



**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ**

Bilgisayar Mühendisliği Bölümü

**ÇOK GÖREVLİ ÖĞRENME
İLE TWEET KULLANICI
ADIVE TARİHİNİ
BELİRLEME**

Sefa Nadir YILDIZ

**Danışman
Dr. Öğr. Üyesi Burcu YILMAZ**

**Ocak, 2019
Gebze, KOCAELİ**



**T.C.
GEBZE TEKNİK ÜNİVERSİTESİ**

Bilgisayar Mühendisliği Bölümü

**ÇOK GÖREVLİ ÖĞRENME
İLE TWEET KULLANICI
ADI VE TARİHİNİ
BELİRLEME**

Sefa Nadir YILDIZ

**Danışman
Dr. Öğr. Üyesi Burcu YILMAZ**

**Ocak, 2019
Gebze, KOCAELİ**

Bu çalışma 04/01/2019 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Bölümü'nde Lisans Bitirme Projesi olarak kabul edilmiştir.

Bitirme Projesi Jürisi

Danışman Adı	Burcu YILMAZ	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Mühendislik Fakültesi	

Jüri Adı	Murat Şeker	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Mühendislik Fakültesi	

ÖNSÖZ

Bu bitirme projesinin hazırlanmasında emeği geçenlere, projenin son halini almasında yol gösterici olan Sayın Dr. Öğr. Üyesi Burcu YILMAZ hocama ve bu çalışmayı destekleyen Gebze Teknik Üniversitesi'ne içten teşekkürlerimi sunarım.

Ayrıca eğitimim süresince bana her konuda tam destek veren aileme ve bana hayatlarıyla örnek olan tüm hocalarıma saygı ve sevgilerimi sunarım.

Ocak, 2019

Sefa Nadir YILDIZ

İÇİNDEKİLER

ÖNSÖZ.....	VI
ŞEKİL LİSTESİ.....	VII
KISALTMA LİSTESİ	X
ÖZET	XI
SUMMARY	XII
1. GİRİŞ	1
2. YÖNTEM.....	5
3. BULGULAR	9
4. TARTIŞMA VE SONUÇ.....	13
KAYNAKLAR.....	14

ŞEKİL LİSTESİ

ŞEKİL 1.1 Kullanıcı adı sınıfı veri seti formatı	2
ŞEKİL 1.2 Tarih sınıfı veri seti formatı	2
ŞEKİL 1.3 Vektör Uzayında Kullanıcı ve Kelimelerin Gösterimi.....	3
ŞEKİL 1.4 Vektör Uzayında Kelime ve Tarih Gösterimi	3
ŞEKİL 1.1 Kullanıcı adı sınıfı veri seti formatı.....	2
ŞEKİL 2.1 Kullanıcı adlarının bulunduğu users_set isimli dizinden her kullanıcı için 100 adet (count=100) tweet veri objesi çekme işlemi.....	5
ŞEKİL 2.2 Tweet objesinden tweet ve tarih bilgisini alma.....	5
ŞEKİL 2.3 Veri setinden kategorilerin ayrılması işlemi.....	6
ŞEKİL 2.4 Shuffle işlemi.....	6
ŞEKİL 2.5 Derin öğrenme modeli python ile tasarımı.....	6
ŞEKİL 2.6 Derin öğrenme modeli eğitim parametreleri.....	7
ŞEKİL 2.7 Verilen tweetin kullanıcı adı veya tarih olarak tahmin edilmesinin 4 örnek için python kodu ile gösterimi.....	7
ŞEKİL 2.8 Derin öğrenme ile kullanıcı adı eğitiminin modelleştirilmesi.....	8
ŞEKİL 2.9 Çok görevli öğrenme ile bir tweete ait kullanıcı adı ve tarihin aynı anda farklı kollarda eğitilmesi örnek model gösterimi.....	8
ŞEKİL 3.1 Veri setinde tweetlerin aylara göre dağılımı.....	9
ŞEKİL 3.2 Kullanıcı adı test sınıfı örneği.....	9
ŞEKİL 3.3 Derin öğrenme ile sadece kullanıcı adı tahmin sonuçları.....	10
ŞEKİL 3.4 Derin öğrenme ile kullanıcı adı tahmini yüzdesel gösterimi.....	10
ŞEKİL 3.5 Ay bazında tarih test sınıfı örneği.....	11
ŞEKİL 3.6 Derin öğrenme ile sadece ay bazında tarih tahmin sonuçları.....	11
ŞEKİL 3.7 Derin öğrenme ile tarih tahmini yüzdesel gösterimi.....	11
ŞEKİL 3.8 Çok görevli öğrenme ile kullanıcı adı ve tarih tahmini.....	12
ŞEKİL 3.9 Çok görevli öğrenme ile kullanıcı adı ve tarih tahmini yüzdesel gösterimi..	12

KISALTMA LİSTESİ

DNN : Derin Sinir Ağları (Deep Neural Networks)
G.T.Ü : Gebze Teknik Üniversitesi

ÖZET

Gelişen teknoloji gün geçtikçe bizlere yeni imkanlar sunmaktadır. Sosyal medya da bu gelişmelerin önemli sonuçlarından biri olmuştur. Sosyal medyanın hayatımıza girişiyle, iletişim daha interaktif bir duruma gelmiş olup, haberlere, güncel olaylara verilen tepkilerin, sesli veya yazılı paylaşımların durumunu incelemek, bu paylaşımların kimler tarafından ve ne zaman yapıldığını öğrenmek son yıllardaki önemli araştırma konularından birisi haline gelmiştir.

