**Yazılım Mühendisliği Güncel Konular I. Aşama Rapor**

1. **Veri Seti**

Çalışmaya başlandığında veri seti oluşturmak için büyüdüğüm şehir olduğu için Kayseri seçilmiştir fakat eksik verilerin çok fazla olması nedeniyle, yine İç Anadoluda, daha merkezi ve daha büyük olduğu için daha fazla veri olacağını düşündüğüm Ankara ile değiştirilmesine karar verilmiştir. Ankarada 4 Kentsel, 1 Trafik, 1 Kırsal ve 1 Sanayi olmak üzere toplam 7 tane istasyon bulunmaktadır, bu istasyonlardan 12 saatte 1 gelen veriler toplanmıştır. Kırsal İstasyon tipinden gelen verilerin ölçüm değerleri diğelerine göre daha farklı olduğu için çalışmadan çıkarılmaya karar verilmiştir.

1. **Veri Setinde Kullanılan Özellikler**

3 istasyondan gelen veriler arasındaki benzerlik göz önüne alınarak tüm veriler birleştirilmiş ve 3 adet yeni özellik eklenmiştir. Bunlar; ölçümün günün hangi bölümünde yapıldığı, ölçümün haftaiçi mi yoksa haftasonu mu yapıldığı ve verinin hangi istasyon tipinden geldiği bilgisidir.

İlk olarak ölçüm veri setinden tamamen boş olan satırlar çıkarılmıştır. Bu düzenleme yapılırken hava sıcaklığında olağan dışı yükseklikte ve düşüklükte ölçümler olduğu farkedilmiştir bu değerler öncelikle boş bırakılıp değerlendirmeye alınmamış, daha sonra ortalama değer ile değiştirilmiştir. Daha detaylı düzenlemeler verilerin karakteristik incelemesi yapıldıktan sonra 3. Aşamada yapılacaktır.

1. **Çalışmanın Amacı**

Bu çalışmada amaçlardan biri insanların çalışmasının veya tatilde olmasının hava kirliliğine etkisinin ölçülmesi ve yeni gelen bir verinin istasyon tipinin tahmin edilmesidir. Şu an için 3 tane sınıf vardır. Eğitim sonrası yeni gelen veri bu 3 sınıftan birine dahil edilecektir.