Ad Soyad: Ertuğrul Sunğur

Öğrenci No: 20120205004

# Veri Madenciliği Ödev 1 Raporu

## **Giriş**

İlk olarak ödevde istenilen modellemelerin test edilebilmesi için ihtiyaç duyulan tüm kütüphaneler import edilmiştir.

Ödevde verilen datasetin, istenilen modellerle eğitilmesi ve istenilen testlerin yapılabilmesi için dataset okunmuş ve ön işlemeleri yapılmıştır. Dataset pathlerinin doğru şekilde kombine edilip ulaşılabilmesi için Python “os” kütüphanesinin fonksiyonlarından yararlanılmıştır. Pathlerin kombine edilmesinden sonra fotoğraflar, cv2 kütüphanesi yardımıyla sırayla okunmuştur. Okunan fotoğraflara resize, 1D array çevirme ve normalize işlemleri uygulanmıştır. İşlenmiş fotoğraflar ve kategoriyleri listelere kaydedilmiştir. Ön işlemenin son aşaması olarak data, train ve test olmak üzere 2 kümeye ayrılmıştır. Ayırma işlemi için sklearn kütüphanesinin fonksiyonlarından yararlanılmıştır.

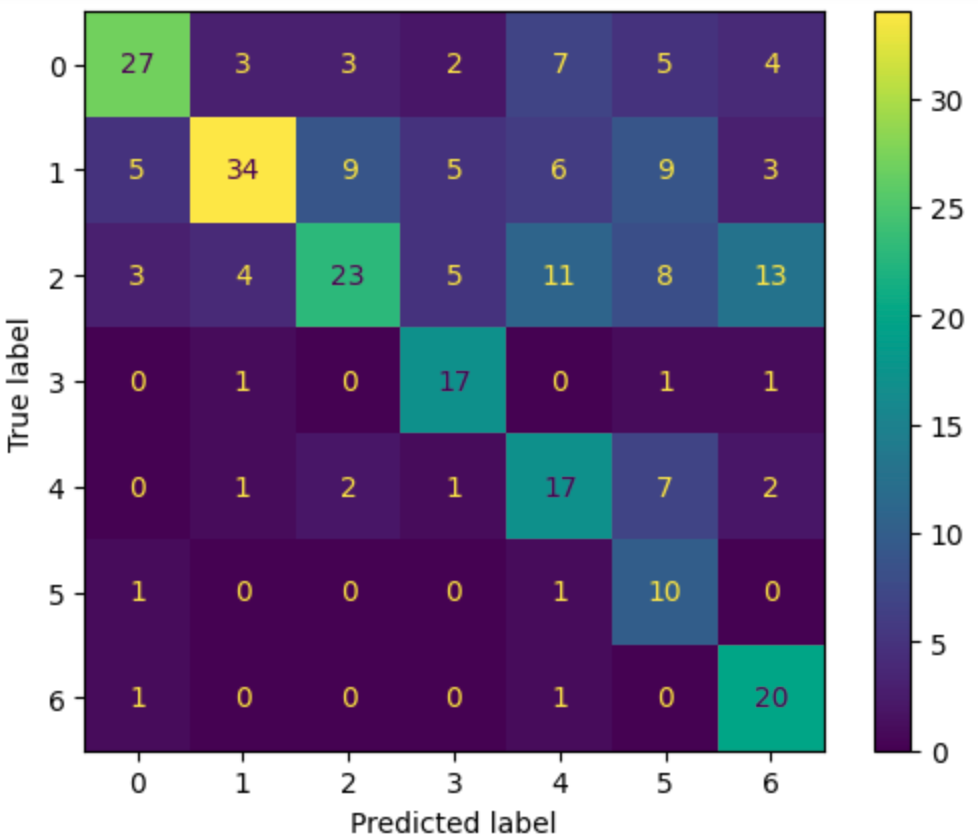
Ön işlemenin tamamlanmasından sonra istenilen her model için ayrı ayrı eğitim yapılmıştır. Ödevde verilen tablonun doldurulabilmesi için her modelin klasifikasyon raporları, confision matrixleri hazırlanmış ve aşağıda sunulmuştur.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Accuracy | F-measure | Precision | Recall | Training Time | Testing Time |
| KNN-3 Euclid | 0.48 | 0.47 | 0.48 | 0.65 | 0.015 | 0.068 |
| KNN-5 Euclid | 0.54 | 0.54 | 0.55 | 0.64 | 0.012 | 0.071 |
| KNN-7 Euclid | 0.48 | 0.49 | 0.48 | 0.61 | 0.010 | 0.101 |
| KNN-3 Manhattan | 0.48 | 0.47 | 0.48 | 0.69 | 0.010 | 0.937 |
| KNN-5 Manhattan | 0.48 | 0.48 | 0.48 | 0.64 | 0.009 | 0.970 |
| KNN-7 Manhattan | 0.43 | 0.44 | 0.43 | 0.66 | 0.011 | 0.928 |
| DT Entropi | 0.41 | 0.40 | 0.41 | 0.40 | 0.490 | 0.003 |
| DT Gini | 0.48 | 0.48 | 0.48 | 0.48 | 0.467 | 0.005 |
| Naive Bayes | 0.56 | 0.55 | 0.56 | 0.60 | 0.087 | 0.041 |
| SVM | 0.59 | 0.58 | 0.59 | 0.59 | 1.402 | 0.322 |

