**T.C.**

**BALIKESİR ÜNİVERSİTESİ**

**MÜHENDİSLİK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

metin, daire, logo, ticari marka içeren bir resim

Açıklama otomatik olarak oluşturuldu

**HABER KATEGORİZASYONU,HABER YORUMLARININ DUYGU ANALİZİ VE OPTİMUM AĞ TASARIMI**

**202113709014 OĞUZHAN YILMAZ**

**BMM4101 YAPAY ZEKA TEKNİKLERİ VİZE ÖDEVİ**

**Danışman: Dr. Öğr. Üyesi Kadriye ERGÜN**

**BALIKESİR, 12-2023**

**İÇİNDEKİLER**

**1-**Çalışmanın tanımı ve amacı, literatür bilgisi

**2-**Seçilen gazetenin haberlerini ve yorumlarını gösteren 20’şer satırlık tablo

**3-**Ödevde yaptığınız çalışma ve seçtiğiniz yöntemlerin sırası ve anlatımı

**4-**Yapay Sinir Ağlarında oluşturduğunuz ağın şekli

**5-**Ağ topolojisi ve varsayımlar

**6-**Optimum ağ tasarımının anlatımı

**7-**Karşılaştırma (optimum ağ ile başlangıçtaki ağ yapısı arasında) ve sonuçların yorumlanması (Sınıflandırıcı başarı oranı hakkında bilgi)

**8-**Sonuç

**9-**Kaynaklar

**1-**

Yapay Sinir Ağları yöntemi ile *sınıflandırma* işlemi kullanarak ,gazete *haberlerinin kategorizasyonu* ve bu haberlere yapılan yorumların *pozitif, negatif ve nötr* duygulardan hangisini barındırdığını belirlemek.

**2-**

Haber başlıkları:

|  |  |
| --- | --- |
| Ankaragücü ile Rizespor yenişemedi | 0,0,0,1,0 |
| Sivasspor Konyaspor deplasmanında kazandı | 0,0,0,1,0 |
| Bursa da otomobil iki tırın arasında sıkıştı 1 ölü 1 yaralı | 0,1,0,0,0 |
| Orta Vadeli Program başarılı bir şekilde çalışıyor | 0,0,1,0,0 |
| Beyoğlu nda 2 katlı ahşap binada yangın | 0,1,0,0,0 |
| Eski Fenerbahçeli Luiz Gustavo nun yeni takımı belli oldu Ülkesine döndü | 0,0,0,1,0 |
| Beşiktaş Lugano maçının hazırlıklarına devam etti | 0,0,0,1,0 |
| Fenerbahçe Spartak Trnava mesaisine başladı | 0,0,0,1,0 |
| Irak ın kuzeyine hava harekatı 13 hedef imha edildi | 0,0,1,0,0 |
| İnşaatın 30 katından düşen 2 işçi yaşamını yitirdi | 0,1,0,0,0 |
| Gazze de can kaybı 18 bini geçti | 0,1,0,0,0 |
| A Milli Takım ın EURO 2024 teki kamp yeri belirlendi | 0,0,0,1,0 |
| Almanca konuşan imamlar yetiştirilmeli | 0,0,1,0,0 |
| Kola nasıl bulundu şaşıracaksınız Meğer morfin yerine | 0,0,0,0,1 |
| Yumurtalarınızı dondurun farkı görün Artık herkes böyle yapıyor | 0,1,0,0,0 |
| Halil Umut Meler e saldırı sonrası hakemler Süper Lig de maçlara çıkmama kararı aldı | 0,0,0,1,0 |
| Erman Toroğlu Bu ahar Ankaragücülü olamaz | 0,0,0,1,0 |
| LeBron James in oğlu Bronny sahalara geri döndü | 0,0,0,1,0 |
| Ertuğrul Doğan ın acı günü laksızl Dayısı Coşkun Şahinkaya vefat etti | 0,0,0,1,0 |
| Beşiktaş tan transfer yalanlaması Kesinlikle itibar etmeyin | 0,0,0,1,0 |

