



**T.C.
FIRAT ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ**

**YAZILIM MÜHENDİSLİĞİ
YMH 418 GÜNCEL
KONULAR DERSİ PROJE
DOSYASI
25.05.2020 – 29.05.2020**

BÖLÜMÜ : YAZILIM MÜHENDİSLİĞİ

NUMARASI : 15542525

ADI ve SOYADI : ÇAĞDAŞ KARACA

Genel Değerlendirme Raporu

İlk olarak İzmir bölgesindeki hava kalitesi verileri covid-19 virüsünün çıkış tarihinden önce olarak en yakın 1 kasım tarihinden günümüze (03.04.2020 tarihine) kadar olan zaman aralığındaki İzmir hava istasyonları genelinde alınan verisetleri incelenmiştir. Bu veri seti içerisinde Aliğa bölgesi bulunmakta olan tershane, yakınında bulunan gemi söküm tesisleri ve Petkim (Petrol ve Kimya tesisleri) firmasının havaya yaymakta olduğu kirlilik baz alınmıştır. Alsancak semtinde bulunan Gümrük ve Liman İşletmelerden kaynaklı oluşan hava kirliliği baz alınmıştır. Alsancak ve Aliğa semtlerinin ortak noktası olan Altinyol da bulunmakta olan araç trafiğinden kaynaklı oluşan egsoz ve sera gazlarının bulunmasından dolayı bu iki semt arası özellikle dikkate alınmıştır. Çiğli semtinde bulunan Organize Sanayinin bulunmasından dolayı bu semtteki datalar veri bilimi projesi için çok önemli bir yere sahiptir. Bornava semtinde Ankara asfaltı üzerinde çeşitli fabrikalar ve çimento fabrikalarının bulunmasından kaynaklı birçok zararlı atık havaya karışmaktadır. Bu veri seti sırasıyla Bayraklı, Şirinyer, Çiğli, Alsancak, Güzelyalı, Bornova ve Karşıyaka semtlerinin Hava İstasyonlarından alınan veriler düzenlenerek oluşturulmuştur. İzmir Liman ve Sanayi şehri olmasından dolayı burda bulunan işletmelerden günde yüzlerce m³ kirli, zararlı gazlar havaya karışmaktadır.

Covid-19 virüsü ne yazık ki tehlikeli bir durum olduğundan buradaki işletmelerin bir çoğu geçici olarak kapanmış yada iş akışı yavaşlatılmış olmasından dolayı hava kirliliği bir nebze azalmıştır. Datalar incelendiğinde yurt dışı çıkış yasağı getirilmesinden ve ithalat, ihracatın kısıtlanması kararından bu yana günümüze kadar olan 2 haftalık zaman zarfında hava kirliliğinde gözle görülür bir azalma gözlemlenmiştir. Bu azalma sonucunda veriseti Veri Madenciliği (Data Mining) algoritmaları kullanılarak semtler arasında gün ve saat baz alınarak oluşan kirlilik oranı ve hava kalitesi oranları argüman halinde incelenip, analiz edilerek projenin ilerleyişi hakkında daha çok bilgiye sahip olunmuştur.

Elde edilen verilere kullanılarak python dilinde yazılan veri görselleştirme aşaması hakkında araştırmalar yapılmış ve Matplotlib kütüphanesi kullanılarak elde edilen veri seti üzerinde görselleştirme işlemi için yapılan araştırmalar sonucu tamamlanmıştır.

Çevre ve Şehircilik Bakanlığı üzerinden elde edilen veri setinde bulunduran hava kirliliğinin ölçümünde ve hesaplanmasında kullanılan ana hava partikülleri hakkında detaylı araştırma yapılmıştır. Yapılan araştırmalara göre elde edilen sonuçlar aşağıdaki gibidir.

Azot dioksit (NO₂)

Havanın içersinde ne zaman bir şey yansa, Azot oksitleri oluşmaktadır. Bunun nedeni de, solumakta olduğumuz havanın temel olarak Azot (78%) ve Oksijen'den (21%) oluşması ve bunların da ortamda enerji (yanan maddelerden) bulunması durumunda birleşmeleri söz konusu olmaktadır.

En yaygın azot oksitler (genel olarak NO_x tanımlanmaktadır) azot oksit (NO) ve Azot dioksit (NO₂)'dir. Azot oksit (NO) kokusuz, renksiz bir gaz olup, içerideki yakıtın yüksek sıcaklıklarda yakılması sonucu elde edilir, örneğin otomobiller ve diğer karayolu araçları, ısıtıcılar ve pişiriciler. Hava ile temasa geçtiklerinde, derhal oksijen ile birleşmeye geçmekte ve sonucunda azot dioksit (NO₂) oluşturmaktadır.

Bu kırmızı kahverengi renkte, alev almayan ve ayırt edici kokuya sahip bir gazdır. Önemli konsantrasyonlarda oldukça zehirli olup sonucunda ciddi gecikmeli etkisi görülen akciğer hasarları ortaya çıkmaktadır. Azot gazına mazur kalmanın diğer sağlık etkileri, nefessiz kalma ve göğüs ağrılarıdır. Azot dioksit kuvvetli bir oksidasyon ajanı olup, hava ile su buharı vasıtası ile reaksiyona girerek korozif nitrik asit oluşturdukları gibi, aynı zaman da toksik organik nitratlar da oluşturmaktadırlar. Bu durumda, ağaçların, balıkların ve hayvansal yaşamın ölümüne neden olan asit yağmurlarının oluşmasına neden olmaktadır. NO₂ aynı zamanda toprak seviyesi ozon ve dumanlı sis oluşmasını sağlayan atmosferik reaksiyonlarda temel rol oynamaktadırlar.

Azot dioksit trafik-alakalı bir kirletici olup, konsantrasyonları genel olarak kırsal kesimlere oranla kentsel kesimlerde daha yüksektir.

Partiküler Maddeler (PM₁₀ ve PM_{2.5})

Partiküller aerodinamik çapları uyarınca tanımlanmaktadırlar, örneğin PM₁₀ (aerodinamik çapları 10 µm den daha küçük olan partiküller) veya PM_{2.5} (aerodinamik çapları 2,5 µm den daha küçük olan partiküller). PM'in temel komponentleri sülfatlar, nitratlar, ammonia, sodium klorit, karbon, mineral tozları ve sudur. Madde, havada askıda bulunan organik ve inorganik maddelerin katı ve sıvı partiküllerinin bir kompleks karışımıdır.

PM₁₀ genellikle toprak kabuklu malzemeleri ile karayolu aracı ve endüstriyel tesis tozu içermektedir. PM_{2.5} ise ikincil derecede oluşmuş aerosoller, yanma partikülleri ve tekrar kondense olmuş organik ve metalik buharlardan oluşmuştur. Partiküler maddelerin asit komponentleri genellikle ince partiküller şeklinde oluşmaktadır.

