İmla Hatalı Cümleleri Düzelten Sistem (Machine Translation)

BIL 495

Yunus ÇEVİK

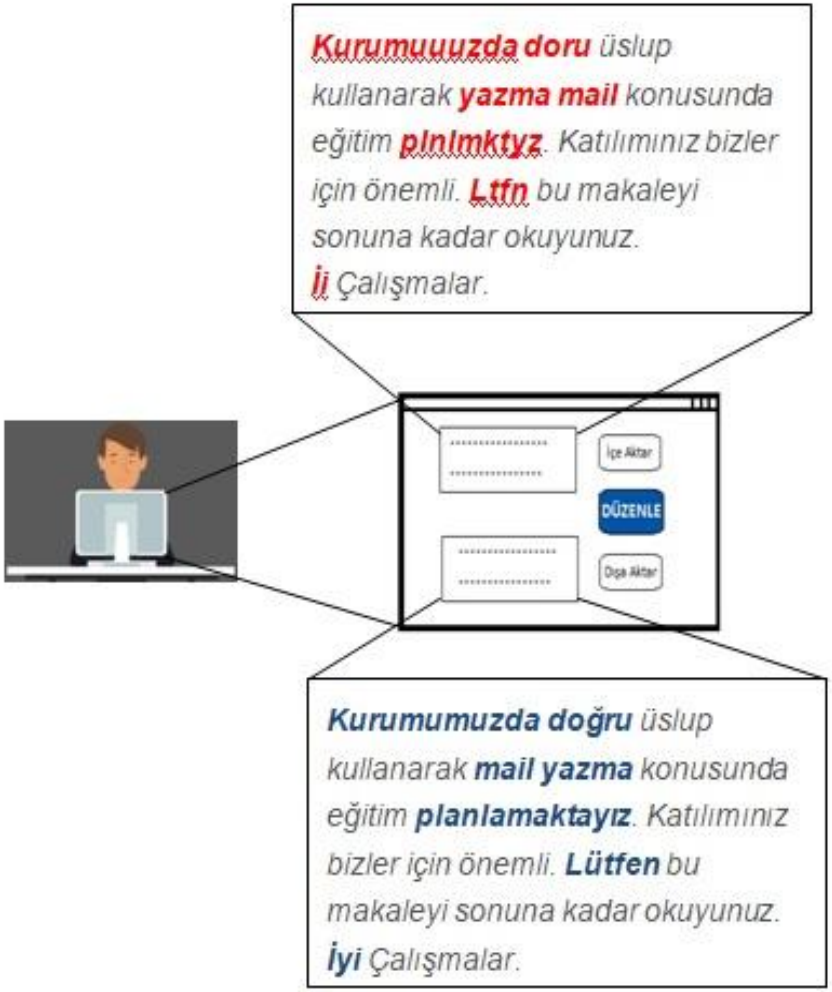
Proje Danışmanı: Dr. Öğr. Üyesi Burcu YILMAZ

Ocak 2019

İçerik

- Projenin Şeması ve Tanımı
- Proje Tasarım Planı
- Doğru İmla Kurallı Dataset
- Yanlış İmla Kurallı Dataset
- Doğru İmla Kurallı Datasetin Bozulması
- Datasetlerin Eğitilmesi
- Karşılaşılan Problemler - Underfitting
- Karşılaşılan Problemler – Overfitting
- Modelin Özet Bilgileri
- 2000 Samples – 200 Epoch – 100 Batch Size
- Bidirectional LSTM İle Eğitme
- Bidirectional LSTM ile Modelin Özet Bilgisi
- Eğitilmiş Verilerin Encode ve Decode İle Tahmini
- Kaynaklar

