenme sebebi yangının gece gündüz farkı olmadan günün herhangi bir saati meydana gelebileceğinden ötürü yalnızca konut nüfusunun ele alınması eksik olacaktır. Bu nedenle bölgelere ait konut nüfusu istihdam ve üniversite bölgelerindeki öğrenci sayıları toplanarak nüfus verisi kullanılmıştır.

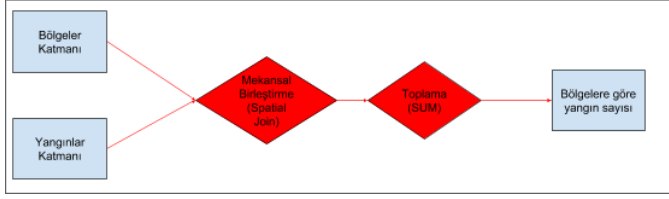
Tablo 1. Sınıflandırma için kullanılan veri seti

Bölgelere göre Nüfus Yoğunluğu	Bölgeler katmanı
Bölgelere göre Yangın Sayısı	Yangınlar katmanı
Bölgelere göre İtfaiye Ana ekip Araç Sayısı	Olay tablosu
Bölgelere göre İtfaiye Yardımcı ekip Araç Sayısı	Olay tablosu
Bölgelere göre İtfaiye Aracı Ulaşım Süresi	Olay tablosu
Bölgelere göre itfaiye istasyonu durumu	İtfaiye istasyonu katmanı (class)

2.2. Veri Önışleme

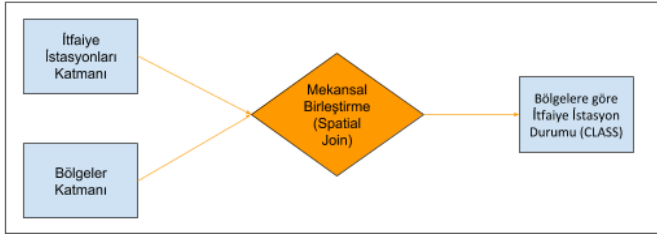
Bu aşamada sınıflandırma için kullanılacak veri setindeki verileri elde etme ve önışleme için coğrafi bilgi sistemlerinin yeteneklerinden faydalanılmıştır. Mekânsal veriler coğrafi bilgi sistemleri aracı ArcGIS Pro programı kullanılarak analiz ve sorgulanarak elde edilmiştir. Örneğin bölgelerin nüfus verileri için bölgelerin içindeki öğrenci sayıları (OGRMAH30), bölgedeki istihdam sayısı (İSTİH30) ve 2030 yılı nüfus tahmini (NUFUS 2030) verileri toplanarak Comma Separated Value

(CSV) formatında her bir bölgenin benzersiz değeri (ID)'si ile eşleştirilerek dışarıya aktarılmıştır. Ardından bölge katmanı ile yangın katmanını mekânsal olarak birleştirme (SPATIAL JOIN) işlemi gerçekleştirilmiştir ve toplama (SUM) fonksiyonunu kullanarak bölgelerin sınırları içerisinde bulunan yangın sayıları elde edilmiştir (Şekil 4).



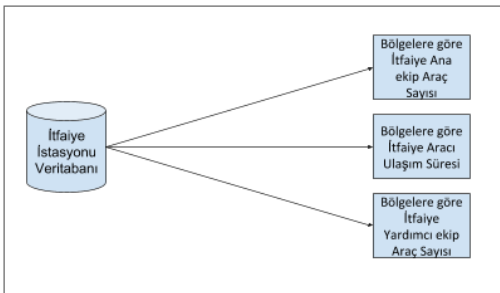
Şekil 4. Bölgelere Yangın Sayıları

Bölgelerin sınırları içerisinde herhangi bir itfaiye istasyonu olup olmama durumunu ise yine mekânsal birleştirme (SPATIAL JOIN) işlemi ile gerçekleştirilmiştir (Şekil 5).



Şekil 5. Bölgelere göre itfaiye istasyonu bulunma durumu

Diğer sözel verileri ise yine olay tablosundan gidiş varış sürelerini birbirinden çıkararak ulaşım süreleri hesaplatıldı. Ana ekip ve yardımcı ekip araç sayılarını da yine veri tabanından çıkarılarak elde ettik. Tüm veriler olaylar için belirlenen ID sütunu kullanılarak birleştirilerek sınıflandırma tablosu elde edilmiştir (Şekil 6).



