**Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme**

**Giriş**

Günümüzde insanlığın en büyük sorunlarından birisi hava kirliliği haline gelmiştir. Hava kirliliği, nüfusun artması, kentsel gelişme ve büyüme, endüstrinin gelişmesiyle giderek artan bir önem arz etmektedir. Genellikle hava kirleticilerinin insanlara, canlılara ve çevreye zararlı etkileri zaman, mekan, etki süresi, konsantrasyon ve diğer ayrıcı niteliklere bağlı olarak karmaşık dağılım şekilleri göstermektedir. Günümüz yapay zeka teknolojilerinden makine öğrenmesi ile hava kirliliğine yönelik objektif ve daha hassas sonuçlar elde edilmiştir. Makine öğrenmesi, karmaşık örüntü algılama ve veriye dayalı karar verebilme özellikleriyle ele alınan problemin çözümünü kendi kendine öğrenebilen bilgisayar algoritmalarının genel adıdır. Varolan veri setleri ve kullanılan makine öğrenmesi yöntemleri ile oluşturulan model, en yüksek performansı elde etmek üzerine kurulmaktadır. Bu nedenle birçok makine öğrenmesi yöntemi geliştirilmiş olup bunlardan bazıları; Destek Vektör Makineleri, Lojistik Regresyon, Lineer Regresyon, Basit Bayes, K-En Yakın Komşu, Rastgele Orman ve Karar Ağacı’dır.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Materyal ve Metot**

**Veri Seti**

Türkiye’nin birçok ilinde olduğu gibi Kastamonu merkezinde zaman zaman hava kirliliği problemi rahatsız edici duruma gelmektedir. Hava kirliliğinin sebeplerinin araştırılması yanında çözüm yollarının ortaya konulması son derece önemlidir. Kastamonu ilinde hava kirliliği ölçümleri düzenli olarak 2002 yılında başlamıştır. Bir kaynaktan çıkan kirleticilerin atmosferdeki dağılımları rüzgar hızı ve yönü, sıcaklık, güneş ışığı oranı, bulutluluk ve yağışlılık gibi meteorolojik koşullara bağlı olarak değişkenlik göstermektedir. Günümüzde de hava kalitesinin düzeyini ölçmek amacıyla kurulan Hava Kalitesi Gözlem İstasyonları vasıtası ile saatlik ortalama olarak **SO2, PM10 ve hava sıcaklığı, hava basıncı, nem, rüzgar yönü ve rüzgar hızı** ölçümleri yapılmaktadır.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Makine Öğrenme Yöntemleri Kullanılarak Hava Kirliliği Tahmini**

***Makine Öğrenmesi Algoritmaları***

* **Rastgele Orman:** Rastgele Orman denetimli bir öğrenme algoritmasıdır. Adından anlaşılacağı üzere rastgele bir orman oluşturur. Oluşturulan orman, genellikle “torbalama” yöntemiyle eğitilmiş karar ağaçları topluluğudur. Torbalama yönteminin amacı, öğrenme modellerinin bir kombinasyonunun genel sonucu arttırmasıdır.
* **K-En Yakın Komşu:** Parameterik olmayan bir tekniktir ve sınıflandırmasında en yakın komşularının sayısı olan k’ı grup üyeliğine göre verileri sınıflandırmak için kullanır.
* **Lojistik Regresyon:** Bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi, temel lojistik fonksiyonunu kullanarak olasılıkları tahmin ederek ölçmektedir.
* **Karar Ağacı:** Bir karar ağacı, her düğümün bir özelliği (niteliği) temsil ettiği, her bağlantının (dal) bir kararı temsil ettiği ve her bir yapının bir sonucu olduğu bir ağaçtır. Her ağaç düğüm ve dallardan oluşur. Her düğüm, sınıflandırılacak olan bir gruptaki özellikleri temsil eder ve her dal, düğümün alabilecğei bir değeri temsil eder.
* **Lineer Regresyon:** Bağımsız değişkenlere dayanan bir hedef tahmin değerini modeller. Çoğunlukla değişkenler ve tahmin arasındaki ilişkiyi bulmak için kullanılır. Farklı regresyon modelleri, bağımlı ve bağımsız değişkenler, kullanılan bağımsız değişkenlerin sayısı arasındaki ilişkiye göre farklılık gösterir.
* **Basit Bayes:** Bayes teoremine dayanan bir denetimli bir makine öğrenmesi yöntemidir. Bu algoritma, koşullu olasılıklara dayalı olarak hedef sınıfta belirli bir değerin olasılığını inceler ve buna dayanarak, hedef sınıfın değerini tahmin eder.

