

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

VERİLER ARASI ANALİTİK KULLANARAK PM10
KONSANTRASYONLARI VE METEOROLOJİK
BİLGİLERDEN HAVA KALİTESİ TAHMİN ETME

MUHAMMED ŞARA

KOCAELİ 2023

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

VERİLER ARASI ANALİTİK KULLANARAK PM10
KONSANTRASYONLARI VE METEOROLOJİK
BİLGİLERDEN HAVA KALİTESİ TAHMİN ETME

MUHAMMED ŞARA

Prof. Dr. Süleyman EKEN

Danışman, Kocaeli Üniversitesi

.....

Doç. Dr. Bilge ÖZBAY

Jüri Üyesi, Kocaeli Üniversitesi

.....

Dr. Öğr. Üyesi Deniz DURAL BALTA

Jüri Üyesi, Sakarya Üniversitesi

.....

Tezin Savunulduğu Tarih: 24.04.2023

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- Bu tez/proje bana ait özgün bir çalışmadır.
- Bilgilerin hazırlanması, verilerin toplanması, analizi ve sunumu dahil olmak üzere çalışmamın tüm aşamalarında bilimsel etik ilke ve kurallara uygun hareket ettiğimi,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgileri alıntıladığımı ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü tarafından belirlenen kriterlere uygun olduğu,
- Kullanılan verilerde herhangi bir tahrifat yapmadım,
- Tezin/Projenin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez/proje çalışması olarak teslim etmediğimi,

Bildiriyorum.

☒ Bu tez çalışmasının hiçbir aşaması herhangi bir kurum/kuruluş tarafından finansal/altyapısal olarak desteklenmemiştir.

Her zaman işimle ilgili yapmış olduğum bu açıklamaya aykırı bir durum oluşması halinde doğabilecek her türlü manevi ve hukuki sonuçları kabul ettiğimi beyan ederim.

.....

(İmza)

.....

Muhammed ŞARA

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin/projemin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanımı bana ait olacaktır.

Tezin/projenin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin/projenin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim kurulu tarafından yayınlanan “*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*” kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü yönetim kurulu kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmişti
- ☐ Enstitü yönetim kurulu gerekçeli kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir
- ☒ Tezim/projem ile ilgili gizlilik kararı verilmemiştir.

.....
(İmza)

.....
Muhammed ŞARA

ÖNSÖZ VE TEŞEKKÜR

Geçmişten günümüze çeşitli bilim alanlarında yapılmış çalışmalar ile insan hayatı ve ekosistemlerin daha iyiye evrilmesi temel amaç olmuştur. Biz de günümüzde eskisinden de fazla miktarda ihtiyacın arttığı bu alanda, bu amaçlar doğrultusunda insanların yaşamlarına dokunarak, daha sağlıklı bir hayat için hava kalitesi tahmin eden bir çalışma hazırladık. Çalışmalarım boyunca değerli bilgi ve tecrübelerini aktardığı için sayın hocam Doç. Dr. Süleyman EKEN 'e teşekkür ederim.

Mayıs- 2023

Muhammed ŞARA



İÇİNDEKİLER

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ.....	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI	ii
ÖNSÖZ VE TEŞEKKÜR.....	iii
İÇİNDEKİLER.....	iv
ŞEKİLLER DİZİNİ.....	vi
TABLolar DİZİNİ.....	vii
SİMGELER VE KISALTMALAR DİZİNİ	viii
ÖZET	ix
ABSTRACT	x
1. GİRİŞ.....	1
1.1. Konunun Önemi.....	2
1.2. Araştırma Problemleri ve Katkıları.....	2
1.3. Tez Yapısı	3
2. İLGİLİ ÇALIŞMALAR	5
2.1. Sağlıklı Yaşam İçin Yapay Zeka	6
2.2. Hava Kalitesi Tahmini	7
2.3. Çevresel Veriler Arası Analitik	10
3. METARYEL VE METOTLAR	12
3.1. Çalışma Alanı ve İklimsel Özellikleri.....	13
3.1.1. Brunei	13
3.1.2. Singapur.....	14
3.1.3. Tayland	14
3.2. Veri Seti Temini.....	15
3.3. Makine Öğrenmesi Yöntemleri.....	15
3.3.1. Neural Network (NN)	15
3.3.2. Random Forest (RF)	16
3.3.3. Adaptive Boosting (AdaBoost).....	18
3.3.4. Gradient Boosting (GB).....	19
3.3.5. XGBoost (XGB)	20
3.3.6. Light Gradient Boosting Machine (LightGBM).....	21
3.3.7. Support Vector Regression (SVR).....	22
4. DENEYSEL TESTLER VE TARTIŞMA.....	24
4.1. Çalışma Ortamı Hazırlıkları.....	24
4.2. Veri Seti Hakkında.....	25
4.3. Veri Ön İşleme.....	28
4.4. Performans Metrikleri.....	30
4.5. Performans Sonuçları ve Tartışma.....	31
4.5.1. Yerelleştirilmiş Hava Kirliliği Tahmini (Görev 1).....	32
4.5.1.1. Brunei Bölgeleri.....	33
4.5.1.2. Singapur Bölgeleri	33
4.5.1.3. Tayland Bölgeleri.....	34
4.5.2. Sınır Ötesi Hava Kirliliği Tahmini (Görev 2).....	34
4.5.2.1. Brunei Bölgeleri.....	35
4.5.2.2. Singapur Bölgeleri	36
4.5.2.3. Tayland Bölgeleri.....	36
4.5.3. Transfer Öğrenme ile Hava Kirliliği Tahmini (Görev 3)	37

4.5.3.1. Brunei Bölgeleri.....	37
4.5.3.2. Singapur Bölgeleri	38
4.5.3.3. Tayland Bölgeleri.....	39
4.5.4. Diğer Çalışmalarla Karşılaştırma	39
4.6. Veri Görselleştirme.....	42
4.6.1. Brunei Bölgeleri.....	42
4.6.2. Singapur Bölgeleri	44
4.6.3. Tayland Bölgeleri	45
5. SONUÇLAR VE ÖNERİLER	47
5.1. Çalışmanın Kısıtları	47
5.2. Gelecek Çalışmalar	47
KAYNAKLAR.....	49
KİŞİSEL YAYIN VE ESERLER.....	52
ÖZGEÇMİŞ.....	53



ŞEKİLLER DİZİNİ

Şekil 2.1.	Birleşik Krallık Hava Kalitesi Tahmini.....	6
Şekil 2.2.	Ulusal Hava Kalitesi İndeksi Kesme Noktaları.....	8
Şekil 2.3.	ABD Hava Kalitesi İndeks Tahmin Tablosu.....	9
Şekil 2.4.	2020 Yılı Boyunca Havası En Kirli 15 İstasyonlar.....	10
Şekil 2.5.	2020 Bazında Hava Kalitesi Durumu (PM10)	11
Şekil 3.1.	Sinir Ağı Yapısı.....	16
Şekil 3.2.	Random Forest Algoritma Yapısı	17
Şekil 3.3.	AdaBoost Topluluk Öğrenme Yöntemi	18
Şekil 3.4.	Gradient Boosting Algoritması	19
Şekil 3.5.	XGBoost'un Standart GBM'yi Optimize Yolları	20
Şekil 3.6.	Seviye Odaklı Büyüme.....	21
Şekil 3.7.	Yaprak Odaklı Büyüme.....	22
Şekil 3.8.	Destek Vektör Regresyon.....	23
Şekil 4.1.	Brunei (2010-2017) PM10 ve Yağmur Değerleri Grafiği.....	28
Şekil 4.2.	Singapur (2010-2017) PM10 ve Yağmur Değerleri Grafiği	28
Şekil 4.3.	Tayland (2010-2017) PM10 ve Yağmur Değerleri Grafiği.....	28
Şekil 4.4.	Brunei Eksik PM10 Doldurma İşlem Örneği	29
Şekil 4.5.	Brunei Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli	43
Şekil 4.6.	Brunei Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli.....	43
Şekil 4.7.	Singapur Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli	44
Şekil 4.8.	Singapur Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli.....	45
Şekil 4.9.	Tayland Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli	46
Şekil 4.10.	Tayland Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli.....	46

TABLolar DİZİNİ

Tablo 4.1.	Hava Kalitesi Veri Seti Yapısı	26
Tablo 4.2.	İstasyon Veri Seti Yapısı.....	26
Tablo 4.3.	Hava Durumu Veri Seti Yapısı	27
Tablo 4.4.	Aynı Bölge İstasyon ID İşlemi.....	28
Tablo 4.5.	Brunei İstasyonlar ve Bilgileri Verisi	29
Tablo 4.6.	Görev 1 İçin Kullanılan Modellere Ait ve Parametre Bilgisi.....	32
Tablo 4.7.	Brunei Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	33
Tablo 4.8.	Singapur Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	33
Tablo 4.9.	Tayland Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	34
Tablo 4.10.	Görev 2 İçin Kullanılan Modellere Ait ve Parametre Bilgisi.....	35
Tablo 4.11.	Brunei Bölgelerine Görev 2'ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	35
Tablo 4.12.	Singapur Bölgelerine Görev 2'ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	36
Tablo 4.13.	Tayland Bölgelerine Görev 2'ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları.....	36
Tablo 4.14.	Görev 3 İçin Kullanılan Modellere Ait ve Parametre Bilgisi.....	37
Tablo 4.15.	Brunei Bölgelerine Görev 3'ye Ait PM10 Tahmininde Kullanılan Transfer Öğrenmesi Sonuçları.....	38
Tablo 4.16.	Singapur Bölgelerine Görev 3'ye Ait PM10 Tahmininde Kullanılan Transfer Öğrenmesi Sonuçları.....	38
Tablo 4.17.	Tayland Bölgelerine Görev3'ye Ait PM10 Tahmininde Kullanılan Transfer Öğrenmesi Sonuçları.....	39
Tablo 4.18.	Görev 1 için PM10 Değer Tahmin Performans Karşılaştırma Tablosu	40
Tablo 4.19.	Görev 2 için PM10 Değer Tahmin Performans Karşılaştırma Tablosu	41
Tablo 4.20.	Görev 3 Transfer Öğrenme ile PM10 Değer Tahmin Performans Karşılaştırma Tablosu.....	41

SİMGELER VE KISALTMALAR DİZİNİ

µg	: Mikrogram
m ³	: Metreküp
°C	: Santigrat

Kısaltmalar

AdaBoost	: Adaptive Boosting (Uyarlanabilir Güçlendirme)
CO	: Karbonmonoksit
CSV	: Comma Separated Values (Virgülle Ayrılmış Değerler)
DMTK	: Distributed Machine Learning Toolkit (Dağıtılmış Makine Öğrenimi Araç Seti)
GBM	: Gradient Boosting Machine (Gradyan Yükseltme Makinesi)
HKİ	: Hava Kalite İndeksi
LightGBM	: Light Gradient Boosting Machine (Hafif Gradyan Artırma Makineleri)
MAE	: Mean Absolute Error (Ortalama Mutlak Hata)
NO ₂	: Azot Dioksit
NN	: Neural Network (Sinir Ağı)
O ₃	: Ozon
PM ₁₀	: Particulate Matter 10 (Parçacık Madde 10)
RF	: Random Forest (Rastgele Orman)
RMSE	: Root Mean Square Error (Kök Ortalama Kare Hatası)
SO ₂	: Kükürt Dioksit
SVM	: Support Vector Machine (Destek Vektör Makinesi)
SVR	: Support Vector Regression (Destek Vektör Regresyon)
XGBoost	: Extreme Gradient Boosting (Aşırı Gradyan Artırma)
YSA	: Yapay Sinir Ağları

VERİLER ARASI ANALİTİK KULLANARAK PM10 KONSANTRASYONLARI VE METEOROLOJİK BİLGİLERDEN HAVA KALİTESİ TAHMİN ETME

ÖZET

Gelişen dünyamızla birlikte artan hava kirliliği dünyamızı ve hayatımızı olumsuz etkilemektedir. Artan nüfus ve tüketim ihtiyacı ile hava kirliliği de artmaktadır. Son çalışmalar, çeşitli bölgesel hava olaylarını inceleyerek hava kirliliği ve kalite tahmin modellemesine odaklanmıştır. Hava kirliliğinin belirlenmesi ve kalitesinin belirlenmesinde birçok parametrenin etkili olması nedeniyle çalışmalarda daha çok makine öğrenmesi yöntemleri tercih edilmeye başlanmıştır. Bu çalışmada, çapraz veri analitiği kullanılarak PM10 konsantrasyonlarından ve meteorolojik bilgilerden hava kalitesi tahmin edilmektedir. Asya bölgesindeki bazı komşu ülkelerden elde edilen günlük hava koşulları, PM10 konsantrasyonu ve hava kalitesi verilerine göre, bölgelerin hava kalitesi tahmin edilmekte ve etki tek tek gözlemlenmektedir. Böylece, başta bu bölgelerdekiler olmak üzere, benzer hava kirliliği sorunu yaşayan veya gelecekte yaşama riski taşıyan tüm toplumların bireysel yaşamları için önemli faydalar sağlanması planlanmaktadır. Mediaeval tarafından verilen “Insight for Wellbeing: (sınır ötesi) Haze Tahmini için Çapraz Veri Analitiği” başlıklı kıyaslama veri setini çalışmamızda kullanmaktayız. Bu veriler üzerinde farklı öğrenme modelleri üç görev üzerinde yoğunlaşarak incelenmiştir. Ayrıca sonuçlar çeşitli grafiklerle görselleştirilip yorumlanmıştır. Bu sayede bölgelerin ve bölge halkının hava kalitesi indeksleri hakkında çıkarımlarda bulunmaktadır.

Anahtar Kelimeler: Evrişimli Sinir Ağı, Hava Kalitesi, Makine Öğrenmesi, Regresyon, Transfer Öğrenimi.

PREDICTION OF AIR QUALITY FROM PM10 CONCENTRATIONS AND METEOROLOGICAL INFORMATION USING CROSS DATA ANALYTICS

ABSTRACT

Increasing air pollution with our developing world affects our world and our lives negatively. Air pollution increases with the increasing population and the need for consumption. Recent studies have focused on air pollution and quality prediction modeling by examining various regional weather events. Since many parameters affect the determination of air pollution and its quality, machine learning methods have started to be preferred mostly in studies. In this work, air quality is estimated from PM10 concentrations and meteorological information using cross-data analytics. According to the daily weather conditions, PM10 concentration and air quality data obtained from some neighboring countries in the Asian region, the air quality of the regions is estimated, and the effect are observed individually. Thus, it is planned to provide significant benefits for the individual lives of all societies that have similar air pollution problems or have the risk of living in the future, especially those in these regions. We use benchmark dataset given by Mediaeval titled “Insight for Wellbeing: Cross-Data Analytics for (transboundary) Haze Prediction”. Different learning models are examined on this data by concentrating on three tasks. Also, results are visualized and interpreted with various graphics. In this way, inferences about the air quality indices of the regions and the people of the region are obtained.

Anahtar Kelimeler: Convolutional Neural Network, Air Quality, Machine Learning, Regression, Transfer Learning.

1. GİRİŞ

Gelişmekte olan ülkelerde uluslararası rekabeti yakalama gerekliliği tarımın beraberinde sanayileşmeye de önem vermeyi getirmiştir. Çevre ve canlı sağlığı, rekabetin neden olduğu maksimum hız ile maksimum ürün ihtiyacına yetişebilmek için halen riske atılabilen ve önemi yeterince anlaşılamamış konulardandır. Bu süreçler sonucu oluşan hava kirliliği, havada bulunan başlıca partikül maddeler (PM), kükürt dioksit (SO₂), nitrojen oksitleri (NO₂) ve ozon (O₃) gibi madde ve moleküllerin ortamda bulunması olarak tanımlanmaktadır. Temel olarak doğal ya da yapay nedenlerle oluşan bu kirliliğin kaynakları tarımda zirai ilaçlamalar, ulaşımda taşıt emisyonları, sanayide çeşitli kimyasallar ve bunların yanma ürünleri, ısıtma sistemlerinde kullanılan yakıtlar, yenilenemez enerji kaynakları, orman yangınları, çöl fırtınaları ve yanardağ patlamalarıyla çıkan duman bulutları kadar geniş bir kapsamda incelenebilir. Artan hava kirliliği, dünyamıza verdiği zararın yanı sıra bireylerin sağlık ve yaşamlarına verdiği zarar da önemli ölçüde artmıştır. Havada bulunan kirletici maddeler özellikle hassas bir yaş grubu olan çocuk ve yaşlılarda bronşit, astım, zatürre ve kalp krizi gibi çeşitli kronik hastalıklara sebep olur veya altta yatan tabloyu kötüleştirir (Şimşek ve Yılmaz, 2020). Başta bireysel önlem alıp karbon ayak izimizi düşürerek hava kalitesinin artırılması bu ve benzeri kirlilik sorunlarıyla mücadelede ilk adım sayılabilir. Hava kalitesi ne kadar yüksek olursa hayat kalitesi de buna doğru orantılı olarak yükselecektir.

Bölgelerdeki hava kirliliğinin hangi düzeyde olduğu, hava kalitesi düzenli olarak ölçülerek belirlenmektedir. Yapılan ölçümler ile bölgelerin temiz hava planları çıkarılmakta, hava kirliliği haritaları oluşturulmakta ve dağılım modelleri yapılabilmektedir.

Son zamanlarda yapılan çalışmalar ile bölgesel olarak çeşitli hava olayları incelenerek hava kirliliği ve kalitesi tahmin modellenmesi üzerine çalışılmaktadır. Hava kirliliğinin ve kalitesinin bulunmasına birçok parametre etki etmesinden kaynaklı yapılan çalışmalarda makine öğrenmesi yöntemleri de çoğunluklu olarak tercih edilmeye başlanmıştır. Biz de çalışmamızda hava kalitesi tahminleme işlemlerinde makine öğrenmesi modellerinden yararlanarak tahminleme işlemleri üzerine durmaya çalışacağız.

1.1. Konunun Önemi

Hava kirliliği artan nüfus ile meydana gelen tüketim ihtiyacı ile artış göstermektedir. Artan hava kirliliği bireylerin sağlıklarını ve refahlarını olumsuz yönde etkilemektedir. Son zamanlarda yapılan çalışmalar ile bölgesel olarak çeşitli hava olayları incelenerek hava kirliliği ve kalitesi tahmin modellenmesi üzerine odaklanılmıştır. Hava kirliliğinin ve kalitesinin bulunmasına birçok parametre etki etmesinden kaynaklı yapılan çalışmalarda makine öğrenmesi yöntemleri de çoğunluklu olarak tercih edilmeye başlanmıştır. Yazacağımız bu tez ile veriler arası analitik kullanılarak PM10 konsantrasyonları ve meteorolojik bilgilerden hava kalitesi tahmini yapılarak çıkan sonuçlara göre hava kalitesi endeks tahminlemesi yapılacaktır. Çalışmalarımızda Asya bölgesinde bulunan bazı komşu ülkelerden alınan günlük hava durumu ve hava kalitesi verilerine göre hava kalitesi tahmini yapılarak bireysel olarak etkisi gözlemlenmiş olunacaktır. Bu bölgelerdeki kişilerin bireysel yaşamları için ve çevresel hava kirliliği hakkında önemli ölçüde fayda sağlanması planlanmaktadır. Kullanılan veriler, 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei, Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen PM10 konsantrasyonlarının günlük okumalarını içermektedir. Bu veriler Dense, XGBoost, Random Forest, AdaBoost, Light Gradient Boosting Machine, Support Vector Regression gibi makine öğrenimi modelleri ile eğitilerek yaklaşımlar incelenecektir. Eğitim ve analizler neticesinde çıkan sonuçlar çeşitli grafikler ile görselleştirilip yorumlanacaktır. Bu şekilde bölge ve bölge halkının hava kalitesi endeksleri hakkında çıkarımlar elde edilmiş olunacaktır.