Şekil 6. Bölgelere göre ana ekip ve yardımcı ekip araç sayıları

2.3. Algoritma Seçimi

Çalışma kapsamında Naive Bayes (NB), Destek Vektör Makineleri (SVM), Decision Tree (J48), K-Nearest Neighbourhood (KNN), Random Forest (RF), Gradient Boosted Trees (GBT) sınıflandırma algoritmaları kullanılmıştır.

Naive Bayes (NB), diğer algoritmalarla karşılaştırmak için temel seviye sınıflandırıcı olarak önerilen iyi bilinen bir istatistiksel öğrenme algoritmasıdır. NB belirli bir sınıf için girdilerin birbirinden bağımsız olduğunu varsayarak sınıf koşulu olasılıkları tarafsız tahmin eder. NB, sınıfları koşullu marjinal yoğunlukları bulmak için ayırıcı sınıflar problemini

azaltır, bu da belirli bir örneğin olası hedef sınıflardan biri olma olasılığını temsil eder (Molina et al., 1994). NB birbirleri ile ilişkili girdiler içermedikçe diğer algoritmalara karşı iyi performans göstermektedir (Hastie vd., 2009.)

En yakın k komşu: Bu algoritma (k-Nearest Neighbors-kNN) sınıflandırılmak istenen yeni bireyin daha önceki bireylerde k tanesine yakınlığına bakıldığı bir algoritmadır (Cover ve Hart, 1967). Sınıflandırma esnasında test örnekleri eğitim örnekleri ile birbirleri ile karşılaştırılır. Komşuluk mesafesi için öklid mesafesi kullanılmaktadır. Tahminler, komşu örneklerin oy çoğunluğuna dayanmaktadır. Yüksek k değerlerine aşırı uyma eğiliminde olmasında dolayı dikkat edilmesi gerekmektedir (Hastie et al., 2009).

Karar ağaçları (Decision Tree): Bu algoritmada bir ağaç yapısı oluşturularak ağacın yaprakları seviyesinde sınıf etiketleri ve bu yapraklara giden ve başlangıçtan çıkan kollar ile de özellikler üzerindeki işlemler ifade edilmektedir (Şeker 2012). Temel algoritmalarından olup aşağıda bahsedilecek olan random forest ve gradient boosted tree bu algoritma yapısından türetilmiştir. Algoritma anlaşılması ve yorumlanması açısından basittir. Hem sayısal hem de sınıfsal verilerin işlenmesi için kullanılabilir.

Destek Vektör Makineleri (SVM): Bu algoritma ikinci dereceden optimizasyon problemini çözerek, yüksek boyutlu değişken uzayda doğrusal olmayan karar sınırlarını tanımlama yeteneğine sahiptir (Hsu et al., 2010). Temel SVM teorisi, iki sınıftan puanlar içeren doğrusal olmayan şekilde ayrılabilir bir veri kümesi için, sınıfları bölen sonsuz sayıda çizginin olduğunu belirtir. İki sınıfı (yani karar sınırını) en iyi şekilde ayıran bir çizginin seçimi, destek vektörleri olarak bilinen sadece eğitim numunelerinin bir alt kümesi kullanılarak gerçekleştirilir. Destek vektörleri arasındaki maksimum sınır mesafesi optimal karar sınırını temsil etmek için alınır. Sınıfların doğrusal olarak ayrıştırılamadığı problemler için SVM, kernel işlevini kullanarak giriş değişkenlerinin örtülü bir dönüşümünü kullanır. Kernel işlevleri, SVM'nin doğrusal bir düzlem kullanarak doğrusal olmayan ayrılabilir destek vektörlerini ayırmasına izin verir (Yu vd., 2012). Çoğu uygulama için performansı optimize etmek için uygun bir çekirdek işlevinin ve çekirdek genişliğinin seçilmesi gerekir (Hsu vd., 2010).

Random Forest: Sınıflandırma işlemi sırasında birden fazla karar ağacı üretmek sınıflandırma oranını yükselten algoritmadır. Rastlantısal olarak seçilen karar ağaçları bir araya gelerek karar ormanını oluşturur. Pek çok veri seti için Destek Vektör Makinelerinden daha doğru sonuçlar üretmektedir. Fazla sayıdaki değişkene ve sınıf etiketine sahip kategorik değişken içeren, kayıp veriye sahip veya dengesiz bir dağılım sergileyen veri setlerinde iyi sonuç verir.