A chart of numbers and a number

Description automatically generated with medium confidence

**Figure 1. KNN- Euclid (k = 3) Confusion Matrix**



**Figure 2. KNN-Euclid (k=5) Confusion Matrix**

**A chart of a number and a number

Description automatically generated with medium confidence**

**Figure 3. KNN-Euclid (k=7) Confusion Matrix**

A chart of a number and a number

Description automatically generated with medium confidence

**Figure 4 KNN-Manhattan (k=3) Confusion Matrix**

A chart of a number and a number

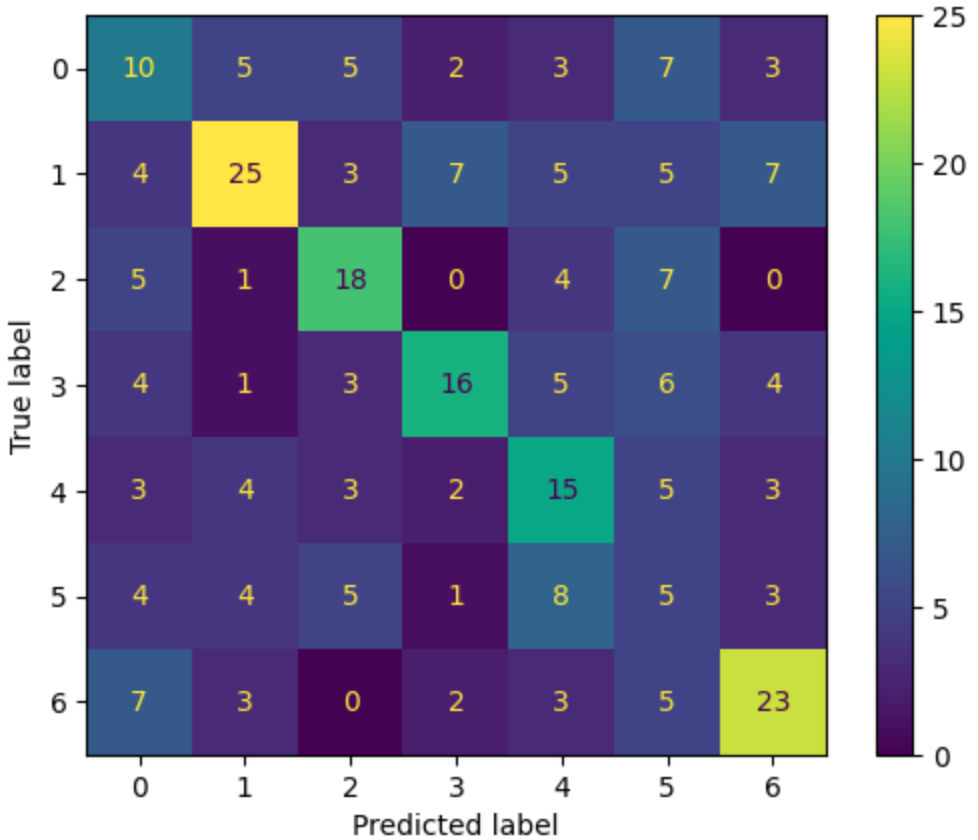
Description automatically generated with medium confidence