1.2. Araştırma Problemleri ve Katkıları

Gün geçtikçe artan nüfus, enerji tüketimi ve şehirleşme ile çeşitli problemleri meydana getirmektedir. Bunlardan birisi de hava kirliliği sorunudur, hava kirliliği çok önemsenmeyen ancak etkisi büyük olan problemlerdendir. Bu nedenle hava kirliliği ve bunun hayatımızdaki önemi araştırmaya değer bir konu olmuştur. Geleceğe yönelik hava tahminleri yapılarak, bireylerin sağlığı, sürdürülebilir kentsel planlama ve ekolojik sistemlerin sürdürülebilmesi gibi konular için de önem arz etmektedir. Tahminleme çalışmaları ile hava kalitesinin değişimi izlenmesi ve kirliliğin analiz edilerek daha bilinçli bir hale gelmesi sağlanabilir. Hava kirliliği ve insan sağlığı üzerinde etkileri

araştırmak için birçok çalışma mevcuttur. Kısa bir araştırma ile denenmiş ve çıkarımlar üretilmiş çalışmalar görülebilmektedir. Gerek sağlık gerekse mühendislik alanlarından birçok araştırmaya da konu olmuştur. Ancak yapılan çalışmalar ile sadece küçük kapsamlı bölgesel ve daha az verilerle yapılmış çalışmalar mevcuttur, yapılan çalışmalar yapıldığı ile kalınarak insanların kişisel yaşamlarını etkileyecek bölgeler herhangi bir görselleştirme ile ayrıntılı analiz yapılmamıştır.

Yapacağımız çalışma ile, gerçek veriler, 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei, Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen PM10 konsantrasyonlarının günlük okumalarını içermektedir. Bu veri içerikleri olarak veri kaynağı olan istasyon bilgileri, hava kirletici partikül değerleri, sıcaklık, nem, rüzgâr yön ve hızı, meteorolojik hareketler ve zamansal koşullar göz önünde bulundurularak analizler yapılacaktır.

Tahminleme işlemleri yaklaşık on makine öğrenmesi modeli algoritması kullanılarak eğitim gerçekleştirilecektir. Eğitim gerçekleştirilen modellerden başarı oranı yüksek olan ve alakasız sonuçlar çıkarmayan modeller tez çalışmasına eklenecektir. Aynı zamanda temin edilene veri setleri üzerinde farklı çalışmalar yapılarak tahminleme başarısı arttırılmaya çalışılacaktır. Bunlar çevre bölgelerin verilerinden yararlanılarak başarı oranını arttırma ve çeşitli transfer learning yani eğitilen modeller referans alınarak tahmin oranındaki değişimler incelenecektir.

Eğitim işlemleri tamamlandıktan sonra tahminleme işleminin interaktif bir şekilde anlık değişimlerin gözlenebileceği bir görselleştirme yapılacaktır. Bu görselleştirme sayesinde geçmiş ve gelecek hava kirliliği değişimi görülecektir. Kişilerin daha kolay analiz yapabilmesi ve verilerle boğulmaması için görseller renklendirilerek uyarı niteliğinde veya tavsiye niteliğinde renkler kullanılacaktır.

1.3. Tez Yapısı

Tez amacı olarak çeşitli hava verileri ile bu veriler arası analitik kullanılarak hava kalitesi tahmini yapılacak bu çalışma ile en ideal modelin bulunması ve bu modellerin karşılaştırılarak çıktılarının incelenmesidir. Hava kirliliği ile ilgili ön yazılar ve konunun

önemi hakkında düşünceler belirtilmiştir. Yapılacak çalışma ile ilgili açıklamalara yer verilmiştir.

Tezin ikinci bölümünde; hava kirliliği hakkında ön araştırmalar yapılarak gelişen teknoloji ve yapay zekâ alanındaki çalışmalara yer verilmiştir. Temel olarak bilinmesi gerek hava kirliliği ile ilgili ön araştırmalara yer verilmiş olup indeks tahminleri ile ilgili bilgiler verilmiştir. Hava kirliliğini gösteren çeşitli grafikler ve görseller ile yapılan çalışmalardan ve riskli bölgelerden bahsedilmiştir.

Tez üçüncü bölümünde; çalışmaların yapılacağı bölgelerin temel coğrafi, sosyoekonomik, demografik özelliklerinden kısa bahsedilmiştir. Kullanılacak olan makine öğrenmesi algoritmaları tarihçesi, yapısı ve modelleri hakkında tek tek bilgi verilerek açıklamalarda bulunulmuştur. Görseller çizilerek olabildiğince çalışma yapıları, algoritmaları hakkında bilgiler verilmiştir.

Tez dördüncü bölümünde; yapılacak olan çalışmaların hazırlıkları ile ilgili bilgilendirmeler verilip kullanılacak olan veri setinin temini hakkında yapılanlar yer almaktadır. Eğitime verilecek olan veri setlerinde bulunan eksiklerin giderilme işlemleri, eğitim öncesi veri ön işleme adımları ayrıntılı olarak ve görsellerle birlikte yer almaktadır. Modele sokulmadan önce verilerden yapılan çıkarımlardan bahsedilmiştir. Modellerin eğitimleri ve bu modellerin üretimden sonra başarı kriterlerine yer verilmiştir. Performans metrikleri incelenen diğer modeller ile karşılaştırma tabloların verilerek tahminleme başarısı en iyi olan modeller belirtilmiştir. En başarılı üç modelin bölgelere göre çıktılarını gösteren Tableau üzerinde görselleştirme çalışmalarına yer verilmiş ve yorumlanmıştır.

Tez beşinci bölümünde; yapılan çalışma kısıtları hakkında bilgilendirmeler yapılmıştır. Yapılan çalışmanın gelecek çalışmalarda üzerine neler katılabileceği neler ile daha ileriye gidilebileceği ve karşılaşılan sorunlar hakkında düşünceler belirtilmiştir.

2. İLGİLİ ÇALIŞMALAR

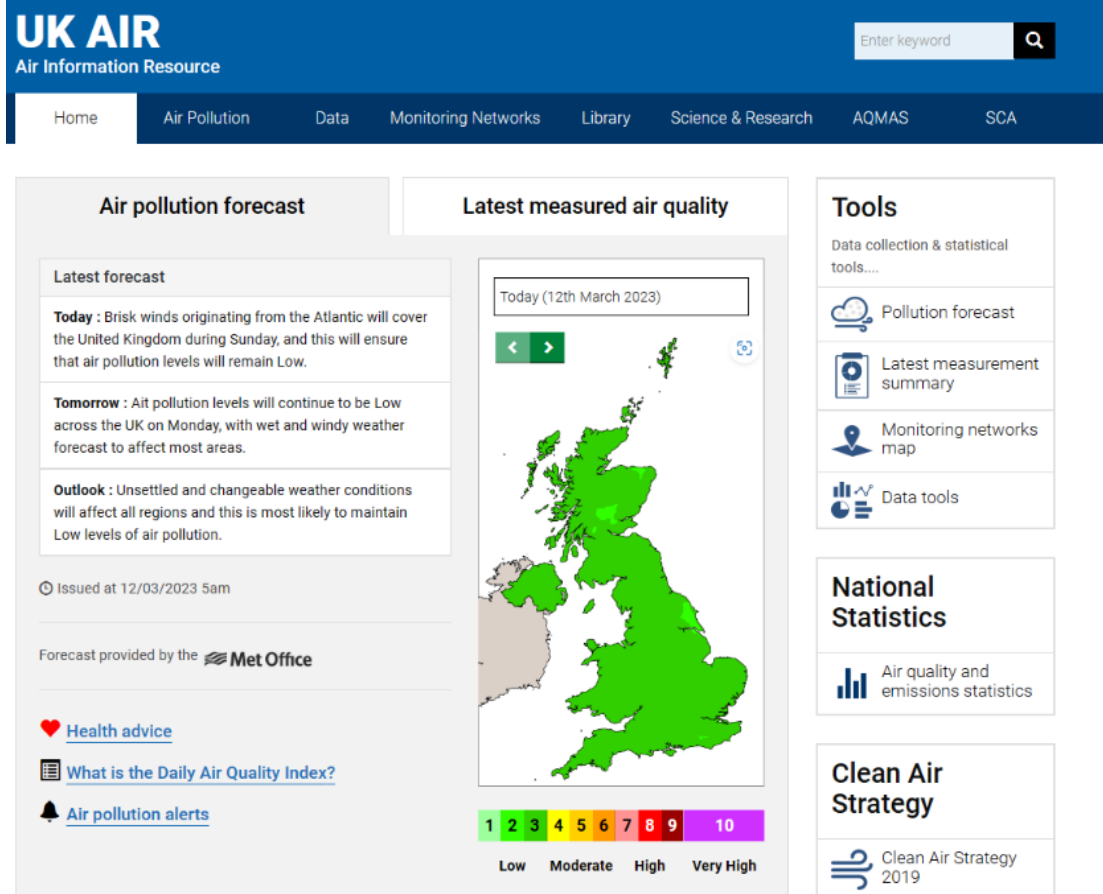
İnsan sağılığını kötü etkileyen ve her geçen yıl artan ölüm oranları yedi milyon civarına çıkmıştır (URL-1). Hava kirliliğı hem üst solunum yolu hem alt solunum yolu hastalıklarının yanı sıra kalp krizi ya da karaciğer yağlanması gibi diğer organ hastalıklarına ve hasarlarına da neden olabilmektedir. Avrupa Çevre Ajansının 2020’de belirttiğı üzere hava kirliliğinin azaltılması çevrenin korunması ile bireylerin yaşam kalitesi ve sağılık hizmetlerinden de tasarruf edilmesi sağılanmaktadır.

Artan sağılık şikayetleri ve olumsuz çevre faktörlerinden kaynaklı olarak son zamanlarda hava kirliliğı alanındaki çalışmalar artmıştır. Çevre, Şehircilik ve İklim Değışikliğı Bakanlığı’nın duyurduğu çeşitli bilgi notları ile farkındalığı arttırmak amaçlanmıştır (URL-2). Bu notlarda yakıt kullanımından, bina izolasyonu, merkezi ısıtılmal binaların bacalarında filtrasyon sistemlerinin kurulması gibi tedbirler yer almaktadır. Avrupa Birliğı (AB) Life Programı kapsamında bölge ve illerin hava kalitesi üzerine çalışmalar yapılmaktadır. Örnek olarak İstanbul Büyükşehir Belediyesi (İBB) ve Dokuz Eylül Üniversitesi ile yapılan ve evrensel ısınma, trafik ve endüstriyel yanma gibi çalışmalar bunu desteklemektedir. Bu çalışma ile 1 km alansal ve 1 saatlik zamansal ve çözünürlükte hazırlanarak PM10, NO₂, CO, SO₂ ve VOC kirleticilerini içermektedir (URL-3).

Gelişen teknoloji ile hava kirliliğinin azaltılarak hava kalitesinin arttırılmasına yönelik çalışmalar yaygınlaşmıştır. Çeşitli sensörlerden alınan veriler ile oluşturulan veri setleri anlamlandırılarak ileriye yönelik tahminler yapıp bunlar yorumlanmaktadır. Bilindiğı üzere yapay zeka teknolojileri bir çok alana girmiş ve başarılı şekilde kullanılmaktadır. Hava kirliliğı analizleri, tahminleri ve raporlama çalışmalarında yapay zeka yöntemlerine oldukça fazla başvurulmaktadır.

Türkiye de hava kalitesi, Çevre, Şehircilik ve İklim Değışikliğı Bakanlığı Ulusal Hava Kalite İzleme Ağı ile anlık olarak takip edilmektedir. Çevre mevzuatlarına göre sanayi tesislerinin baca gazı emisyonları gibi çevreye zarar veren durumlar anlık olarak takip edilmekte ve işletmelerin yükümlölüklerini yerine getirip getirmediikleri takip edilmektedir.

İngiltere hava kalitesi bilgi arşivi web sitesinde, İngiltere hava kirliliği 24 saat öncesine kadar günlük olarak güncellenen tahminler anlık olarak izlenmektedir. Yapılan bu tahminler farklı kaynaklardan faydalanılarak yapılmaktadır. Kirletici kontraksiyonları, iki gün öncesine ait hava durumu tahminleri ve ozon gibi veriler kullanılmakta olup bunlar aracılığıyla tahminleme yapılmaktadır. Şekil 2.1’de Birleşik Krallık hava kalite tahmin bilgisi verilmektedir (URL-4).



Şekil 2.1.Birleşik Krallık Hava Kalitesi Tahmini

Tez kapsamında da yapay zeka modelleri kullanılarak hava kalitesi ile kişilerin sağlıklı yaşamları için çeşitli log analizleri yapılacaktır. Aynı zamanda görselleştirme çalışmaları ile geçmiş ve gelecek günler için hava kalitesi tahminlemeleri raporlanacaktır.

2.1. Sağlıklı Yaşam İçin Yapay Zeka

Sağlık hizmetlerinde ve yönetiminde yapay zeka uygulamaları gelişen teknoloji ve bilgi çağıyla ün kazanmaktadır. Yapay zeka uygulamalarının sağlık hizmetlerinde de

kullanılması için bilimsel çalışmalar yürütülmektedir. Artan nüfus yoğunluğu ve buna bağlı olarak gelişen sağlık problemleri yapay zeka teknolojileri kullanılarak insan bazı hataları minimize edip, maliyetten tasarruf sağlayarak kalitenin ve performansın artırılması sağlanmaktadır. Buna örnek olarak ise geçtiğimiz pandemi sürecinde birçok alanda yapay zeka uygulamalarına başvurularak iş yükü azaltılmış ve öngörülemeyen olaylar daha kolaylıkla tahminlenebilmiştir (URL-5).

Türkiye’de sağlık sektöründe yapay zeka uygulamaları için Microsoft, Oracle gibi sektörde önde isim olan teknoloji firmalarının kullanılmaktadır. Sağlık Bakanlığı bu firmaların ürünlerini kullanarak, bu ürünlerin sahip oldukları yapay zekâ yöntemleri ile bireylerin MHRS (Merkezi Hekim Randevu Sistemi) kullanım oranları, aile hekimliği performans raporları, ameliyat, tanı ve hastaneye yatış raporları gibi eNabız değerlendirme raporlarını oluşturmaktadır (Akalin, 2021).

İngiltere’de covid-19 için Kapasite Planlama ve analiz Sistemi (Capacity Planning and Analysis System (CPAS)) geliştirilmiş ve çoğu sağlık kuruluşlarında kullanılmaktadır (Qian ve diğ., 2021).

Yapay zeka yöntemleri kullanılarak bireylerin yaşamlarını sağlıklı bir şekilde sürdürebilmeleri için çeşitli uygulamalar geliştirilmektedir. Bu uygulamalar ile bireylerin günlük yedikleri yiyeceklere göre diyet planlaması, atılana adımlara göre yakılan kalori miktarları ve analizi, yapılan sportif faaliyetlere göre sağlıksal analiz raporları çıkarılabilmektedir. Yaygın olarak kullanılan akıllı saatler ve buna benzer kişilerin yanlarında kolaylıkla yanlarında taşıyabileceği cihazlar ile bu mümkün kılınmaktadır. Kullanılan teknolojiler ile kişilerin farkındalıkları arttırılarak sağlıklı bir yaşam sürmeleri sağlanmaktadır.

2.2. Hava Kalitesi Tahmini

Hava Kalitesi İndeksi HKİ (Air Quality Index - AQI), belirli bir bölgedeki hava kalitesinin sağlık durumuna etkisini tahmin etmektedir. Temel olarak havada bulunan Partikül madde (PM_{2.5} ve PM₁₀), Azot Dioksit (NO₂), Kükürt Dioksit (SO₂), Ozon (O₃) ve Karbon Monoksit (CO) emisyonlarının ölçümüne dayanmaktadır. Bu sınıflandırma sistemi çoğu ülkede yaygın olarak kullanılmakla birlikte havadaki

kirleticilerin konsantrasyonlarına göre iyi, orta, kötü ve tehlikeli olarak derecelendirilmektedir. Ölçümler sonucu gösterge değerlerinin yüksekliği bireylerin sağlık durumlarını olumsuz yönde etkilediği ve ciddi sağlık problemlerine neden olacağı için bu değerlere dikkat edilmesi son derece önem ifade etmektedir. Şekil 2.2’de Ulusal hava kalite indeksine ait kesme noktaları verilmiştir (URL-6).

İndeks	HKİ	SO ₂ [µg/m ³]	NO ₂ [µg/m ³]	CO [µg/m ³]	O ₃ [µg/m ³]	PM10 [µg/m ³]
		1 Sa. Ort.	1 Sa. Ort.	8 Sa. Ort.	8 Sa. Ort.	24 Sa. Ort.
İyi	0 – 50	0-100	0-100	0-5500	0-120	0-50
Orta	51 – 100	101-250	101-200	5501-10000	121-160	51-100
Hassas	101 – 150	251-500	201-500	10001-16000	161-180	101-260
Sağlıksız	151 – 200	501-850	501-1000	16001-24000	181-240	261-400
Kötü	201 – 300	851-1100	1001-2000	24001-32000	241-700	401-520
Tehlikeli	301 – 500	>1101	>2001	>32001	>701	>521

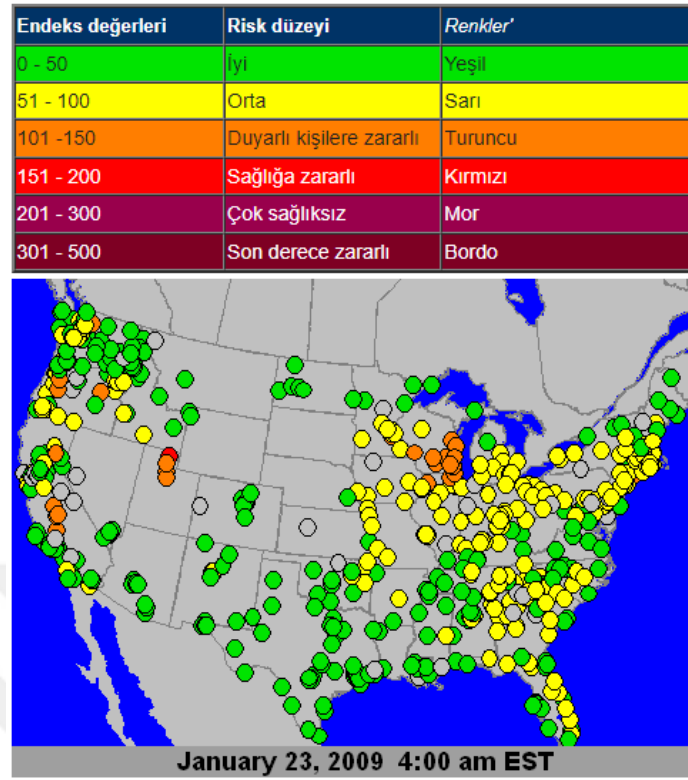
Şekil 2.2. Ulusal Hava Kalitesi İndeksi Kesme Noktaları

Örnek olarak Amerika Birleşik Devletleri’ne ait Şekil 2.3’te hava kalitesi indeks tahmin tablosu verilmiştir (URL-7).