Bu çalışmada min-max normalizasyon metodu uygulanmış ve veri seti değerleri 0 ile 1 arasında normalize edilmiştir. Min-Max normalizasyon yöntemi için Denklem 1 kullanılır.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

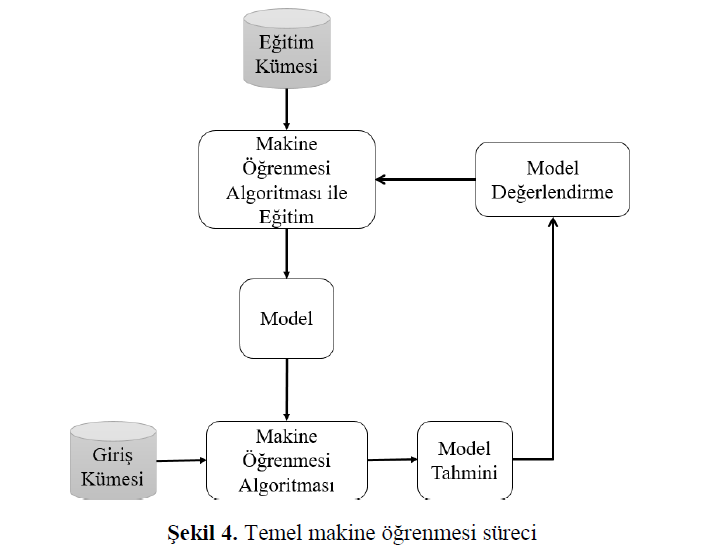
Bu çalışma, Tablo 3'te gösterilen doğruluk kriterlerini kullanır. Bu performans ölçütleri arasında R2, model doğruluğu için belirleyici faktördür. Bu katsayının yüksek bir değeri, iyi bir tahmin ilişkisini gösterir. MSE, RMSE ve MAE hata ölçüleri olduğundan, düşük puanlar performansla ters orantılı olan yüksek performansın ölçüleridir. Örneğin, RMSE sıfırsa iyi çalışır. Zaman serisi, 𝑟𝑡 zaman aralığı ile gözlenir ve 𝑝𝑡 tahmini süre t'yi gösterir. Hata, denklem 2'de 𝑒𝑡 olarak ifade edilir.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Makine Öğrenmesi Adımları**

Makine öğrenmesi yöntemleri ise Pyhton ortamında kodlanmıştır. Python dili, yapay zeka ve makine öğrenmesi uygulamalarında oldukça popüler ve kullanışlı bir dildir.



**Araştırma Sonuçları ve Tartışma**

Makine öğrenimi ile yapılan testlerde R2, MSE, RMSE ve MAE, modellerin performansını değerlendirmeye dayalıdır. Hava kirliliği değerlendirme analizinin ortalama değerleri Tablo 4'te gösterilmiştir. Rastgele Orman ve Karar Ağacı yöntemleri en yüksek R2, en düşük MSE ve MAE oranına sahiptir. En başarılı model parametreleri, R2 değerinin 1'e ve MSE değerinin sıfıra yaklaştığı modeller tarafından temsil edilmektedir. Makine öğrenimi modelleri arasında en iyi sonuçları Random Forest ve Decision Tree değerleri verdi. Doğruluk 0 olduğunda doğrusal regresyon yöntemi çok kötü performans gösteriyor. Ayrıca deneysel sonuçlar, makine öğrenme yöntemlerinin (Random Forest() ve Decision Tree()) yüksek doğruluk elde ettiğini gösterdi.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Sonuç**

Bu çalışma, çeşitli meteorolojik parametreler ile hava kirliliği arasındaki ilişkiyi araştırmıştır. Bu amaçla hava kirliliği beş farklı makine öğrenme modeli ile değerlendirilmiş ve model sonuçları karşılaştırılmıştır. Min-Max (Min-Max) normalleştirme tekniği öğrenme yöntemleri ile birlikte kullanılmıştır. Analizde kullanılan makine öğrenmesi yöntemleri için 2015-2018 yıllarında toplam 2.500 veri hacminin p kısmına karşılık gelen 1.750 verilik bir eğitim aşaması; Test aşaması, kalan 0'a karşılık gelen 750 veri ile gerçekleştirildi. Elde edilen sonuçlara göre, hava kirliliği tahmininde en yüksek performansı rastgele orman ve karar ağacı makine öğrenmesi yöntemleri gösterdi. En kötü sonuç lineer regresyon yöntemiydi. Çalışma sonucunda lojistik regresyon, hava kirliliğini değerlendirmede lineer regresyon makine öğrenmesi yöntemlerine göre daha başarılı olmuştur. Ayrıca geçmiş yıllara ait hava kirliliği ve konsantrasyonu verileri kullanılarak gelecek yıllar için uygun önlemlerin alınabileceği sonucuna varılmıştır.

**Rapor**

<https://dergipark.org.tr/tr/download/article-file/764199>