Proje Şeması ve Tanımı

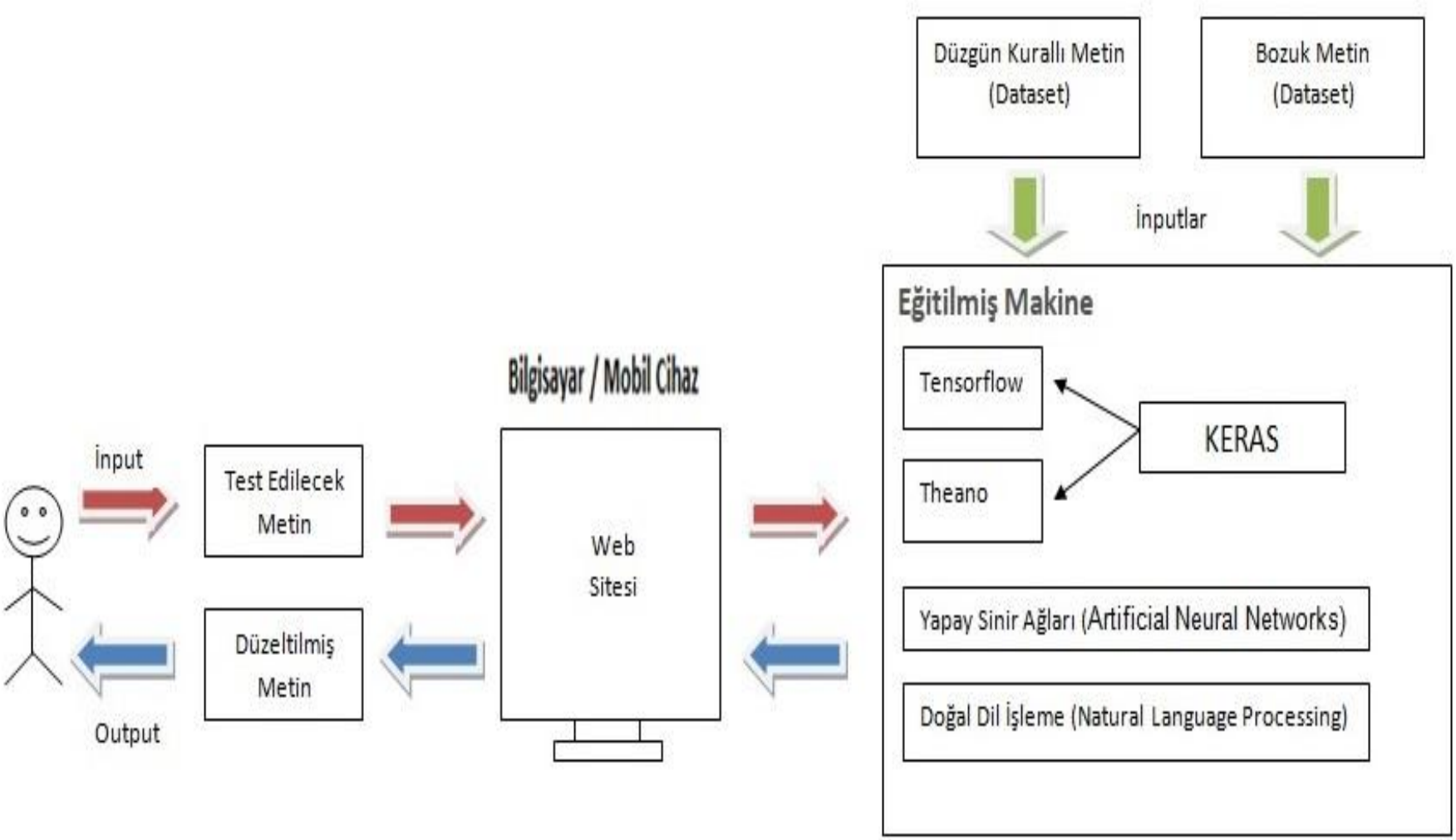


- Günlük hayatta birçok kelime kısaltılarak ve imla kurallarından saparak kullanılır. Bu duruma örnek olarak mesajlaşma dilinde "merhaba" yerine "mrhb" veya "oğlum" yerine "olum" kullanılması veya bilgisayarda on parmak yazı yazılırken yanlış bir harfe basmak verilebilir.

  Bu durum resmi yazılarda imla bozukluğundan dolayı anlam bozukluğuna ve anlam karmaşasına yol açar.

- İmla Hatalı Cümleleri Düzelten Sistem ise bu tür problemleri Türkçe imla kurallarına bağlı olarak tekrardan düzenler ve bu şekilde oluşabilecek sorunların giderilmesine yardım eder.

Proje Tasarım Planı



Doğru İmla Kurallı Dataset

– 2015 yılına ait makale, dergi, gazete gibi düzgün kaynaklardan Python da **crawler** yardımı ile bilgiler çekilerek doğru imla kurallı bir dataset oluşturulmuştur.

rightDataset.csv

1 Şirketlerin EFF'den geçer not alması için devletlerin kullanıcılar üzerinde bulunduğu taleplerin açıklanması, kullanıcı gizliliği, devletlerin içerik kaldırma taleplerinin paylaşımı, verilerin saklanması konusunda açık ilkeler ve sektörün kabul ettiği en iyi uygulama kurallarını uyma kategorilerinden geçer not alması gerekiyor.

2 Buna göre Apple başta olmak üzere Adobe, Dropbox, Yahoo ve Wordpress 5 yıldız alırken, Google ve Microsoft ise sadece üç kategoride geçer not alabildi.

3 Google devletlerin içerik kaldırma gibi taleplerini paylaşma konusunda pek de cömert davranmazken, diğer yandan da kullanıcıların verlerini nasıl kendinde sakladığını da tam olarak belirtmiyor.