Hava kalitesi tahmin başarısının artırılması için yapay zeka uygulamalarına da başvurulmaktadır. Alınan veriler makine öğrenmesi algoritmalarına koyularak kalite tahminin doğruluk oranını arttırmaya çalışılmaktadır.

Şehirlerde meydana gelen hava kirliliğinin önlenmesi için çeşitli noktalarda bulunan sensör ve cihazlar ile hava kirliliğinin ölçülmesi ve bu sonuçlara göre gerekli tedbirlerin alınması sağlanmaktadır. Alınan bu veriler ile hava kalitesinin tahmin başarısının artırılması için matematiksel modeller kullanıldığı gibi yapay zeka uygulamaları da kullanılmaktadır.

Veljanovska ve Dimoski (2018) tarafından yapılan çalışma ile Üsküp şehrinin hava kalite indeksini belirlemede karar ağacı, en yakın komşu, yapay sinir ağları ve destek vektör makineleri sınıflandırma algoritmalarını kullanarak elde ettiği başarı sonuçlarını kıyaslamışlardır (Veljanovska ve Dimoski, 2018).



řekil 2.3. ABD Hava Kalitesi İndeks Tahmin Tablosu

Adams ve ark. (2013) tarafından yapılan  alıřmada Kanada Ontaryo g l   evresinde havada bulunan Partik l Maddeler PM_{2.5} ve NO₂ deęerlerini yapay sinir aęları ile hesaplayarak ve  ıkan sonu ları geleneksel y ntemle karřılařtırmıřlardır (Adams ve dię., 2013).

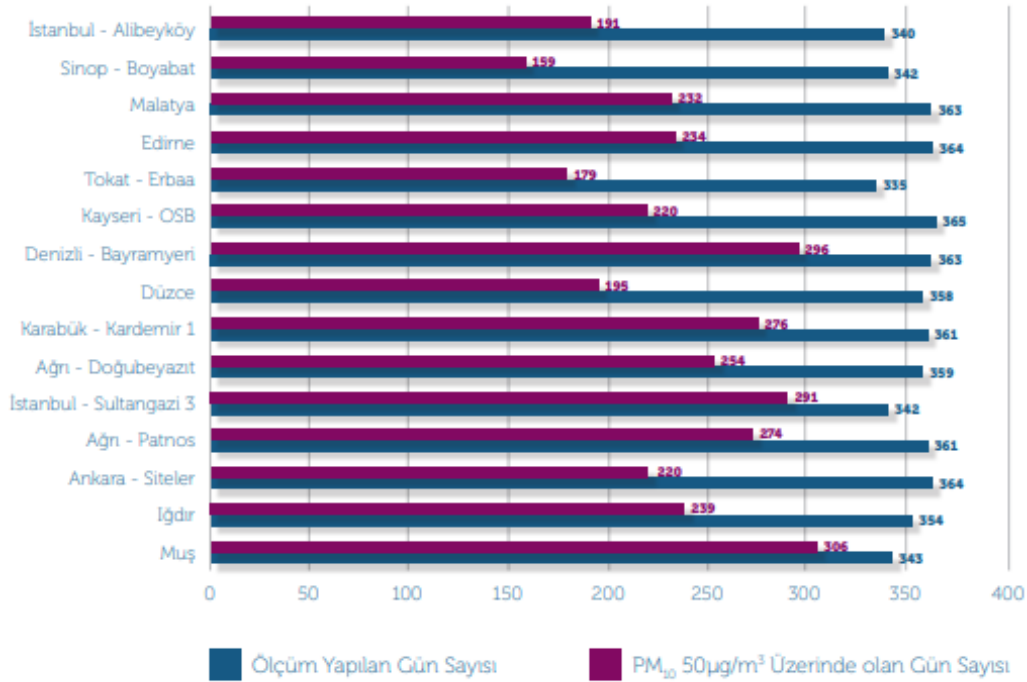
Ghaemi, Alimohammadi ve Farnaghi (2018) yaptıkları  alıřmalarda, LaSVM tabanlı online bir algoritma ile hava kirlilięi tahmini yapılmıřtır. Yapılan  alıřmada coęrafi veriler ve parametrelerle birlikte kirletici konsantrasyonu ile meteorolojik veriler kullanılmıřtır.  alıřma sonucunda ise hava kalite indeksi sisteminin performansı sık kullanılan geleneksel algoritma sonu ları ile yapılan kıyaslama sonucunda  onerilen modelin zamansal ve mek nsal hava kirlilięi tahmininde kullanılmasının yararları ortaya konulmuřtur (Ghaemi ve dię., 2018).

Wang ve Yang (2020), geliřtirdikleri bir yapay sinir aęı modeli geliřtirerek b y k veri kullanımında tahmin doęruluęunu ama lamıřtır. Yapılan  alıřma ile hava kirlilięi  zerine etkisi bulunan temel altı kirleticinin verisi alınarak yapay sinir aęına girdi olarak verilmiřtir. Geliřtirilen modelin b y k veri kullanımı ve yapay sinir aęlarının  ęrenme

özelliği sayesinde hava kalitesindeki gelişimleri ile ortaya çıkarabilecek bir olası ve yüksek doğruluk oranına sahip olduğu görülmektedir (Wang ve Yang, 2020).

2.3. Çevresel Veriler Arası Analitik

Gün içerisinde insanların yaşadıkları alanlarda maruz kaldığı birçok çevresel etken ve olaylar meydana gelmektedir. Bunlar gerek iklim değişikliğinden meydana gelen kuraklık, sıcak hava, aşırı nem gibi hava durumu olayları, gerekse ne kadar görünmese de çeşitli fabrika, sanayi tesisleri, orman yangınlarında, taşıtlarda ve kentlerde ortaya çıkan zararlı kimyasal gazların havada bulunmasıdır. Kronik hastalığı bulunan bireyler öncelikli olarak ve diğer bireyler in sağlığına olumsuz etkisi oldukça yüksektir. Özellikle bölgesel olarak değişen hava olayları ve hava kalitesi değerleri bireylerin üzerindeki etkisi kişisel olarak farklılık göstermektedir. Çevresel analitik çalışmalarına örnek olarak, Türkiye’de ağır sanayi ve sanayi bölgesinde yaşayan bireylerin maruz kaldığı hava kirliliği ile sanayinin olmadığı kesimlerde yaşayan bireylerin ise maruz kaldığı kirlilik aynı olmamaktadır. Şekil 2.4’te yer alan grafikte 2020 yılı boyunca ülkemiz bazı illerinden hava kirliliği durumuna göre en yüksek on beş istasyon bilgisi verilmektedir (URL-8).



Şekil 2.4 2020 Yılı Boyunca Havası En Kirli 15 İstasyonlar

Şekil 2.5'te ise örnek olarak Türkiye haritası üzerinde hava kalitesi yani PM10 değerleri durumunun Dünya Sağlık Örgütü (DSÖ) limit bilgisine göre değerleri verilmiştir. Yeterli ölçüm yapılan il sayısı 51, yeterli ölçüm yapılan istasyon sayısı 152 ve Dünya Sağlık Örgütü sınır değeri (PM10): 20 $\mu\text{g}/\text{m}^3$ olarak değerlendirilmiştir (URL-8).



Şekil 2.5. 2020 Bazında Hava Kalitesi Durumu (PM10)

3. METARYEL VE METOTLAR

Birçok alanda kullanılmaya başlanan makine öğrenmesi, yapay zekanın alt çalışma alanı olmakla birlikte karmaşık ve anlamsız verilerin doğru, hızlı ve akıcı bir şekilde karar vermeyi sağlamaktadır. Gelişmiş görevlerde insan için zor ve zaman alıcı olan olayların makine öğrenmesi algoritmaları yöntemleri ile kolaylık sağlanmaktadır (URL-9). Makine öğrenmesi yöntemleri 1980 yıllarında ortaya çıkarak teknolojinin gelişmesinde önemli bir yer edinmiş ve birçok sık rastlanan problemlerin çözülmesine rol oynamıştır. Günümüzde de sık kullanılan makine öğrenmesi temel olarak iki amaca hitap etmektedir. Bunlardan ilki, verileri geliştirilen modellere göre sınıflandırmak iken ikincisi ise çalışmamızın ağırlıklı ana konusu olan bu sınıflandırma modellerine dayalı gelecekteki sonuçlar için tahminler yapmaktır.

Makine öğrenmesi birçok alanda kullanılmaktadır bunlar; mühendislik, sağlık, çevre, finans, sosyal medya, eğitim, enerji gibi tahmin, kümeleme ve sınıflandırma alanlarında özetlenebilir (URL-10). Genel olarak, veri seti oluşturulması, özellik belirleme ve kullanılacak algoritma seçimi, eğitimin gerçekleştirilmesi, son aşama olarak da eğitim verisinden farklı olarak oluşturulan verilerle testin yapılması ile makine öğrenmesi gerçekleştirilir.

Hava kirliliği problemi her ülke ve vatandaşları için bir tehdit niteliği taşımakla birlikte geliştirilen yöntemler ile bir nebze olsa bile bu tehditlerden korunulabilir. Hava kirliliği tahmini yapılırken çalışmaların yapılacağı bölgelerin temel coğrafi, sosyoekonomik, demografik özellikleri de belli oranda etki etmektedir. Çalışmamızda ele alınan bölgelerin hakkında tek tek bilgi verilerek genel bilgi edinilecektir. Havadaki kirliliğin tahmininin makine öğrenmesi yöntemlerince yapıldığında yararlılığı açıkça görülmektedir. Yazacağımız bu tez ile Asya bölgesinde bulunan üç ülke, Brunei, Singapur ve Tayland referans alınarak günlük hava durumu ve hava kalitesi verilerine göre bölgelerin hava durumu ve hava kalitesi tahminleri makine öğrenmesi yöntemleri kullanılarak PM10 ve PM2.5 değerlerinin bireyler üzerinde etkisi gözlemlenmiş olacaktır. Hata oranı en az olan yöntem belirlenerek bölgelerdeki kişilerin bireysel yaşamları için önemli ölçüde fayda sağlanması planlanmaktadır. Aynı zamanda dinamik bir görselleştirme araçlarından yararlanılarak çalışmada elde edilecek tahminleme sonuçlarının daha anlaşılır hale getirilmesi sağlanacaktır.

3.1. Çalışma Alanı ve İklimsel Özellikleri

Çalışma kapsamında Brunei, Singapur ve Tayland bölgeleri için kısa bilgilendirme yapılarak hava kirliliği konusunda etkili olan bazı durumlar hakkında ön bilgi edinilmiştir.

3.1.1. Brunei

Brunei, güneydoğu Asya'da yer alan küçük bir ülkedir. Borneo adasının kuzeyinde, Malay Yarımadası'nın güneybatısında yer almaktadır. Toplam yüzölçümü yaklaşık 5.765 km²'dir. Ülke, güneydoğu Asya'daki en küçük ülkelerden biridir ve nüfusu yaklaşık 450.000'dir (URL-11).

Brunei'nin coğrafyası, engebeli ormanlık arazilerden oluşmaktadır. En yüksek noktası, ülkenin batısında yer alan Bukit Pagon Dağı'dır. Ülkenin doğusunda yer alan Ulu Temburong Milli Parkı, Brunei'nin biyolojik çeşitliliğini korumak için tasarlanmıştır.

Sosyoekonomik olarak, Brunei, dünyanın en zengin ülkelerinden biridir. Petrol ve doğal gaz endüstrileri ülke ekonomisinin belkemiğini oluşturmaktadır. Brunei, Birleşmiş Milletler İnsani Gelişme Endeksi'nde "çok yüksek insani gelişme" kategorisinde yer almaktadır.

Brunei'nin resmi dili Malayca'dır ve İslam ülkenin resmi dini olarak kabul edilmektedir. Brunei Sultanlığı, mutlak monarşi ile yönetilmektedir ve siyasi partiler yasaklanmıştır.

Brunei'nin iklimi, tropikal muson iklimi etkilidir. Bu iklim türü, yıl boyunca yüksek sıcaklık ve nemle birlikte bol miktarda yağmur getirir. Brunei'nin iklimi, bitki örtüsü ve biyolojik çeşitlilik üzerinde önemli bir etkiye sahiptir. Yağışlar, ülkenin doğal ekosistemlerinin zenginliğini ve yeşilliklerini korumasına yardımcı olur. Muson mevsimi, Nisan ile Ocak ayları arasında uzun yağmur dönemini temsil ederken, Şubat ile Mart ayları arasındaki kısa muson dönemi ise daha az yağış getirir. Sıcaklıklar yıl boyunca nispeten sabit kalır ve genellikle yüksek sıcaklıklara sahiptir, özellikle Mart ve Eylül ayları arasında en yüksek sıcaklıklar gözlenir. Brunei'nin iklimi, bitki örtüsü ve biyolojik çeşitlilik üzerinde önemli bir etkiye sahiptir. Yağışlar, ülkenin doğal ekosistemlerinin zenginliğini ve yeşilliklerini korumasına yardımcı olur.

3.1.2. Singapur

Singapur, Güneydoğu Asya'da yer alan bir ada ülkesidir. Yüzölçümü sadece 728,3 km²'dir ve nüfusu yaklaşık 5.4 milyondur. Coğrafi olarak, Singapur Malezya'nın güneyinde ve Endonezya'nın Riau Adaları'nın kuzeyinde yer almaktadır (URL-12).

Singapur, dünya genelinde en yoğun nüfusa sahip ülkelerden biridir. Bu nedenle, ülke oldukça gelişmiş bir ekonomiye sahiptir ve dünya çapında bir finans merkezi olarak kabul edilir. Ülke ayrıca hızlı bir endüstriyel gelişim yaşamıştır ve dünya genelinde önemli bir nakliye merkezidir.

Singapur'un iklimi, tropikal iklimler arasında yer alır ve yıl boyunca sıcak ve nemlidir. Yağışların büyük çoğunluğu muson sezonunda (Ekim-Nisan) düşer. Bu nedenle, Singapur'da hava sıcaklıkları genellikle 24 ila 32 derece Celsius arasında değişir ve nem oranı yüksektir.

3.1.3. Tayland

Tayland da Güneydoğu Asya'da yer alan bir ülkedir. Coğrafi olarak Myanmar, Laos, Kamboçya ve Malezya ile komşudur. Tayland, yaklaşık 513.000 km²'lik bir alan kaplar ve yaklaşık 69 milyonluk bir nüfusa sahiptir (URL-13).

Temel coğrafi özellikleri arasında, ülkenin kuzeyinde dağlık bölgelerin bulunduğu, güneyde ise düzlüklerin ve kıyı şeridinin yer aldığı söylenebilir. Ülke aynı zamanda birçok nehir ve gölü de barındırır.

Sosyo-ekonomik olarak, Tayland Orta gelirli bir ülkedir. Ülkenin ekonomisi, turizm, tarım, sanayi ve hizmet sektörleri ile çeşitlendirilmiştir. Ayrıca, son yıllarda bilgi ve iletişim teknolojileri gibi yüksek katma değerli sektörlerle de yatırım yapılmaktadır. Nüfusun çoğunluğu tarım sektöründe çalışırken, turizm de önemli bir istihdam kaynağıdır.

Tayland'ın iklimi, tropikal bir iklime sahiptir. Muson yağmurları, Mayıs ve Ekim ayları arasında etkilidir. Ayrıca, ülke Güneydoğu Asya'nın diğer bölgeleri gibi tayfunlar ve sellerle de mücadele etmek zorundadır.

3.2. Veri Seti Temini

Tez çalışmasında kullanılacak veri seti, yarışma kapsamında tedarik edilen ve kullanılan verilerden yararlanılmıştır. Temin edilen veri setinde Singapur, Tayland ve Brunei ülkelerinde bulunan bölgelerden alınan hava ölçüm rapor sonuçları bulunmaktadır. Kullanılacak hava tahmin verileri 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei, Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen PM10 konsantrasyonlarının günlük okumalarını içermektedir. Veriler ayrıca günlük (ve bazı durumlarda saatlik) sıcaklık, yağış, nem ve rüzgâr hızı/yönü değerlerinin atmosferik parametrelerini içerir. Ancak, yapılan yarışma kapsamından dolayı verilerde bazı eksik değerlerle birlikte tüm parametreler her istasyonda/ülkede mevcut değildir.

Bu veriler Neural Network, XGBoost, Random Forest, AdaBoost, Gradient Boosting, Light Gradient Boosting Machine, Support Vector Regression gibi makine öğrenimi modelleri ile eğitilerek yaklaşımlar incelenecektir. Model başarısı ve çıktıları sonrası elde edilen sonuçlar incelenerek karşılaştırma yapılacaktır. Eğitim ve analizler neticesinde çıkan sonuçlar çeşitli grafikler ile görselleştirilip yorumlanacaktır. Bu şekilde bölge ve bölge halkının hava kalitesi hakkında çıkarımlar elde edilmiş olacaktır.

3.3. Makine Öğrenmesi Yöntemleri

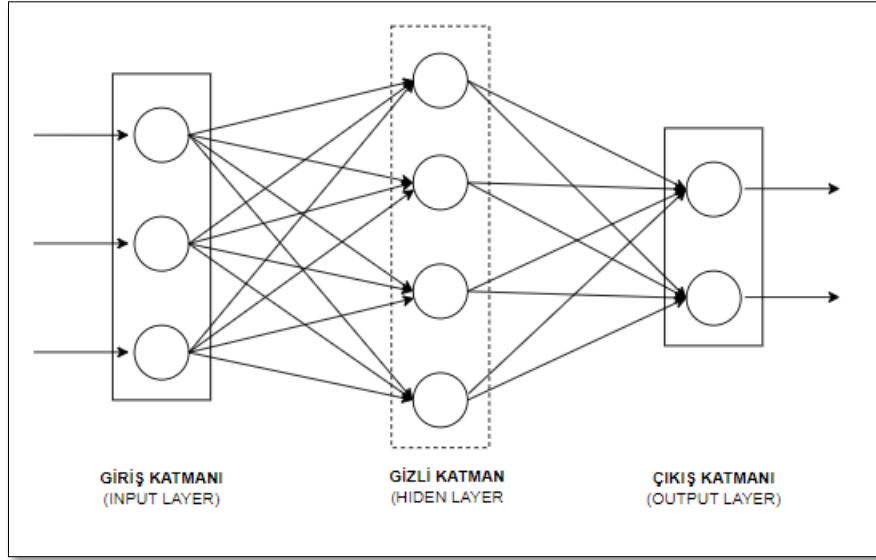
Tez çalışmamızda Neural Network, XGBoost, Random Forest, AdaBoost, Gradient Boosting, Light Gradient Boosting Machine, Support Vector Regression gibi özellikle başarı oranı yüksek olan ve en çok tercih edilen makine öğrenmesi tekniklerini aşağıda açıklanmıştır.

