**CSE 619-Final Project**

**Son Gönderim Tarihi: 12.01.2021 Saat 23:00**

## Kullanılacak VeriSeti: Wine-Quality Dataset. Bu veri setinde 2 farklı sınıfa (White, Red) ait veriler bulunmaktadır. Her kayıtta bir örneğe dair 11 öznitelik ve de 12. öznitelik olarak alınan örneğin kalitesi (nümerik) yer almaktadır.

## Öznitelikler: "fixed acidity";"volatile acidity";"citric acid";"residual sugar"; "chlorides"; "free sulfur dioxide"; "total sulfur dioxide"; "density"; "pH"; "sulphates"; "alcohol"; "quality"

**İstenilen:** Bu veri setini kullanarak aşağıda belirtilen maddeleri python ortamında gerçeklemeniz ve her madde sonucunda elde ettiğiniz çıktıları yorumlamanız beklenmektedir. Proje neticesinde hazırlayacağınız raporda akademik yazım kurallarına uymanız beklenmektedir. Proje sonucunda raporu ve projeyi hazırlayacağınız Jupiter Notebook çalışmanızı paylaşmanız beklenmektedir.

1. Her özniteliğin dağılımını boxplot ile çizerek yorumlayınız.
2. Correlasyon değerlerini hesaplayarak analiz ediniz.
3. Her özniteliğin Fisher Uzaklığını (sınıflar arası) hesaplarak analiz ediniz.
4. PCA ile veri setini dönüştürüp her Eigen vektöre üzerindeki izdüşümünü bir öznitelik olarak değerlendirin ve her özniteliğin Fisher Uzaklığını hesaplayın ve 3. Maddede elde edilen Fisher Uzaklıkları ile kıyaslayınız. Eigen değerleri (Eigen Value) ile hesapladığını Fisher uzaklıkları arasında bir ilişki var mıdır? Yorumlayınız.
5. Verinin en önemli iki Eigen vektör üzerindeki izdüşümünü scatter plot ile çizerek yorumlayınız. Scatter plot işleminde her sınıfa (red/white) ait özneği farklı renkte gösterin. Aynı işlemi en önemsiz 2 Eigen vektör için tekrarlayınız.
6. Veri setini K-Means, K-Centers, SOM ve DBSCAN algoritması ile gruplayınız ve sonuçları gösteriniz. K-Means, K-Centers ve DBSCAN sınıflandırmayı 2 boyutta ve en önemli 2 eigen vectörü üzerindeki izdüşümünü dikkate alarak gerçekleyiniz ve sonuçları görsel olarak gösteriniz. Bu işlemlerde “quality” özniteliğini dikkate almayınız ve K değeri olarak max quality değerini ele alınız. SOM algoritmasında orijinal veri setini dikkate alınız (quality değerini çıkararak).