Haber yorumları:

|  |  |  |
| --- | --- | --- |
| 3 4 sene önce 200tl banknotu esnafa verince bozuk yokmu diyor mırın kırın ediyordu Şimdi 200tl veriyorsun üzerine para alamıyorsun | Notr | 0,1,0 |
| 3 4 sene önce ekonomi iyiydi diyorsun yani İnşaAllah afet felaket salgınlardan musibetlerden uzak oluruz da tekrar ekonomi toparlanır | Positive | 0,0,1 |
| 3 4 sene önce ekonomi ondan önceki 3 4 seneye göre kötüydü Afet mafet değil tepe taklak gidiyor gemi batıyor | Notr | 0,1,0 |
| Türk ekonomisi 2010 dan beri hiç iyiye gitmedi ama şimdi deniz de bitti Feto darbe yaptı ama ülkenin yarısını reis onlara vermişti onlara çöktüler o da bitti | Negative | 1,0,0 |
| battığımızın resmidir | Negative | 1,0,0 |
| Ben ekonomistim diyerek meydanlarda boy gösterip soğan fiyatını bile kontrol edemeyen Dünya liderimiz tatlı uykusundan uyanmadı mı acaba | Negative | 1,0,0 |
| ne yapsın o dünya lideri senin gibilere güvenip yola mı çıksın Adam ülke etrafındaki girdaptan çıkmaya çalışıyor senin derdin hala soğan iyi ki kurtuluş savaşı zamanında sizin zihniyetiniz yoktu aç karnına savaş mı olur deyip satardınız ülkeyi | Negative | 1,0,0 |
| ekonomik olarak iyi yoldayız | Positive | 0,0,1 |
| Türk ekonomisi iyiye gitmiyor | Negative | 1,0,0 |
| Değer etmeyen paradanmı bahsediyoruz en yüksek paranla bile iyi bir kilo peynir alınmıyor | Negative | 1,0,0 |
| Türk lirası tamamen değersizleşti maalesef 200 liraya yarım kilo kıyma vermiyorlar bu ne rezillik kepazeliktir Millet parayı siyah çöp torbalarıyla spor çantalar taşımaya başladı geçen gün bankada bizzat gördüm Para sayma makinaları para sayıyor sürekli ama işte değeri yok paranın | Negative | 1,0,0 |
| 10 yıl ülke sabredip beklemeli | Notr | 0,1,0 |
| Geçen sene cüzdana sığmayan maaş bu sene artık çantaya sığmaz oldu Yeter artık 1000₺ banknot çıkarın Çuvalla gezmeye az kaldı | Negative | 1,0,0 |
| ooo çok iyi maaş alıyormuşsun | Positive | 0,0,1 |
| Türk ekonomisi iyiye gidiyor | Notr | 0,1,0 |
| Atılan adımların sonuçlarını bekliyoruz | Negative | 1,0,0 |
| 500’lük banknotların çıkması için daha ne olması gerekiyor | Positive | 0,0,1 |
| Köprü Yol Hastahane Havaalanları Hızlı Tren tümü yap işlet modeli yapılmıştır Devletin Bütçeside Yandaşa Yoldaşa İhalelere Rant olarak gitmiştir çoğu paranında akibetide belli değildir | Notr | 0,1,0 |
| Türk ekonomisi dışarıdan aldığı yatırımlar ile gelişimine devam edebilir | Notr | 0,1,0 |
| HAFİZE GAYE YAZINCA NE OLUYOR Kİ ÜLKENİN EN BÜYÜK BANKNOTUNDA | Notr | 0,1,0 |

**3,4,5 -**

Haberleri çekmek için bir python scripti yazdım bu script haber sitesi üzerinde ana sayfada yayınlanan güncel haberlerin baslıklarını çekip “haberbaslik.csv” adlı oluşturduğum dosyaya kayd ediyor.

“haberbaslik.csv” ve haberbaslik\_cekme.py dosyası klasörde bulunmaktadır.

Haber başlıklarını tokenize etmek için “tokenize\_islem.py” adlı python scripti yapdım ve bu baslıkları tokenize ederken internetten bulmuş olduğum "turkce-stop-words.txt" ile cümlelerdeki gereksiz kelimeleri çıkartım. "turkce-stop-words.txt" klasörde bulunmaktadır.

