Yıldız Türlerinin Sınıflandırılması için Yapay Sinir Ağı Kullanımı

1. Proje Hakkında

Bu projede, yapay sinir ağı kullanılarak yıldız türlerinin doğru bir şekilde sınıflandırılması amaçlanmıştır. Kullanılan veri seti, yıldızların sıcaklık, parlaklık, kütle ve yarıçap gibi fiziksel özelliklerini içermektedir. Bu veriler, yıldızların farklı kategorilere (örneğin kırmızı cüce, beyaz cüceler, süper dev vb gruplara) ayrılmasını sağlamaktadır. Proje, bu özellikleri kullanarak yıldızların hangi türe ait olduğunu belirlemek için yapay sinir ağlarının kullanılması hedeflenmektedir. Bu süreçte farklı eğitim ve test yöntemleri uygulanmış, modelin en iyi sonuçları veren parametreleri optimize edilmiştir.

1. Veri Seti Hakkında

Bu çalışmada, yıldızları belirli özelliklerine göre sınıflandırmayı hedefleyen bir veri seti kullanılmıştır. Veri seti, toplamda **240 örnek** ve **6 sınıf** içermektedir. Her bir örnek yıldızların fiziksel ve spektral özelliklerini içermektedir. Veri setinde aşağıdaki sütunlar bulunmaktadır örnek, bir yıldıza ait fiziksel ve spektral özellikleri temsil etmektedir.

**Veri Setindeki Sütunlar:**

|  |  |
| --- | --- |
| Sütun İsmi | Açıklama |
| Temperature (K) | Yıldızın yüzey sıcaklığı (Kelvin cinsinden). |
| Luminosity (L/Lo) \* | Yıldızın parlaklığı (Güneş'e oranla). |
| Radius (R/Ro) \* | Yıldızın yarıçapı (Güneş'e oranla). |
| Absolute magnitude (Mv) | Yıldızın mutlak büyüklüğü. |
| Star color | Yıldızın rengi. |
| Spectral Class | Yıldızın spektral sınıfı (M, K, G, F, A, B, O). |
| Star type | Yıldızın sınıfını ifade eden kategorik değişken (hedef değişken). |

**\*** Her yıldızın parlaklığı ve yarıçapı Güneş'in değerlerine göre hesaplanır.

Lo = 3.828 x 10^26 Watt

Ro = 6.9551 x 10^8 m

**Veri Hazırlama Süreci:**

* Star color sütunundaki tutarsızlıklar giderilmiş ve veriler birleştirilmiştir. Örneğin, yellowishwhite ve whiteyellow gibi benzer renkler yellowwhite olarak gruplandırılmıştır.
* Veri setindeki *Temperature, Luminosity ve Radius* sütunları çok geniş bir değer aralığına sahiptir. Bu aralığı daha yönetilebilir hale getirmek için logaritmik dönüşüm uygulanmıştır.
* Logaritmik dönüşüm sonrası *Temperature, Luminosity, Radius ve Absolute magnitude değerlerine* Min-Max normalizasyonu ile [0, 1] aralığına ölçeklendirilmiştir.
* *Spectral Class* sütunundaki kategorik veriler, sayısal değerlere dönüştürülmüştür. Kodlama şu şekildedir: (M → 0, K → 1, G → 2, F → 3, A → 4, B → 5, O → 6)
* *Star color* sütunundaki kategoriler, One-Hot Encoding yöntemi ile ayrı sütunlara dönüştürülmüştür. Bu işlem, her bir renk için bir sütun oluşturmuş ve değerlerin 0 veya 1 olarak kodlanmasını sağlamıştır.

**Veri Görselleştirme:**

* Her bir eğitim senaryosu sonunda sonuçlar konfüzyon matrisi şeklinde görselleştirildi.
* Her bir eğitim ve test aşamasının kayıp ve doğruluk değerleri grafikleştirildi.
* Tasarlanan ağ yapısı görselleştirilmiştir.

1. Modelleme

Bu proje kapsamında, sınıflandırma problemini çözmek için bir yapay sinir ağı modeli geliştirilmiştir. Modelin performansını değerlendirmek amacıyla farklı yöntemler uygulanmıştır. Kullanılan yöntemler ve sonuçları aşağıda detaylı şekilde açıklanmıştır

Her eğitim senaryosu için model, 64, 32 ve 16 nöronlu üç gizli katmandan oluşan bir yapıya sahiptir ve çıkış katmanında sınıf sayısı kadar nöron bulunmaktadır.