Gradient Boosting Trees: Random Forest algoritmasına benzeyen bu algoritma karar ağaçları ile çalışmaktadır. Temel fark, son tahminin tüm ağaçların doğrusal bir toplamı olması ve her ağacın amacının, önceki ağaçların artık hatasını en aza indirmesidir. XGBoost yani gradient boosting trees algoritması, tüm senaryolarda ölçeklendirilebilen bu algoritmanın verimli bir şekilde uygulanmasını sağlar (Chen and Guestrin, 2016).

Tahminleme Değerlendirmesi

Algoritmaların performansları en başarılı algoritma belirlenirken Doğruluk Oranı, Ortalama Mutlak Hata (MAE), Kök Kareler Hata Ortalaması (RMSE) ve Kappa değerleri ile

performans ölçümü ile gerçekleştirilmiştir. Veri seti %66'sı eğitim %33'ü test verisi olarak ayrılmıştır. Doğruluk Oranı (Accuracy Rate); modelin başarısını ölçmek için en çok kullanılan yöntem doğruluk oranı ölçümüdür (Lu and Weng 2007). Doğruluk oranı doğru olarak sınıflandırılan verilerin tüm verilere oranı olarak hesaplanmaktadır. Doğruluk oranı aşağıdaki şekilde formül ile hesaplanmaktadır (Şekil 5).

$$\text{Doğruluk} = \frac{(T_p + T_n)}{(T_p + F_p + F_n + T_n)}$$

T_p+T_n : Doğru sınıflandırılan veri sayısı F_p+F_n : Yanlış sınıflandırılan veri sayısı

Şekil 5. Doğruluk Oranı Ölçüm Formülü

Ortalama Mutlak Hata (MAE); bir dizi tahmindeki hataların ortalama büyüklüğünü ölçer. Ortalama mutlak hata tahmin ile gözlemlenen arasındaki farkların mutlak değerlerinin doğrulama örneğinin ortalamasını ifade eder.

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_t|$$

Kök kareler hata ortalaması (RMSE); hata ortalama büyüklüğünü ölçen, ikinci dereceden bir puanlama kuralıdır. Tahmin ile gözlemlenen değerler arasındaki fark her bir kareler alınır ve daha sonra örnek üzerinde ortalananır. Her iki değerlendirme aracı bir dizi tahmindeki hataların varyasyonunu teşhis etmek için birlikte kullanılabilir. RMSE her zaman MAE'den daha büyük veya ona eşit olacaktır. Aralarındaki daha büyük fark, örneklemedeki bireysel hataların varyansının daha büyük olmasıdır.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}$$

Kappa testi, iki veya daha fazla gözlemci arasındaki uyumun güvenilirliğini ölçen istatistiksel bir yöntemdir [Congalton ve Green 1998]. Kappa katsayısı -1 ile +1 arasında değişir. Eğer gözlenen uyum şansa bağlı olarak uyumdan daha büyük veya eşitse, o zaman $K \geq 0$; gözlenen uyum şansa bağlı olarak uyumdan daha az ise, o zaman $K < 0$ olmaktadır. Kappa katsayısının yorumlanabilir aralığı 0 ile +1 arasındadır ve negatif ($K < 0$) değerler güvenilirlik açısından anlam taşımamaktadır. Kappa skoru 0,4'ün üzerinde ise makul bir değer olarak görülür. Kappa değeri şu şekilde hesaplanır:

$$K = \frac{(P_o - P_c)}{(1 - P_c)}$$

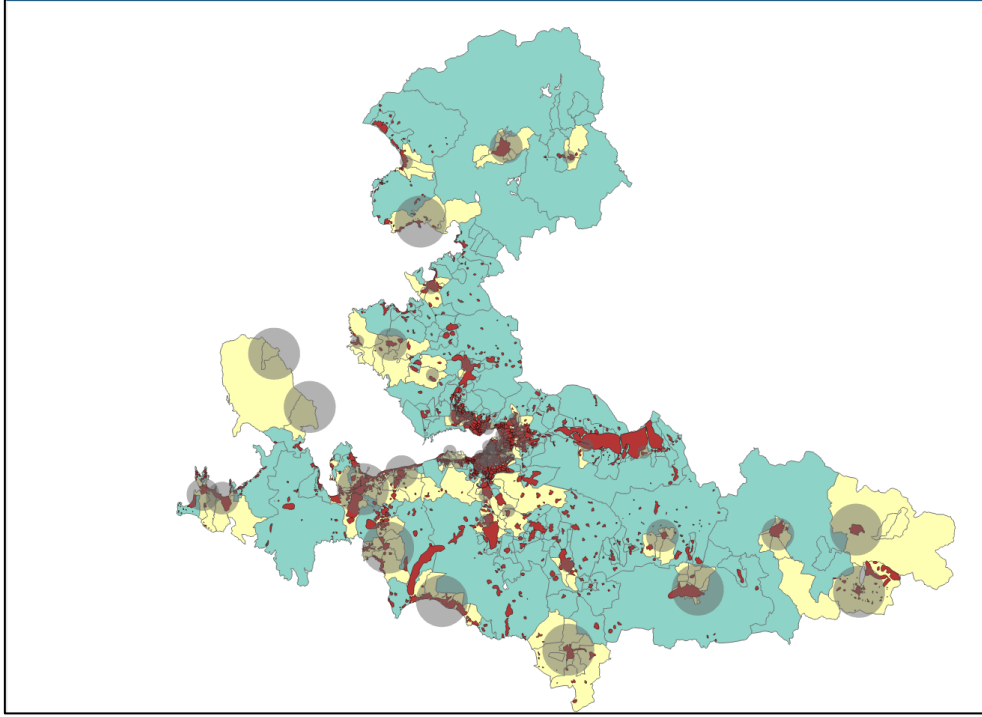
(P_o kabul edilen oran, P_c beklenen oran)

Bu bölümde beş farklı makine öğrenme algoritmasının sonuçları karşılaştırılmıştır. Elde edilen sonuçlar bölgelerin itfaiye istasyonu hizmet alma durumunun sınıflandırmasını sağlamıştır. Algoritmaları karşılaştırmalı sonuçları incelendiğinde Random Forest algoritması %93.8402 oranında doğrulukla sınıflandırma işlemini gerçekleştirmiştir (Tablo 2) ve bölgelerin hizmet alma durumu sonuçları mekansal olarak harita üzerine işaretlenmiştir. Test verileri harita üzerine işaretlendikten sonra öneri istasyonlara yine itfaiye istasyonlarının 5 dk.'lık erişim mesafesi göz önüne alınarak tampon bölgeler çizilerek kapsama alanları ile karşılaştırılmıştır. (Şekil 7).

Tablo 2. Algoritmaların sonuçlarının karşılaştırılması

	Doğru Sınıflandırma Oranı	Kappa İstatistiği	Ortalama Mutlak Hata (MAE)	Karekök ortalama hata (RMSE)	Görelî mutlak hata	Görelî kare kök hatası
Naive Bayes	87.3606	0.0055	0.1731	0.3267	%147.558	%135.5245
KNN	87.8563	-0.0071	0.1225	0.348	%104.4204	%144.3503
Decision Tree	93.8042	0	0.1162	0.2411	%99.1015	%99.9988
SVM	93.8042	0	0.062	0.2489	%52.8235	%103.2484
Random Forest	93.8042	0	0.1154	0.2479	%98.3982	%102.831
Gradient Boosting Trees (Logit Boost)	93.6803	-0.0024	0.1137	0.2428	%96.9376	%100.6932

Tablo 2. Algoritmaların sonuçlarının karşılaştırılması



Şekil 7. Sınıflandırılan itfaiye istasyonlarının 5 dk. erişim mesafesi haritası

3. Sonuç

Bu sonuçlara göre makine öğrenmesi yöntemi kullanılarak itfaiye istasyonu hizmet alma durumuna göre sınıflandırılabilir olduğu kapsama alanı ile %85,43 oranında uyumluluk sağladığı görülmüştür. Bu çalışmada mekânsal veriler kullanılarak makine öğrenme yöntemi ile itfaiye istasyonu hizmet durumu sınıflandırılması gerçekleştirilmiştir. Çalışma da veri hazırlama aşamasında itfaiye istasyonlarının kapsama problemi incelenerek çözülmüş ve veri seti hazırlanmıştır. Veri ön işleme aşamasında verilerin analiz için hazırlanmasında coğrafi bilgi sistemleri kullanılarak mekânsal veriler uygun formatta hazırlanmıştır. Ardından veriler beş farklı makine öğrenmesi algoritması ile test edilmiş ve sonuçlar karşılaştırılmıştır. Gelecek çalışmalarda makine öğrenmesi algoritması için belirlenen değişkenlere trafik yoğunluğu, yol genişliği gibi veriler eklenerek daha iyi sonuçlar elde edilebilir.