Bu çalışma çok görevli öğrenme ile aynı anda bir tweetin kullanıcı adı ve tarihinin belirlenmesi ve derin öğrenmeyle bir tweetin kullanıcı adı ve tarihinin ayrı ayrı eğitilmesine göre başarı oranının nasıl geliştiğinin analizlerini kapsamaktadır. Model eğitilirken 500 kullanıcı ve her kullanıcının 100 tweeti olacak şekilde toplam 50.000 adet tweet içeren bir veri seti kullanılmıştır. Veri setindeki tüm tweetler 2018 yılına aittir. Tarih belirlenmesi yapılırken, ocak ayından aralık ayına kadar 12 farklı sınıflandırma yapılmış olup ay bazında tarih sonucu elde edilmektedir. Yapılan eğitimlerin test sonuçlarına bağlı olarak, sadece kullanıcı adı belirlenmesi ortalama %56 oranında, sadece ay bazında tarih belirlenmesi ortalama %53 oranında doğruluk sonucu vermiştir.

Çok görevli öğrenme ile kullanıcı adı ve ay bazında tarih belirlenmesi testler sonucunda %63 oranında doğruluk sonucu vermiştir. Yapılan tüm testler sonucunda çok görevli öğrenmenin, ayrı ayrı yapılan eğitimlere göre başarı oranını ortalama %8 oranında arttırdığı tespit edilmiştir. Sonuç olarak bir tweetin tarihini ve kullanıcı adını belirlerken çok görevli öğrenme sınıflandırma görevlerinin performansını ortaklaşa öğrendiği görülerek, ayrı ayrı eğitilmesine göre daha başarılı sonuçlar vermiştir.

SUMMARY

Developing technology offers us new opportunities day by day. Social media has been one of the important results of these developments. With the introduction of social media into our lives, communication has become more interactive. Examining the state of the responses to news and current events has become one of the important research topics in recent years.

This study includes multi tasking learning to determine the user name and date of a tweet and the analysis of how the success rate develops according the training of the user name and date of a tweet separately with deep learning. While the model was being trained, a data set with a total of 50,000 tweets was used, with 500 users and 100 tweets per user. All tweets in the data set belong to year of 2018. While determining the date, 12 different classifications made from January to December and the date result is determined by month class. Based on the test results of the trainings, only the user name was determined as 56% and the date determination as month was 53%.

Multitasking learning resulted in an accuracy of 63% as a result of tests with user name and date as month. As a result of all tests conducted, it was determined that multitasking learning increased the success rate by 8% on average according to the individual trainings. As a result, when determining the date and user name of a tweet, it was seen for multitasking learning learned the performance classification tasks in a common way and showed more successful results than individual training.

1. GİRİŞ

Son yıllarda sosyal medya iletişimin yeni türlerinden biri olmuştur ve popülerliği her geçen gün artmaktadır. Özellikle Facebook, Twitter, Instagram gibi sosyal platformlar bilginin oluşması, yayılması ve tartışılmasını hiç olmadığı kadar hızlandırmıştır. Sosyal medya sayesinde bilgi paylaşımı interaktif bir boyuta dönüşmüştür. Devletler, şirketler, topluluklar kararlarını, planlarını, izleyecekleri politikaları sosyal medya da oluşan algıya göre şekillendirmektedirler. Bu nedenle bir tweetin kime ait olduğu, hangi tarihte atıldığı bu kurum ve kuruluşların alacağı kararlar için önemli hale gelmiştir.

Bu proje de çok görevli öğrenme ile aynı anda bir tweetin kullanıcı adı ve tarihinin belirlenmesi ve derin öğrenmeyle bir tweetin kullanıcı adı ve tarihinin ayrı eğitilmesine göre başarı oranının nasıl geliştiğinin analizlerini kapsamaktadır.

Çok görevli öğrenme, birden çok öğrenilmesi gereken görevin aynı anda öğrenildiği, görevler arasındaki ortaklıkları ve farklılıkları kullanan makine öğrenmesinin bir alt dalıdır. Çoklu görev öğrenme ile sınıflandırma görevlerinin performansı ortaklaşa öğrenilerek gelişmesi amaçlanmaktadır.

Derin öğrenme, yapay sinir ağı adı verilen, algoritma olarak beynin yapısı ve fonksiyonundan esinlenen makine öğrenmesinin bir alt dalıdır. Derin öğrenme de yapay sinir ağı modeli bir görüntü, metin ya da ses verilerini sınıflandırma yapmayı öğrenir.

Bu çalışma için öncelikli olarak Twitter platformundan farklı kullanıcılara ait 2018 yılındaki tweetler çekilmiştir. Veri seti aşağıdaki 2 farklı formatta hazırlanmıştır.

Verilen bir tweet cümlesinden kullanıcı adı belirlenebilmesi için kullanıcı adı sınıfı veri seti aşağıdaki json formatında hazırlanmıştır..

```
“@user1” : [“tweet1”, “tweet2”, ....., “tweet100”]  
“@user2” : [“tweet1”, “tweet2”, ....., “tweet100”]  
.....  
“@user500”: [“tweet1”, “tweet2”, ....., “tweet100”]
```

ŞEKİL 1.1 Kullanıcı adı sınıfı veri seti formatı

Şekil 1.1’e göre kullanıcı adı veri setinde, her kullanıcının tweetleri tek satırda iki köşeli parantez içinde virgül ile ayrılarak sıralanmıştır.

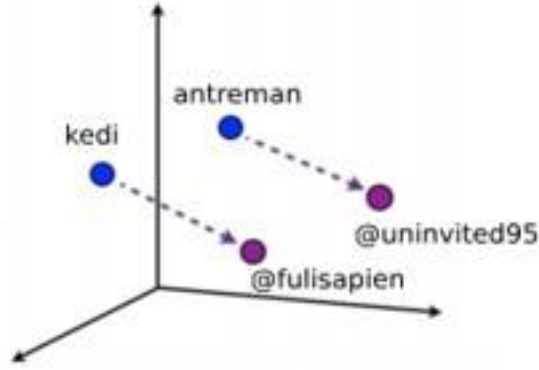
Verilen bir tweet cümlesinden ay bazında tarih belirlenebilmesi için tarih sınıfı veri seti aşağıdaki json formatında hazırlanmıştır.

```
“Ocak” : [“tweet1”, “tweet2”, “tweet3”, .....]  
“Şubat”: [“tweet1”, “tweet2”, “tweet3”, .....]  
.....  
“Aralık”: [“tweet1”, “tweet2”, “tweet3”, .....]
```

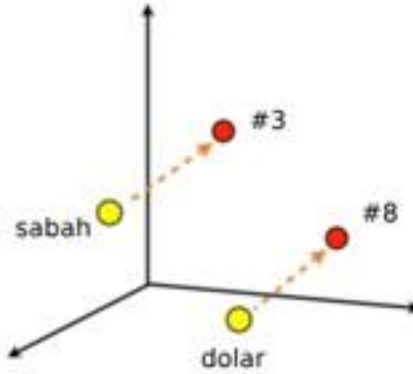
ŞEKİL 1.2 Tarih sınıfı veri seti formatı

Şekil 1.2’ye göre ay bazında tarih veri setinde, o aya ait tüm tweetler tek satırda iki köşeli parantez içinde virgül ile ayrılarak sıralanmıştır.

