

# HALİÇ ÜNİVERSİTESİ

## Mühendislik Fakültesi

## Bilgisayar Mühendisliği Bölümü

**BİM 399 Endüstri Stajı II**

**Staj Raporu**



**Ad Soyad : Emin BAŞTÜRK**

**Öğrenci Numarası : 57133075450**

|  |  |
| --- | --- |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **Iris Veri Seti**  Iris veri seti, İngiliz istatistikçi, öjenikçi ve biyolog Ronald Fisher’ın 1936 tarihli makalesinde kullandığı çok değişkenli bir veri kümesidir. Iris veri seti, üç Iris türünün (Iris setosa, Iris virginica ve Iris versicolor) her birinden 50 gözlem birimi olmak üzere toplamda 150 gözlem birimi içerir. Her gözlem birimi için dört farklı özellik ölçülmüştür: çanak yaprak uzunluğu (sepal length), çanak yaprak genişliği (sepal width), taç yaprak uzunluğu (petal length) ve taç yaprak genişliği (petal width). Özellikler santimetre cinsinden ölçülmüştür.  **Iris Veri Setinin Elde Edilmesi**    **Min – Max Normalizasyonu**  Değişkenlerin ortalaması (μ = 0) ve standart sapması (σ = 1) üzerinden ölçeklendirilen değişkenleri [0,1] aralığına getirilen, minimum ve maksimum değerin sırasıyla 0 ve 1 olduğu normalizasyon yöntemidir. Birbirinden farklı pek çok verinin, normalize edilerek kendi içinde karşılaştırılmasını sağlar.  **Min – Max Normalizasyonunun Matematiksel Formülü**  **Min – Max Normalizasyonunun Veri Setine Uygulanması** | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **Gerçek Veri ve Min – Max Normalizasyonu Uygulanan Veri** | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **PCA (Principal Component Analysis)**  Çok değişkenli verinin ana özelliklerini daha az sayıda değişken/bileşen ile temsil etmektedir. Başka bir deyişle PCA’nın amacı küçük miktarda bir bilgi kaybını göze alıp değişken boyutunu azaltmaktır.  **Varyans**  Verilerin dağılımının belirler.  **Kovaryans Matrisi**  İki değişkenin birlikte değiştiği seviyeyi gösterir.  **Özvektör ve Özdeğerler**  Özvektörler maksimum varyans yönünü gösterir. Ona karşılık gelen özdeğerler ise karşılık gelen özvektörün önemini gösterir.  **PCA Yöntemininin Uygulanma Adımları**   1. Ortalamanın X’ten çıkarılması. 2. Cov (X, X) değerinin hesaplanması. 3. Kovaryans matrisinin özvektörlerinin ve özdeğerlerinin hesaplanması. 4. Özvektörlerin, özdeğerlerine göre büyükten küçüğe sıralanması. 5. İlk k özvektörünün seçilmesi ve orijinal n boyutlu veri noktalarının k boyutlarına dönüştürülmesi | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **PCA Modülünün Oluşturulması** | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **PCA Modülünün Oluşturulması**  **LDA (Linear Discriminant Analysis)**  Amaç, bir veri kümesinin özelliklerinin bir doğrusal birleşimini bularak veriyi sınıflara ayırmaktır.  **Matematiksel Model**  1- Sınıf içi dağılım matrisi  2- Sınıflar arası dağılım matrisi  **Özdeğer ve Özvektör**  Aşağıdaki denklem için özdeğer ve özvektör hesaplanmalı  **LDA Yöntemininin Uygulanma Adımları**   1. ve değerinin hesaplanması. 2. denkleminin özdeğerlerinin hesaplanması. 3. Özvektörlerin, özdeğerlerine göre büyükten küçüğe sıralanması. 4. İlk k özvektörünün seçilmesi ve orijinal n boyutlu veri noktalarının k boyutlarına dönüştürülmesi | |

|  |  |
| --- | --- |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **LDA Modülünün Oluşturulması** | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **PCA Yöntemi Uygulanan Min-Max Normalizasyonlu Verinin Gösterilmesi** | |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **LDA Yöntemi Uygulanan Min-Max Normalizasyonlu Verinin Gösterilmesi** | |

|  |  |
| --- | --- |
| **Proje** | **Iris Veri Setinin Min-Max Yöntemine Göre Normalize Edilmesi ve PCA ile LDA Yöntemlerinin Uygulanarak Elde Edilen Sonuçların Karşılaştırılması** |
| **PCA ve LDA Yöntemleriyle Elde Edilen Sonuçların Karşılaştırılması**  LDA: İki sınıf arasındaki farkı en üst düzeye çıkaran yönü bulur.  PCA: Verilerdeki varyansı en üst düzeye çıkaran yönü bulur.  LDA, Bağımsız ve bağımlı değişken arasındaki ilişkinin gücüne dayalı olarak değişken azaltma için kullanılır.  PCA, Bağımsız değişkenlerin doğrusallığına bağlı olarak değişkenleri azaltmak için kullanılır.  PCA gözetimsiz öğrenme algoritmasıdır. LDA gözetimli öğrenme algoritmasıdır. Gözetimsiz öğrenme, gözetimli öğrenmeden farklı olarak, verileri sebep-sonuç ya da giriş-çıkış şeklinde etiketlenmeden, veri içerisinde var olan ilişkilerin ve yapıların öğrenmesidir.  PCA kümeleme problemlerinde kullanılırken, LDA sınıflandırma problemlerinde görülür.  Öncesinde minimum maximum yöntemi ile normalize edilen verinin üzerine uygulanan bu yöntemler [0,1] arasına yerleştirilmektedir.  Grafiklerde görüldüğü üzere her iki yöntem de başarılı bir şekilde uygulanmıştır. Verileri gözlemlediğimizde setosa ve virginica iris çiçeklerinde görülen veriler ile versicolor iris çiçeğinde görülen veriler arasında bir zıtlık bulunmaktadır. Aynı şekilde iki yöntem arasında da bu veriler üzerinden karşılaştırma yaptığımızda bir zıtlık görülmektedir. Bu zıtlığın sebebi PCA’nın veri noktaları arasındaki mesafeyi maksimize etmeye çalışırken LDA’nın sınıflar arasındaki mesafeyi maksimize etmeye çalışmasındandır. | |