Daha ileri bir ayrıştırma ise partiküllerin orijinleri üzerinden birincil veya ikincil olarak sınıflandırılmaları olacaktır. Birincil partiküller atmosfere direkt olarak yayılırken, ikincil partiküller ise diğer kirleticilerin reaksiyonları sonucunda oluşanlardır. Kırsal çevrede oluşan ikincil partiküller genellikle, içersinde kükürt dioksit ve azot oksitler bulunan reaksiyonlar sonucunda oluşan sülfatlar ve nitratlar şeklindedir.

Kirletici	Ana Kaynağı	Etkisi
Kükürtdioksit(SO ₂)	Fosil Yakıt Yanması, Taşıt Emisyonları	Solunum Yolu Hastalıkları, Asit Yağmurları
Azotoksitler(NO _X)	Taşıt Emisyonları, Yüksek Sıcaklıkta Yakma Prosesler	Göz Ve Solunum Yolu Hastalıkları, Asit Yağmurları
Partikül Madde (PM)	Sanayi, Taşıt Emisyonları, Fosil Yakıt Yanması, Tarım Ve İkincil Kimyasal Reaksiyonlar	Kanser, Kalp Problemleri, Solunum Yolu Hastalıkları, Bebek Ölüm Oranlarında Artış
Karbonmonooksit(CO)	Eksik Yanma Ürünü, Taşıt Emisyonları	Kandaki Hemeogloblin İle Birleşerek Oksijen Taşıma Kapasitesinde Azalma, Ölüm
Ozon (O ₃)	Trafikten Kaynaklanan Azot Oksitler ve Uçucu Organik Bileşiklerin (VOC) Güneş Işığıyla Değişimi	Solunum Sistemi Problemleri, Göz Ve Burunda İritasyon, Astım, Vücut Direncinde Azalma

Tablo 1.Hava da tespit edilen maddeler, ana kaynakları ve maddelerin etkileri

İSTASYONLAR	Ulusal Sınır Değeri Aşan Saat Sayısı	DSÖ Sınır Değeri Aşan Saat Sayısı	Ölçüm Yapılmayan Saat Sayısı	Veri Alımı (%)	Yıllık Ortalama (µg/m ³)
İzmir - Alsancak İBB	-	-	8760	0,00	-
İzmir - Bornova İBB	0	0	237	97,29	17
İzmir -Çiğli İBB	0	0	1417	83,82	40
İzmir - Güzelyalı İBB	0	0	371	95,76	9
İzmir - Karşıyaka İBB	-	-	8760	0,00	-
İzmir -Sirinyer İBB	0	0	181	97,93	59

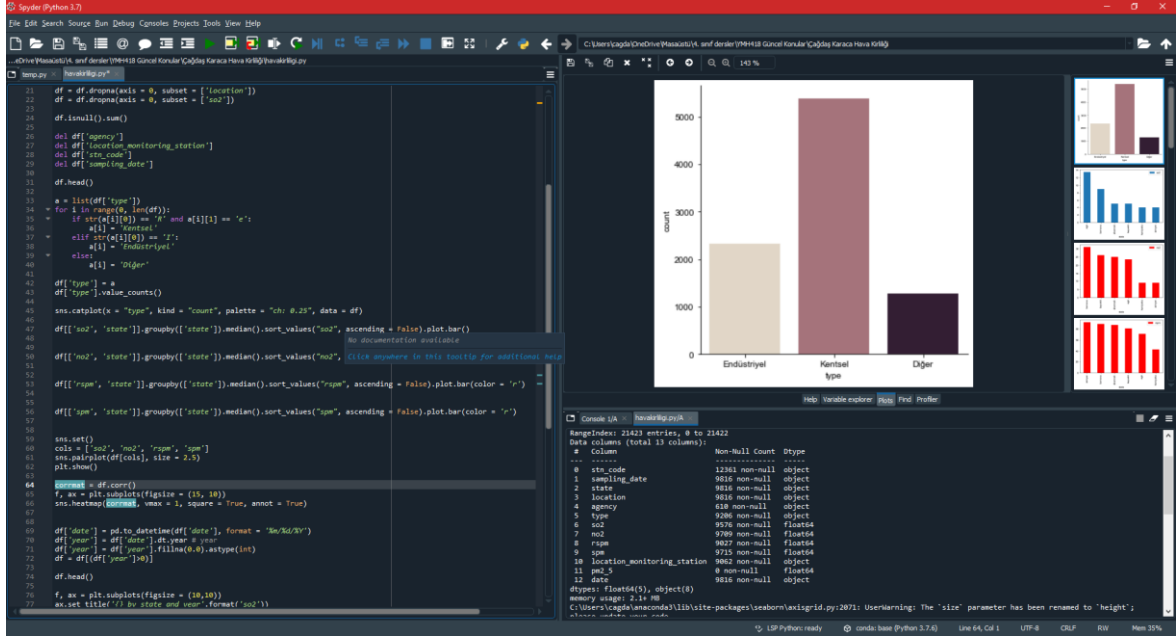
Tablo 2. İstasyonlardaki ölçümlere göre NO2 sınır değerlerini aşan, ölçüm yapılmayan gün sayıları, veri alımı yüzdeleri ve yıllık ortalamalar

Tablo 2 de bulunan veriler Çevre Mühendisleri Odasının 2018 yılı Hava Kirliliği Raporuna bakılarak elde edilmiştir. Alsancak ilçesi için projede kullanılmak üzere gerekli veriler Meteoroloji Genel Müdürlüğünden elde edilmiştir.

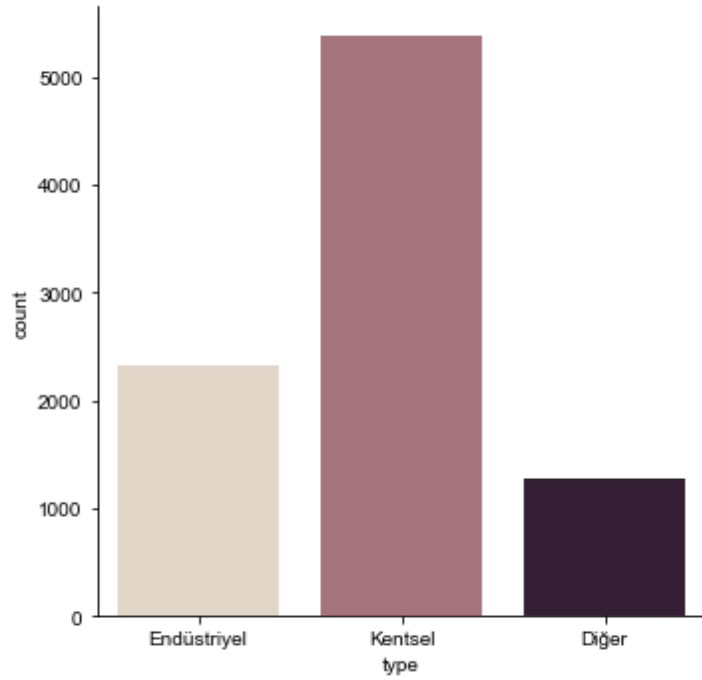
A	B	C	D	E	F	G	H	I	J	K	L	M	N
stn_code	sampling_date	state	location	agency	type	so2	no2	rspm	spm				
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,8							
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	4,8							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,3							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,4							
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	6,1							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	6,9							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	8,2							
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	6,5							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,1							
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	4,7							
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,7				133			
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,7				82			
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	5,2				111			
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,6				118			
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,1				135			
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,2				80			
351	03/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	5,9				179			
351	03/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	6,5				58			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	6,0				99			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,9				270			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,6				97			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	4,6				167			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,8				145			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,9				75			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	4,8				212			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,9				93			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,0				61			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	5,3				255			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,1				197			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	5,2				148			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	4,8				77			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	5,7				125			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	7,2				330			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	8,0				93			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	11,0				287			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	10,8				241			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	13,7				85			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	9,4							
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	8,8				283			
351	02/04/2020	izmir	izmir Bayraklı		Konut, Kirsal Alan ve Diğer Alanlar	9,2				108			
351	02/04/2020	izmir	izmir Bayraklı		Endüstriyel alan	9,3				234			