Model hazırlanmadan önce veri setinde bulunan tweetlerdeki kelimeleri vektör uzayında ifade etmek için tahmin temelli olan word2vec vektörü kullanılmıştır. Tweetler word2vec vektörüne yerleştirildiklerinde 3 boyutlu uzayda aşağıdaki modellemelerdeki gibi gözükcektir.



ŞEKİL 1.3 Vektör Uzayında Kullanıcı ve Kelimelerin Gösterimi



ŞEKİL 1.4 Vektör Uzayında Kelime ve Tarih Gösterimi

ŞEKİL 1.4’de #3, #8 ifadeleri ayların sayısal karşılığı olarak ifade edilmiştir. Örneğin ocak ayı #1, aralık ayı #12’dir.

Kelimeleri vektöre yerleştirme işlemi yapılırken, tekrarlayan kelimeler, tüm noktalama işareti, gereksiz boşluklar uzaklaştırılmıştır. Nasıl yapıldığına dair ayrıntılı bilgiler yöntem bölümünde verilmiştir.

Eđitime bařlamadan nce test setinde bařarı oranını arttırıp ezberlemeyi azaltmak iin veri seti karıřtırılmıřtır. Model oluřturulurken Google'ın aık kaynak kodlu zellikle derin đrenme iin kullanılan makine đrenmesi ktphanesi olan tensorflow kullanılmıřtır. Model inřa edilirken her hedef sınıfın olası tm hedef sınıflar zerindeki olasılıklarını hesaplamak iin softmax fonksiyonu kullanılmıřtır.

Derin đrenme mimarisi olarak karmařık iliřkileri modelleyebilmek iin ileri beslemeli ađ yapısında sahip olan DNN kullanılmıřtır. Bu sayede DNN mimarileri nesne ilkellerinin katmanlı bir bileřimi olarak ifade edildiđi bileřimsel modeller retmiřtir.

Model hazırlandıktan sonra test iin veri setinde bulunan 500 kullanıcının her birinin rastgele 10 adet tweeti seilerek 5000 verilik ve her defasında bu 10 adet tweet farklılařtırılarak 10 adet test sınıfı hazırlanmıřtır.

Tahmin iřlemi yapılırken test setleri, tm kelimelerin cmlesini ve listesini alan ve verileri tensorflow'a beslenebilecek biimde geri dndren bir fonksiyona gnderilmektedir.

Ayrı 10 farklı test sınıfı iin eđitim yapıldıktan sonra , sadece kullanıcı adı belirlenmesi ortalama %56 oranında, sadece ay bazında tarih belirlenmesi ortalama %53 oranında dođruluk sonucu verdiđi gzlenmiřtir.

Aynı anda ok grevli đrenme ile eđitim yaparken eřitli grevlere zel ıktı katmanlarının elde edilmesi ve gizli grevlerin tm grevler arasında paylařması iin modelde sabit parametre paylařım yntemi kullanılmıřtır.

Aynı anda eđitim yapılırken grevlerin arasında bir paylařım olması ve ortaklařa đrenme olduđu iin kullanıcı adı ve tarih belirlenmesi %63 oranında bařarı sađlamıřtır.

2. YÖNTEM

Bu projede öncelikli olarak twitter developer account hesabı açılmış olup, twitterdan tweet verisi çekebilmek için twitter hesabına kod ile erişimi sağlamak için lazım olan consumer key, access key bilgileri için başvuru yapılmıştır. Onay alındıktan sonra twitterdan veri çekmek için kullanılan tweepy kütüphanesi kullanılarak python dili ile bir script yazılmıştır. Bu scriptte öncelikli olarak access key, access secret, consumer key, consumer secret bilgileri ile twitter hesabına bağlantı sağlanmıştır. Bu script kodunda tweet bilgilerine ulaşmak istenen hesapların kullanıcı adları users dizinine elle girilmesi gerekmektedir. Bir kişinin tweet bilgilerini çekebilmek için o kişinin hesabı herkese açık olmalıdır veya benim takip ettiğim kişiler arasında olmalıdır.

```
126
127 for user in users_set:
128     public_tweets = api.user_timeline(screen_name = user, count = 100)
129
```

ŞEKİL 2.1 Kullanıcı adlarının bulunduğu users_set isimli dizinden her kullanıcı için 100 adet (count=100) tweet veri objesi çekme işlemi

Tweet objeleri çekildikten sonra bu objeler içinden tweet ve tarih bilgisi alınmalıdır. Bunun için obje sınıfına ait text ve created_at methodları kullanılmıştır.

```
140
141     for tweet in public_tweets:
142         tweet.text
143         tweet.created_at.isoformat()[1]
144
```

ŞEKİL 2.2 Tweet objesinden tweet ve tarih bilgisini alma

Twitterdan çekilen bu bilgiler very setini oluşturmak için ŞEKİL 1.1 ve ŞEKİL 2.2'deki formatlara uygun olacak şekilde dosyaya yazılmıştır.