Yazmış olduğum nokta\_silme.py import ederek haber başlıklarını noktasız ve satır boşluklarından arındırdım. Sonra yazdığım tokenize\_islem.py import ederek yorumları tokenizeleştirdim.

Nokta\_silme.py ve tokenize\_islem.py klasörde bulunmaktadır.

metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Daha sonra bu tokenize ettiğim veriyi word2vec yöntemi ile vektörleştirdim

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Vektörleştirdiğim veriyi ysa test için kullanmak amacıyla “testhaber\_input.csv” kaydettim.kitap, ekran görüntüsü, metin, kalıp, desen, düzen içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu verilerin kategorilerinide “testhaber\_target.csv” olarak tutum.

metin, ekran görüntüsü, kitap, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Şu ana kadar yaptığımız işlemlerle test verimizin input ve target dosyalarını oluşturduk. Şimdi kaggle da bulduğum veri setini tokenizeleştirme ve vectorleştirme adımlarından geçirelim.

<https://www.kaggle.com/code/mervetncr/turkish-text-classification-fasttext-and-bert/notebook>

Bu veri seti 7 haber kategorisi içeriyordu 5 kategorisini kullandım. Klasördeki haber\_veriseti.csv dosyası olarak kayıtlı.

Bu veri setini yazmış olduğum haber\_tokenize.py dosyasını kulanarak verinin haber kısmını tokenize ederek “tokenize\_haber\_veriseti.csv” olarak , kategorilerini LabelBinarizer kullanarak binary şeklinde “target\_haber.csv” olarak kaydettim.

metin, elektronik donanım, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Daha sonra “tokenize\_haber\_veriseti.csv” yi word2vec ile vectöre dönüştürüp “input\_haber.csv” kaydettim.

Ellimizdeki 2 test ,2 eğitim verisi olmak üzere 4 veri dosyası ile matlabda ysa oluşturalım.

metin, ekran görüntüsü, diyagram, paralel içeren bir resim

Açıklama otomatik olarak oluşturuldu

Matlaba giriyoruz daha sonra veri setlerini import ediyoruz .Ardından bu verilerin transpozlarını aldım ve nntool aracını çalıştırdım.

Bu verileri nntool networku içine import ettim

metin, ekran görüntüsü, yazılım, bilgisayar simgesi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Daha sonra new diyerek ağı tasarlama kısmına geçtim, burada input ve target verilerini tanımladım ağın katman sayısını hangi algoritma kullanılacağını seçtikten sonra create değerek ağı oluşturdum

metin, ekran görüntüsü, yazılım, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu

View kısmından seçerek ağın şeklini gözlemleyebiliriz

metin, yazı tipi, yazılım, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazılım, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

Oluşturmuş olduğumuz ysa ağını eğitmek için ağda train kısmına geldim burada input ve target verilerini atadıktan sonra train işlemini başlattım

metin, ekran görüntüsü, yazılım, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazılım, web sayfası içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, diyagram, ekran görüntüsü, paralel içeren bir resim

Açıklama otomatik olarak oluşturuldu

Train işlemi sonrası regression grafiği yukarıdaki gibidir.

Eğitmiş olduğum ysa test etmek için simulate kısmına gelerek test inptu ve target verilerini seçiyoruz

metin, ekran görüntüsü, yazılım, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Ortaya çıkan test\_outputs ve test\_errors dosyaları aşağıdaki gibidir.

metin, ekran görüntüsü, yazılım, web sayfası içeren bir resim

Açıklama otomatik olarak oluşturuldu

Matlab dosyası “202113709014\_Oğuzhan\_Yılmaz\_haber”adıyla klasörde bulunmaktadır.

* Yorumları çekmek için bir python scripti yazdım bu script haber sitesi üzerinde bir habere yapılan güncel yorumları çekip “yorum.csv” adlı oluşturduğum dosyaya kaydediyor.