3.3.1. Neural Network (NN)

Neural Network (Sinir Ağı), bilgisayar sistemlerine insan beyninin özelliklerinden esinlenilerek ve öğrenme yoluyla işlemeyi öğreten bir makine öğrenim yöntemidir. İnsan beyninin temel birimi olan nöronlardan ilham alınarak modellenmiştir, yapay sinir ağının yapı taşı basit sinyal işlemeyi gerçekleştiren bir algılayıcıdır ve birbirine bağlanmış ağ yapısı şeklindedir. İnsan beyninin yetenekleri olan öğrenme, hatırlama ve

genelleme gibi biyolojik sinir ağılarını taklit ederek oluşmuştur. Yapay sinir hücrelerinin birbiri ile bağlanması sonucu yapay sinir ağıları oluşmaktadır. Sinir ağıları öğrenme işlemler sırasında örnekler kullanılarak gerçekleştirilmektedir. Yapay sinir ağıları giriş (input) katmanı, ara (gizli) katmanlar ve çıkış (output) katmanı olarak üç ana katmanda incelenir ve karar verilirken bazı kurallara tabi tutulmaktadır (McCulloch ve Pitts, 1943).

Şekil 3.1’de görüldüğü gibi, bilgiler ağı girdi katmandan iletilmektedir. Ardından bilgiler ara katmanlarda işlemlere tabi tutulduktan sonra çıktı katmanına gönderilmektedir. Buradaki tabi tutulan işlem ile gelen bilgiler ağıın ağırlık değerleri kullanılarak çıktılara dönüştürmesidir. Oluşturulan ağıın girdileri için doğru ve tutarlı çıktılar üretilmesi için ağırlıkların doğru değerlerinin olması gerekmektedir. Sinir ağıının temel olarak mimarisi ve öğrenme kuralları, kullanılacak olan öğrenme stratejisine göre belirlenmektedir. Danışmanlı (Supervise), Danışmansız (Unsupervised), Pekiştirmeli/ Destekleyici (Reinforcement) öğrenme ve Karma (Hybrid) öğrenme olarak kendi içinde gruplanmaktadır.



Şekil 3.1. Sinir Ağı Yapısı

3.3.2. Random Forest (RF)

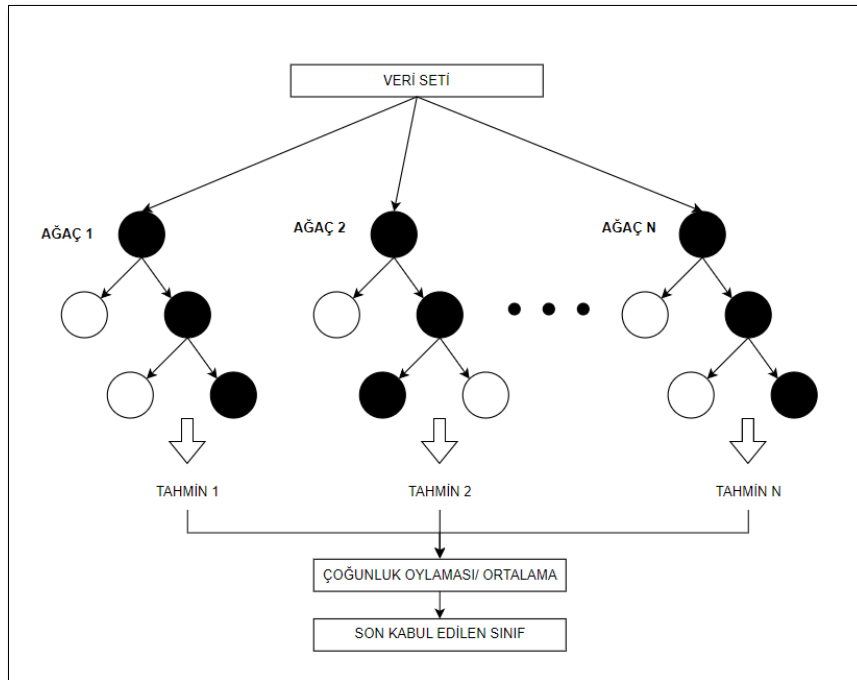
Makine öğrenmesi algoritmalarından Random Forest (Rastgele Orman) algoritması, birden çok karar ağacı üzerinden her bir karar ağacını farklı bir gözlem örneği üzerinde eğitmek sureti ile farklı modeller oluşturularak bu algoritma ile hem sınıflandırma hem

de regresyon maksatlı analizler yapılabilir. Ancak sıklıkla sınıflandırma ile kullanılmaktadır. RF ile son adıma gelindiğinde tahmin sonucu en başarılı olan çözüm olarak seçilir. Farklı karar ağaçlarının üzerinden değerlendirme yapılan ve sonuçların ortalaması alınan bu algoritma sayesinde aşırı öğrenmenin önüne geçilmesi ve böylelikle RF başarısının yüksek olması sağlanmaktadır.

RF algoritmasının sağladığı kullanım kolaylığı, esnekliği sayesinde sınıflandırma ve regresyon çalışmalarında kullanılması yaygınlaşmıştır.

Şekil 3.2’de görsel baz alınarak kısaca RF algoritmasının temel adımları şu şekildedir (Biau ve Scornet, 2016).

- İşlem yapılacak veya çıkarım elde edilecek veri setinin hazırlığı yapılır, bu aşamada gerekli olması halinde veri, veri ayıklama işlemlerine tabi tutulur.
- Seçilecek olan her bir örnek için yeni karar ağacı oluşturulur, oluşturulan karar ağaçlarından tahmin sonuçları elde edilmiş olunur.
- İşlem sonrası oluşan tahmin sonuçlarından yapılacak çalışma sınıflandırma ise mod, regresyon ise ortalama işlemi yapılır.
- Son olarak ise en çok oylama yapılan işlem tahmin sonucu olarak seçilir.

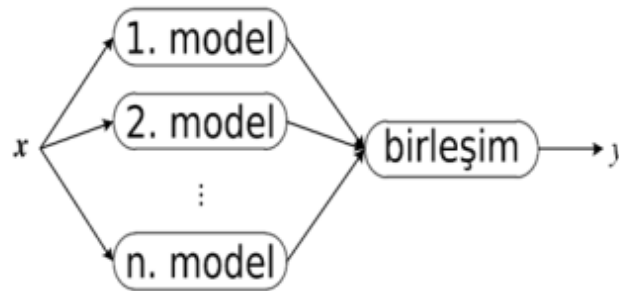


Şekil 3.2. Random Forest Algoritma Yapısı

3.3.3. Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) algoritmasından bahsetmeden önce Boosting (Arttırma) metodundan bahsetmek gerekirse: Boosting, işlem yapılacak çalışmada sınıflar dengesiz olduğu durumda zayıf sınıflandırma performansını artırabilmeyi sağlayan topluluk tabanlı bir öğrenme metodudur. Boosting metodu tahmin edicileri kümülatif olarak eğitmekte ve tahminleyici modellerinde ise genellikle karar ağaçları kullanmaktadır. En çok kullanılan modelleri ise; AdaBoost, Gradient Boosting, XGBoost, LightGBM ve CatBoost'tur. Yapıtığımız çalışmada da AdaBoost, Gradient Boosting, XGBoost, LightGBM metotları kullanılmıştır.

AdaBoost yöntemi, Yoav Freund ve Robert Schapire tarafından 1997 yılında önerilmiş ve formüle edilmiş bir makine öğrenmesi algoritmasıdır (Freund ve Schapire, 1997). AdaBoost, boosting algoritmasının ilk algoritması olarak bilinmekte ve bilişim dünyasında değerli ödüllerden biri olan Gödel ödülünü almıştır. AdaBoost algoritması "esemble learnin" yani topluluk algoritması yöntemi olarak da bilinmektedir, aynı zamanda ilk boosting metodunun pratik uygulamasıdır (Zhou, 2012). Çözülecek problem için eğitim verisinden oluşturulacak tek bir model ile var olan öğrenme yöntemlerinden farklı olarak, problemi çözmek ve birçok modelin eğitilmesi topluluk yöntemlerine dayanmaktadır. Şekil 3.3'te görüldüğü üzere topluluk yöntemleri ile problemler için eğitim verisi kullanılarak birçok model oluşturulur ve oluşturulan modellerin sonuçları sınıflandırma ve regresyon amacıyla birleştirilmektedir (Zhou, 2012). Adaboost yöntemi ile geleneksel olarak zayıf sınıflandırıcıların yanlış olarak oluşturduğu sınıflandırmaya odaklanarak, yanlış şekilde sınıflandırılan eğitim örneklerine bakarak daha iyi bir şekilde sınıflandırma oluşturulmuştur. Bu şekilde her sınıflandırıcı yönteminin oluşturulan bakış açısına katkısı sağlamaktadır.



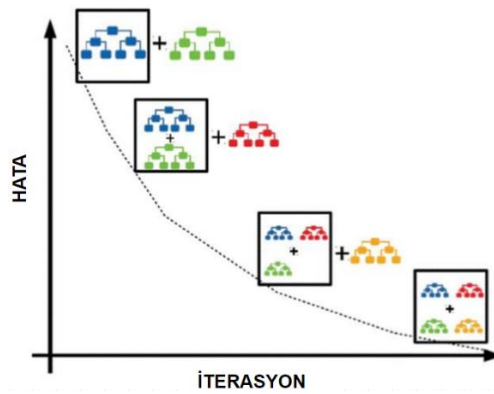
Şekil 3.3. AdaBoost Topluluk Öğrenme Yöntemi

AdaBoost, yüksek düzeylerde yaptığı son sınıflandırmaya karar verme aşamasında RF algoritmasına benzemektedir. AdaBoost ile RF arasında küçük farklılıklar söz konusudur, örneğin karar ağaçlarının derinliği 1'dir yani 2 yaprak. Her karar ağacının yaptığı tahminler, model tarafından yapılan son tahmin üzerinde farklı etkilere sahip olmaktadır.

3.3.4. Gradient Boosting (GB)

Gradient Boosting (GB) algoritmasının temelinde karar ağacı yatmaktadır. Güçlü bir makine öğrenmesi tekniği olan GB, Friedman tarafından 2001 yılında tanıtılmıştır. Sınıflandırma ve regresyon problemleri için bir makine öğrenme tekniği olarak tanımlanmakta ve tahminleme için yaygın kullanılan yöntemlerden bir tanesidir. Hata oranını belli tekrarlar ile minimuma indirmesiyle yapay sinir ağına benzerlik göstermektedir. Açıklanabilirliğinin kolay olması GB'in yapay sinir ağlarına sağladığı en büyük avantajlardan birisi olmaktadır.

GB algoritmasında ana prensip ise oluşturulan her yeni karar ağacının bir önceki ağaçta hesaplanan hataları en aza indirmesine dayanmaktadır. Kök noktasından başlanılarak veriler yorumlanır ve aşağıya doğru dallanma sağlanır bu sayede yapraklar oluşturulmuş olunur. İlk oluşturulan karar ağacı ile bir tahmin elde edilir ve elde edilen tahmin ile hedef arasındaki fark hesaplanır. Oluşturulan her iterasyonda, hesaplanan fark ile yeni bir ağaç yapısı oluşturulmuş olunur. Son olarak ise tahmin ile hedef arasındaki farkı sıfırlamak amaçlanmış olunur (Marksfels, 2018). Şekil 3.4'te de GB'i algoritma yapısını daha iyi anlatan görsel çalışması yer almaktadır ve iterasyon ilerledikçe hatanın nasıl azaldığını göstermektedir.



Şekil 3.4. Gradient Boosting Algoritması

3.3.5. XGBoost (XGB)

XGBoost (XGB), Karar Ağacı (Decision Tree) temelli ve aynı zamanda Eğim Arttırmalı (Gradient Boosting) bir makine öğrenmesi sistemidir. İlk olarak 2016 yılında Washington Üniversitesi'nde iki araştırmacı Tianqi Chen ve Carlos Guestrin tarafından yayımlanmış yenilikçi bir makine öğrenme algoritmasıdır (Chen ve Guestrin 2016). Yayımlanan makalenin ardından sadece akademik alanda kalmayıp Kaggle yarışmalarının da gözdesi haline gelmiştir.

XGBoost kullanımının yaygınlaşmasının en önemli özellikleri aşırı öğrenmenin önünü kapatması, boş olan verilerin yönetilebilmesi, yüksek tahmin gücü ve tüm bunlarla birlikte hızlı çalışmasıdır. Aynı zamanda işlem sırasında oluşturduğu ağaç yapısı ile maksimum derinlik değeri kullanılmaktadır. Avantajlarına bakıldığında GB algoritmasının optimize edilmiş bir türü olarak görülmektedir. İki algoritma kıyaslanacak olursa GB kayıp fonksiyonluların hesaplanmasında birinci dereceden fonksiyon kullanırken XGB ikinci dereceden fonksiyon kullanarak bu hesaplamayı gerçekleştirmektedir. XGB bu sayede ve paralel çalışma özelliği ile diğer algoritmalara göre de sonuca daha kısa sürede ulaşılmasını sağlamaktadır. Şekil 3.5'te görsel anlatım ile algoritmanın avantajları gösterilmektedir (Morde, 2023).



Şekil 3.5. XGBoost'un Standart GBM'yi Optimize Yolları

XGB, arama işlemi sırasında başladığı düğümden gidebileceği en derin düğüme ulaşmaya çalışmaktadır. En sona kadar ilerleme tamamlandıktan sonra geriye doğru sarar ve derin olanlara öncelik verilerek işlemini sürdürülmeye devam ettirmiş olur. İşlemler sırasında donanım kaynaklarının iyi kullanılması için her iş yapısı kendi

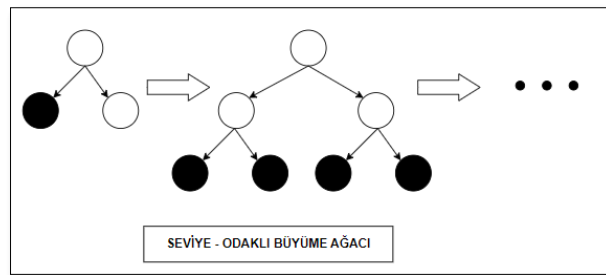
içerisinde bir tampon oluşturmakta ve bu tamponlarda eğitim istatistiklerini tutarak önbellemenin doluluğunu göz önünde bulundurmaktadır.

3.3.6. Light Gradient Boosting Machine (LightGBM)

LightGBM, 2017 yılında Microsoft DMTK (Distributed Machine Learning Toolkit) projesi kapsamında geliştirilmiş boosting algoritmasıdır. Tanıtıldığı zaman “LightGBM: A Highly Efficient Gradient Boosting Decision Tree” makalesine göre yapılan çalışmada LightGBM’in diğer modellere göre 20 kat daha hızlı olduğu görülmektedir (Ke ve diğ., 2017). Boosting algoritmalarından farklı olarak LightGBM’in avantajları ise; büyük olan verileri işleyebilmesi, yüksek işlem hızı, paralel öğrenme, yüksek tahmin oranı, daha az kaynak tüketimi (RAM) ve GPU öğrenimini desteklemesidir. Ayrıca LightGBM diğer metotlardan farklı olarak değişkenlerin sayısı ve veri örnekleri ile ilgilenen Gradyan Tabanlı Tek Yönlü ve Özel Değişken Paketi’dir.

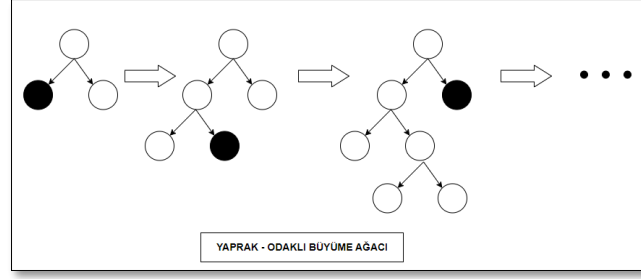
LightGBM, sürekli (continuous) değerlere sahip olan değişkenleri kesikli (discrete) hale getirerek hesaplama maliyetini azaltmaktadır. Aynı zamanda histogram tabanlı bir algoritmadır. Gerçekleştirilen çalışmalarda karar ağaçlarının eğitim süreleri yapılan hesaplama ve bölünme sayısı ile doğru orantılı olmaktadır, LightGBM sayesinde hem eğitim süresinin hem de kaynak kullanımı azaltılmaktadır (Ke ve diğ., 2017).

Seviye odaklı (level-wise or depth-wise) veya yaprak odaklı (leaf-wise) olarak iki strateji karar ağaçları öğreniminde kullanılmaktadır.



Şekil 3.6. Seviye Odaklı Büyüme

Seviye odaklı stratejide ağaç yapısı Şekil 3.6’da görüldüğü üzere büyürken ağacın dengesi korunmaktadır. Bir diğer strateji olan yaprak odaklıda ise, kaybı azaltan yapraklardan bölünme işlemi devam etmektedir. Her iki yöntem ile LightGBM algoritması diğer boosting algoritmalarından ayrılmaktadır.



Şekil 3.7. Yaprak Odaklı Büyüme

Yaprak odaklı strateji daha hızlı olup daha az hata oranına sahiptir ancak veri sayısının az olduğu durumlarda modelin aşırı öğrenmeye (over fitting) yatkın olmasına neden olmaktadır, bu durum Şekil 3.7’de de görselleştirilmiştir. Aşırı öğrenmeden dolayı büyük verilerde kullanılması için daha uygun olarak görülmektedir. Aşırı öğrenme, yaprak sayısı ve ağaç derinliği gibi parametrelerin optimizasyonu ile engellenebilir.