4 Bu da şirketin bu noktalardan geçer not alamamasına yol açıyor.

5 Microsoft'ta da durum çok farklı değil.

6 Ancak raporda en kötü notu alan şirket Whatsapp oluyor.

7 Facebook 'un kendisi 4 yıldız alırken, yine Facebook'un altında bir platform olan Whatsapp'ta durum hiç iç açıcı değil; zira Whatsapp'ın geçer not aldığı tek madde veri güvenliği... Bunun dışında Whatsapp'ın pek de şeffaf olmadığı görülüyor.

8 Diğer bir deyişle Whatsapp'ta içeriklerin devletlerle paylaşılıp paylaşılmadığı belli değil; diğer yandan Whatsapp'ın konuşma kayıtlarının ne kadarını sakladığı ise tam bir muamma... Merkezi Amerika Birleşik Devletleri'nde bulunan, kısa adıyla EFF diye bilinen Electronic Frontier Foundation, amacı siber dünyada vatandaş hürriyetlerini özellikle keyfi polis baskınlarından, ve kusurlu yetki kullanımı, ya da gereksiz kanunlarla getirilen kısıtlamalardan korumak olan uluslararası bir sivil toplum örgütü olarak biliniyor.

9 Hâlâ deneme sürümü ücretsiz olarak yüklenebilen işletim sistemini test eden kullanıcılara ücretsiz kullanım imkânı da sağlanacak.

10 Microsoft'tan yapılan açıklamaya göre Windows 10'u test eden kullanıcılar son sürüme ücretsiz olarak güncelleştirme yapabilecek.

Yanlış İmla Kurallı Dataset

- Oluşturulan dataset (Doğru imla kurallı) de bulunan kelimelerin veya cümlelerin yapısının bozulması ile yeni bir dataset (Bozuk imla kurallı) oluşturulmuştur.

- Bu bozma işlemi kelime uzunluğunun %15' i oranında kelimedeki harflerin yerinin değiştirilmesi, tekrar eden yeni bir harf eklenmesi, sessiz harflerin atılması veya ilk ünlü harf hariç diğer ünlü harflerin çıkarılması ile yapılmıştır.

Doğru İmla Kurallı Datasetin Bozulması

```
def checkText(lists):
    for text in lists:
        if (1 < len(text)-1):
            switcher_function(random.randrange(0,6),text)
        else:
            writeTXT(text)

def switcher_function(argument, text):
    if(argument == 0):
        writeTXT(replaceAlphabet(text))
    elif(argument == 1):
        writeTXT(addAlphabet(text))
    elif(argument == 2):
        writeTXT(dropVowelAlphabet(text))
    elif(argument == 3):
        writeTXT(dropConsonantAlphabet(text))
    else:
        writeTXT(text)
```

```
def replaceAlphabet(text):
    textList = list(text)
    piece =round(len(text) * 0.15)
    rand = random.sample(range(0, len(text)-1), piece)
    for i in rand:
        if textList[i].isalpha():
            textList[i] = sameAlphabet(textList[i])
    return ''.join(textList)

def sameAlphabet(char):
    if(char == "a"):
        sAlphabet = "qwsz"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
    elif(char == "A"):
        sAlphabet = "QWSZ"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
    elif(char == "b"):
        sAlphabet = "vghn"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
    elif(char == "B"):
        sAlphabet = "VGHN"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
    elif(char == "c"):
        sAlphabet = "xdfv"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
    elif(char == "C"):
        sAlphabet = "XDFV"
        retVal = sAlphabet[random.randrange(0,len(sAlphabet))]
        return retVal
```

Yanlış İmla Kurallı Dataset

wrongDataset.csv

1 Şirketlerin EFF'den ggeçer not alması için devletlerin kulanıcılar üzerinde bulunduğu taleplerin açıklanması, kullanıcı gizliliği, devltlrn içerik kalrırma taleplerinin paylaşımı, verlrn saklanması konusunda aşık ilkeler ve sektörün kabul ettiği en iyi uyglm kurallarını uym katfgorilerimden geçer not alması gerekiiyyor.

2 Buna göre Apple başta olmak üzege Adobe, Dropbx, Yahoo ve Wordpress 5 yıldız alırkenn, Google ve Micrsft ise sadece üç kategprüde geçer not alabildi.

3 Googlle devletlerin içerik kaödırma ggibi taeplerini payylaşma konusunea pek de cömrt davrnmzkn, diğdr yandan da kullanıcıların ferlerini nasıl kendnd saakladdığını da tam olarak belirrtmmiyor.

4 Bu da şirketiin bu noktalardan geçsr not alamaöasjna yol açyr.

5 Çidrosoft'ta da durum çok farklı değil.

