



**T.C.
FIRAT ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ**

**YAZILIM MÜHENDİSLİĞİ
YMH 418 GÜNCEL
KONULAR DERSİ PROJE
DOSYASI
11.05.2020 – 15.05.2020**

**BÖLÜMÜ : YAZILIM MÜHENDİSLİĞİ
NUMARASI : 15542525
ADI ve SOYADI : ÇAĞDAŞ KARACA**

IV. Aşama - Model Kurma

Veri seti üzerinde Weka programı kullanılarak K-means ile kümeleme işlemi yapıldı. Fold değeri 10 olarak tanımlanmıştır. Kümeleme işlemi için veri seti üzerinbu işlem sonrası elde edilen sonuçlar aşağıda gösterilmiştir.

Yineleme sayısı: 14

Küme içindeki kare hataları toplamı: 25941.494474992956

İlk başlangıç noktaları (rastgele):

Küme 0: 01/20/2020, izmir, 19,0,118,344

Küme 1: 01/21/2020, izmir, 14,0,112,0

Dünya genelinde eksik değerler ortalama / mod ile değiştirildi

Son küme sentroidleri:

| | Küme # | | |
|-----------|-----------------------|---------------|----------------|
| Öznitelik | Tam Veri (25879.0) | 0 (9702.0) | 1 (16177.0) |
| Tarih | 04/02/2020 | 01/20/2020 | 01/21/2020 |
| Durum | izmir | izmir | izmir |
| so2 | 11.8472 | 13.996 | 10.5584 |
| no2 | 2.6968 | 1.9048 | 3.1718 |
| rspm | 74.9012 | 85.8207 | 68.3523 |
| spm | 91.2781 | 220.6888 | 13.6652 |

Model oluşturmak için geçen süre (tam eğitim verisi): 0.06 saniye

=== Eğitim setinde model ve değerlendirme ===

Kümelenmiş Örnekler

0 9702 (% 37)

1 16177 (% 63)

Elde edilen verilere bakılacak olursak spm değerine göre rastgele seçilen küme değerleriyle diğer veriler karşılaştırıldığında %63 oranında 01/21/2020 tarihinde elde edilen verilerle diğer verilerin kümelenmesi başarımlı verilmiştir. %37 oranla küme 0 daki verilerle veri setindeki diğer verilerle kümelenmiştir.

Python'da oluşturulan k means kümeleme algoritmasının yapay sinir ağı ile oluşturularak fold değeri yine 10 olarak seçilerek veri seti üzerinde yapılan kümeleme ile weka üzerinden elde edilen sonuçlar karşılaştırılırsa;

Kat puanı (RMSE): 0.044751691281048715

Final, örnek dışı skor (RMSE): 0.044751691281048715

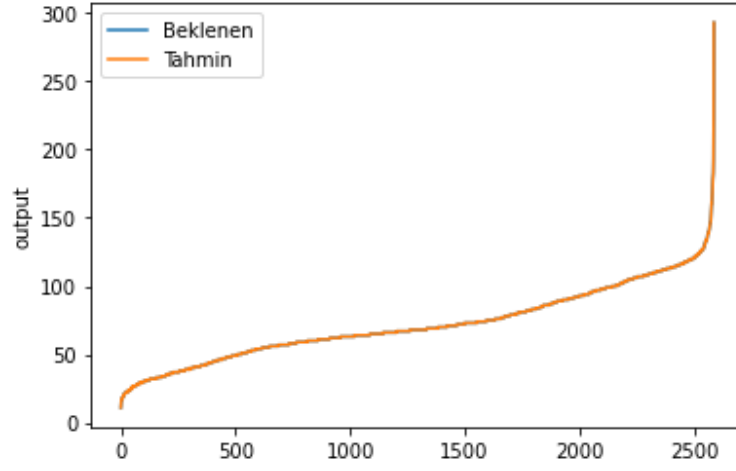
('Ortalama Eğitim Başarısı', 0.8591075)

('Ortalama Doğrulama Başarısı', 0.9255894919236501)

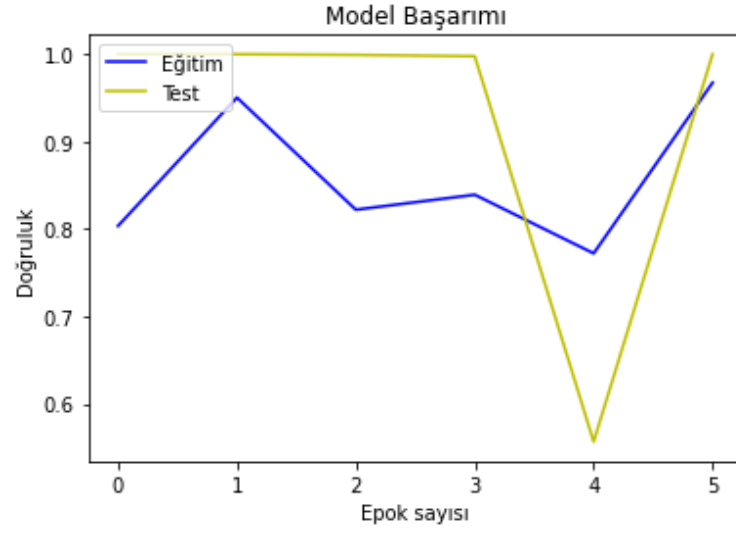
('Ortalama Eğitim Kaybı', 6.956798865816162)

('Ortalama Doğrulama kaybı', 0.05066158099921711)