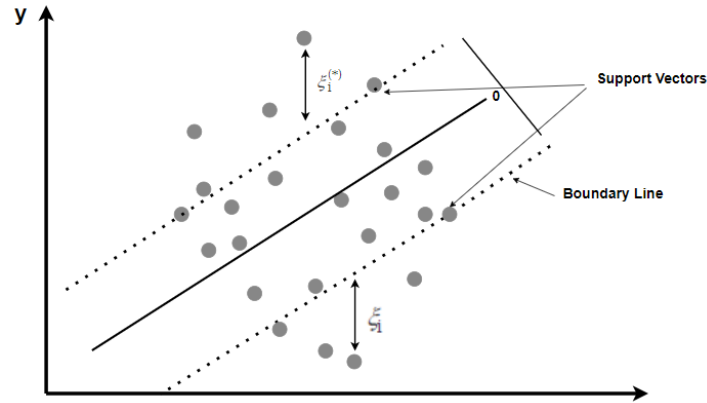
3.3.7. Support Vector Regression (SVR)

Support Vector Machine (Destek Vektör Makineleri) sınıflandırma ve regresyon problemlerinde kullanılmasına rağmen genel olarak sınıflandırma problemlerinde tercih edilen gözetimli makine öğrenme yöntemlerinden biridir. SVM örüntü tanıma ve sınıflandırma problemlerinin çözümü için Vapnik tarafından geliştirilmiştir (Cortes ve Vapnik, 1995). Tasarlanan SVM’lerin çıkarılan ilk versiyonları, ikili veri kümesini doğrusal vektörler yardımıyla sınıflara ayırmak için kullanılan güçlü sınıflandırıcılardır. Ardından sonraki zamanlarda geliştirilen SVM modelleri ise genel olarak algoritmanın çalışmasında çekirdek fonksiyonlarından yararlanarak iç içe geçmiş şekilde birden fazla sınıfı da zahmetsiz olarak ayırabilmektedir. Bu yöntemden dolayı birçok makine öğrenmesi çalışmalarında SVM tercih sebebi olmuştur. Aynı zamanda hem doğrusal hem de doğrusal olmayan sınıflandırma işlemlerini gerçekleştirebilirler. Yapılan işlemlerden sınıflandırma işlemi için, tam olarak ayrışılabilir veriler kullanılması halinde bütün veriler bir hiper düzlem ile sınıflandırılabilir. Ancak verilerde tam ayrıştırılamayan veri olması halinde, genellikle aynı boyutlarda tek düzlem ile sınıflandırmaktadır, bu yüzden farklı çekirdek fonksiyonları kullanmaya yönelinmektedir (Metlek ve Kayaalp, 2020).

Sınıflandırmada olduğu gibi, destek vektör regresyonu (SVR), çalışma sırasında çekirdeklerin kullanımı, seyrek çözüm ve marjın vektör kontrolü ve destek vektörlerinin

sayısı ile karakterize edilmektedir. SVR'nin SVM'den daha az popüler olmasına rağmen, gerçek değer fonksiyonu tahmininde etkili bir araç olduğu kanıtlanmıştır. SVR denetimli öğrenme yaklaşımı olarak düşük ve yüksek yanlış tahminleri eşit şekilde cezalandıran simetrik bir kayıp fonksiyonu kullanarak eğitmektedir. Vapnik'in e-duyarsız yaklaşımı kullanılarak, tahmini fonksiyonun etrafında simetrik olarak minimum yarıçap ile esnek bir alan oluşturulur, öyle ki belirli bir e eşliğinden daha düşük hataların mutlak değerleri alınarak, tahminin hem üstünde bulunan hem de altında bulunanlar göz ardı edilmektedir. Bu şekilde, hat dışındaki noktalar görmezden gelinir, ancak oluşturulan hat içindeki, fonksiyonun üstündeki veya altındaki noktalar görmezden gelinmez. SVR'nin en önemli avantajlarından biri ise, hesaplama sırasında karmaşıklığının girdi uzayının boyutsallarına bağlı olmamasıdır. Buna ek olarak, yüksek tahmin doğruluğu ve başarısı ile mükemmel bir genelleme yeteneğine sahiptir.

SVM algoritması, çalışma sırasında veriye ilişkin herhangi bir birleşik dağılım fonksiyonları bilgisine ihtiyaç duymadığı için bu dağılımdan bağımsız öğrenme algoritmalarıdır. Algoritma başarısından dolayı denetimli bir mimariye sahip olan çoğu uygulamalarda tercih edilmektedir.



Şekil 3.8. Destek Vektör Regresyon

SVM'ler algoritmaları elde bulunan verilerden yeni kalıpları tanımlamak amacıyla büyük miktarda veriyi analiz ederler. Şekil 3.8'de gösterildiği üzere bir veri kümesini en iyi iki sınıfa ayıran bir hiper düzlem bulma amacı ile oluşturulmaktadır. Destek vektörleri, iki sınırı (düzlem/çizgi) en iyi bir şekilde ayıran sınır çizgileridir ve karar yüzeyine (veya hiper düzleme) en yakın olan veri noktalarıdır.

4. DENEYSEL TESTLER VE TARTIŞMA

4.1. Çalışma Ortamı Hazırlıkları

Gerçekleştirilecek çalışma konusu olarak veriler arası analitik kullanılarak hava kalitesi tahmini yapılacaktır. Hava kalitesine çevre şartları önemli bir etken teşkil etmektedir. Bu sebeple çevreyi ve insan sağlığını tehdit eden hava kirliliği verileri bu nedenler gözetilerek temin edilecektir. Veri seti için ön araştırma gerçekleştirilerek, konum bazlı olarak elde edilebilecek en doğru veriler üzerinde çalışılabilmesi adına çalışmalara başlanmıştır. Yapılan araştırmalar neticesinde çalışma hava kalitesi tahmini için Mediaeval tarafından verilen “Insight for Wellbeing: (sınır ötesi) Haze Tahmini için Çapraz Veri Analitiği” başlıklı kıyaslama veri setini kullanıyoruz (Sweeney ve diğ., 2021). Bu çalışmada sağlanan veri kümesinden, çalışma konumuz için farklı türde nitelikler çıkardık; zamansal bilgiler, coğrafi veriler, ölçüm yapılan sensör verileri gibi anlamsal özellikler 2010 ve 2019 arası yıllar için toplanmış olup bu kriterler göz önünde tutularak çalışmalar yürütülecektir. Hava kalitesi analizi için tedarik edilen veriler ve gerçekleştirilen ön işleme adımlarından sonra bu verilere makine öğrenmesi yöntemleri uygulanarak çıkarılan sonuçlar analiz edilmiş olunacaktır.

Dünya genelinde çoğu ülkede, hava kirliliği inkâr edilemeyecek derece önemli bir sorun haline gelmiş ve insan refahını etkileyen başlıca problemlerden biri olmuştur. Bu tarz problemlerin en aza indirgenmesi için çeşitli makine öğrenmesi yöntemlerine başvurulmaktadır. Bu çalışmada da kullanılacağı gibi en çok veri kaynaklı yapılan çalışmalarda makine öğrenmesi yöntemleri tercih edilmeye başlanmıştır. Bu tez ile belirtilen veri seti kullanılarak Asya bölgesinde bulunan Singapur, Taylan’da ve Brunei ülkelerinin çeşitli bölgelerinde alınan günlük hava durumu ve hava kalitesi verilerine göre bölgelerin hava durumları ve hava kalitelerinin tahmini yapılarak bireysel olarak etkisi gözlemlenmiş olunacaktır. Yapılacak çalışma ile bu bölgelerdeki veriler kullanılarak hava kalitesi tahmini edilecektir. Kişilerin bireysel yaşamları için önemli ölçüde fayda sağlanması planlanmaktadır ve aynı zamanda gerçekleştirilen alt yapı ile istenilen ülkeler içinde kolaylıkla entegrasyon sağlanılacaktır. Kullanılacak veriler 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei, Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen PM10 konsantrasyonlarının günlük okumalarını içermektedir. Temin edilen ve ön işlemlerden geçirilecek olan

veriler Neural Network, XGBoost, Random Forest, AdaBoost, Gradient Boosting, Light Gradient Boosting Machine, Support Vector Regression gibi makine öğrenimi modelleri ile eğitilerek yaklaşımlar incelenecektir. Model başarısı ve çıktıları sonrası elde edilen sonuçlar incelenerek karşılaştırma yapılacaktır. Eğitim ve analizler neticesinde çıkan sonuçlar Tableau ve Python kütüphanelerinden yararlanılarak çeşitli grafikler ile görselleştirilip yorumlanacaktır. Bu şekilde bölge ve bölge halkının yaşamsal hava kalite endeksleri üzerine çıkarımlar elde edilmiş olunacaktır. Çalışmalar boyunca insan sağlığını tehdit eden PM10 verisi ağırlıklı olmak üzere hava tahmin ve diğer PM2.5 üzerinde durulmuş olunacaktır.

Hazırlanan veriler ve algoritmalar Google Collaboration vasıtası ile Python programlama dili ve kütüphaneleri kullanılarak eğitim gerçekleştirilmiş olup sonuçlar incelenmiş ve gerekli revizyonlar yapılmıştır (URL-14). Aynı zamanda görselleştirme ve analiz işlemleri için popüler olan ve yaygın olarak kullanılan Tableau aynı zamanda dinamik bir arayüz sağlayarak veri seti hakkında ve modellerin tahmin sonuçlarının analizlerini ayrıntılı ve profesyonel şekilde gerçekleştirmektedir (URL-15).

4.2. Veri Seti Hakkında

Temin edilen veri setinde Singapur, Tayland ve Brunei ülkelerinde bulunan bölgelerden alınan hava ölçüm rapor sonuçları bulunmaktadır. Kullanılacak hava tahmin verileri 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei, Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen PM10 konsantrasyonlarının günlük okumalarını içermektedir. Veriler ayrıca günlük (ve bazı durumlarda saatlik) sıcaklık, yağış, nem ve rüzgâr hızı/yönü değerlerinin atmosferik parametrelerini içerir. Ancak, verilerde bazı eksik değerlerle birlikte tüm parametreler her istasyonda/ülkede mevcut değildir.

Her bölge için veri seti şablon yapıları aynı olmakla birlikte sadece istasyonlar ve ölçüm bilgi verileri farklıdır. Bu yöntem aynı zamanda verilerin eğitimi ve analizi kısmında kolaylık sağlamaktadır.

Tedarik edilen tüm veriler CSV formatında saklanarak, klasör ve dosya yapıları şeklinde test ve eğitim verisi olarak hazırlanmıştır. Her üç ülke için toplanan veri setleri temel

olarak her bölge için hava kirliliği, meteorolojik veriler ve istasyonlar olmak üzere üç ana veri seti dosyası şeklinde bulunmaktadır. Ön işleme ile birleştirme işleminden sonra eğitim içi hazırlanacaktır. Bu veriler ve içerikleri ayrıntılı olarak şu şekildedir:

Hava Kalitesi Verisi (Airquality):

Tablo 4.1.'de gösterildiği üzere bölgelerin hava kalite değerlerinin ölçüldüğü istasyon numarası, yıl, ay, gün ve PM10 değer bilgisi bulunmaktadır.

Tablo 4.1. Hava Kalitesi Veri Seti Yapısı

Name	Description	Unit
ID	İstasyon Numarası	
Y	Yıl	YYYY
M	Ay	MM
D	Gün	DD
PM10	Partikül Madde 10	Hava kalitesi izleme istasyonları tarafından ölçülür. Bu veriler Temel Gerçek olarak kabul edilir. $\mu\text{g}/\text{m}^3$

İstasyonlar Verisi (Station):

Tablo 4.2.'de gösterildiği üzere her bölgelere ait verilerin ölçümlerin yapıldığı istasyon bilgileri ve bu istasyonlara ait kendi bilgileri; numarası, adı, tipi, bölgesi ve koordinat bilgisi bulunmaktadır.

Tablo 4.2. İstasyon Veri Seti Yapısı

Name	Description
ID	İstasyon Numarası
NAME	İstasyon Adı
TYPE	İstasyon Tipi
DISTRICT	İstasyon Konumu
LATITUDE	
LONGITUDE	

Hava Durumu Verisi (Weather):

Tablo 4.3.' de gösterildiği üzere ölçümlerin gerçekleştirildiği istasyonlara ait istasyon bazlı ve zamansal alınan hava durumu verileri; sıcaklık, nem, yağmur ve 24 saatlik rüzgâr hız (her saati içeren) ve yönü (her saati içeren) hakkında bilgilerden oluşmaktadır.

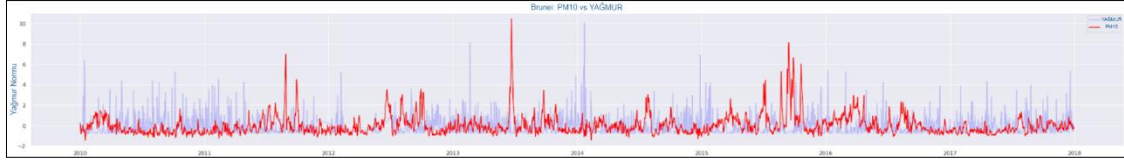
Tablo 4.3. Hava Durumu Veri Seti Yapısı

Name	Description	Unit
ID	İstasyon Numarası	
Y	Yıl	YYYY
M	Ay	MM
D	Gün	DD
TEMP	Sıcaklık	°C
HUM	Nem	%
RAIN	Yağış	Milimetre (mm)
WD1 ~ WD24	Rüzgâr yönü (saat 1 ~ saat 24)	° Gerçek Kuzey
WS1 ~ WS24	Rüzgâr hızı (saat 1 ~ saat 24)	Düğüm

Tüm bu bilgiler göz önünde bulundurularak bölgesel olarak analizler yapılacaktır. Aynı zamanda kullanılacak olan bu verilerden çıkarımlar ile modellerin performansı artırılması için verilerde bulunan eksikler, hatalı işlemler ve çeşitli ön işlemler (ortalama, istasyon bazlı yakınlık durumuna göre) yapılarak daha anlamlı hale getirilmiştir.

Veri setinin istatistiksel özellikleri, veri ön işleme ve analiz sürecinde büyük önem arz eder. Verileri eğitimden önce daha iyi anlamak ve bazı çıkarımlar yapabilmek için Python programlama dili kullanılarak görselleştirmeler ile incelenmiştir.

Şekil 4.1., Şekil 4.2. ve Şekil 4.3'te, farklı yıllarda bölgelere göre alınan meteorolojik ve PM10 parametrelerine ait grafiksel gösterimler verilmiştir. Kırmızı PM10, Mor ise yağmur grafiğini temsil etmektedir.



Şekil 4.1. Brunei (2010-2017) PM10 ve Yağmur Değerleri Grafiği



Şekil 4.2. Singapur (2010-2017) PM10 ve Yağmur Değerleri Grafiği



Şekil 4.3. Tayland (2010-2017) PM10 ve Yağmur Değerleri Grafiği

4.3. Veri Ön İşleme

Temin edilen veriler eğitim ve tahminleme için kullanılmadan önce çeşitli ön işlemlerden geçirilmiştir. Veri setinde bazı zamanlar bazı bölge ve istasyondan alınan verilerin boş olduğu veya kayıp olarak verilmiş olup bu boşluklar doldurulmuştur. Eksik veya boş bulunan alanlar/veriler eğitim sonrası tahminleme hata oranını arttırmaktadır. Veri setinde buluna boş olan veriler yakın çevre istasyonunda aynı zaman içerisinde alınan veriler ile ortalama bir değer üretilerek doldurulmuştur. Bu sayede eğitim sonrası tahminleme hata oranının minimuma indirgenmesi sağlanmıştır. Tablo 4.4'te görüldüğü üzere ölçüm yapılan istasyonlar sınıflandırılarak aynı bölgedeki istasyonlar ve hangi ölçüm için kullanıldığı Tablo 4.5'te gösterildiği gibi ID olarak numaralandırılmıştır.

Tablo 4.4. Aynı Bölge İstasyon ID İşlemi

ID	İstasyon ID
0	302B, AWS9
1	101B, AWS1
2	401B, AWS12
3	201B, AWS6

Tablo 4.5. Brunei İstasyonlar ve Bilgileri Verisi

ID	AD	TÜR	SEMT	ENLEM	BOYLAM	
0	101B	Anggerek Air Quality Monitoring Station	Hava Kalitesi İzleme İstasyonu	Brunei-Muara	4.933330	114.933330
1	201B	Bukit Bendera Air Quality Monitoring Station	Hava Kalitesi İzleme İstasyonu	Tutong	4.810257	114.233096
2	302B	Mumong Air Quality Monitoring Station	Hava Kalitesi İzleme İstasyonu	Belait	4.575119	114.233096
3	401B	Taman Batang Duri Air Quality Monitoring Station	Hava Kalitesi İzleme İstasyonu	Temburong	4.578603	115.121521
4	AWS1	Brunei International Airport	Meteoroloji İstasyonu	Brunei-Muara	4.939831	114.923664
5	AWS6	Sinaut Weather Station	Meteoroloji İstasyonu	Tutong	4.813889	114.740278
6	AWS9	Sungai Liang Weather Station	Meteoroloji İstasyonu	Belait	4.676944	114.489167
7	AWS12	Pekan Bangar Weather Station	Meteoroloji İstasyonu	Temburong	4.712222	115.068611

Örnek olarak Şekil 4.4'te boş olan PM10 sütunları aynı gün yakın bölgedeki istasyonların ölçtüğü değerlerin ortalaması alınarak doldurulmuştur.

	ID	Y	M	D	PM10
0	101B	2010	1	1	NaN
1	201B	2010	1	1	NaN
2	302B	2010	1	1	19.0
3	401B	2010	1	1	NaN
4	101B	2010	1	2	NaN

→

	ID	Y	M	D	PM10
0	1	2010	1	1	19.0
1	3	2010	1	1	19.0
2	0	2010	1	1	19.0
3	2	2010	1	1	19.0
4	1	2010	1	2	17.0
...
11683	2	2017	12	30	11.2
11684	1	2017	12	31	13.9
11685	3	2017	12	31	17.1
11686	0	2017	12	31	17.7
11687	2	2017	12	31	10.8

11688 rows x 5 columns

Şekil 4.4. Brunei Eksik PM10 Doldurma İşlem Örneği

PM10 eksik veri doldurma işleminde olduğu gibi diğer hava sıcaklığı, nem, yağmur, durumları aynı günlerdeki istasyonların ortalama değerleri alınarak doldurulmuştur. Aynı şekilde saat 1'den 24'e kadar rüzgâr yönü ve hız değerlerinde boş bulunan değerler için de aynı işlemler yapılmıştır. Örnek olarak Brunei verisi verilmiş olup tüm bölgeler için bu eksik veriler doldurulmuştur.

Eksik veriler giderildikten sonra oluşturulan veriler makine öğrenmesi modellerine verilmek üzere birleştirme işlemlerine geçilmiştir. Aynı istasyon ve tarih işlemlerine göre PM10 değerleri ve hava durumlarını bulunduran yeni CSV dosyaları oluşturulmuştur. Oluşturulan yeni düzenli veri setlerinin bulunduğu veriler eğitime hazır hale getirilmiştir.