Ayrı ayrı derin öğrenme ile eğitim yapmadan önce kullanıcı adı ve tarih sınıflarındaki kategorileri belirlemek için, veri seti formatında bulunan “.” işaretinden önceki kısımlar sınıf kategorileri olarak belirlenmiştir. Örneğin ŞEKİL 1.1’de gösterilen kullanıcı adı sınıfı için @user1, @user2 örnek sınıf kategorileridir. ŞEKİL 1.2’de gösterilen tarih sınıfı için ocak, şubat örnek sınıf kategorileridir.

```
19
20 tbl = dict.fromkeys(i for i in xrange(sys.maxunicode)
21 if unicodedata.category(unichr(i)).startswith('P'))
22
23 categories = list(data.keys())
```

ŞEKİL 2.3 Veri setinden kategorilerin ayrılması işlemi

Kategori ayrımı sonucu kullanıcı adı sınıfında 500 farklı kullanıcı olduğu için 500 kategori, tarih sınıfında ay bazında sınıflandırma olduğu için 12 kategori elde edilmiştir.

Eğitim sırasında gürültü kirliliğini engellemek ve hafıza da yer kaplamaması için tüm noktalama işaretleri, fazla boşuklar, tekrar eden cümleler, emojiler , yeni satır karakterleri uzaklaştırılmıştır. Ayrıca eğitim sırasında modelin ezberleme yapmasını engellemek için eğitim başlamadan önce very seti parçalara ayrılmadan önce veri seti karıştırma manasına gelen shuffle yöntemi uygulanmıştır.

```
89
90 random.shuffle(training)
91 training = np.array(training)
92
```

ŞEKİL 2.4 Shuffle işlemi

Derin öğrenme mimarisi inşa edilirken tensorflow kütüphanesi ve ilişkileri modelleyebilmek için ileri beslemeli ağ yapısında sahip olan DNN kullanılmıştır.

```
103
104 net = tflearn.input_data(shape=[None, len(train_x[0])])
105 net = tflearn.fully_connected(net, 8)
106 net = tflearn.fully_connected(net, 8)
107 net = tflearn.fully_connected(net, len(train_y[0]), activation='softmax')
108 net = tflearn.regression(net)
109 model = tflearn.DNN(net, tensorboard_dir='tflearn_logs')
110
```

ŞEKİL 2.5 Derin öğrenme modeli python ile tasarımı

Eğitimler sonucunda linerazasyonu azaltmak ve daha doğru değerler elde etmek için girdilerin parçalar halinde işlenmesini ifade eden batch size 8 olarak belirlenmiştir. Batch size değerine göre model eğitilirken verilerin tamamı aynı anda eğitime katılamadığından model için en uygun ağırlık değerlerini hesaplamak adına eğitim adımları anlamına gelen epoch değeri 1000 olarak belirlenmiştir.

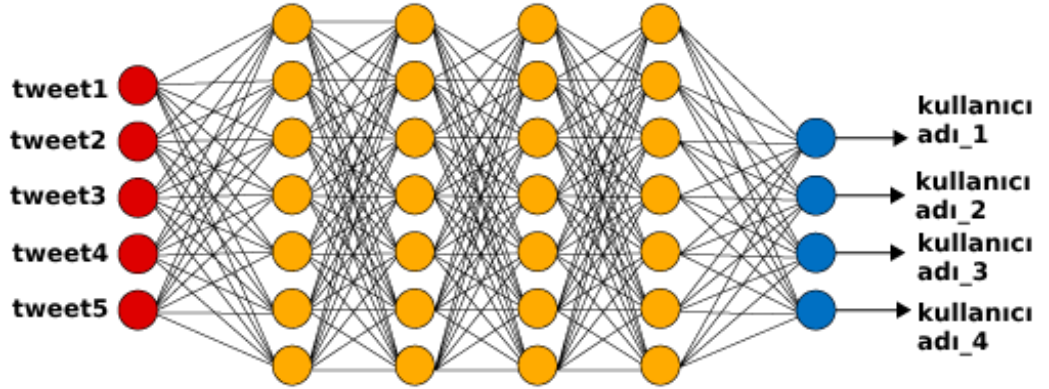
```
114  
115 model.fit(train_x, train_y, n_epoch=1000, batch_size=8, show_metric=True)  
116 model.save('model.tflearn')  
117
```

ŞEKİL 2.6 Derin öğrenme modeli eğitim parametreleri

Model eğitildikten sonra test seti verilirken tahmin sonuçları için tüm kelimelerin cümlesini ve listesini alan ve verileri tensorflow'a beslenebilecek şekilde geri döndüren bir get_tf_record isiminde bir fonksiyon yazılmıştır. Aşağıda ŞEKİL 2.7’de örnek olarak 4 tweet için tahmin olayının kod ile nasıl yapıldığı gösterilmiştir.

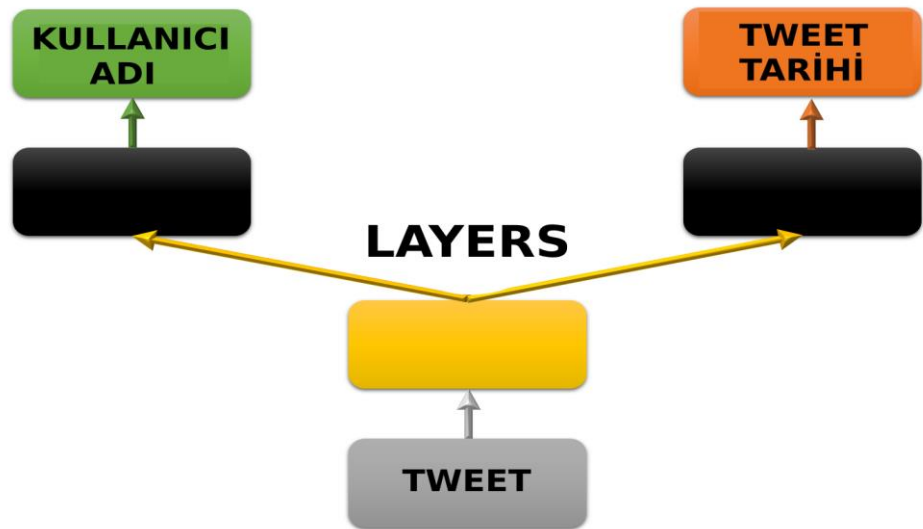
```
161  
162 tweet_1 = "Kredi kartıyla milli piyango bileti almak " #@zehirli_oksijen  
163 tweet_2 = "EserYenenlerShow bu akşam 23.15'te TV8'de. eseryeneler hayirli olsun Eser'im :)" #@acunilicali  
164 tweet_3 = "ekonomikanaliz arzovaone Hocam basiniz sag olsun. Allah rahmet etsin." #@bmylz  
165 tweet_4 = "muzik Bu sene Grammy'yi 'Rock' kategorisinde bu parca alir: https://t.co/QRnnKX29Sc" #@kaansezyum  
166 print(categories[np.argmax(model.predict([get_tf_record(tweet_1)]))])  
167 print(categories[np.argmax(model.predict([get_tf_record(tweet_2)]))])  
168 print(categories[np.argmax(model.predict([get_tf_record(tweet_3)]))])  
169 print(categories[np.argmax(model.predict([get_tf_record(tweet_4)]))])  
170
```