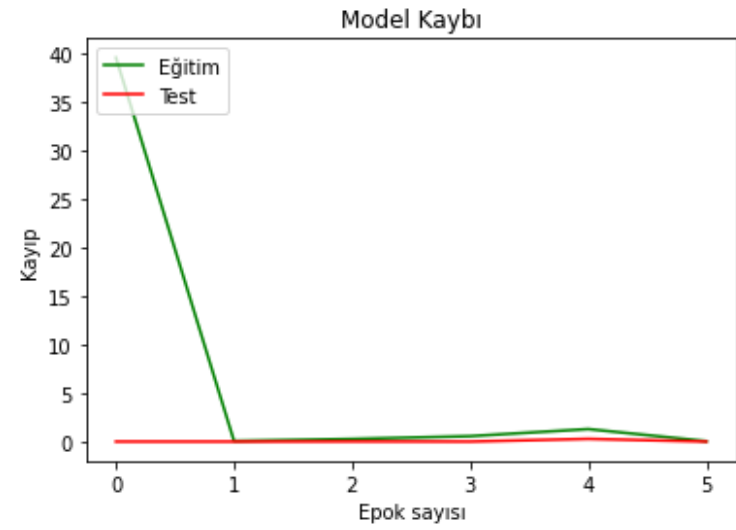
Ortalama eğitim başarımlı yapay sinir ağı ile oluşturulan model ile ortalama %85 ile başarılı sonuçlandırılmıştır. Oluşturulan modelin ortalama eğitim kaybı ise %7 dir. Oluşturulan modele göre elde edilen grafik sonuçları aşağıda gösterilmiştir.



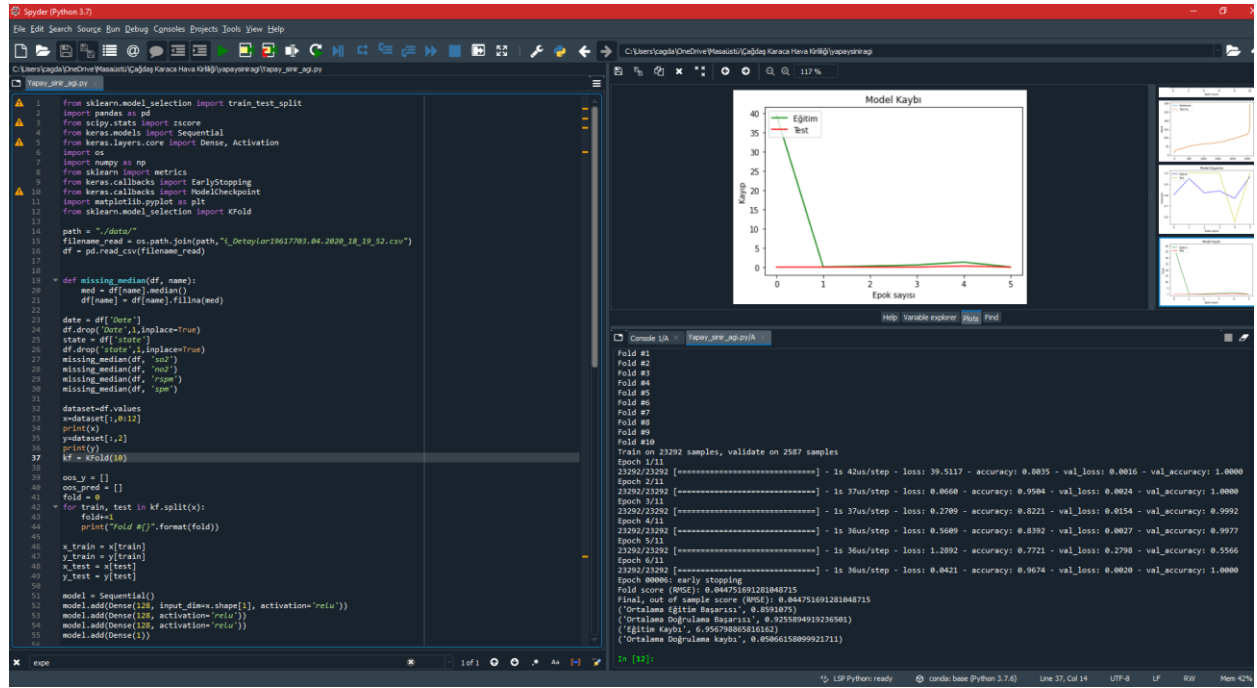
Görsel 1. Kümeleme işleminde tahmin ve beklenen değerlerin görselleştirilmesi



Görsel 2. Modelin eğitim ve test verilerine göre başarımlarının görselleştirilmesi



Görsel 3. Modelin eğitim ve test verilerine göre kayıp değerlerinin görselleştirilmesi



Görsel 4. Sonuçların Ekran Görüntüsü

Sonuç:

Python'da oluşturulan algoritmanın başarımlar değeri wekaya göre daha yüksektir. Bunun nedeni simple k means algoritması ile yapay sinir ağlarında kullanılan modelin el ile oluşturulmasından kaynaklanmaktadır. Wekada kullanılan algoritmadaki model programın üreticileri tarafından en optimize çalışacak şekilde programlanmıştır. Python'da ise biz belirli bir veri seti üzerinde en iyi başarımları elde edecek bir model oluşturduk. Buradaki başarımların weka ya göre daha yüksek olmasındaki en büyük neden budur.

Saygılarımla,
15542525 Çağdaş Karaca