“yorum.csv” ve yorum\_cekme.py dosyası klasörde bulunmaktadır

* Yazmış olduğum nokta\_silme.py import ederek haber başlıklarını noktasız ve satır boşluklarından arındırdım.Sonra yazdığım tokenize\_islem.py import ederek yorumları tokenizeleştirdim ve “tokenize\_yorumlar.csv” ye kaydettim.

Nokta\_silme.py ve tokenize\_islem.py klasörde bulunmaktadır.

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Tokenize olmuş bu verileri haberlerde yapmış olduğum gibi word2vec kullanarak vectörleştirdim ve “input\_test\_yorum.csv” olarak kaydettim ve bu veriyi manuel olarak Positive(0,0,1) ,Notr(0,1,0), Negative(1,0,0) olacak şekilde etiketleyip “target\_test\_yorum.csv” ye kaydettim.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Açıklama otomatik olarak oluşturuldu

Böylece yorum duygu analizi için test verilerimiz hazırlamış olduk .

Ysa eğitmek için internetten bulmuş olduğum yorum veri setini kullanacağım. Dosya klasöründe “yorum\_veriseti.csv” olarak bulabilirsiniz.

<https://huggingface.co/datasets/winvoker/turkish-sentiment-analysis-dataset/tree/main>

Bu veri seti 49.226 yorum içeriyordu ben ysa eğitmek için 12.000 yorum kullandım. Klasördeki “yorum\_veriseti.csv” dosyası olarak kayıtlı.

Bu veri setini yazmış olduğum yorum\_tokenize.py dosyasını kullanarak verinin haber kısmını tokenize ederek “tokenize\_yorum\_veriseti.csv” olarak , kategorilerini LabelBinarizer kullanarak (Positive(0,0,1) ,Notr(0,1,0), Negative(1,0,0)) binary veri şekline çevirerek “target\_haber.csv” olarak kaydettim.

metin, ekran görüntüsü, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Daha sonra “tokenize\_yorum\_veriseti.csv” yi word2vec ile vektöre dönüştürüp “input\_yorum.csv” kaydettim.

Ellimizdeki 2 test ,2 eğitim verisi olmak üzere 4 veri dosyası ile matlabda ysa oluşturalım.

Matlaba giriyoruz daha sonra veri setlerini import ediyoruz .Ardından bu verilerin transpozlarını aldım ve nntool aracını çalıştırdım.

metin, ekran görüntüsü, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bu verileri nntool networku içine import ettim

metin, ekran görüntüsü, yazılım, bilgisayar simgesi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Daha sonra new diyerek ağı tasarlama kısmına geçtim, burada input ve target verilerini tanımladım ağın katman sayısını hangi algoritma kullanılacağını seçtikten sonra create değerek ağı oluşturdum

metin, ekran görüntüsü, yazılım, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu

View kısmından seçerek ağın şeklini aşağıdaki gözlemleyebiliriz

metin, yazı tipi, sayı, numara, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazılım, ekran, görüntüleme içeren bir resim

Açıklama otomatik olarak oluşturuldu

Oluşturmuş olduğumuz ysa ağını eğitmek için open diyoruz ardından ağda train kısmına geldim burada input ve target verilerini atadıktan sonra train işlemini başlattım

metin, ekran görüntüsü, ekran, görüntüleme, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazılım, bilgisayar simgesi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Train işlemi tamamlandıktan sonra regression grafiği aşağıdaki gibidir.

metin, ekran görüntüsü, diyagram, paralel içeren bir resim

Açıklama otomatik olarak oluşturuldu

İstersek performance veya training state diyerek eğitmiş olduğumuz yapay sinir ağının eğitim sırasındaki performansını ve eğitim durumunu görüntüleyebiliriz.

metin, ekran görüntüsü, ekran, görüntüleme, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, ekran görüntüsü, çizgi, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Eğitmiş olduğum ysa test etmek için simulate kısmına gelerek test inptu ve target verilerini seçiyoruz

metin, ekran görüntüsü, ekran, görüntüleme, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

Ortaya çıkan test\_outputs ve test\_errors dosyaları aşağıdaki gibidir.

metin, ekran görüntüsü, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Test verisi ile test sonuç verisini tablolarını kıyasladığımızda 22 test verisinin 17 tanesini doğru kategorize ettiği sonucuna ulaştık.

metin, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, ekran görüntüsü, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Matlab dosyası “202113709014\_Oğuzhan\_Yılmaz\_yorum” adıyla klasörde bulunmaktadır.