ŞEKİL 2.7 Verilen tweetin kullanıcı adı veya tarih olarak tahmin edilmesinin 4 örnek için python kodu ile gösterimi



ŞEKİL 2.8 Derin öğrenme ile kullanıcı adı eğitiminin modelleştirilmesi

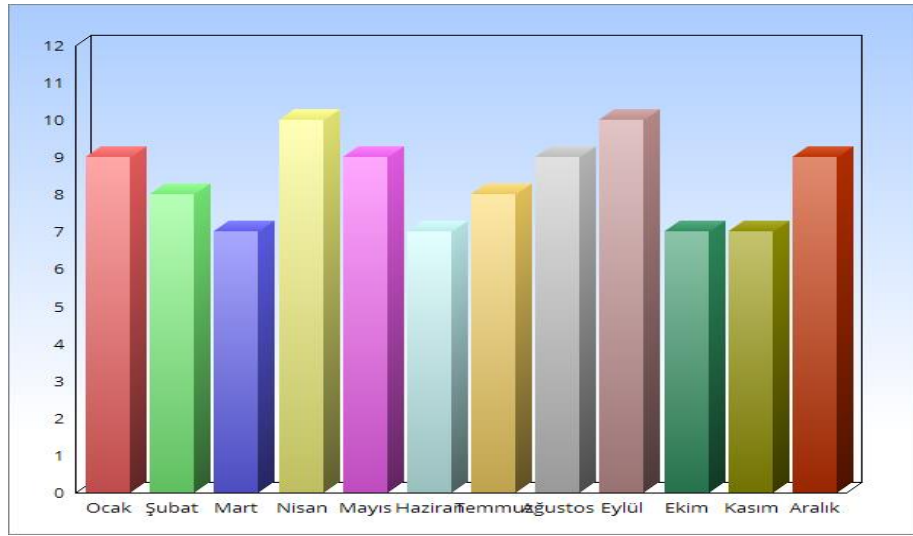
Kullanıcı adı ve tarihi ayrı ayrı eğitilip sonuçlar alındıktan sonra (sonuçlar bulgular bölümde gösterilecektir.) kullanıcı adı ve tarih belirleme görevleri için özel çıktı katmanlarının uygulanması ve görevlerin tüm görevler arasında paylaşılması için sabit parametre paylaşım methodu kullanılmıştır. Bu method ile aşırı uydurma riskinin büyük ölçüde azaldığı görülmüştür. Aslında paylaşılan parametrelere yüklenme riskinin görevlere göre sayısıdır. Başarılmak istenen görev özel parametrelerin yani çıkış katmanlarının aşırı donatılmasından daha küçük bir görev olarak gözükmektedir.



ŞEKİL 2.9 Çok görevli öğrenme ile bir tweete ait kullanıcı adı ve tarihin aynı anda farklı kollarda eğitilmesi örnek model gösterimi

3. BULGULAR

Bu proje de model eğitimleri sonucu elde edilen bulgular 500 kullanıcı ve her kullanıcıya ait 100 tweet içeren 50.000 adetlik veri seti üzerinde, her kullanıcıya ait 10 farklı tweet örneğinin bulunduğu 10 farklı sınıftaki test verisi ile sağlanmıştır. Veri setinde bulunan tweetlerin aylara göre yüzdesel dağılımı aşağıdaki gibidir.



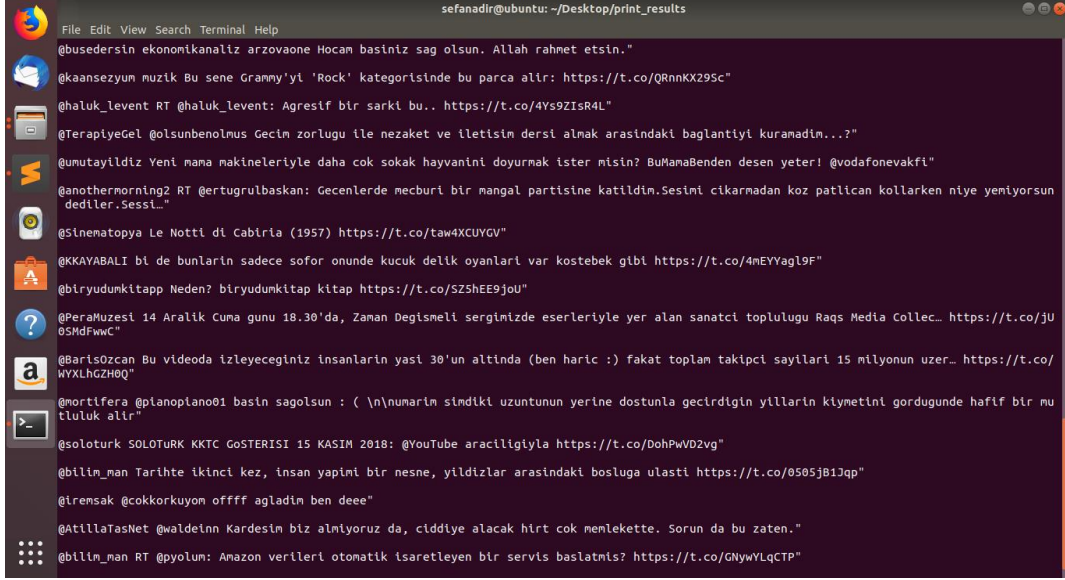
ŞEKİL 3.1 Veri setinde tweetlerin aylara göre dağılımı