Görsel 1. Projede Kullanılan Veri Setinin Bir Kısmı

Verilerin Görselleştirilmesi için araştırmalar yapılarak verilerin gruplandırılması ve görselleştirilmesi python kullanılarak elde edilmiştir. Veri setindeki istenilen maddelerin ve bu maddeler arasındaki bağıntı, hesaplamalar baz alınarak yapılan araştırmalar sonucu bu veriler üzerinde yapılan projeler incelenmiş, alınan veri seti üzerinde değişiklikler yapıp, normalizasyon işlemleri uygulandıktan sonra proje de kullanılacak veri setlerinden çalıştırılan kod sayesinde aşağıdaki veriler elde edilmiştir.

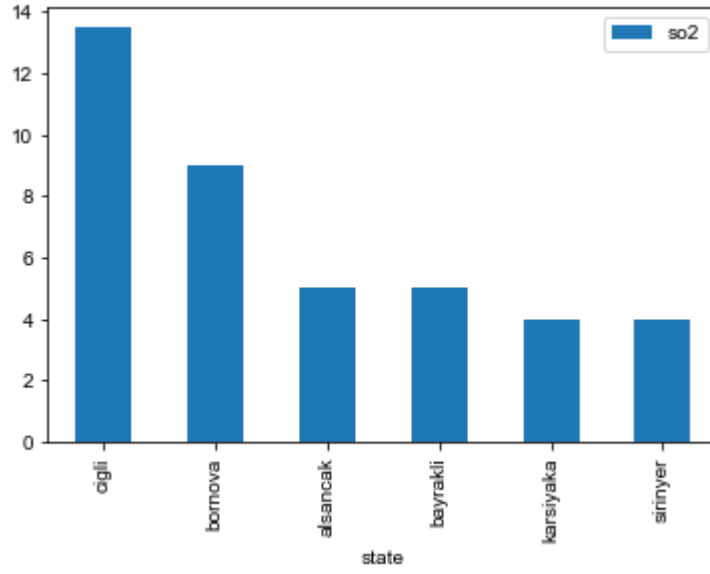


Görsel 2. Semtlere Göre Verilerin Görselleştirilmesi



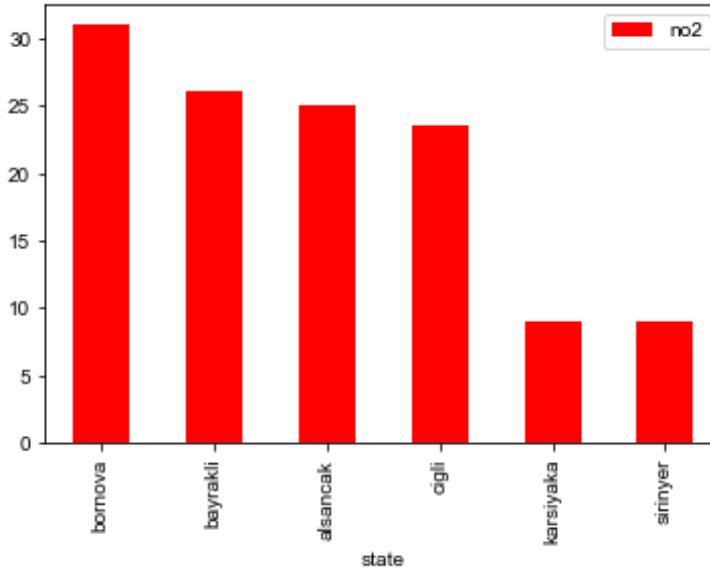
Görsel 3. Hava kirliliğine neden olan temel etkenlerin sınıflandırılması

Veriler oluşturulan csv dosyasından program üzerinde okunup kirliliğin sebep olduğu başlıca faktörlere göre sınıflandırılmıştır ve istatistiksel olarak sıralanmıştır. Genel veri setine göre kentsel kirlilik endüstriyel kirliliğe nazaran daha fazladır. Veri setinde bulunan ilçelerin tamamındaki verilerden elde edilerek bu grafik oluşturulmuştur.



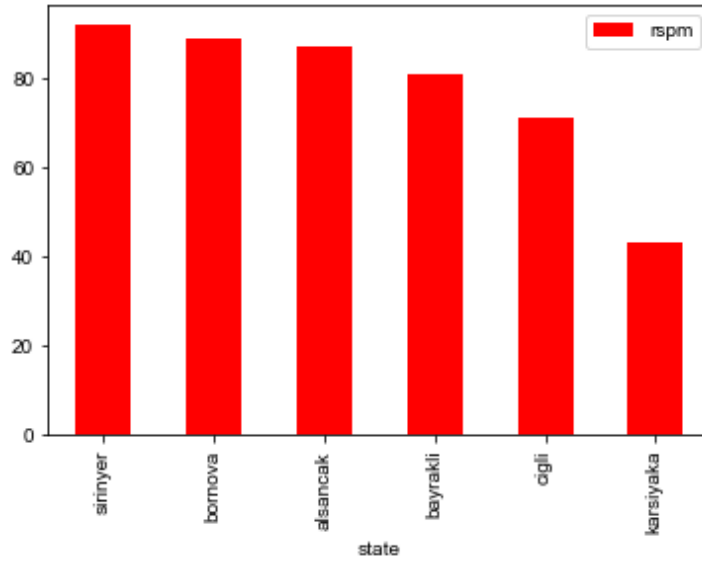
Görsel 4. Sülfür dioksit değerinin ilçelere göre sınıflandırılması

So2 değeri ilçelere göre 1990 yılından bu yana kadar olan değerlerin toplamı alınarak oluşturulmuştur. Havada bulunan sülfür dioksit gazının çiğlide bulunan Organize Sanayi Bölgesinden meydana geldiği gözlemlenmektedir. Çiğli ve Aliğa bölgesinde bulunan Petkim (Petrol ve Kimya tesisleri) havada sülfür gazının oluşmasındaki en büyük etkindir.



