**YMH418 - Yazılım Mühendisliği Güncel Konular - Rapor 9**

1. **Giriş**

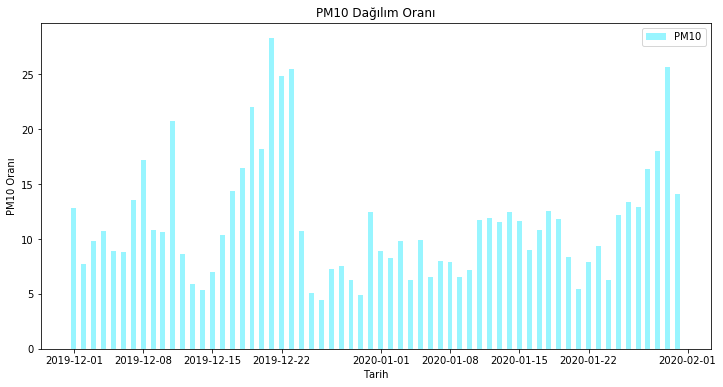
Bu hafta proje için önceki haftalarda yapılan tüm raporların genel bir değerlendirilmesi yazıldı. Bununla birlikte proje için ufak bir yol haritası değişikliği yapıldı. Bu rapor içerisinde genel rapor ve B planı içerecek çalışmalar yazılmıştır.

1. **Genel Rapor**

Dönem başından beri üzerinde durduğumuz veri bilimi ve veri analizi konuları için Çevre ve Şehircilik Bakanlığı tarafından bize sunulan hava verileri sitesi üzerinden bilgiler topladık.

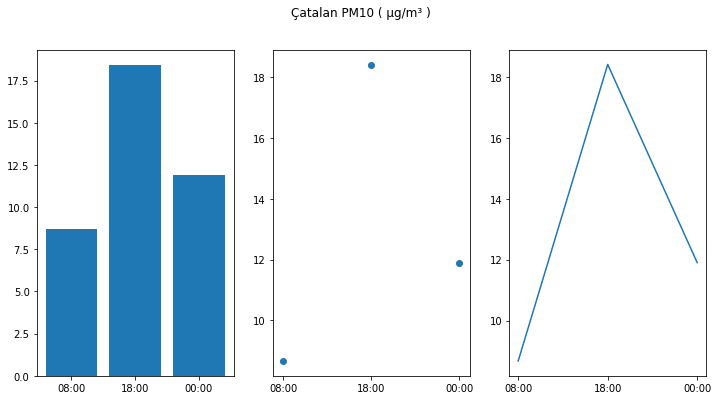
Birinci aşama olan veri seçimi için uygun bir lokasyon ve seçilen lokasyon içerisinde bulunan hava ölçüm istasyonları seçildi. Seçilen ölçüm istasyonları kırsal ve kentsel olarak düşünüldü. Adana-Çatalan ve Adana-Valilik olarak seçildi. Verilerin tarih aralığı ilk olarak Aralık 2019-Ocak 2020 olarak seçildi. Daha sonra istasyon içindeki tutulan veriler hava kirliliği oluşturan gazlar olarak seçildi. Bunlar sırasıyla Partiküler madde, Kükürt dioksit, Ozon, Azot dioksit ve Karbon monoksit olarak düşünüldü. Veriler elde edildikten sonra boş alanlar bulunuyordu. Proje için uygun dil olan Python sayesinde KNN algoritması ile dolduruldu.

İkinci aşamada ise doldurulan verilerden sonra gerekli görselleştirmeler yapıldı. Bu görselleştirmeler Matplotlib kütüphanesi sayesinde gerçekleştirildi. Her bir veri grubunu kendi arasında seçilen tarih aralığı içinde görselleştirildi. Şekil 1 üzerinde örnek olarak Adana-Çatalan bölgesinin Partiküler madde dağılımı gösterildi.



Şekil 1. Adana-Çatalan PM10 Dağılımı

Üçüncü aşama olan Karakteristik analiz aşamasında ise verilerin ilk olarak ayrılması yapıldı. Ayrılma işlemini proje için insanların işe gidiş saati, işten dönüş saati ve herkesin evde olduğu saat olarak düzenlendi. Bu verilerin ayrılma işlemini ise numpy kütüphanesi sayesinde gerçekleştirildi. Ayırma işleminden sonra ikinci aşamada öğrenilen Matplotlib sayesinde görselleştirme yapıldı (Şekil 2).



Şekil 2. Adana-Çatalan Karakteristik Analiz PM10

Bununla birlikte hava izleme sitesinde öğrenilen bilgiler doğrultusunda Partiküler madde analizi saatlik değil daha çok günlük veriler üzerinden analiz yapıldığı öğrenildi. Veri seti düzenlendikten sonra Matplotlib sayesinde görselleştirildi. Son olarak korelasyon matrisi oluşturuldu. Korelasyon matrisi ise veri setinde bulunan alanların aralarındaki ilişkiyi gösteren matris olarak bilinmektedir.

Dördüncü aşamada ise model kurulması için öncelikle Yapay Zekâ dersinde öğrenilen bir bilgiye dayanarak veri seti genişletildi. Veri seti Ocak 2015-Aralık 2019 arasına çekildi. Veri seti değişkenlik gösterdiği için tek bir saat üzerinden model kurulması tercih edildi. Bunun için en uygun saat 18:00 olarak düşünüldü. Temizleme işlemi için numpy sayesinde sadece 18:00 verileri bırakıldı. İlk seçilen veriler eğitim, 2020 Ocak ayı ise test olarak düşünüldü. Son olarak çok katmanlı yapay sinir ağı tercih edildi.

Beşinci aşamada ise kurulan modelin çalıştırılması ve tahmin edilmesi hedeflendi. Geçilen zamana göre daha herhangi bir test yapılmamıştır.

1. **B Planı**

Bu plan içerisinde model kurma aşamasında karşılaşılan sorunlar için düşünüldü. Önceki haftalarda kurulan modelde sadece bilinen verilere göre eğitilen değer soruluyordu. B planı içerisinde ise 2015 Ocak -2020 Mayıs arası tüm verileri sisteme yükleyip daha sonra eksik verileri KNN sayesinde doldurup eğittiğim alana göre ilerleyen tarihlerde nasıl bir grafik oluşturulması düşünüldü. Her bir alan için değil daha çok Ozon üzerinden testler yapıldı. Ozon gazı insanların kullandığı motosiklet kullanırken, fabrika çalışırken vb. dışarıya verdiği zehirli gazlardır. Bu plan ile birlikte önceki model tamamen github içerisine yüklenecektir.

1. **Sonuç**

Sonuç olarak Yazılım Mühendisliği Güncel Konular dersi için geliştirdiğim tüm aşamalar veri bilimi ve veri analizi için uygun katkılar sağlamıştır. Proje için her iki model hem bilinen veriler hem de bilinmeyen veriler üzerinden testler yapılmıştır.