Derin öğrenme ile sadece kullanıcı adı tahmini yapılırken alınan sonuçların bir kısmı ŞEKİL 3.2 ve ŞEKİL 3.3’de gösterildiği gibidir.

```
@zehirci oksijen Kredi kartıyla milli piyango bileti almak "  
@acunilicali EserYenelerShow bu aksam 23.15'te TV8'de. eseryeneler hayirli olsun Eser'im :) https://t.co/8TjBAfItol"  
@bmlyz ekonomikanaliz arzovaone Hocam basiniz sag olsun. Allah rahmet etsin."  
@kaansezyum muzik Bu sene Grammy'yi 'Rock' kategorisinde bu parca alır: https://t.co/QRnnKX29Sc"  
@haluk levent RT @haluk levent: Agresif bir sarki bu.. https://t.co/4Ys9ZIsR4L"  
@TerapiyeGel @olsunbenolmus Gecim zorlugu ile nezaket ve iletisim dersi almak arasindaki baglantiyi kuramadim...?"  
@glshybs Yeni mama makineleriyle daha cok sokak hayvanini doyurmak ister misin? BuMamaBenden desen yeter! @vodafonevakfi"  
@anothermorning2 RT @ertugrulbaskan: Gecenlerde mecburi bir mangal partisine katildim.Sesimi cikarmadan koz patlican kollarken niye yemiyorsun dediler.Sessi.."  
@Sinematopya Le Notti di Cabiria (1957) https://t.co/taw4XCUYGV"  
@terskaplumbaa bi de bunlarin sadece sofor onunde kucuk delik oyanlari var kostebek gibi https://t.co/4mEYyag19F"  
@biryudumkitapp Neden? biryudumKitap kitap https://t.co/SZ5hEE9joU"  
@PeraMuzesi 14 Aralik Cuma gunu 18.30'da, Zaman Degismeli sergimizde eserleriyle yer alan sanatci toplulugu Raqs Media Collec...  
https://t.co/jU0SMdFwWC"  
@BarisOzcan Bu videoda izleyeceginiz insanlarin yasi 30'un altinda (ben haric :) fakat toplam takipci sayilari 15 milyon uzer...  
https://t.co/WYXlhGZH0Q"  
@mortifera @pianopiano01 basin sagolsun : ( \n\nnumarim simdiki uzuntunun yerine dostunla gecirdigin yillarin kiymetini gordugunde hafif bir mutluluk alır"  
@soloturk SOLOTuRK KKTC GoSTERISI 15 KASIM 2018: @YouTube araciligiyla https://t.co/DohPwVD2vg"  
@bilim man Tarihte ikinci kez, insan yapimi bir nesne, yildizlar arasindaki bosluga ulasti https://t.co/0505jB1jqP"  
@iremsak @cokkorkuyom offff agladim ben deee"  
@AtillaTasNet @waldeinn Kardesim biz almiyoruz da, ciddiye alacak hirt cok memlekette. Sorun da bu zaten."  
@LaleAkarun RT @pyolum: Amazon verileri otomatik isaretleyen bir servis baslatmis? https://t.co/GNywYLqCTP"  
@yazilimci adam Gutenberg Firefox'ta asiri yavas. Bildigin makine otuyor. Uzunca yazi yazmiyorum sirf bu yuzden."
```

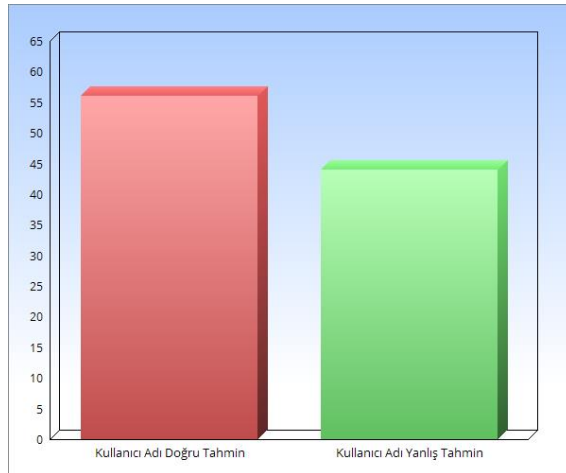
ŞEKİL 3.2 Kullanıcı adı test sınıfı örneği

ŞEKİL 3.2’de doğruluğu karşılaştırmak için tweetlerin kime ait olduğu sol taraflarına örnek olarak yazılmıştır. ŞEKİL 3.2’de verilen test verisinin sonuçları aşağıdaki gibidir.



ŞEKİL 3.3 Derin öğrenme ile sadece kullanıcı adı tahmin sonuçları

ŞEKİL 3.2 ve ŞEKİL 3.2 incelendiği zaman @anothermorning2, @Sinematopya ve çoğu kullanıcının tweetlerini doğru tahmin ettiği görülmektedir. Ancak ŞEKİL 3.2’de bulunan @yazilimci_adam adlı kullanıcının tweeti @bilim_man olarak sonuç vermiştir. Buradan bu iki kullanıcının benzer konulardan sıkça bahsettiği yorumu çıkarılmıştır. Yapılan eğitimler sonucunda kullanıcı adı tespitinde %56 oranında başarı oranı yakalanmıştır.