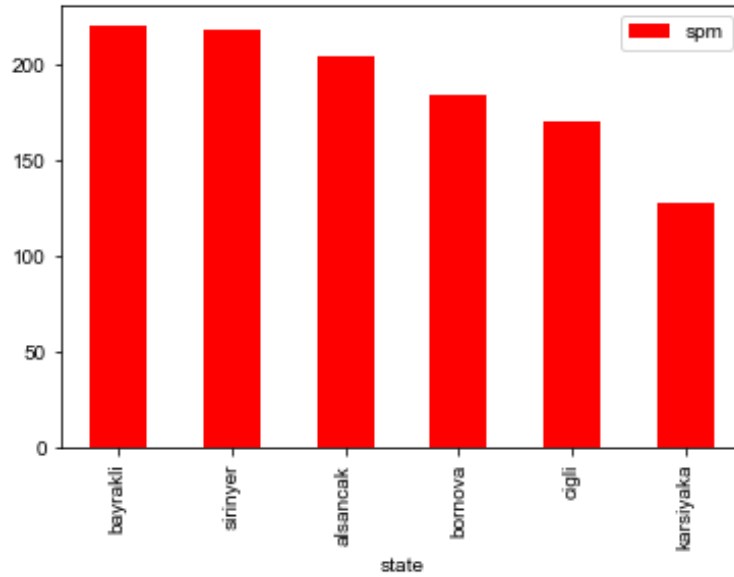
Görsel 5. Azot dioksit değerinin ilçelere göre sınıflandırılması

Bu grafikte ise No2 değerinin ilçelere göre oranı verilmiştir. No2 Kırsal kesimlerde daha yüksek orandadır. Bornova, Bayraklı gibi ilçeler deniz kenarından daha uzakta olmasından ve bu ilçelerde diğer bölgelere nazaran sanayi ve endüstriyel firmalar daha fazla bulunmaktadır.



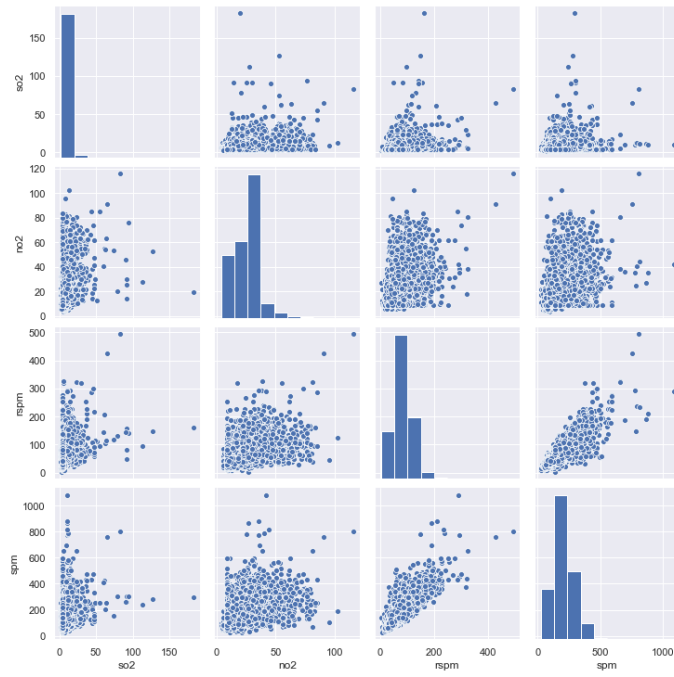
Görsel 6. Atmosferik partikül madde değerinin ilçelere göre sınıflandırılması

Solunabilir Askıda Partikül Madde (RSPM) anlamına gelen rspm daha çok kırsal alanlarda diğer sanayi şehirlerine yakın olan hava akımının daha çok olduğu bölgelerde meydana gelmektedir. Bu nedenle Şirinyer semti Hava alanı, otobanlar ve sanayi şehirlerinin merkezinde bulunmaktadır.

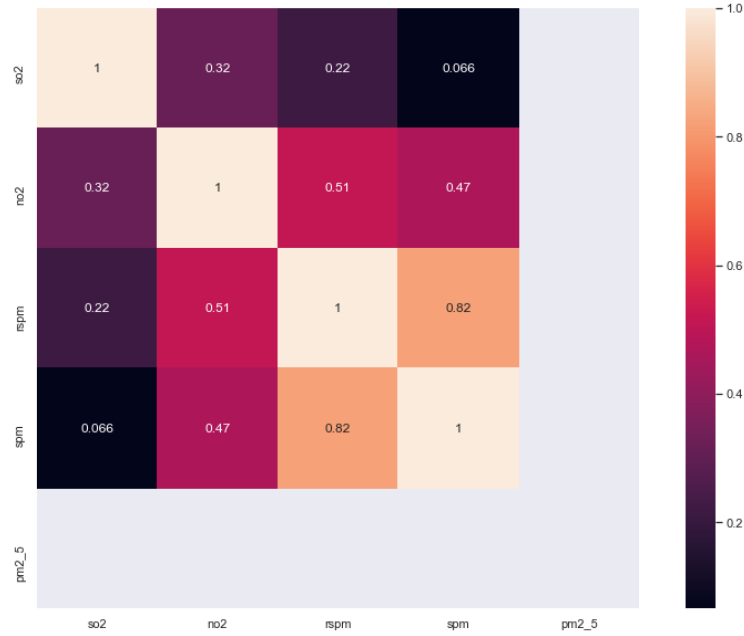


Görsel 7. Havada asılı duran parçacık madde miktarının ilçelere göre dağılımı

Hava kirleticileri gaz halindeki kirleticiler, kokular ve SPM'den (toz, duman, buğu ve duman gibi asılı parçacıklı madde) oluşur. Bunların kentsel alanlarda ve yakınında yoğunlaşması çevrede ciddi kirliliğe neden olmaktadır. İnsan kaynaklı hava kirliliğinin en büyük kaynakları enerji üretimi, ulaşım ve çok fazla enerji kaynağı kullanan endüstrilerdir. Bu nedenle çevre şehirlerde meydana gelen havadaki partiküllerin toplandığı şehirler en yüksek değere sahip olan semtlerdir.