**Figure 5 KNN-Manhattan (k=5) Confusion Matrix**

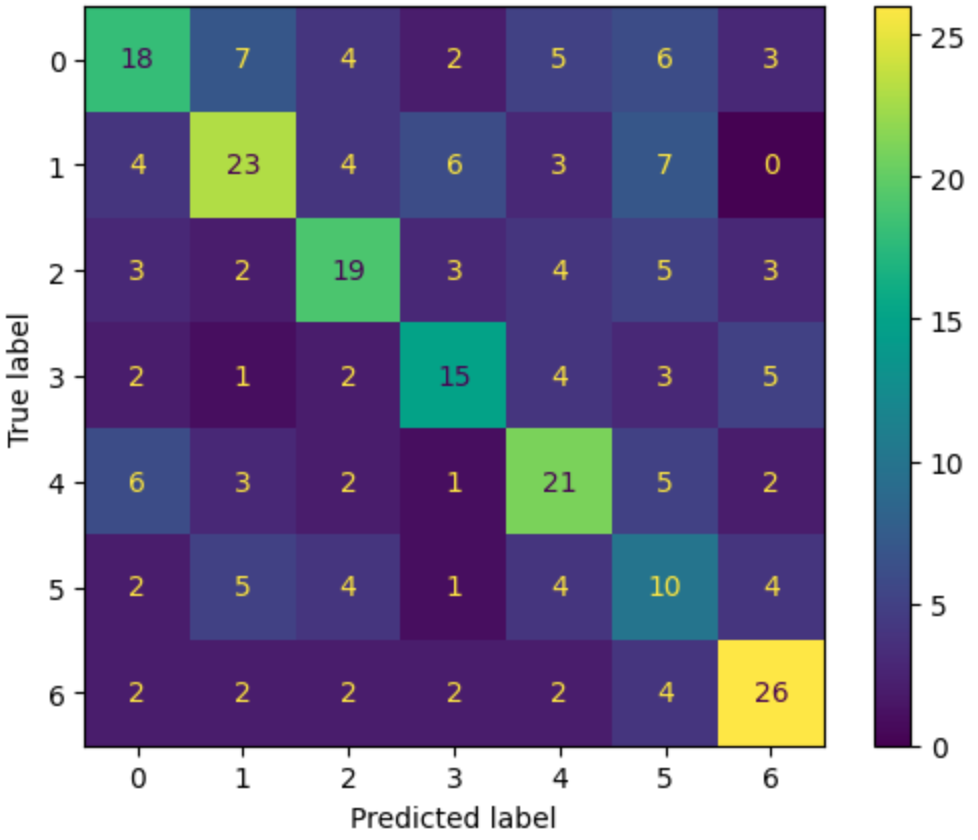
A chart of a number of colored squares

Description automatically generated with medium confidence

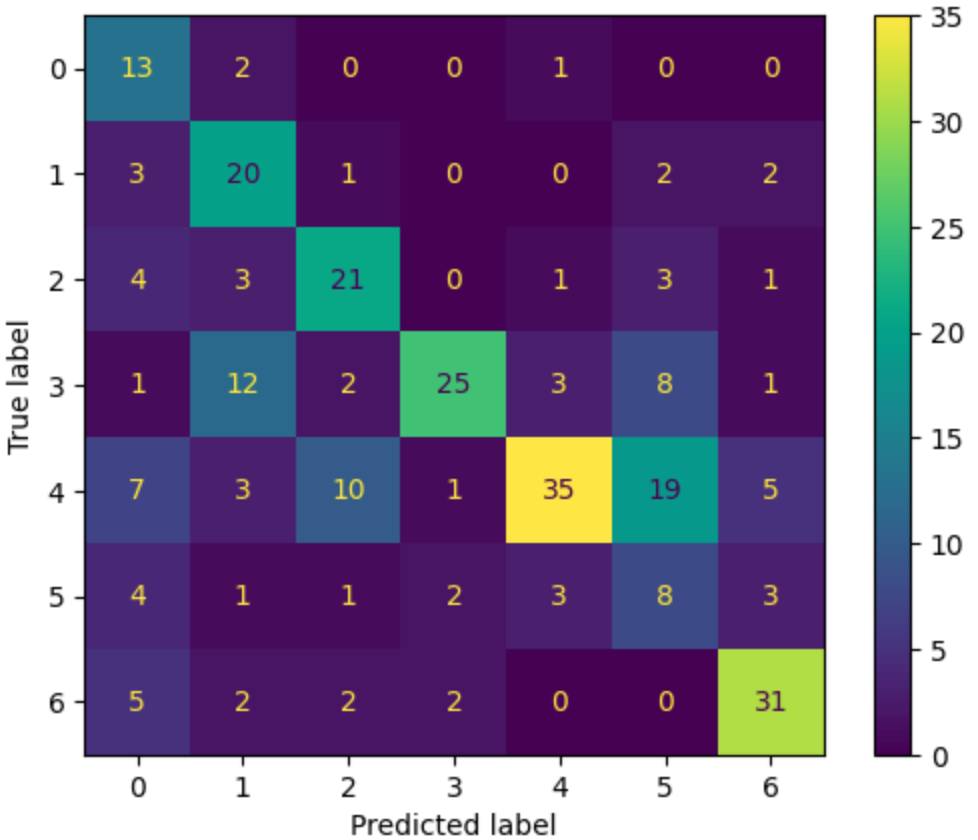
**Figure 6. KNN-Manhattan (k=7) Confusion Matrix**