Eğitime girdi olarak bölge numarası (ID), yıl (Y), ay (M), gün (D), hava durumu verileri sıcaklık (TEMP), nem (HUM), yağmur (RAIN), rüzgâr yönü (saat 1 ~ saat 24) ve rüzgâr hızı (saat 1 ~ saat 24) verileri, çıktı olarak ise PM10 değeri verilerek bölge, meteorolojik şartlar ve tarih bilgisine göre hava kirliliği tahminlemesi yapılacaktır. Nihai algoritma sonrası çıktı olarak günlük hava durum bilgisi ve PM10 değerlerini verecektir.

4.4. Performans Metrikleri

Düzenlenen hava tahmin verileri makine öğrenimi modelleri ile eğitilerek performans değerlendirmesi yapılacaktır. Model başarısı ve çıktıları sonrası elde edilen sonuçlar incelenerek karşılaştırma yapılması için performans metriklerine ihtiyaç duyulmaktadır. Makine ve derin öğrenme yöntemlerinin performans değerlendirmelerinde sıklıkla kullanılan Mutlak Hata (MAE) ve Kök Ortalama Kare Hatası (RMSE) yöntemleri kullanılmıştır. Bu performans metrikleri neticesinde çıkan sonuçlar görselleştirilip yorumlanacaktır. Bu şekilde bölge ve bölge halkının hava kalitesi hakkında çıkarımlar elde edilmiş olunacaktır.

Kök Ortalama Karesel Hata (Root Mean Squared Error-RMSE), yaygın olarak kullanılan bir performans ölçütüdür, MSE sonucu çıkan değerın karekökünü alınmasıyla elde edilen ve daha küçük değerler karşılaştırmada kolaylık sağlamaktadır. RMSE yüksek tahmin hatalarına karşı hassas olduğundan küçük hataların tolere

edilebilmekte ve büyük hataların ise yüksek maliyete sebep olduğu durumlar için kullanılabilir.

Ortalama Mutlak Hata (Mean Absolute Error-MAE) tahmin edilen değerlerde ortalamanın üzerinde büyük hata(lar) elde edilmişse kullanılabilir. MAE yönlerini dikkate almadan hataların ortalama büyüklüğünü ölçer

4.5. Performans Sonuçları ve Tartışma

Gerçekleştirilen çalışmalar ve yapılan incelemeler sonucunda makine öğrenmesi modellerin performans değerleri incelenerek hangi modellerin en uygun olduğuna karar verilmiştir. Neural Network, XGBoost, Random Forest, AdaBoost, Gradient Boosting, Light Gradient Boosting Machine, Support Vector Regression makine öğrenimi modelleri kullanılması kararlaştırılarak uygulanmıştır. Modeller ile gerek eğitim algoritma gerekse modelin sahip olduğu parametreler ile daha hassas değerler verilerek denemeler yapılmıştır.

Tüm istasyonlarda geçerli olabilecek tahmin modeli geliştirmek için bu modeller ile yapay sinir ağları kullanılmaya çalışılmıştır. Modellerin başarımı açısından önemli olan yapay sinir ağı geliştirilir iken istasyon tahmin modeli ile benzer şekilde model kullanılacak hiper parametre konfigürasyonlarına önem verilmiştir.

Gerçekleştirilen çalışmalar ve yapılan incelemeler sonucu Neural Network, XGBoost, Random Forest, AdaBoost, Gradient Boosting, Light Gradient Boosting Machine, Support Vector Regression modellerinin performans değerleri incelenerek hangi modellerin en uygun olduğuna karar verilmiştir. Çalışmaların yapıldığı modeller ile gerek eğitim algoritma gerekse modelin sahip olduğu parametreler ile daha hassas değerler verilerek denemeler yapılmıştır. İşlemler sonucu modellerin eğitim sonucu en uygun sonuçlar çıkarması sağlanmıştır. Aynı zamanda karar verilen modeller tüm bölgeler için ayrı ayrı kullanılarak değerlendirilmiştir. Bunlarla ilgili ayrıntılı bilgi ve analizler yapılmıştır.

Veri setinden yapılan ön işlemler sonrasında PM10 ve hava durumu tahmini için Brunei, Singapur ve Tayland bölgelerindeki veriler elde edilmiştir. Kullanılan veriler 2010 ile 2019 arasındaki farklı dönemleri kapsayan Asya ülkelerindeki (Brunei,

Singapur ve Tayland) birden fazla hava istasyonu tarafından kaydedilen meteorolojik veriler ve PM10 konsantrasyonlarının günlük okumalarını içermektedir. Bu veriler temel olarak istasyon verileri, hava durumu verileri ve hava kalitesi verileri olarak üç bölümden oluşmuş ve bunlar oluşturulan modelleri eğitmek ve değerlendirmek için eğitim ve test olmak üzere ikiye ayrılmıştır. Veri setinin %80'i modelleri eğitmek için kullanılırken %20'si modelleri test etmek için ayrılmıştır.

4.5.1. Yerelleştirilmiş Hava Kirliliği Tahmini (Görev 1)

Tez çalışmamız ve yarışma görevi olarak bu görev de, bir ülkedeki her bir hava kalitesi istasyonunun bulunduğu yerde, test dosyalarındaki her boşlukta ilk 3 gün için PM10 değerini tahmini yapılmıştır. Ancak biz 3 gün ile kalmayıp yaklaşık 3 ay gibi bir tahmin de de bulunacağız. Bunu yapmak için ise yalnızca her ülkeden yerelleştirilmiş eğitim verilerine dayalı farklı tahmin modeli geliştirmeleri gerekmektedir. Bu görev, önümüzdeki takip eden 3 gün için hava kirliliğini tahmin etmenin doğruluğunu araştırarak ve her ülkenin yalnızca kendi hava durumu ve hava kirliliği verilerine bağlı olması durumunda bu hedefe ne kadar iyi ulaşılabileceğini değerlendirmek için tasarlanarak eğitimler gerçekleştirilmiştir. Eğitim aşamasında kullanılan bazı modellerin parametre bilgileri Tablo 4.6'da Görev 1 İçin Kullanılan Modellere Ait ve Parametre Bilgisi olarak verilmiştir.

Tablo 4.6. Görev 1 İçin Kullanılan Modellere Ait ve Parametre Bilgisi

Kullanılan Model	Parametreler
RF	n_estimators=500, max_leaf_nodes=16
XGBoost	n_estimators=1000, learning_rate=0.01
GB	n_estimators=500, learning_rate=0.05
ADABOOST	n_estimators=500, learning_rate=1.0
LightGBM	n_estimators=500, learning_rate=0.05

Çalışma sonucu çıktıları ve model performans metrikleri daha ayrıntılı olarak aşağıda bölgesel başlıklar halinde verilmiştir.

4.5.1.1. Brunei Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri eğitim sonrası test setlerinden yararlanılarak elde edilen performans metrik değerleri Brunei Bölgeleri için Tablo 4.7’de PM10 tahmin sonuçları verilmiştir.

Tablo 4.7. Brunei Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	7.224	11.589
RF	6.184	10.235
XGBoost	7.467	11.492
GB	7.432	11.470
ADABOOST	6.797	9.807
LightGBM	7.510	11.490
SVR	6.870	11.293

Tablo 4.7’de gösterilen performans metrikleri incelendiğinde RF, ADABOOST ve SVR kullanılması diğer modellerden daha iyi performans göstermektedir. LightGBM modelinde ise en düşük sonuç alındığı görülebilir.

4.5.1.2. Singapur Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri eğitim sonrası test setlerinden yararlanılarak elde edilen performans metrik değerleri Singapur Bölgeleri için Tablo 4.8’de PM10 tahmin sonuçları verilmiştir.

Tablo 4.8. Singapur Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	8.387	11.683
RF	8.152	11.326
XGBoost	8.781	11.893
GB	8.509	11.609
ADABOOST	8.287	10.740
LightGBM	8.896	11.984
SVR	8.098	11.314

Tablo 4.8’de gösterilen performans metrikleri incelendiğinde SVR, RF ve ADABOOST kullanılması diğer modellerden daha iyi LGBM is en düşük performansı göstermektedir

4.5.1.3. Tayland Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri eğitim sonrası test setlerinden yararlanılarak elde edilen performans metrik değerleri Tayland Bölgeleri için Tablo 4.9’da PM10 tahmin sonuçları verilmiştir.

Tablo 4.9. Tayland Bölgelerine Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	11.196	14.692
RF	11.108	14.784
XGBoost	12.314	16.260
GB	12.353	16.292
ADABOOST	13.566	16.812
LightGBM	12.801	16.832
SVR	11.137	14.350

Tablo 4.9’da gösterilen performans metrikleri incelendiğinde RF, SVR ve NN kullanılması diğer modellerden daha iyi performans göstermektedir. ADABOOST modelinde ise en düşük sonuç alındığı görülmüştür.

4.5.2. Sınır Ötesi Hava Kirliliği Tahmini (Görev 2)

Tez çalışmamızda ve yarışma kapsamında ulaşmak istediğimiz hava kirliliği tahmini çalışması için ve aynı zamanda verilerin temin edildiği yarışma kapsamında ikinci görev ile tahmin edilecek değerler açısından bir önceki alt görevle aynıdır, ancak yapılan çalışmalar ile aynı ülkeden veya komşu ülkelerden temin edilebilen diğer veri kaynaklarını dikkate alınmıştır. Bu görev, komşu ülkelerdeki sis ve hava durumu dikkate alındığında tahmin doğruluğundaki gelişmeleri gözlemlenerek sınır ötesi pus etkilerini ele almaya çalışılmıştır. Eğitim aşamasında kullanılan bazı modellerin parametre bilgileri Tablo 4.10’da Görev 2 İçin Kullanılan Modellere Ait ve Parametre Bilgisi olarak verilmiştir.

Tablo 4.10. Görev 2 İçin Kullanılan Modellere Ait ve Parametre Bilgisi

Kullanılan Model	Parametreler
RF	n_estimators=500, max_leaf_nodes=16
XGBoost	n_estimators=1000, learning_rate=0.01
GB	n_estimators=500, learning_rate=0.05
ADABOOST	n_estimators=500, learning_rate=1.0
LightGBM	n_estimators=500, learning_rate=0.05

Çalışma sonucu her üç bölge için yapılan çalışmalar sonucu elde edilen çıktılar aşağıdaki gibi başlıklar halinde verilmiş ve her model için performans metrikleri tablo ile görselleştirilmiştir.

4.5.2.1. Brunei Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri eğitim sonrası test setlerinden yararlanılarak elde edilen performans metrik değerleri Brunei Bölgeleri için Tablo 4.11’de PM10 tahmin sonuçları verilmiştir.

Tablo 4.11. Brunei Bölgelerine Görev 2’ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	6.624	10.891
RF	6.386	10.482
XGBoost	7.495	11.501
GB	7.455	11.562
ADABOOST	6.401	9.641
LightGBM	7.415	11.454
SVR	6.860	11.212

Tablo 4.11’de gösterilen performans metrikleri incelendiğinde RF, ADABOOST ve NN kullanılması diğer modellerden daha iyi performans göstermektedir. XGBoost modelinde ise en düşük sonuç alındığı görülmüştür.

4.5.2.2. Singapur Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri eğitim sonrası test setlerinden yararlanılarak elde edilen performans metrik değerleri Singapur Bölgeleri için Tablo 4.12’de PM10 tahmin sonuçları verilmiştir. Tablo 4.12’de gösterilen performans metrikleri incelendiğinde GB, LGBM ve XGB ve kullanılması diğer modellerden daha iyi performans göstermektedir. NN modelinde ise en düşük sonuç alındığı görülmüştür.

Tablo 4.12. Singapur Bölgelerine Görev 2’ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	8.393	11.655
RF	8.159	11.372
XGBoost	8.628	11.838
GB	8.395	11.623
ADABOOST	7.827	10.763
LightGBM	8.553	11.771
SVR	8.163	11.394

4.5.2.3. Tayland Bölgeleri

Tez çalışması kapsamında kullanılan makine öğrenmesi yöntemleri diğer bölgelere uygulanan NN, RF, GB, ADABOOST, LightGBM ve SVR eğitim sonrası test setlerinden yararlanılarak elde edilen sonuçlar MAE ve RMSE performans metrik hesaplamaları kullanılarak Tayland Bölgeleri için Tablo 4.13’te PM10 tahmin sonuçları verilmiştir.

Tablo 4.13. Tayland Bölgelerine Görev 2’ye Ait PM10 Tahmininde Kullanılan Makine Öğrenmesi Algoritma Sonuçları

Kullanılan Model	MAE	RMSE
NN	11.398	14.612
RF	11.960	15.374
XGBoost	12.397	16.121
GB	11.698	15.346
ADABOOST	12.896	15.995
LightGBM	12.327	16.028
SVR	11.185	14.375

Tablo 4.13'te gösterilen performans metrikleri incelendiğinde SVR, NN ve GB kullanılması diğer modellerden daha iyi performans göstermektedir. ADABOOST modelinde ise en düşük sonuç alındığı görülmüştür.

4.5.3. Transfer Öğrenme ile Hava Kirliliği Tahmini (Görev 3)

Bu görev ve çalışmada ise PM10 değeri için yapılan tahminleme çalışmaları çözümlerinde transfer öğrenme tekniklerini göz önünde bulundurarak görev 1 veya görev 2'yi tekrar kullanılarak eğitilmiştir. Her üç bölge için ayrı ayrı olarak eğitilen modeller kullanılarak ve aynı zamanda ilgi bölge dışındaki diğer bölgelere ait veriler ile desteklenilerek başka bir ülkenin hava tahmin sonucunu iyileştirmede katkı durumu hakkında bilgi sahibi olunmuş ve performans metrik çıktıları alt başlık olarak her bölge için tablo oluşturulmuştur. Eğitim aşamasında kullanılan bazı modellerin parametre bilgileri Tablo 4.14'te Görev 3 İçin Kullanılan Modellere Ait ve Parametre Bilgisi olarak verilmiştir.

Tablo 4.14. Görev 3 İçin Kullanılan Modellere Ait ve Parametre Bilgisi

Kullanılan Model	Parametreler
RF	n_estimators=500, max_leaf_nodes=16
XGBoost	n_estimators=1000, learning_rate=0.05
GB	n_estimators=500, learning_rate=0.05
ADABOOST	n_estimators=500, learning_rate=1.0
LightGBM	n_estimators=500, learning_rate=0.1

Çalışma sonucu çıktıları ve model performans metrikleri aşağıda başlıklar halinde verilmiştir.

4.5.3.1. Brunei Bölgeleri

Önceki görevlerde yapılan tahminleme çalışmaları çözümlerinde transfer öğrenme tekniklerini göz önünde bulundurarak Brunei bölgesi PM10 tahmini sonucunu iyileştirmek için Singapur bölgesinin eğitilen modeli ve Tayland bölgesinin verileri dahil edilerek eğitimler gerçekleştirilmiştir. Çalışması kapsamında kullanılan transfer öğrenmesi yöntemi ile eğitim sonrası test setlerinden yararlanılarak elde edilen PM10

tahminleme performans metrik deęerleri Brunei Bۆlgeleri iin Tablo 4.15'te PM10 olarak verilmiřtir.

Tablo 4.15 Brunei Bۆlgelerine Gۆrev 3'ye Ait PM10 Tahmininde Kullanılan Transfer ۆğrenmesi Sonuları

Kullanılan Model	MAE	RMSE
NN	8.497	11.780
RF	8.225	11.444
XGBoost	8.697	11.880
GB	8.488	11.646
ADABOOST	7.909	10.902
LightGBM	8.782	11.948
SVR	8.363	11.691

Tablo 4.15'te gۆsterilen performans metrikleri incelendięinde RF, ADABOOST ve SVR kullanılması dięer modellerden daha iyi performans gۆstermektedir. LightGBM modelinde ise en dۆřuk sonu alındıęı gۆrۆlmüştür.

4.5.3.2. Singapur Bۆlgeleri

Önceki gۆrevlerde yapılan tahminleme alıřmaları özۆmlerinde transfer ۆğrenme tekniklerini gۆz önünde bulundurarak Singapur bۆlgesi iine Tayland'ın modeli ve Brunei'in verisi uygulanmıřtır. alıřması kapsamında kullanılan transfer ۆğrenmesi yöntemi ile eęitim sonrası test setlerinden yararlanılarak elde edilen PM10 tahminleme performans metrik deęerleri Singapur Bۆlgeleri iin Tablo 4.16'da. PM10 olarak verilmiřtir.

Tablo 4.16 Singapur Bۆlgelerine Gۆrev 3'ye Ait PM10 Tahmininde Kullanılan Transfer ۆğrenmesi Sonuları

Kullanılan Model	MAE	RMSE
NN	12.067	16.045
RF	11.376	15.077
XGBoost	12.126	16.146
GB	12.088	16.110
ADABOOST	11.647	15.169
LightGBM	12.199	16.126
SVR	10.622	14.209

Tablo 4.16’da gösterilen performans metrikleri incelendiğinde RF, ADABOOST ve SVR kullanılması diğer modellerden daha iyi performans göstermektedir. LightGBM modelinde ise en düşük sonuç alındığı görülmüştür.

4.5.3.3. Tayland Bölgeleri

Önceki görevlerde yapılan tahminleme çalışmaları çözümlerinde transfer öğrenme tekniklerini göz önünde bulundurarak Tayland bölgesi için Brunei’in modeli ve Singapur’un verisi uygulanmıştır. Çalışması kapsamında kullanılan transfer öğrenmesi yöntemi ile eğitim sonrası test setlerinden yararlanılarak elde edilen PM10 tahminleme performans metrik değerleri Tayland Bölgeleri için Tablo 4.17’de PM10 olarak verilmiştir.

Tablo 4.17 Tayland Bölgelerine Görev 3’ye Ait PM10 Tahmininde Kullanılan Transfer Öğrenmesi Sonuçları

Kullanılan Model	MAE	RMSE
NN	7.560	11.822
RF	7.050	10.958
XGBoost	7.314	11.259
GB	7.359	11.351
ADABOOST	7.033	8.238
LightGBM	7.384	11.343
SVR	7.709	12.086

Tablo 4.17’de gösterilen performans metrikleri incelendiğinde FR, ADABOOST ve XGBoost kullanılması diğer modellerden daha iyi performans göstermektedir. NN modelinde ise en düşük sonuç alındığı görülmüştür.

4.5.4. Diğer Çalışmalarla Karşılaştırma

Makine öğrenmesi modeller ile gerçekleştirilen çalışmalar ve yapılan incelemeler sonucunda performans değerleri incelemiştir. Tez kapsamında gerçekleştirilen çalışma ile aynı veri setini kullanarak yapılan diğer çalışmaların performans metrikleri kıyaslanmıştır. Birebir yapılan çalışmalar bulunmadığından yapılacak olan karşılaştırma sadece PM için yapılan çalışmalar ile tablo halinde oluşturulmuştur, burada model

performans ve başarımları kıyaslaması yapılarak en doğru tahmin elde edilen sonuçlar üzerine durulmuştur.