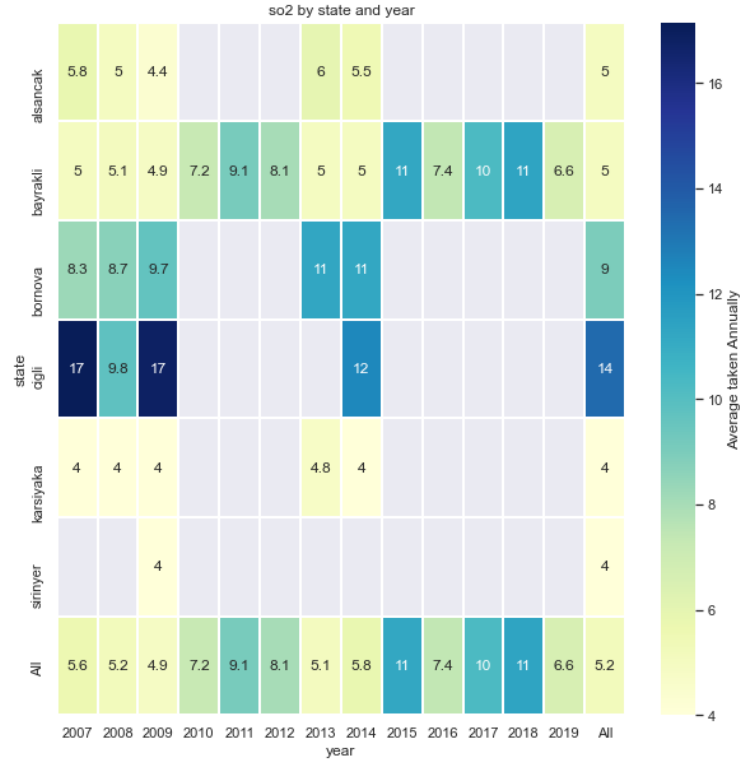


Görsel 8. Havada ölçülen SO₂, NO₂, RSPM, SPM ve CO₂ değerlerinin semtlere göre dağılımı

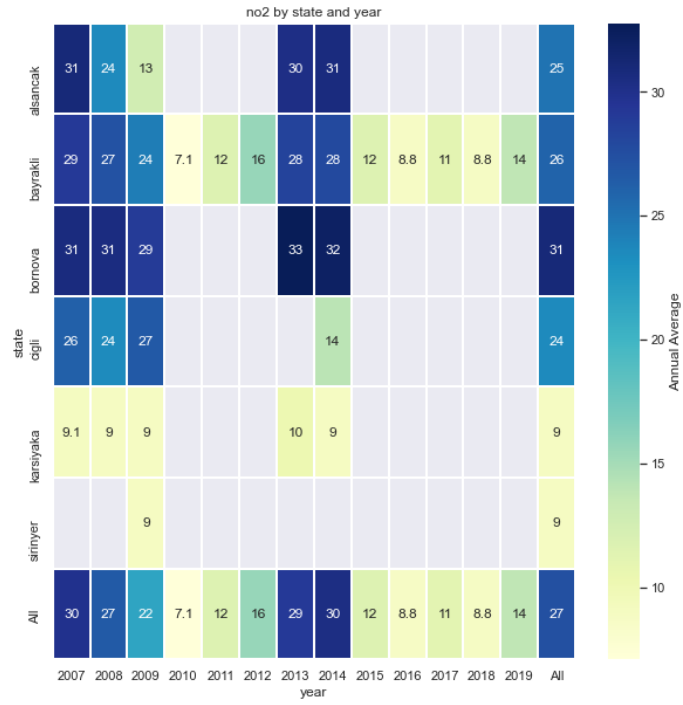


Görsel 9. Kirlilik oranına göre renklendirme grafiği

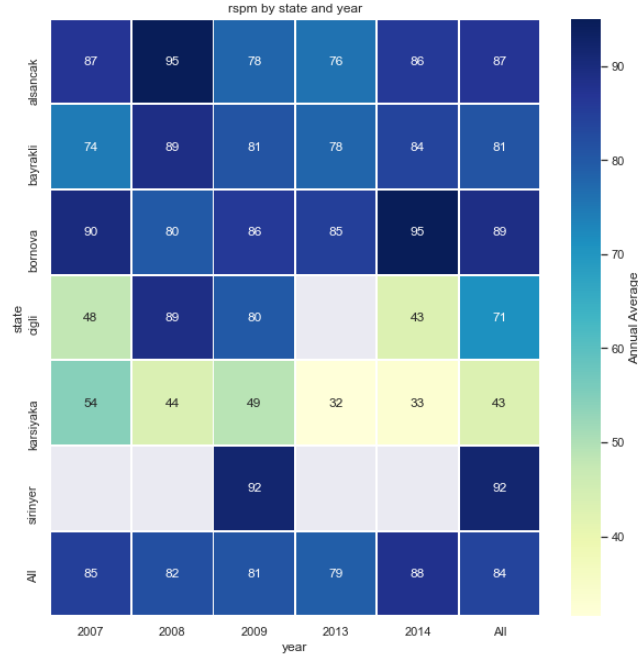
Veri setinde bulunan moleküllerin ortalama değerleri hesaplanarak kirlilik oranının ve bu oranın seviyesine göre renklendirilmesi yapılmıştır. Bu renklendirmelere göre aşağıda bulunan yıllara göre kirliliğin ne denli olduğu incelenebilmektedir.



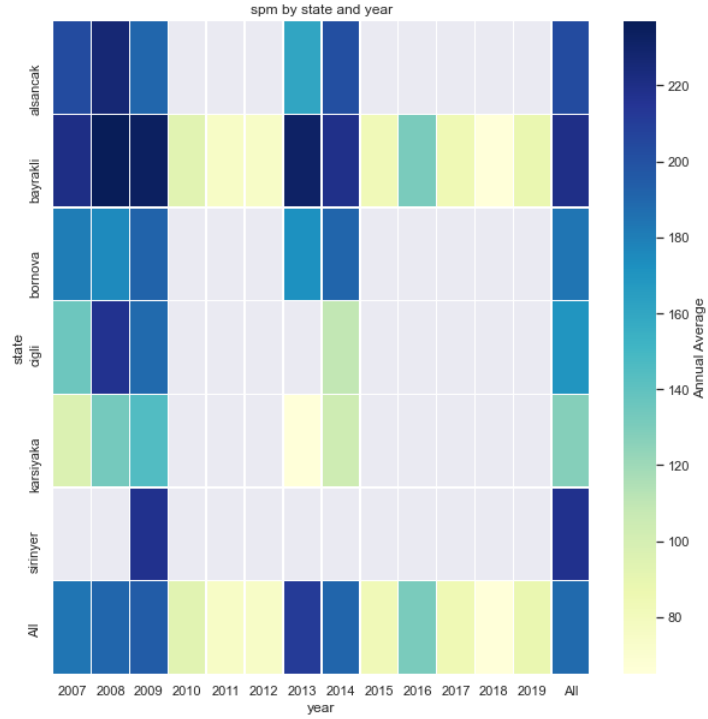
Görsel 10. So2 değerinin ilçelere ve yıllara göre dağılımı ve sınıflandırılması



Görsel 11. No2 değerinin ilçelere ve yıllara göre dağılımı ve sınıflandırılması



Görsel 12. Rspm değerinin ilçelere ve yıllara göre dağılımı ve sınıflandırılması



Görsel 13. Spm değerinin ilçelere ve yıllara göre dağılımı ve sınıflandırılması

Elde edilen görselleştirme işlemlerinden sona weka programı ile veri seti üzerinde çeşitli sınıflandırma ve kümeleme algoritmaları kullanılarak veri madenciliği uygulanmıştır. Weka programı kullanılarak veri seti üzerinde sınıflandırma işlemi yapılması için veriseti programda çalıştırılmıştır. Elde edilen veriler bölge bölge ayrılarak tüm ilçeler için SimpleKmeans Algoritması kullanılarak aşağıdaki veriler elde edilmiştir.