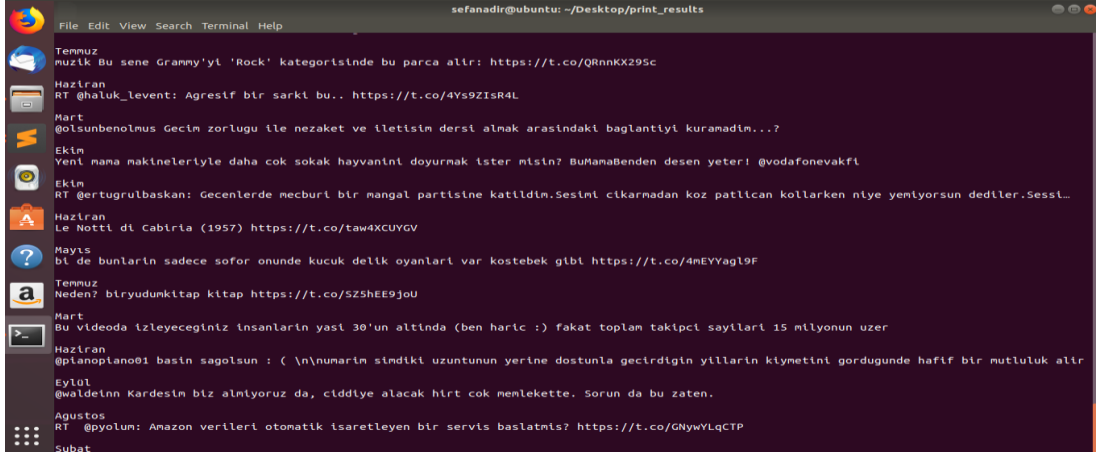


ŞEKİL 3.4 Derin öğrenme ile kullanıcı adı tahmini yüzdesel gösterimi

Derin öğrenme ile sadece ay bazında tarih tahmini yapılırken alınan sonuçların bir kısmı ŞEKİL 3.5 ve ŞEKİL 3.6’da gösterildiği gibidir.

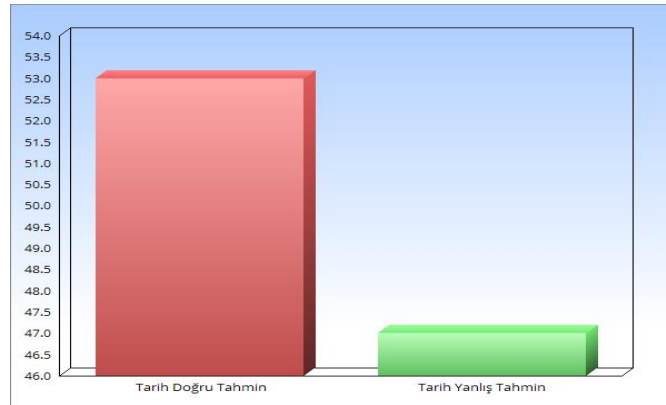
1	Temmuz	ekonomikanaliz arzovaone Hocam basınız sag olsun. Allah rahmet etsin.
2	Ekim	muzik Bu sene Grammy'yi 'Rock' kategorisinde bu parca alır: https://t.co/QRnnKX29Sc
3	Haziran	RT @haluk_levent: Agresif bir sarki bu.. https://t.co/4Ys9ZIsR4L
4	Mart	@olsunbenolmus Gecim zorlugu ile nezaket ve iletisim derisi almak arasindaki baglantiyi kuramadim...?
5	Ekim	Yeni mama makineleriyle daha cok sokak hayvanini doyurmak ister misin? BuMamaBenden desen yeter! @vodafonevakfi
6	Nisan	RT @ertugrulkaskan: Gecenlerde mecburi bir mangal partisine katildim.Sesimi cikarmadan koz patlican kollarken niye yemiyorsun dediler.Sessi...
7		
8	Ekim	Le Notti di Cabiria (1957) https://t.co/taw4XCUYGV
9	Mayıs	bi de bunların sadece sofor onunde kucuk delik oyanlari var kostebek gibi https://t.co/4mEYyag19F
10	Mayıs	Neden? biryudumkitap kitap https://t.co/SZShEE9joU
11	Ekim	Bu videoda izleyeceginiz insanların yasi 30'un altinda (ben haric :) fakat toplam takipci sayilari 15 milyonun uzer
12	Haziran	@pianopiano01 basin sagolsun : (\n\numarim simdiki uzuntunun yerine dostunla gecirdigin yillarin kiymetini gordugunde hafif bir mutluluk alır
13		@waldeinn Kardesim biz almiyoruz da, ciddiye alacak hirt cok memlekette. Sorun da bu zaten.
14	Eylül	RT @pyolum: Amazon verileri otomatik isaretleyen bir servis baslatmis? https://t.co/GNywYLqCTP
15	Ağustos	Gutenberg Firefox'ta asiri yavas. Bildigin makine otuyor. Uzunca yazı yazmıyorum sirf bu yuzden.
16	Haziran	

ŞEKİL 3.5 Ay bazında tarih test sınıfı örneği



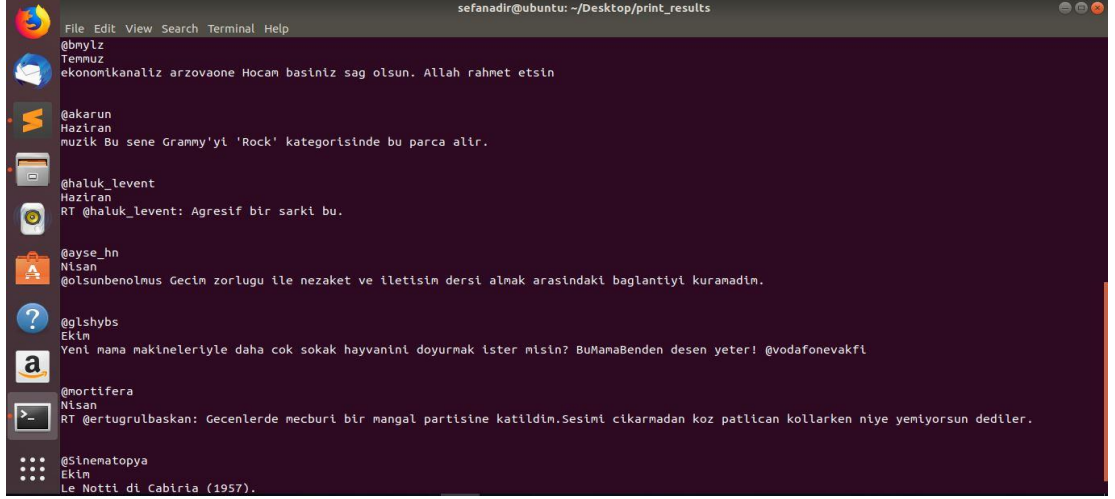
ŞEKİL 3.6 Derin öğrenme ile sadece ay bazında tarih tahmin sonuçları