Bunlar Tablo 4.18’de verilmiş olup numaralandırılan çalışmalarda kullanılan makalelerin eşleşmesi şu şekildedir;

- 1- Çalışma: Veriler Arası Analitik Kullanılarak PM10 Konsantrasyonları ve Meteorolojik Bilgilerden Hava Kalitesi Tahmini Etme (Şu anki Tez Çalışmamız)
- 2- Çalışma: LSTM Kullanarak Hava Kalitesi Tahmini ve Veri İşleme Teknikleri İçin Bir Yaklaşım (Ton-Thien ve diğ., 2021).
- 3- Çalışma: Haze için Veriler Arası Analitiğin Zaman Serisi Öngörüsüne Doğru Tahmin (Akbar ve diğ., 2021).
- 4- Çalışma: Refah İçin İçgörüler: İstifleme Topluluğu Modelini Kullanarak PM10 Değerlerini Tahmin Etme (Nguyen ve Duong, 2021).
- 5- Çalışma: Sınır Ötesi Pus Tahmini: Haze Tahmini için Çapraz Veri Analitiğinin Zaman Serisi Tahminine Doğru (Akbar ve diğ., 2022).

Tablo 4.18. Görev 1 için PM10 Değer Tahmin Performans Karşılaştırma Tablosu

GÖREV 1	MODEL	BRUNEI		SINGAPUR		TAYLAND	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
1- Çalışma (TEZ)	RF	6.184	10.235	8.152	11.326	11.108	14.784
	SVR	6.870	11.293	8.098	11.314	11.137	14.350
	ADABOOST	6.797	9.807	8.287	10.740	13.566	16.812
2- Çalışma	Bi-LSTM	-	3.625	-	5.821	-	10.624
	LSTM	-	3.629	-	-	-	-
	Stacked LSTM	-	3.921	-	-	-	-
3- Çalışma	ARIMA/ SARIMA	9.418	-	7.437	-	7.473	-
4- Çalışma	Single variable LSTM	9.6541	4.11	4.7474	5.81	5.7	7.4
	Bi-LSTM	13.9908	14.0457	21.3828	21.3828	16.9081	17.0003
	LSTM	14.4873	14.5318	21.612	21.6594	17.0171	17.0769
	GRU	14.4799	14.4799	21.4659	21.5168	17.042	17.1219
	Stacking	0.0244	0.16	0.17	0.21	0.154	0.2
5- Çalışma	ARIMA/ SARIMA	9.418	-	7.437	-	8.595	-

Tablo 4.18’da görüldüğü üzere yapılan karşılaştırma performans metrikleri incelendiğinde çalışmamızda kullanılan RF, SVR ve ADABOOST modelleri ve 4- Çalışma da kullanılan Stacking modelinde elde edilen sonuçlar diğer çalışmalarda kullanılan modellere göre daha başarılı olduğu görülmektedir.

Tablo 4.19. Görev 2 için PM10 Değer Tahmin Performans Karşılaştırma Tablosu

GÖREV 2	MODEL	BRUNEI		SINGAPUR		TAYLAND	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
1- Çalışma (TEZ)	RF	6.386	10.482	8.159	11.372	11.960	15.374
	ADABOOST	6.401	9.641	7.827	10.763	12.896	15.995
	SVR	6.860	11.212	8.163	11.394	11.185	14.375
2- Çalışma	Bi-LSTM	-	10.967	-	10.248	-	9.762
3- Çalışma	VAR	9.419	44.97	7.436	29.07	7.443	27.74
5- Çalışma	VAR	4.284	25.08	6.888	25.27	7.443	27.74

Tablo 4.19’da görüldüğü üzere Görev 2 için yapılan karşılaştırma performans metrikleri incelendiğinde çalışmamızda kullanılan RF, AdaBoost ve SVR modelleri ve 5- Çalışma da kullanılan VAR (Vector Auto-Regression) modelinde elde edilen sonuçlar diğer çalışmalarda kullanılan modellere göre daha başarılı olduğu görülmektedir.

Tablo 4.20. Görev 3 Transfer Öğrenme ile PM10 Değer Tahmin Performans Karşılaştırma Tablosu

GÖREV 3	MODEL	BRUNEI		SINGAPUR		TAYLAND	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
1- Çalışma (TEZ)	RF	8.225	11.444	11.376	15.077	7.050	10.958
	ADABOOST	7.909	10.902	11.647	15.169	6.505	10.773
	SVR	8.363	11.691	10.622	14.209	7.709	12.086
5- Çalışma	ARIMA/SARIMA	-	-	-	3.105	-	7.732

Tablo 4.20’de görüldüğü üzere Görev 3 için yapılan Transfer Öğrenme ile tahminleme için karşılaştırma performans metrikleri incelendiğinde tekrardan çalışmamızda kullanılan LightGB ve GB modelleri ve 5- Çalışma da kullanılan ARIMA/ SARIMA modelinde elde edilen sonuçlar diğer çalışmalarda kullanılan modellere göre daha başarılı olmuştur.

4.6. Veri Görselleştirme

Taz kapsamında yapılan çalışmalar sonucunu makine öğrenmesi modellerinin tahmin sonuçlarının daha iyi analiz edebilmek adına çeşitli değerlendirmeler yapılmıştır, bunlardan birisi de ayrıntılı analiz için görselleştirilmesidir. Görselleştirme çalışmaların daha ayrıntılı ve profesyonel olarak yapmak için Tableau kullanılmıştır. Tableau özellikle kullanım kolaylığı ile öne çıkan, birçok ihtiyacı karşılamak için halihazırda modülleri bulunan ve ek ihtiyaçlar dahilinde hızlıca çözümler geliştiren oldukça genç markalardan birisi olmuştur (URL-16). Görselleştirme çalışmaları Tableau Public kullanılarak ücretsiz sürüm üzerinden gerçekleştirilmiştir.

Görselleştirme çalışmalarında hava kirliliğine neden olan PM10 değeri üzerinde durularak sırasıyla Brunei, Singapur ve Tayland bölgelerinde bulunan alt bölgelerin hava kirliliği ve bunların bireyler ve çevre üzerine etkisi incelenmiştir. Sağlıklı bir yaşam için geçmiş değerlere bakılara gelecek günler için endeks tahmini yapılmıştır.

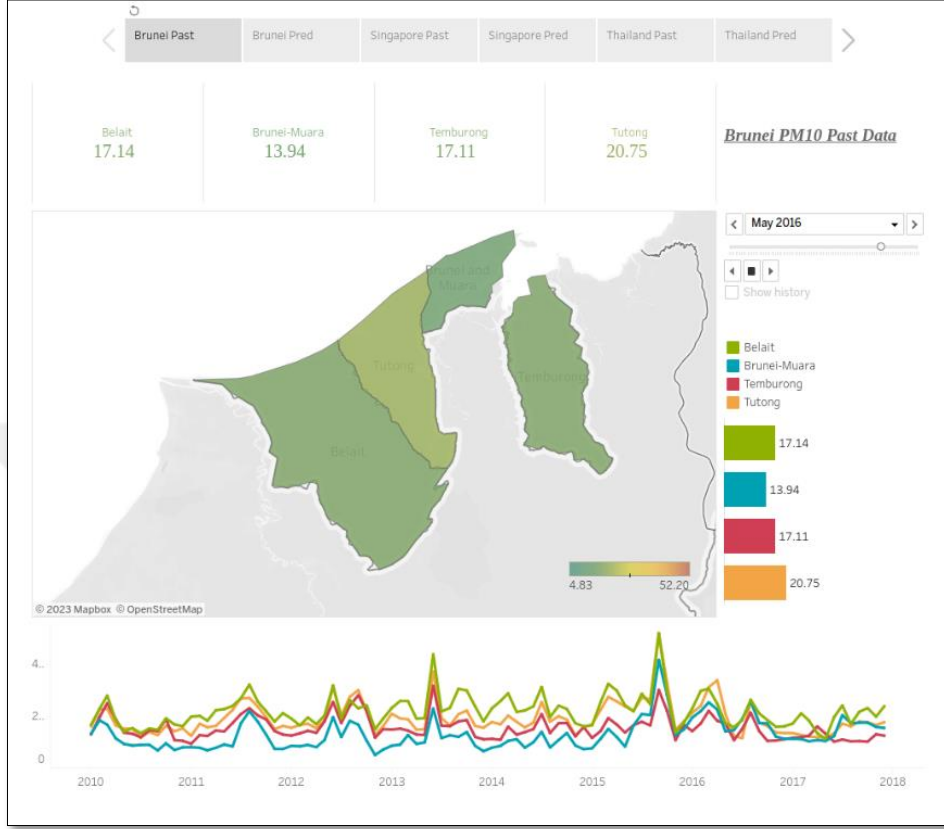
Veri setleri içerisinde bulunan veriler 2010 ve 2019 tarihleri arasındaki toplanan veriler olduğu için tahmin etmesi planlanan zamanlar 2018 tarihinden sonraki üç ay zaman dilimini kapsayacak şekilde tahminleme yapılmıştır. Ekran görüntüsü olarak verilen görseller aynı zamanda dinamik bir yapı şeklinde yapılmıştır. İstenilen bölgenin istenilen tarih ile geçmiş ve gelecek hava tahmin rapor bilgisi sunulmaktadır. Görselleştirilen veriler renklendirme işlemleri ile hem ayırt edilmesi hem de bölge de ki tehlikeli durum uyarısı niteliği taşımaktadır.

4.6.1. Brunei Bölgeleri

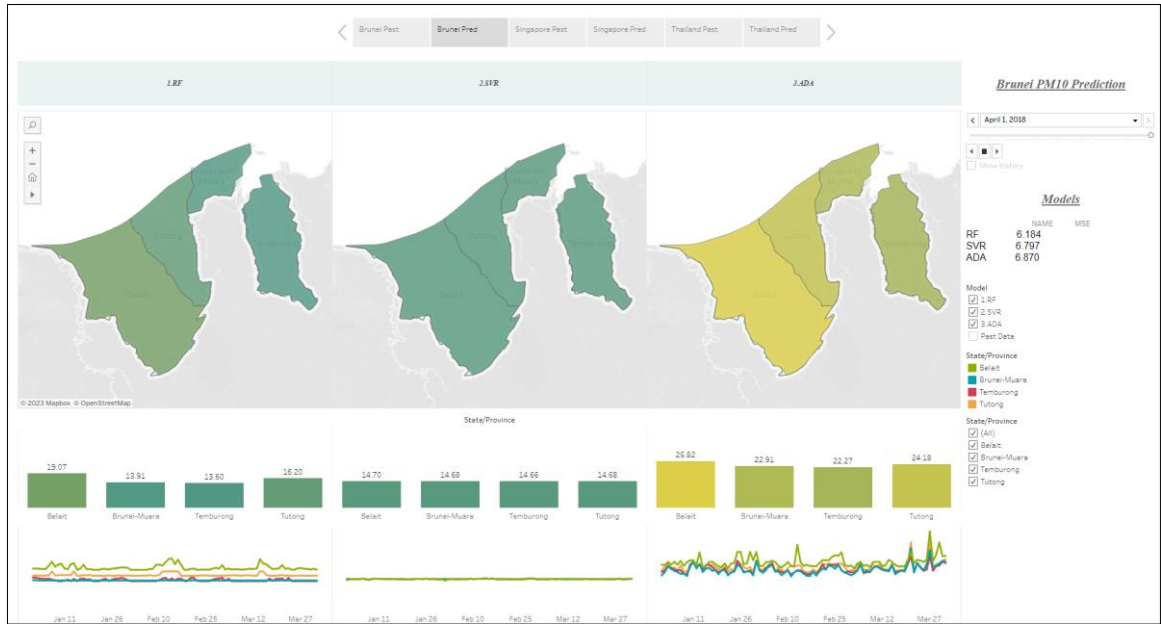
Brunei bölgelerine ait veriler Şekil 4.5'te ki gibi görselleştirilerek eğitim sonrası tahminleme ile geçmiş veriler hakkında daha net çıkarımlar yapılabilmelektir. Brunei için Belarit, Brunei Muara, Temburang ve Tutong eyaletleri olarak harita 4 bölüme ayrılmış ve değerlerin ortalamaları alınarak oluşturulmuştur. Dinamik olarak farklı tarihler ve bölgeler üzerinde incelemelerde bulunabilmektedir.

Brunei bölgeleri için Şekil 4.6'da 30 Mart 2018 tarihindeki ortalama hava kirliliği tahmin bilgisi verilmektedir. Belarit, Brunei Muara, Temburang ve Tutong eyaletleri için harita 4 bölüme ayrılarak bu alanlar üzerinden ayrışık tahminleme yapılmıştır. Aynı

zamanda dinamik olarak farklı modellerin bölgeler üzerindeki tahmin sonuçları da görülebilmektedir.



Şekil 4.5. Brunei Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli

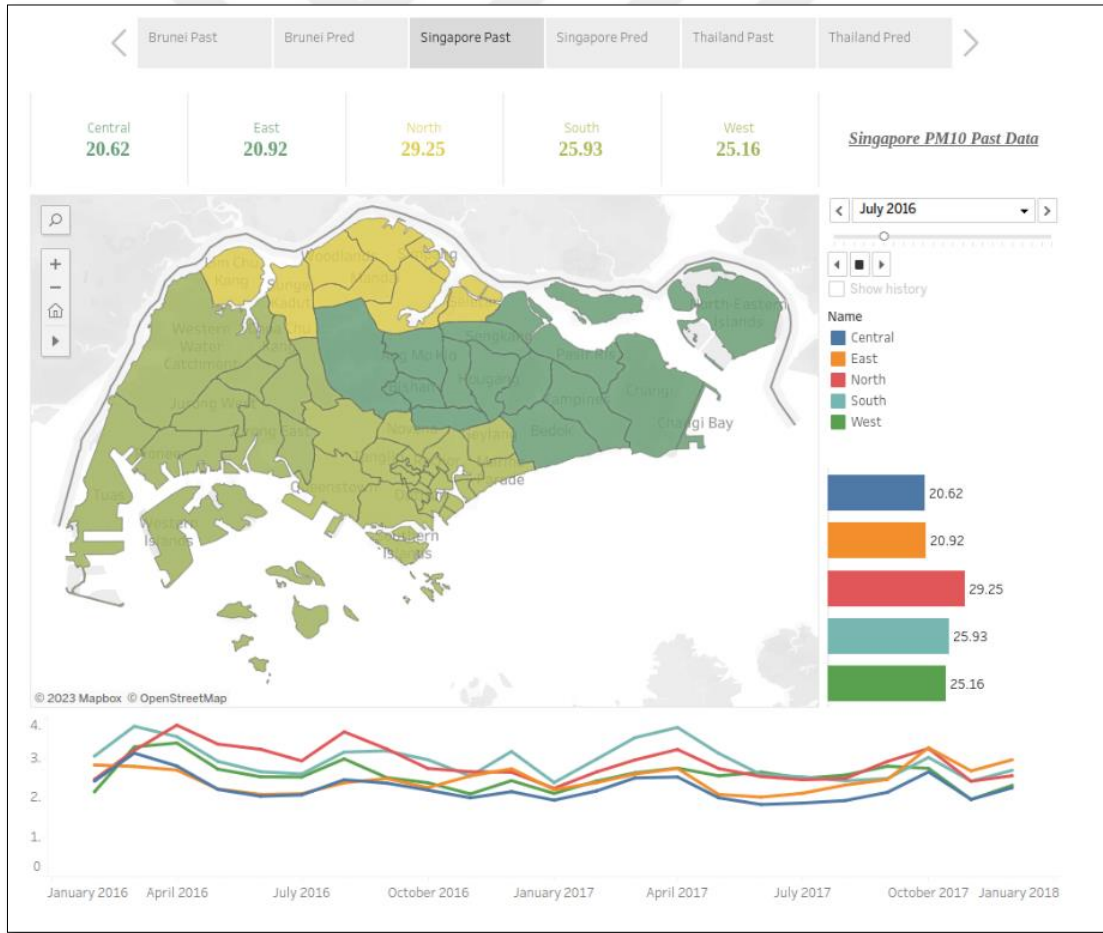


Şekil 4.6. Brunei Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli

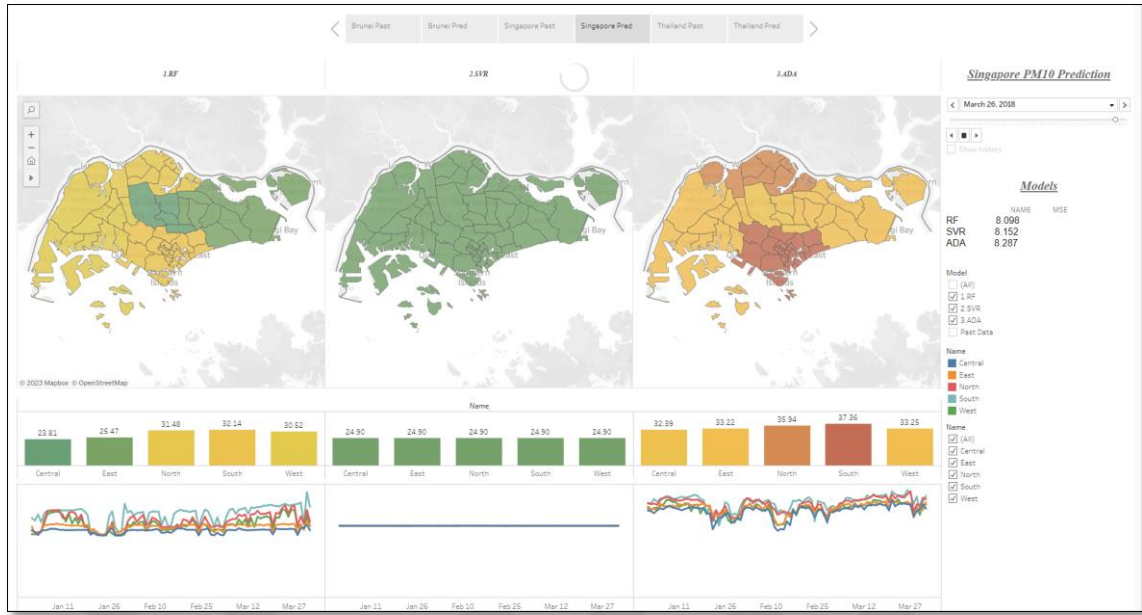
4.6.2. Singapur Bölgeleri

Singapur bölgelerine ait veriler Şekil 4.7’de ki gibi görselleştirilerek eğitim sonrası tahminleme ile geçmiş veriler hakkında daha net çıkarımlar yapılabilmektedir. Singapur için Central, East, North, South ve West etiketi ile haritada 5 bölüme ayrılmış ve değerlerin ortalamaları alınarak oluşturulmuştur. Dinamik olarak farklı tarihler ve bölgeler üzerinde incelemelerde bulunabilmektedir.

