

Hava kirliliği tehlikeli boyutlara ulaşırken hava kirliliği ile mücadele etmek elzem olmaktadır. Bu amaçla sürekli ölçümler yapılmaktadır. Kriter olarak ölçülmesi gereken kirleticiler ise, Karbon monoksit (CO), Kürek dioksit (SO_2), Ozon (O_3), Partikül madde (PM), Azot oksitler (NO_x) olarak belirtilmektedir.

Günümüzde yapay zekada yaşanan gelişmeler sadece bilgisayar bilimcileri değil diğer bilim dallarında çalışan araştırmacıların da ilgisini çekmeye başlamıştır. Hava kirliliği tahmininde yapay zekanın kullanılması literatürde önemli bir yer tutmaya başlamıştır.

Literatürde, O_3 konsantrasyonlarının modellenmesi için makine öğrenmesi ve derin öğrenme tabanlı farklı çalışmalar bulunmaktadır. Makine öğrenmesi doğrusal olmayan ve yüksek boyutlu veri kümeleri üzerinden kararlı ve performansı yüksek bilgi çıkarımı yapmaktadır (Bilgin, 2021; Yıldırım vd., 2021). Makine öğrenmesi yöntemlerinden çok katmanlı algılayıcı (CKA) (Paoli vd., 2011; Chattopadhyay vd., 2019; Yang vd., 2021; Bekesiene vd., 2021; Makarova vd., 2021), destek vektör makineleri (DVM) (Chelani, 2010; Tanaskuli, 2019; Mehdipour ve Memarianfard; 2019), lineer regresyon (Alipio, 2020; Allu vd., 2020; Matasović vd., 2021), Xgboost (Ding vd., 2020; Liu vd., 2020), rastgele orman (Liu vd., 2020; Ma vd., 2021) yöntemleri ile yapılmış çalışmalar mevcuttur.

Büyük veri analizi ve Grafik İşleme Biriminin (GPU) kullanılmasından bu yana, derin öğrenme büyük ilgi görmekte ve makine öğrenmesinin uygulandığı her alana uygulanmaktadır (Çağıl ve Yıldırım, 2020; Darendeli ve Yılmaz, 2021). Derin öğrenme yöntemleri ile yapılan çalışmalarda derin sinir ağları (DSA) (Wang vd., 2020; Felix vd., 2021), oto kodlayıcı (Nghiem vd., 2021), özyinelemeli sinir ağları (Adnane vd., 2021), Uzun Kısa Süreli Bellek (LSTM) (Alghieith vd., 2021; Ekinci vd. 2021; Zhang vd., 2021), konvolüsyonel sinir ağları (CNN) (Eslami vd., 2020; Sayeed vd., 2021) kullanılmıştır.

Bu çalışmanın amacı saatlik O_3 konsantrasyonlarını modellemede makine öğrenmesi ve derin öğrenme yaklaşımlarını etkinliğini değerlendirmektir. Bu amaçla kirliliğe sebep olan parametrelerden (PM_{10} , SO_2 , NO , NO_2 ve O_3) oluşan zaman serisi veri kümesi kullanılarak Xgboost, YSA ve Uzun Kısa Süreli Bellek (LSTM) yöntemleri karşılaştırılmıştır. Yapılan deneyler sonucunda ozon seviyesini tahmin etmede LSTM yönteminin diğer iki makine öğrenmesi yöntemine kıyasla daha başarılı olduğu gözlemlenmiştir.

Makalenin geri kalan kısmı ikinci bölümde veri kümesi ve uygulanan yöntemlerin anlatıldığı materyal ve yöntemler kısmıdır. Üçüncü bölümde veri önişleme, kullanılan hata metrikleri, modellerin tasarımını ve elde edilen deneyel sonuçlar ayrıntılı şekilde verilmiştir. Son bölümde ise sonuçlar ve öneriler yer almaktadır.

2. Materiał ve Yöntemler (Materials and Methods)

2.1. Veri kumesi (Dataset)

Marmara bölgesinde özellikle Kocaeli ve Sakarya illerinin sanayilerinin gelişmesi ile birlikte bu illerde hava kirliliği oldukça yüksektir. Bu çalışmada, Kocaeli ve Sakarya ile birlikte sanayinin yoğun olmadığı Çanakkale illeri için T.C. Çevre ve Şehircilik Bakanlığı Hava Kalitesi İzleme Ağrı'ndan¹ sürekli ölçümler yapılarak elde edilen verilerden oluşan bir veri kumesi oluşturulmuştur.

İstasyonda ölçülen meteorolojik parametreler 10 µm'nın altındaki parçacıkları ifade eden PM_{10} , azot oksitlerinden NO , NO_2 , NOX , SO_2 ve O_3 şeklindedir. Bu parametreler içerisinde O_3 konsantrasyonunu tahmin etmek için PM_{10} , SO_2 , NO , NO_2 ve O_3 parametrelerine dayalı bir tahmin yapılması hedeflenmiştir. Kocaeli, Sakarya ve Çanakkale illeri için 2018 Kasım ile 2021 Kasım arası saatlik ölçülen zaman serisi değerleri kullanılmıştır. 4 saatlik bir pencere boyutu ile (yani 4 zaman noktası) 5. saat için O_3 konsantrasyonlarının tahmini gerçekleştirilmiştir.

2.2. Yöntemler (Methods)

2.2.1. Xgboost (Xgboost)

Xgboost (Chen vd., 2016) karar ağıacı temelli topluluk öğrenimi algoritmasıdır. Algoritmanın çalışma mantığı, değişkenlere farklı ağırlıklar vererek elde edilen ağaç topluluğundan çıkarımlar yapmaktadır. İlk etapta tüm değişkenler eşit ağırlığa sahiptir. Ağaç topluluğu büyümeye başladıkça, problem bilgisine bağlı olarak ağırlıklar düzenlenmektedir. Yanlış sınıflandırılan gözlemlerin ağırlığı yükseltilirken, doğru sınıflandırılan gözlemlerin ağırlığı düşürülmektedir. Bu sayede ağaçlar zor durumlar karşısında kendini düzenleyebilme yeteneği kazanmaktadır. Fazla uyumu azaltan ve genel performansı artıran çeşitli düzenlemeler içermektedir (Ekinci vd., 2020). Bu özelliğinden dolayı “düzenli artırma” tekniği olarak da isimlendirilmektedir. Xgboost algoritması çeşitli düzenlemeler ile doğruluğu artıran, paralel işleme ile hızlı sonuçlar verebilen, eksik değerlerin kullanımı için standart bir yapıya sahip olan, yükseltme işleminin yineleme aşamalarının her birinde çapraz doğrulama yapan bir algoritmadır.

2.2.2. Yapay Sinir Ağları (Artificial Neural Networks)

Beyin insan vücudunun yapı taşı olup girdileri sinyal şeklinde alan, işleyen ve çıkış sinyallerini gönderen biyolojik sinir ağıdır. Beyin temel birimi nöronudur. Beyin 200 milyar nöronandan oluşmaktadır. Nöronlar dentrit, soma, akson ve sinapsis olmak üzere dört temel kısımdan oluşmaktadır. Nöron, dentritlerden sinyal

¹ <http://sim.csb.gov.tr/Services/AirQuality/>