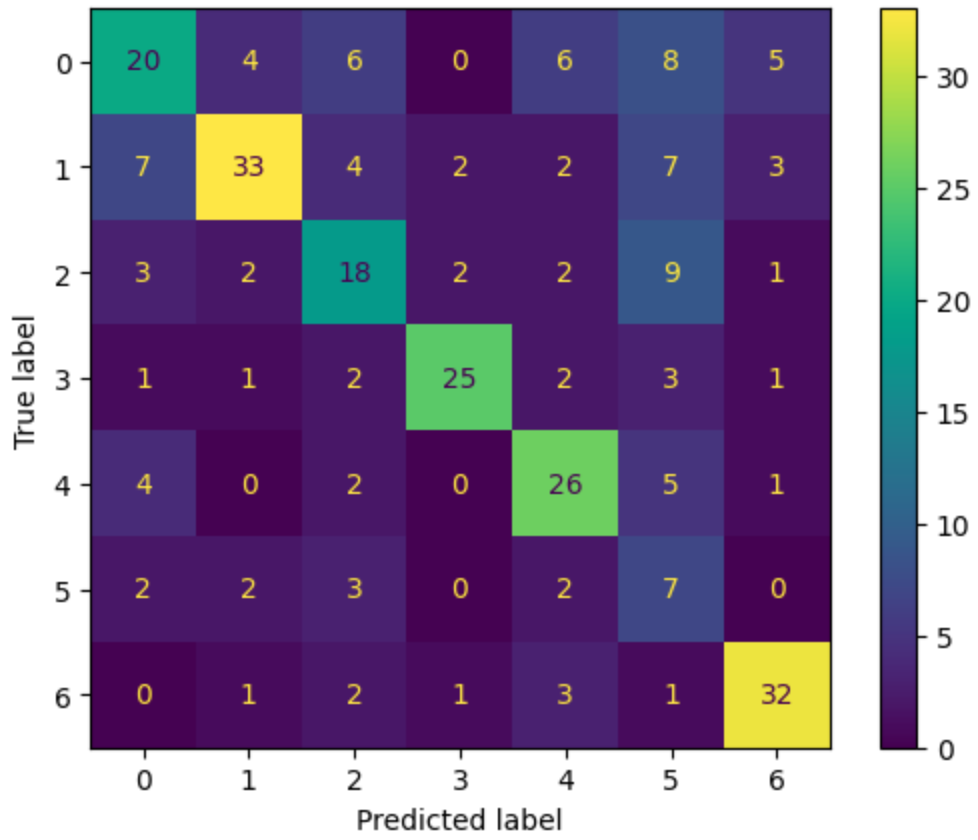
**Figure 7. Decision Tree (entropy) Confusion Matrix**



**Figure 8. Decision Tree (gini) Confusion Matrix**



**Figure 9. Naive Bayes Confusion Matrix**



**Figure 10. SVM Confusion Matrix**

A line of dots on a white background

Description automatically generated

**Figure 11. Visualized Decision Tree (gini)**

## **Sonuç**

Klasifikasyon tablosu incelendiğinde, en yüksek başarının SVM metoduyla elde edildiği gözükmektedir. KNN manhattan yönteminin test süresinin, KNN euclid yöntemi ile aynı başarı oranına sahip olmasına rağmen manhattan ile yapılan metodda test süresinin çok daha yüksek olduğu sonucuna ulaşılmıştır. En iyi test süresinin ise Decision Tree algoritmalarında olduğu gözükmektedir.