SimpleKmeans İle Sınıflandırma

Alsancak ilçesi için elde edilen veriler:

Küme 0: 2019/5 / 7,17,6.918762,5,98.287879

Küme 1: 2018/3 / 18,26,6.918762,4,98.287879

Attribute	Full Data	Cluster#	
		0	1
	(2200.0)	(273.0)	(1927.0)
=====			
date	2020/4/1	2020/4/10	2020/4/1
pm10	31.753	37.9653	30.8729
no2	6.9188	10.3269	6.4359
so2	4.5847	5.0981	4.512
co	98.2879	7.9969	111.0795

Kümelenmiş Örnekler

0 273 (% 12)

1 1927 (% 88)

0. Küme değeri 2019/5/7,17,6.918762,5,98.287879 1. Küme değeri 2018/3/18, 26, 6.918762, 4,98.287879 olarak alınmış ve elde edilen verilere göre 2020/4/10 tarihinden 2020/4/1 tarihi arasındaki pm10,no2,so2,co değerleri kullanılarak oluşturulan modele göre küme değerleri %88 oranla 1. Kümeye ait olduğu anlaşıyor. Yani 2020/4/1 tarihindeki elde edilen veriler 2018/3/18 tarihinde elde edilen verilerle aynı kümede bulunmaktadır.

Bayraklı ilçesi için elde edilen veriler:

Küme 0: 2019/1 / 11,48,15.285714,3,4.465116

Küme 1: 2015/8 / 14,54,15.285714,1,4.465116

Attribute	Full Data	Cluster#	
		0	1
	(2216.0)	(2198.0)	(18.0)
=====			
date	2020/4/1	2020/4/1	2020/3/15
pm10	52.6083	48.2597	583.6111
no2	15.2857	15.2857	15.2857
so2	4.4861	4.4348	10.7492
co	4.4651	4.4651	4.4651

Kümelenmiş Örnekler

0 2198 (% 99)

1 18 (% 1)

Burada elde edilen verilere bakıldığında 2019/1/11 tarihinde elde edilen veriler 2020/04/1 tarihinde elde edilen verilerin sınıfına ait olduğu anlaşılmaktadır.

Bornova ilçesi için elde edilen veriler:

Küme 0: 2018/10 / 1.125,1,3,1

Küme 1: 2016/2 / 11,29,22.371163,1,5

Attribute	Full Data	Cluster#	
		0	1
	(2170.0)	(1938.0)	(232.0)
date	2020/4/1	2020/4/1	2020/4/9
pm10	44.7906	44.3716	48.2909
no2	22.3712	14.8117	85.5188
so2	4.6589	4.8514	3.0507
co	5.5005	5.3481	6.7737

Kümelenmiş Örnekler

0 1938 (% 89)

1 232 (% 11)

Rastgele seçilen küme değerleri arasından 2020/4/1 tarihindeki değerler ile 2018/10/1 tarihinde elde edilen veriler ile aynı kümede bulunmaktadır.

Gazimir ilçesi için elde edilen veriler:

Küme 0: 2018/4 / 13,46,2,16.847695,8,3.068966

Küme 1: 2017/2 / 6,44,2,16.847695,17,3.068966

Attribute	Full Data	Cluster#	
		0	1
	(2083.0)	(1753.0)	(330.0)
date	2020/4/1	2020/4/1	2020/2/16
pm10	38.8363	36.557	50.9442
o3	2	2	2
no2	16.8477	15.5844	23.5583
so2	6.7219	5.0677	15.5091
co	3.069	2.8647	4.154

Kümelenmiş Örnekler

0 1753 (% 84)

1 330 (% 16)

2018/4/13 tarihinde elde edilen veriler ile 2020/4/1 tarihinde elde edilen veriler kümesiyle %84 benzerlik gösterdiği gözlemlenmiştir.

Çiğli ilçesi için elde edilen veriler:

Küme 0: 2014/9 / 25,22,10.516611,7,3.134775

Küme 1: 2019/5 / 29,18,11,5,3.134775

Attribute	Full Data	Cluster#	
		0	1
	(2206.0)	(330.0)	(1876.0)
date	2020/4/1	2020/4/1	2020/4/2
pm10	35.3831	31.3464	36.0932
no2	10.5166	3.2371	11.7971
so2	6.8696	6.6152	6.9144
co	3.1348	1.8893	3.3539

Kümelenmiş Örnekler

0 330 (% 15)

1 1876 (% 85)

Seilen 1. Kme deęeri 2019/5/29 tarihindeki veriler ile 2020/4/2 tarihinde elde edilen veriler %85 oranında benzerlik göstermektedir ve bu deęere bakılarak aynı kmede oldukları belirlenmiştir.

Gzelyalı ilçesi için elde edilen veriler:

Kme 0: 2014/4 / 25,36,7,1,1

Kme 1: 2019/6 / 10,28,4,4,1

Attribute	Full Data (2206.0)	Cluster#	
		0 (1494.0)	1 (712.0)
date	2020/4/1	2020/4/2	2020/4/1
pm10	37.7799	32.0301	49.8448
no2	7.1364	6.6955	8.0616
so2	5.3797	4.2641	7.7206
co	2.9478	1.8716	5.2061

Kmelenmiş rnekler

0 1494 (% 68)

1 712 (% 32)

2014/4/25 tarihinde elde edilen veriler ile 2020/4/2 tarihinde elde edilen veriler %68 oranında aynı kmede bulunmaktadır.

Karşıyaka ilçesi için elde edilen veriler:

Kme 0: 2018/10 / 3,20,8.529511,9,6.053908

Kme 1: 2019/4 / 1,16,8.529511,3,6.053908

Attribute	Full Data (2042.0)	Cluster#	
		0 (546.0)	1 (1496.0)
date	2019/10/1	2019/4/8	2019/10/1
pm10	26.3601	31.754	24.3915
no2	8.5295	9.4159	8.206
so2	5.2303	9.2857	3.7501
co	6.0539	7.7754	5.4256

Kmelenmiş rnekler

0 546 (% 27)

1 1496 (% 73)

2019/4/1 tarihinde elde edilen veriler ile 2019/10/1 tarihinde elde edilen veriler %73 oranında benzerlik gösterdiklerinden aynı kmede bulunmaktadır.

K-Means ile Kmeleme

Veri seti zerinde Weka programı kullanılarak K-means ile kmeleme işlemi yapıldı. Fold deęeri 10 olarak tanımlanmıştır. Kmeleme işlemi için veri seti zerinbu işlem sonrası elde edilen sonuçlar aşıęıda gsterilmiştir.