**6-** Optimum ağ tasarımı, yapay sinir ağını eğitirken doğru sonuçlar elde etmek için ve ysa genellemesini artırmak için önemlidir. Bende bu sebepten ötürü ağımı tasarlarken bunu göz önüne alarak katman ve katmandaki nöron sayılarını deneme yanılma yoluyla en iyi sonuç verecek şekilde ayarlamaya çalıştım . Aşağıda bulunan yapay sinir ağını kullandım.

metin, yazı tipi, sayı, numara, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

**7-** Başlangıçtaki ağ yapım resim1 deki gibiydi ve başarı oranı 64.37 çıktı .Daha sonra resim2 deki ağ yapısını kullanınca başarı oranı 79.27 yükseldi ve sonuçlara bakınca ağ yapısının doğruluk oranı üzerindeki etkisini gözlemledim

resim1:

metin, diyagram, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

resim2:

metin, yazı tipi, sayı, numara, yazılım içeren bir resim

Açıklama otomatik olarak oluşturuldu

**8-**Sonuç olarak bu projede internetten veri çekmeyi ,verileri tokenize etmeyi, verileri vektörleştirmeyi ve kategorizeleştirmeyi, matlabda yapay sinir ağı oluşturmayı ve düzenlemeyi, bu yapay sinir ağını eğitmeyi ve test etmeyi öğrendiğimi söyleyebilirim.

**9-Kaynaklar:**

[**https://www.kaggle.com/code/mervetncr/turkish-text-classification-fasttext-and-bert**](https://www.kaggle.com/code/mervetncr/turkish-text-classification-fasttext-and-bert)**Erişim Tarihi:13.12.2023**

[**https://colab.research.google.com/drive/13PjeOgL52ECAJjkkLDiU0T7HDeUuUxmn**](https://colab.research.google.com/drive/13PjeOgL52ECAJjkkLDiU0T7HDeUuUxmn)

**Erişim Tarihi:13.12.2023**

**<https://huggingface.co/datasets/winvoker/turkish-sentiment-analysis-dataset/tree/main>**

**Erişim Tarihi:13.12.2023**

[**https://github.com/ahmetax/trstop**](https://github.com/ahmetax/trstop) **Erişim Tarihi:13.12.2023**

[**https://medium.com/kaveai/web-scraping-453e96a86195**](https://medium.com/kaveai/web-scraping-453e96a86195) **Erişim Tarihi:13.12.2023**

[**https://www.youtube.com/watch?v=7zhP8coO\_fc&t=745s**](https://www.youtube.com/watch?v=7zhP8coO_fc&t=745s) **Erişim Tarihi:13.12.2023**

[**https://www.youtube.com/watch?v=CtCeuJaRgLQ**](https://www.youtube.com/watch?v=CtCeuJaRgLQ) **Erişim Tarihi:13.12.2023**

[**https://pythonspot.com/tokenizing-words-and-sentences-with-nltk/**](https://pythonspot.com/tokenizing-words-and-sentences-with-nltk/) **Erişim Tarihi:13.12.2023**

[**https://www.youtube.com/watch?v=hzxbFqVz9eI**](https://www.youtube.com/watch?v=hzxbFqVz9eI) **Erişim Tarihi:13.12.2023**

[**https://medium.com/@muhammedbuyukkinaci/word2vec-nedir-t%C3%BCrk%C3%A7e-f0cfab20d3ae**](https://medium.com/@muhammedbuyukkinaci/word2vec-nedir-t%C3%BCrk%C3%A7e-f0cfab20d3ae)

**Erişim Tarihi:13.12.2023**

[**https://www.youtube.com/watch?v=eHqWvpKjLpU&t=81s**](https://www.youtube.com/watch?v=eHqWvpKjLpU&t=81s) **Erişim Tarihi:13.12.2023**

[**https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelBinarizer.html**](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelBinarizer.html)

**Erişim Tarihi:13.12.2023**