ŞEKİL 3.5 ve ŞEKİL 3.6 incelendiği zaman bazı sonuçların doğru bazılarının ise yanlış olduğu ama çoğunlukla doğru sonuçlar içerdiği görülmektedir. Yanlış gelen sonuçların ise ŞEKİL 3.1 de yaklaşık olarak veri setinde fazla yüzdeye sahip aylar olduğu gözlemlenmiştir. Yapılan eğitimler sonucunda ay bazında tarih tespitinde %53 oranında başarı oranı yakalanmıştır.



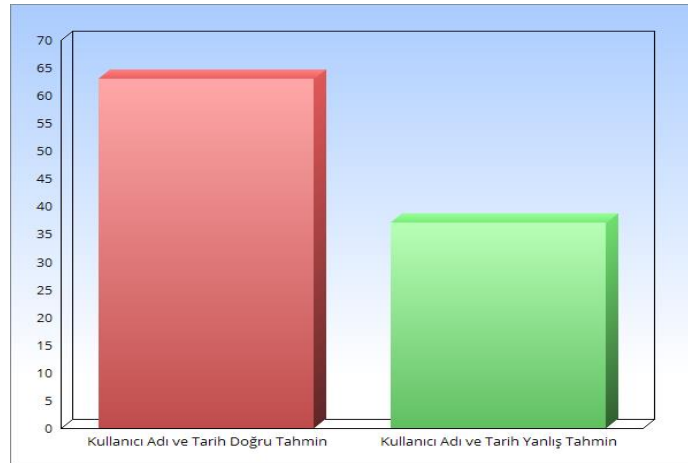
ŞEKİL 3.7 Derin öğrenme ile tarih tahmini yüzdesel gösterimi

Çok görevli öğrenme ile yapılan eğitim sonuçları ŞEKİL 3.8'deki gibi gözükmetedir. Önce kullanıcı adı sonra ay bazında tarih ve bu bilgilerin hangi tweete ait olduğu sonuç olarak gözükmetedir.



ŞEKİL 3.8 Çok görevli öğrenme ile kullanıcı adı ve tarih tahmini

ŞEKİL 3.2 ve ŞEKİL 3.5 incelendiğinde çok görevli öğrenme ile kullanıcı adı ve tarihinin bazı sonuçlarda birisi doğru birisi yanlış gözükürken bazılarında ikisi de doğru çıkmıştır. Tüm test sınıfları ile denendiğinde de hem kullanıcı adı hem de tarihin aynı anda daha çok doğru çıktığı gözlemlenmiştir. Yapılan tüm eğitimler sonucunda ise %63 oranında doğruluk gözlenmiştir.



ŞEKİL 3.9 Çok görevli öğrenme ile kullanıcı adı ve tarih tahmini yüzdesel gösterimi

4. TARTIŞMA VE SONUÇ

Proje sonunda derin öğrenme ile sadece kullanıcı adı ve ay bazında tarih tahmini için %60 oranında başarı beklenmekteydi. Testler ve eğitimler sonucunda sadece kullanıcı adı eğitimi %56, sadece ay bazında tarih tahmini %53 oranında başarı sonu göstermiştir. Çoklu görev öğrenme ile aynı anda kullanıcı adı ve tarih eğitimi sonucunda da %60 başarı oranı beklenmekteydi. Testler ve eğitimler sonucunda çoklu görev öğrenme sağlanan başarı oranının ortalama %63 olduğu gözlemlenmiştir. Burada çok görevli öğrenmenin sınıflandırma görevlerinin performansını ortaklaşa öğrenerek başarıyı arttırdığı gözlemlenmiştir. Çok görevli öğrenme de kullanılan sabit parametre paylaşım methodunun gizli görevleri tüm görevler arasında paylaşması ve her göreve özel çıktı katmanları uygulamasıyla beklenen %60 başarı oranının geçilmesini sağlamıştır. Çalışmanın geliştirilmesi için veri seti arttırılabilir, veri setinde daha ayrıntılı çalışma yaparak hafıza da yer kaplayacak gereksiz tüm bilgiler daha fazla temizlenebilir. Tahminlerin iterasyonu arttırılarak başarı oranını artması sağlanabilir. Eğitim modellerine ek katmanlar ekleyerek ve çeşitli ağlarda eğitim yapılarak daha uygun bir ağ bulunabilir. Veri setinde tespit edilen gürültüler düzenlenebilir. Girdiye çeşitli işlemler uygulanabilir. Bu yöntemler ile sonraki çalışmalarda daha iyi sonuçların elde edilmesi mümkündür.

KAYNAKLAR

- [1]<https://www.technologyreview.com/s/513696/deep-learning/>
- [2]<https://dzone.com/articles/comparison-between-deep-learning-vs-machine-learn>
- [3]<https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network>
- [4]<https://www.linkedin.com/pulse/derin-öğrenme-uygulamlarında-başarım-iyileştirme-necmettin-çarkacı>
- [5]<https://medium.com/jatana/report-on-text-classification-using-cnn-rnn-han-f0e887214d5f>
- [6]https://github.com/brightmart/text_classification
- [7]<https://academic.oup.com/nsr/article/5/1/30/4101432>
- [8]<https://blog.fastforwardlabs.com/2018/07/24/ff08-launch.html>
- [9]<https://ieeexplore.ieee.org/document/6406646>
- [10]<https://openreview.net/forum?id=B1nZ1weCZ>
- [11]<https://www.endustri40.com/yapay-zeka-makine-ogrenimi-ve-derin-ogrenme-arasindaki-farklar/>