Yineleme sayısı: 14

Kme iindeki kare hataları toplamı: 25941.494474992956

Kme 0: 01/20/2020, izmir, 19,0.118.344

Kme 1: 01/21/2020, izmir, 14,0.112,0

Öznitelik	Küme #		
	Tam Veri (25879.0)	0 (9702.0)	1 (16177.0)
Tarih	04/02/2020	01/20/2020	01/21/2020
Durum	izmir	izmir	izmir
so2	11.8472	13.996	10.5584
no2	2.6968	1.9048	3.1718
rspm	74.9012	85.8207	68.3523
spm	91.2781	220.6888	13.6652

Kümelenmiş Örnekler

0 9702 (% 37)

1 16177 (% 63)

Elde edilen verilere bakılacak olursak spm değerine göre rastgele seçilen küme değerleriyle diğer veriler karşılaştırıldığında %63 oranında 01/21/2020 tarihinde elde edilen verilerle diğer verilerin kümelenmesi başarımlı verilmiştir. %37 oranla küme 0 daki verilerle veri setindeki diğer verilerle kümelenmiştir.

Projede yapay zeka dersinde edinilen lstm ile sınıflandırma işlemi proje üzerinde uygulanmıştır. Veri seti üzerinde okunan değerlerden state Date ve State değişkenleri kaldırılıp okunması sağlanmıştır. SimpleKmeans ile sınıflandırma işlemi üzerinde başarılı sonuç değeri doğru bulunmadığı tespit edilmiş olup üzerinde geliştirmeler yapılmıştır. Model üzerinde fl skor başarımlı ve doğruluk değerleri hesaplanmıştır. Bu işlemin kodu şu şekildedir.

```
import numpy as np
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense,Activation,LSTM,BatchNormalization,Dropout,Flatten
from keras import backend as K
```

```
veri = pd.read_csv('C:/Users/cagda/OneDrive/Masaüstü/Çağdaş Karaca Hava
Kirliliği/yapaysiniragi/data/i_Detaylar19617703.04.2020_18_19_52.csv')
label_encoder=LabelEncoder().fit(veri.Date)
labels=label_encoder.transform(veri.Date)
classes=list(label_encoder.classes_)
veri=veri.drop(["Date","state"],axis=1)
```

```
nb_classes=len(classes)
```

```
scaler= StandardScaler().fit(veri.values)
veri=scaler.transform(veri.values)
```

```
X_train,X_valid,y_train,y_valid=train_test_split(veri,labels,test_size=0.2)
```

```
y_train=to_categorical(y_train)
y_valid=to_categorical(y_valid)
```

```

X_train=np.array(X_train).reshape(20703, 4,1)
X_valid=np.array(X_valid).reshape(5176, 4,1)

model=Sequential();

model.add(LSTM(512,input_shape=(4,1)))
model.add(Activation("relu"))
model.add(BatchNormalization())
model.add(Flatten())
model.add(Dropout(0.25))
model.add(Dense(2048,activation="relu"))
model.add(Dense(1024,activation="relu"))
model.add(Dense(nb_classes,activation="softmax"))
model.summary()

def recall_m(y_true,y_pred):
    true_positives=K.sum(K.round(K.clip(y_true*y_pred,0,1)))
    possible_positive=K.sum(K.round(K.clip(y_true,0,1)))
    recall=true_positives/(possible_positive+K.epsilon())
    return recall
def precision_m(y_true,y_pred):
    true_positives=K.sum(K.round(K.clip(y_true*y_pred,0,1)))
    predicted_positives=K.sum(K.round(K.clip(y_pred,0,1)))
    precision=true_positives/(predicted_positives+K.epsilon())
    return precision

def f1_m(y_true,y_pred):
    precision=precision_m(y_true, y_pred)
    recall=recall_m(y_true, y_pred)
    return 2*((precision*recall)/(precision+recall+K.epsilon()))

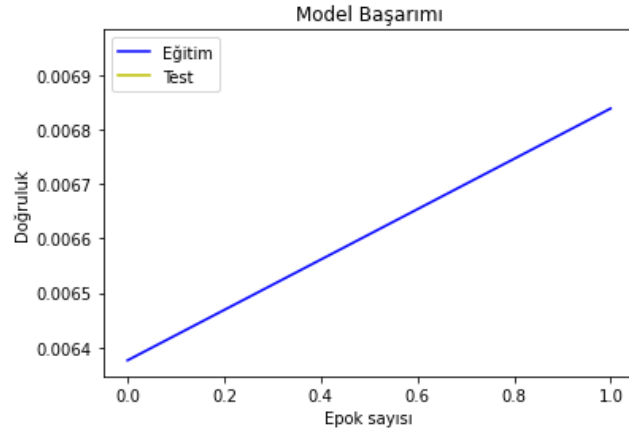
model.compile(loss="categorical_crossentropy",optimizer="adam",metrics=["accuracy",f1_m,precision_m,
recall_m])

score=model.fit(X_train,y_train,epochs=15,validation_data=(X_valid,y_valid))

print(("Ortalama Eğitim Başarısı",np.mean(model.history.history["accuracy"])))
print(("Ortalama Doğrulama Başarısı",np.mean(model.history.history["val_accuracy"])))

import matplotlib.pyplot as plt
plt.plot(model.history.history['accuracy'],color="b")
plt.plot(model.history.history['val_accuracy'],color="y")
plt.title("Model Başarımı")
plt.ylabel("Doğruluk")
plt.xlabel("Epok sayısı")
plt.legend(["Eğitim", "Test"],loc="upper left")
plt.show()

```

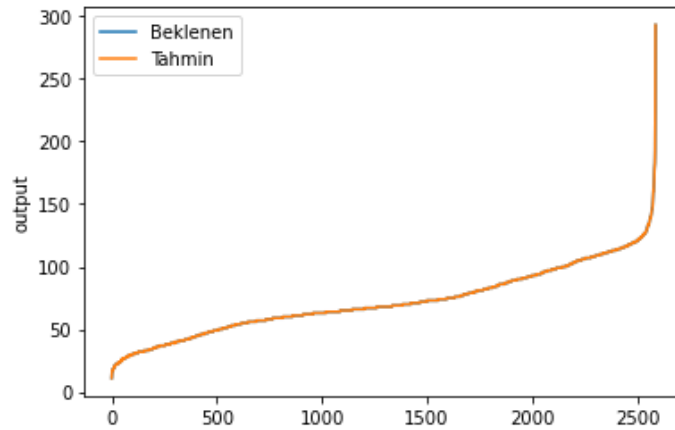


Görsel 14. LSTM ile sınıflandırma işlemi

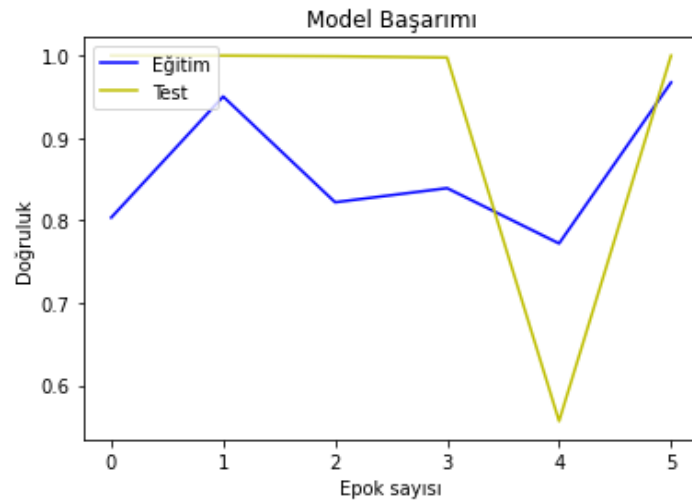
Python'da oluşturulan k means kümeleme algoritmasının yapay sinir ağı ile oluşturularak fold değeri yine 10 olarak seçilerek veri seti üzerinde yapılan kümeleme ile weka üzerinden elde edilen sonuçlar karşılaştırılırsa;