Eğitim sürecinde model her eğitim senaryosu için:

* 50 epoch boyunca eğitildi.
* batch size = 16 kullanıldı.
* Öğrenme katsayısı 0.001 olarak belirlendi.
* Optimizasyon için *Adam* algoritması tercih edildi.
* Kayıp fonksiyonunda *Categorical Crossentropy* tercih edilmiştir.
* Aktivasyon fonksiyonu olarak gizli katmanlar için *reLU*, çıkış katmanı için s*oftmax* uygulandı.

Sonuçlar her bir rassal ayrım için ayrı ayrı değerlendirilmiştir

* Her bölme için konfüzyon matrisleri oluşturulmuş ve başarı oranları hesaplanmıştır.
* Eğitim ve Test Doğruluğu grafiklerle görselleştirilmiştir,

3.1 Modelin %66-%34 Bölümmüş Veri Seti ile Testi

Veri seti %66’ü eğitim ve %34’ü test olmak üzere 2 parçaya bölündü. Modelleme kısımında belirtilen model ile eğitim gerçekleştirildi.

Model, eğitim verisinde ~%97, test verisinde ortalama **%96,096** doğruluk oranı elde etti. Yapılan gözlemlerde eğitim ve test verisinde kayıp oranı her epok için düşüş göstererek iyi bir genelleme yeteneği sergilemiştir. Her aşama için konfüzyon matrisleri oluşturulmuş ve başarı oranları hesaplanmıştır. *(Şekil 1).* Ayrıca Eğitim ve Test Doğruluğu grafiklerle görselleştirilmiştir, *(Şekil 2)*

***1. Eğitim 2. Eğitim***

ekran görüntüsü, diyagram, metin, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu ekran görüntüsü, metin, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

***3. Eğitim 4. Eğitim***

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu ekran görüntüsü, diyagram, metin, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

***5. Eğitim***

ekran görüntüsü, diyagram, dikdörtgen, kare içeren bir resim

Açıklama otomatik olarak oluşturuldu

*(Şekil 1) Konfüzyon matrisleri*

***1. Eğitim***

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

***2. Eğitim***

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

***3. Eğitim***

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***4. Eğitim***

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

***5. Eğitim***

metin, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

(*Şekil 2) Kayıp ve doğruluk değerleri grafiği*

3.2 Modelin 5 ve 10 Fold Cross-Validation Testi

Veri seti, 5 ve 10 eşit parçaya bölünmüş ve her bir parça sırasıyla test verisi olarak kullanılarak çapraz doğrulama gerçekleştirilmiştir. Bu yöntemde her katmanın doğruluk ve kayıp değerleri hesaplanmış ve ortalama başarı oranı raporlanmıştır.

Model 5 fold cross-validation’ da eğitim verisinde ortalama %10,8 test verisinde ortalama %13,6 kayıp vermiştir. Model train verisinde ortalama %97 test verisinde ise ortalama **%96,10** doğruluk elde etmiştir. Her fold için konfüzyon matrisleri toplamı oluşturulmuş ve başarı oranları hesaplanmıştır. *(Şekil 3).*

Model 10 fold cross-validation’ da ise eğitim verisinde ortalama %12,6 test verisinde ortalama %14,9 kayıp vermiştir. Model train verisinde ortalama %96,9 test verisinde ise ortalama **%96,95** doğruluk elde etmiştir. Konfüzyon matrisleri toplamı oluşturulmuş ve başarı oranları hesaplanmıştır. *(Şekil 4).*

metin, ekran görüntüsü, diyagram, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, diyagram, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

*(Şekil 3) 5 Cross-validation için toplam (Şekil 4) 10 Cross-validation için toplam*

*konfüzyon matrisi konfüzyon matrisi*

3.3 Modelin Eğitim Verisi ile Testi

Bu yaklaşımda eğitim verisi, modelin hem eğitimi hem de testi için kullanılmıştır. Bu durum, modelin öğrenme sürecine dair içgörüler sunar.

Yapılan gözlemlerde modelin kaybının her epok için kayıbın düştüğü gözlemlenmiştir. *(Şekil 5)* Aynı zamanda model **%99** doğruluk elde etmiştir. *(Şekil 6)* Bu başarı oranı oldukça yüksek olup, model veriyi ezberleme eğilimi göstermiştir.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

*(Şekil 5) Doğruluk ve kayıp grafikleri*

ekran görüntüsü, metin, diyagram, kare içeren bir resim

Açıklama otomatik olarak oluşturuldu

*(Şekil 6) konfüzyon matrisi*