Singapur bölgeleri için Şekil 4.8’de 30 Mart 2018 tarihindeki ortalama hava kirliliği tahmin bilgisi verilmektedir. Singapur için Central, East, North, South ve West etiketi ile harita üzerinde 5 bölüme ayrılarak bu alanlar üzerinden ayrışık tahminleme yapılmıştır. Aynı zamanda dinamik olarak farklı modellerin bölgeler üzerindeki tahmin sonuçları da görülebilmektedir.



Şekil 4.7. Singapur Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli



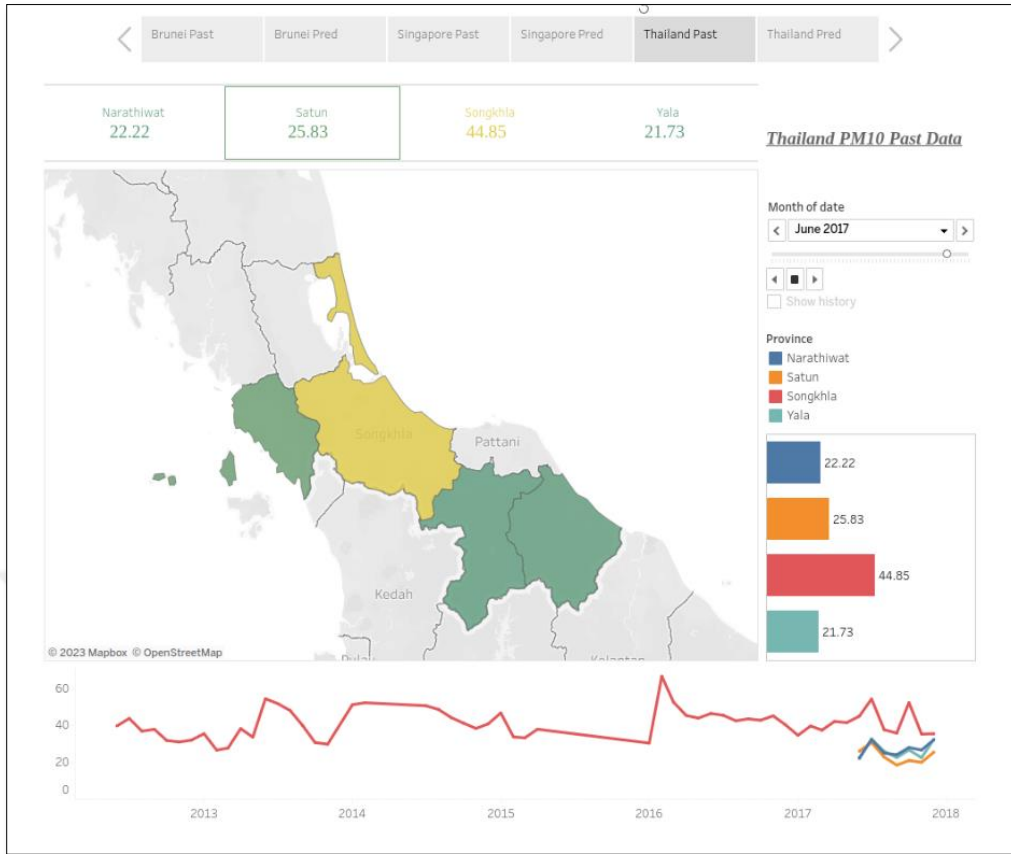
Şekil 4.8. Singapur Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli

4.6.3. Tayland Bölgeleri

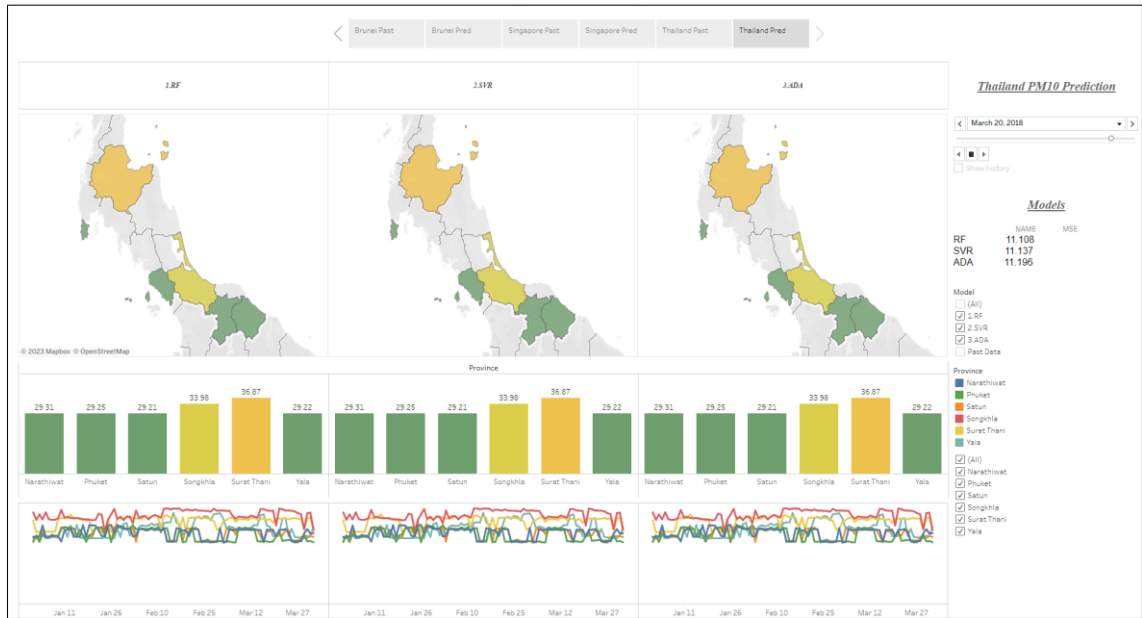
Tayland bölgelerine ait veriler Şekil 4.9'daki gibi görselleştirilerek eğitim sonrası tahminleme ile geçmiş veriler hakkında daha net çıkarımlar yapılabilir. Tayland Narathiwat, Phuket, Satun, Songkhia, Surat Thai ve Yala etiketi ile haritada 6 bölüme ayrılmış ve değerlerin ortalamaları alınarak oluşturulmuştur. Dinamik olarak farklı tarihler ve bölgeler üzerinde incelemelerde bulunabilmektedir.

Tayland bölgeleri için Şekil 4.10'da 30 Mart 2018 tarihindeki ortalama hava kirliliği tahmin bilgisi verilmektedir. Tayland için Narathiwat, Phuket, Satun, Songkhia, Surat Thai ve Yala etiketi ile haritada 6 bölüme ayrılarak bu alanlar üzerinden ayrışık tahminleme yapılmıştır. Aynı zamanda dinamik olarak farklı zamansal aralıklarda ve eğitilen diğer modellerin bölgeler üzerindeki tahmin sonuçları da görülebilmektedir. Tahmin çıktıları renklendirilerek de daha kolay anlaşılması ve yorumlanması açısından kolaylık sağlanmıştır.

Bu şekilde, diğer bölgelerde olduğu gibi Tayland'da bulunan bölgelerin hava kirliliği seviyeleri hakkında daha detaylı bilgiler elde edebilir ve farklı zaman dilimleri ve bölgeler üzerindeki değişiklikleri gözlemleyebiliriz. Bu verilerin görselleştirilmesi, hava kirliliği durumunu daha anlaşılır hale getirir ve ilgili bölgeler üzerinde daha iyi planlamalar yapmamıza yardımcı olur.



Şekil 4.9. Tayland Bölgeleri İçin PM10 Geçmiş Veri Gösterge Paneli



Şekil 4.10. Tayland Bölgeleri İçin PM10 Gelecek Tahmin Durumu Gösterge Paneli

5. SONUÇLAR VE ÖNERİLER

5.1. Çalışmanın Kısıtları

Hava kirliliğine ve kalitesine etki eden birçok etken ve sebepler bulunmaktadır. Çalışmamız kapsamında analiz ve incelemelerin yapılabilmesi için ihtiyaç duyulan verilerin istenilen şekilde temin edilmesi gerekmektedir. Araştırmalar sonucu hava kalitesi tahmini için Mediaeval tarafından verilen “Insight for Wellbeing: (sınır ötesi) Haze Tahmini için Çapraz Veri Analitiği” başlıklı kıyaslama verileri kullanılarak hava kalitesi tahmini ve buradaki veri setleri kullanılarak analizler gerçekleştirildi. İlgili çalışma sadece Brunei, Singapur ve Tayland bölgelerini kapsayan bir veri seti çalışması içermektedir. Ancak yapılan çalışma sadece bu alan ile sınırlı kalmayıp kullanılan makine öğrenmesi modelleri ve görselleştirme çalışmaları değişen veri setleri ve değişen veriler için de uyum sağlamaktadır. Yapılan çalışma ile insan sağlığı için hava kirliliğinin en doğru şekilde bulunması ve bu yönde insanların bilinçlendirilmesi neticesine ulaşıldı.

Çalışma kapsamında yapılan tahminler sonucunu nelerin etkilediği görülmüş olup ekstra farklı hava kirlleticileri veya olaylarının da eklenerek başarımlarının artırılması sağlanabilirdi. Bu konularda veri bağımlı olduğundan dolayı modeller üzerinde değişiklikler yapılarak ve var olan veri seti ön işlemlere tabi tutularak kendi içerisinde öngörülerde bulunulmuştur.

5.2. Gelecek Çalışmalar

Yapılan çalışmalar ile elimizde bulunan veriler ile hava tahmini ve kirliliği ile ilgili analizler yapılmış olup bölgesel olarak gelecek zamanlar için tahmin çalışmaları yürütülmüştür. Aynı zamanda kişilerin bu bölgelerde yaşam durumlarına göre sağlıklarının tehdit altında olup olmadıkları görseller üzerinden verilmiştir. Sonraki çalışmalar için hava kalitesini etkileyen başka etmenler ve bu etmenler ile birleştirilerek sonraki zamanlar için tahminler kuvvetlendirilebilir. Tez çalışmasında yapıldığı gibi geçmiş yıllara göre indeksleme yapılarak değerlerin ağırlık ortalamaları alınarak gelecek planlaması yapılabilmektedir. Hava kirliliğine etki eden faktörle gün geçtikçe artmakta ve bu etkenlerin tespiti de teknolojinin gelişmesi ile hassas tespiti

kolaylaşmaktadır. Hava durumları, iklim değışiklikleri, hayvan ve bitki popölasyon etkileri bu analizlere katılabilir. Bu sayede daha net tahminler ve güncel sonuçlar elde edilebilir.

Tez kapsamında analiz çalışmaları incelenilerek elde edilen sonuçlara bakılarak artış sağlanan bölgelerde artışa neden olan etkenler bulunup azaltılarak bireylerin sağlıklı yaşamlarına katkı sağlanabilir. Aynı zamanda alt yapısı kurulan makine öğrenmesi modelleri ve görselleştirmeler ile farklı ülke ve bölgelerin analizleri daha kolay yapılabilir. PM10, sıcaklık, nem, yağış, rüzgâr yönü ve rüzgâr hızları yanı sıra O₃, NO₂, CO, SO₂ gibi havayı ve çevreyi etkileyen değerler de analiz yapılarak çıktılara göre tedbirler alınabilir.

Geliştirilen görselleştirme panel arayüzü sayesinde farklı site ve sayfalarda yayınlanarak anlık takip edilebilir veya mobil uygulama ile anlık hava tahmin ve sağlık durum raporu çıkarılabilir. Yeni algoritmalar eklenerek modeller geliştirilebilir ve arayüz olarak daha kullanışlı eklentiler üzerine çalışılabilir.

KAYNAKLAR

- Adams, M. D., De Luca, P. F., Corr D. & Kanaroglou, P. S. (2013). Air Quality Health Index Mapping: A Data Driven Modelling Approach. *In Proceedings of the 13th International Conference on Environmental Science and Technology Athens*, (s. 5-7). Greece.
- Akalın, B., & Veranyurt, Ü. (2021). Sağlık hizmetleri ve yönetiminde yapay zekâ. *Acta Infologica*, 5(1), 231-240.
- Akbar, A., Tahir, M. A., Raf, M. (2021). Towards Time Series Forecasting of Cross-Data Analytics for Haze Prediction. *MediaEval*.
- Akbar, A., Tahir, M. A., & Rafi, M. (2022, September). Transboundary Haze Prediction: Towards Time Series Forecasting of Cross-Data Analytics for Haze Prediction. *In 2022 International Conference on Emerging Trends in Smart Technologies (ICETST)* (pp. 1-6). IEEE.
- Biau, G., Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
- Chen, T., & Guestrin, C. (2016). Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.
- Cortes, C., Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20,1-25.
- Freund, Y., Schapire, R. E. (1997). A Decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.
- Ghaemi, Z., Alimohammadi, A., & Farnaghi, M. (2018). LaSVM-based big data learning system for dynamic prediction of air pollution in Tehran. *Environmental monitoring and assessment*, 190, 1-17.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (NIPS 2017, 03 17). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in neural information processing systems*, 30.
- Sweeney, L., Matran-Fernandez, A., Halder, S., de Herrera, A. G. S., Smeaton, A., & Healy, G. (2021). Overview of the EEG pilot subtask at MediaEval 2021: predicting media memorability. *arXiv preprint arXiv:2201.00620*.
- Marksfeld, A. E. (2018). Machine Learning in Sociodemographic Segmentation of a Telco Company Customers. *Czech Technical University, Department of Theoretical Computer Science, Prague*.
- McCulloch, W., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.

- Metlek, S., Kayaalp, K. (2020). *Destek Vektör Makineleri*. Ankara/Türkiye: İKSAD.
- Morde, V. (2019). *XGBoost Algorithm: Long May She Reign*. Towardsdatascience: <https://towardsdatascience.com/https-mediumcom-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d> , (Ziyaret tarihi: 15 Mart 2023).
- Nguyen, H.-V., Duong, T. T. N. (2021). Insights for Wellbeing: Predicting PM10 Values Using Stacking Ensemble Model. MediaEval. 2021.
- Qian, Z., Alaa, A. M., & van der Schaar, M. (2021). CPAS: the UK's National Machine Learning-based Hospital Capacity Planning System for COVID-19.
- Ton-Thien, M.-A., Nguyen, C. T., M. Le, M., Q.Duong, D. (2021). Air Quality Estimation Using LSTM and An Approach for Data Processing Technique. MediaEval. 2021.
- URL-1: <https://www.bbc.com/turkce/haberler-dunya-58664094>, (Ziyaret tarihi: 3 Mart 2023).
- URL-2: <https://181.csb.gov.tr/hava-kirliligi-ile-ilgili-bilgi-notu-duyuru-365102>, (Ziyaret tarihi: 15 Mart 2023).
- URL-3: <https://havakalitesi.ibb.gov.tr/Icerik/istanbul-hava-kalitesi/azaltim-calismalari>, (Ziyaret tarihi: 15 Mart 2023).
- URL-4: <http://www.airquality.co.uk>, (Ziyaret tarihi: 15 Mart 2023).
- URL-5: <https://www.frontiersin.org/articles/10.3389/fmed.2021.704256/full>, (Ziyaret tarihi: 15 Mart 2023).
- URL-6: <https://webdosya.csb.gov.tr/db/zonguldak/webmenu/webmenu36983.pdf>, (Ziyaret tarihi: 15 Mart 2023).
- URL-7: https://tr.wikipedia.org/wiki/Hava_kalitesi_indeksi, (Ziyaret tarihi: 15 Mart 2023).
- URL-8: <https://ttb.org.tr/userfiles/files/KaraRapor2021.pdf?ysclid=lff9rer7ag177490425>, (Ziyaret tarihi: 17 Mart 2023).
- URL-9: https://tr.wikipedia.org/wiki/Makine_öğrenimi#cite_note-Alpaydin2020-1, (Ziyaret tarihi: 15 Mart 2023).
- URL-10: <https://bootcamp.cvn.columbia.edu/blog/machine-learning-applications/#:~:text=Machine%20learning%20is%20a%20crucial,healthcare%2C%20retail%2C%20and%20education>, (Ziyaret tarihi: 17 Mart 2023).
- URL-11: <https://data.worldbank.org/country/brunei-darussalam>, (Ziyaret tarihi: 24 Nisan 2023).

- URL-12: <https://data.worldbank.org/country/singapore>, (Ziyaret tarihi: 24 Nisan 2023).
- URL-13: <https://data.worldbank.org/country/thailand?view=chart>, (Ziyaret tarihi: 24 Nisan 2023).
- URL-14: <https://colab.research.google.com/>, (Ziyaret tarihi: 20 Mart 2023).
- URL-15: <https://www.tableau.com/why-tableau/what-is-tableau>, (Ziyaret tarihi: 20 Mart 2023).
- URL-16: <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>, (Ziyaret tarihi: 20 Mart 2023).
- Veljanovska , K., & Dimoski, A. (2018). Air Quality Index Prediction Using Simple Machine Learning Algorithms. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, 025- 030.
- Wang, W., & Yang, S. (2020, September). Research on air quality forecasting based on big data and neural network. *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, 180-184. IEEE.
- Yılmaz, V. & Şimşek, T., (2020). Hava kalitesi ile covid-19 arasındaki ilişki: G-7 ülkeleri üzerine bir uygulama. *journal of Turkish Studies*, 1353-1366.
- Zhou, Z. H (2012). Ensemble Methods: Foundations and Algorithms. *New York, USA, Chapman and Hall/CRC*.

KİŞİSEL YAYIN VE ESERLER

- Eken, S., **Şara, M.**, Satılmış, Y., Karşlı, M., Tufan, M. F., Menhour, H., & Sayar, A. (2020). A reproducible educational plan to teach mini autonomous race car programming. *The International Journal of Electrical Engineering & Education*, 57(4), 340-360.
- Karsli, M., Satılmış, Y., **Şara, M.**, Tufan, F., Eken, S., & Sayar, A. (2018, May). End-to-end learning model design for steering autonomous vehicle. In *2018 26th Signal processing and communications applications conference (SIU)* (pp. 1-4). IEEE.
- Satılmış, Y., Tufan, F., **Şara, M.**, Karşlı, M., Eken, S., & Sayar, A. (2019). CN based traffic sign recognition for mini autonomous vehicles. In *Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology-ISAT 2018: Part II* (pp. 85-94). Springer International Publishing.

ÖZGEÇMİŞ

Muhammed ŞARA, 2019 yılında Kocaeli Üniversite Bilgisayar Mühendisliği bölümünden mezun oldu. Üniversite hayatı boyunca ve mezun olduktan sonra OTOKAR, STM, AUTONOM gibi firmalarda yarı ve tam zamanlı olarak yapay zekâ ve otonom sistemler üzerine çalışmalar gerçekleştirdi. 2020 yılında kendi girişimi kurmak için ayrılarak VISIGHT Teknoloji firmasını kurdu. Bu firmada yapay zekâ teknolojilerini kullanarak çeşitli doğrulama ve tanıma cihazları geliştirmektedir. 2021 yılında Kocaeli Üniversitesi Bilişim Sistemleri Mühendisliği bölümünde yüksek lisans eğitimine başladı.