Kat puanı (RMSE): 0.044751691281048715
Final, örnek dışı skor (RMSE): 0.044751691281048715
(Ortalama Eğitim Başarısı', 0.8591075)
(Ortalama Doğrulama Başarısı', 0.9255894919236501)
(Ortalama 'Eğitim Kaybı', 6.956798865816162)
(Ortalama Doğrulama kaybı', 0.05066158099921711)

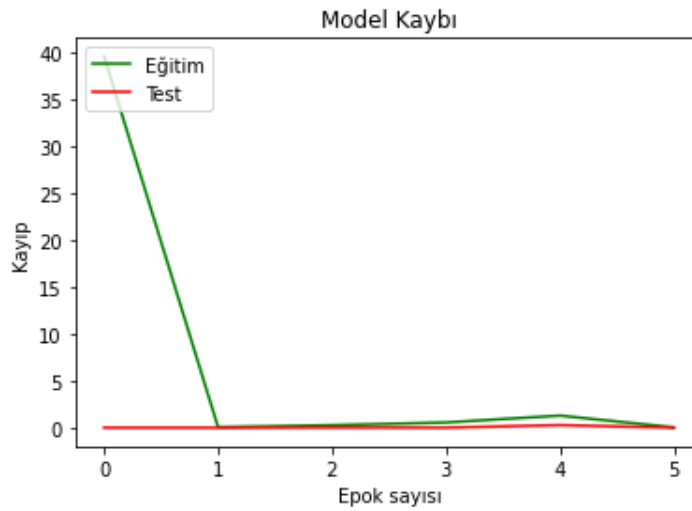
Ortalama eğitim başarımız yapay sinir ağı ile oluşturulan model ile ortalama %85 ile başarılı sonuçlandırılmıştır. Oluşturulan modelin ortalama eğitim kaybı ise %7 dir. Oluşturulan modele göre elde edilen grafik sonuçları aşağıda gösterilmiştir.



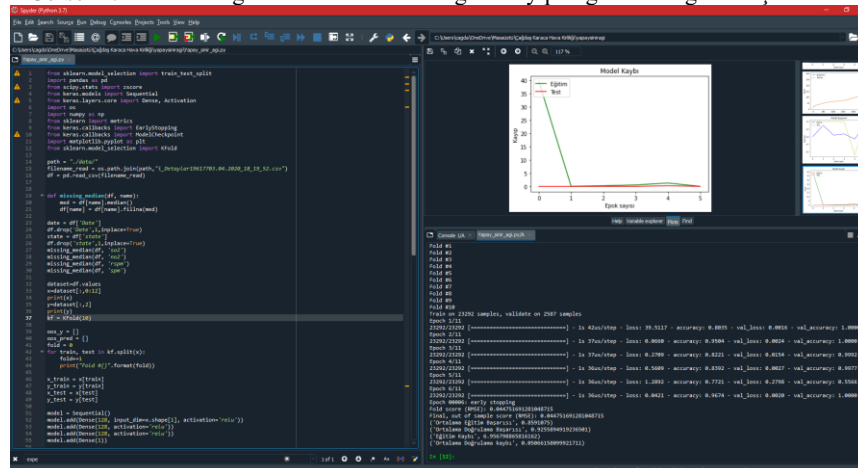
Görsel 15. Kümeleme işleminde tahmin ve beklenen değerlerin görselleştirilmesi



Görsel 16. Modelin eğitim ve test verilerine göre başarımlarının görselleştirilmesi



Görsel 17. Modelin eğitim ve test verilerine göre kayıp değerlerinin görselleştirilmesi



Sonuç:

Python'da oluşturulan algoritmanın başarımları değeri wekaya göre daha yüksektir. Bunun nedeni simple k means algoritması ile yapay sinir ağlarında kullanılan modelin el ile oluşturulmasından kaynaklanmaktadır. Wekada kullanılan algoritmadaki model programın üreticileri tarafından en optimize çalışacak şekilde programlanmıştır. Python'da ise biz belirli bir veri seti üzerinde en iyi başarımları elde edecek bir model oluşturduk. Buradaki başarımları weka ya göre daha yüksek olmasındaki en büyük neden budur. Lstm kullanılarak oluşturulan modelde başarımları oranı 0,0065046693 olarak hesaplanmıştır. Veri setinin yapay sinir ağları ile işlenip modellenmesi işlemi tamamlanmış ve bu işlem başarımları 0,8591075 olarak hesaplanmıştır.

Olası Risk ve B Planı

Proje üzerinde uygulanan veri madenciliği uygulamaları, sınıflandırma ve yapay sinir ağları ile oluşturulan veri bilimi uygulamaları yapılmıştır. Veri seti üzerinde kullanılan verilerin doğruluğu bakımından yanlış yada null veri girildiğinde sınıflandırma işleminde programın bazı noktalarda hatalı sonuç elde edebileceği düşünüldüğünde eksik verilerin tamamlanması için araştırmalar yapılacaktır. Hava kirlilik raporlarının Covid-19 virüsü tedbirleri kapsamında alınan yönetmelik ve kanunlar çerçevesinde son zamanlarda kirlilik oranı farkedilebilir derecede azalmıştır. Fakat virüsün hayatımızdan yavaş yavaş çıkmasına bağlı olarak bu tedbirlerin gevşetilmesine bağlı olarak farklı faktörlere göre hava kirliliği oranının artacağı düşünülmektedir. Bu artışın incelenmesine göre bugüne kadar yapılmış tüm aşamalar tekrarlanarak kirlilik düzeyinde artış olup olmadığı tespit edilecektir. Yıllara göre havanın ne kadar temiz ne kadar kirli olduğunu yukarıdaki istatistiklere bakılarak anlaşılabilir. Bu oranlara bakılarak dışarı çıkılıp çıkılmayacağının bilgisi alınabilir. Gerçek zamanlı veriler elde edilerek bu grafiğin oranları baz alındığında kullanıcının dışarıya çıkmadan ne gibi önlemler alabileceği yada hava kirliliğine neden olan firma yada tesislerin bacalarından çıkan dumanların çevreye ne denli zararı/ etkisi olduğu anlaşılabilir. Bu sonuçlar ile firma yada tesislere uyarıda bulunularak hava kirliliğinin önüne geçilebilmektedir.

Saygılarımla,

15542525 Çağdaş Karaca