Referanslar

Alonso-Betanzos, A., Fontenla-Romero, O., Guijarro-Berdiñas, B., Hernández-Pereira, E., Andrade, M. I. P., Jiménez, E., ... & Carballas, T. (2003). An intelligent system for forest fire risk prediction and fire fighting management in Galicia. *Expert systems with applications*, 25(4), 545-554.

Amatulli, G., Rodrigues, M. J., Trombetti, M., & Lovreglio, R. (2006). Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences*, 11(Song, C., Kwan, M. P., Song, W., & Zhu, J. (2017). A comparison between spatial econometric models and random forest for modeling fire occurrence. *Sustainability*, 9(5), 819.1(G4).

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). ACM.

Cortez, P., & Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data.

Congalton, R.G., Green, K., 1998. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, first edn. Lewis Publications, Boca Raton p. 137.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27.

Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 22-33.

De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251. doi:10.1890/0012-9658(2007)88[243:BTFFEMA] 2.0.CO;2

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity & Distributions* 17, 43–57. doi:10.1111/J.1472-4642.2010.00725.X

Guyon, I., 2008. Practical feature selection: from correlation to causality. In: Fogelman-Soulié, F., Perrotta, D., Piskorski, J., Steinberger, R. (Eds.), *Mining Massive Data Sets for Security – Advances in Data Mining, Search, Social Networks and Text Mining, and their Applications to Security*. IOS Press, Amsterdam, pp. 27–43.

Guyon, I., 2009. A practical guide to model selection. In: Marie, J. (Ed.), *Proceedings of the Machine Learning Summer School*. Canberra, Australia, January 26 - February 6, Springer Text in Statistics, Springer p.37.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, Inference and Prediction, 2nd edn. Springer, New York, USA p. 533.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2010. A Practical Guide to Support Vector Classification Department of Computer Science, National Taiwan University, Taipei, Taiwan 16.

- Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, Inference and Prediction, 2nd edn. Springer, New York, USA p. 533.
- O'Connor, C. D., Calkin, D. E., & Thompson, M. P. (2017). An empirical machine learning method for predicting potential fire control locations for pre-fire planning and operational fire management. *International journal of wildland fire*, 26(7), 587-597.
- Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine learning*, 30(2-3), 195-215.
- Kuncheva, L., 2004. Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons p. 376.
- Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* 28, 823–870
- Marsland, S., 2009. Machine Learning: An Algorithmic Perspective. Chapman & Hall/CRC (406 pp.)
- Molina, R., P´erez de la Blanca, N., Taylor, C.C., 1994. Modern statistical techniques. In: Michie, D., Spiegelhalter, D.J., Taylor, C.C. (Eds.), Machine Learning. Neural and Statistical Classification. Ellis Horwood, New York, pp. 29–49.
- Iliadis, L. S. (2005). A decision support system applying an integrated fuzzy model for long-term forest fire risk estimation. *Environmental Modelling & Software*, 20(5), 613-621.
- Şeker Ş. E. (2012). Karar Ağacı Öğrenmesi. Bilgisayar kavramları internet sitesi: <http://bilgisayarkavramlari.sadievrenseker.com/2012/04/11/karar-agaci-ogrenmesi-decision-tree-learning/>
- Tzeng, G. H., & Chen, Y. W. (1999). The optimal location of airport fire stations: a fuzzy multi-objective programming and revised genetic algorithm approach. *Transportation Planning and Technology*, 23(1), 37-55.
- Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Elsevier/Morgan Kaufman, San Fransisco, USA p. 525
- Valinski D. 1955. "A Determination of the Optimum Location of Fire-Fighting Units in New York City," *Journal of Operations Research Society of America*, 3(4) 494-512.
- Yu, L., Porwal, A., Holden, E.J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput. Geosci.* 45, 229–239.