6 Ancak raporda en kötü not aln şirkt Dhatsapp olluyor.

7 Facebook 'un kends 4 yıldız alırken, yüne Facehpok'un altında bir platfrm olan Whatsapp'ta durm hiç iç açıcı ddeğil; zir Whasapp'ın geçer not aldığı tek madde ver güvenliği... Bjnun dışında Whatsapp'ın pek de şefff olmdğ görüllüüyor.

8 Diğr bir deyişle WWhatsapp'tta içerrikklerin devltlrl paylllaşılıp paylaşılmadığı belli reğil; diğer yandan Whatsapp'ın konuşma kayıtlarının ne kadarını sakşadığı ise tam bir muamma... Merkezi AAmerika Birleşik Devpetleri'nde bılunan, kısa adıylla EFF diye biliinen Elevtrknic Rrontier Fouundatioon, amacı siber dünyd vatandaş hirrijetlerini özelllikle keyf polis baskılarından, ve kkusurlu ydtki kulllanımı, ya da geremsiz kanunlarla geyirilen kıssıtlamalaardan korumak oln uluslararası bir siviil topplum örgütü ooarak bilnyr.

9 Hâlâ deneme sürümü ücrtsz olarak yüklenebilen işletim sistemini test eeen kullnclr ücretsiz kullanım imkânı da sağlajaczk.

10 Micrsft'tn yapln açıklamaya gör Windws 10'u test eden kullnclr son sirüme ücretsiz olarak güncdlpeştirme yapabileceeek.

Datasetlerin Eğitilmesi

- Hazırlanan datasetlerin eğitimi için Google Colab kullanılmıştır.

- Googe Colab’ ın özellikleri;
  - CPU: Intel(R) Xenon(R) CPU @ 2.30 GHz
  - GPU: NVIDIA Tesla K80
  - Memory: 12.71 GB

Datasetlerin Eğitilmesi

- Eğitim aşamasında bilinmesi gereken önemli terimler:

- **Eğitim Tur (Epoch) Sayısı:** Model eğitilirken verilerin tamamı aynı anda eğitime katılmaz. Belli sayıda parçalar halinde eğitimde yer alırlar. Eğitim adımlarının her birine “epoch” denilmektedir.

- **Batch Sayısı:** Model tasarlanırken batch parametresi olarak belirlenen değer; modelin aynı anda kaç veriyi işleyeceği anlamına gelmektedir.

- Epoch sayısı arttıkça modelin başarımı gözle görülür oranda artmaktadır.

- Batch size küçük olması iyileştirme (reguralization) etkisi yaratmaktadır. Modele veri büyük gruplar halinde verildiğinde ezberleme daha fazla olur.

Datasetlerin Eğitilmesi

```
Number of Wrong Samples: 4
Number of Right Samples: 4
Number of unique input tokens: 45
Number of unique output tokens: 46
Max sequence length for inputs: 321
Max sequence length for outputs: 331
```

```
encoder_input_data = np.zeros((len(wrongSentences), max_encoder_seq_length, num_encoder_tokens), dtype='float32')
decoder_input_data = np.zeros((len(wrongSentences), max_decoder_seq_length, num_decoder_tokens), dtype='float32')
decoder_output_data = np.zeros((len(rightSentences), max_decoder_seq_length, num_decoder_tokens), dtype='float32')
```

np.zeros(): Elemanlarının hepsi 0 olan dizi oluşturmamızı sağlar.

Datasetlerin Eğitilmesi

- Listeler **one-hot encode** yapısına uygun olarak karakter tabanlı vektörler oluşturulmuştur.

```
H => [[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
E => [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
L => [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
L => [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
O => [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]
```

LSTM: Uzun/Kısa Süreli Bellek Ağları. Uzun vadeli bağımlılıkları öğrenebilen özel bir RNN türüdür. Çok çeşitli problemlerde çok iyi çalıştıkları için günümüzde çok yaygın şekilde kullanılıyorlar.

Karşılaşılan Problemler - Underfitting

```
Train on 2 samples, validate on 2 samples
Epoch 1/100
2/2 [==============================] - 3s 2s/step - loss: 2.7841 - acc: 0.2024 - val_loss: 1.4712 - val_acc: 0.0604
Epoch 2/100
2/2 [==============================] - 1s 303ms/step - loss: 2.7453 - acc: 0.1178 - val_loss: 1.3415 - val_acc: 0.0514
Epoch 3/100
2/2 [==============================] - 1s 304ms/step - loss: 2.4715 - acc: 0.0937 - val_loss: 1.4148 - val_acc: 0.0483
Epoch 4/100
2/2 [==============================] - 1s 287ms/step - loss: 2.7610 - acc: 0.0906 - val_loss: 1.2402 - val_acc: 0.0347
                                                    .
                                                    .
                                                    .
Epoch 97/100
2/2 [==============================] - 1s 297ms/step - loss: 1.8737 - acc: 0.2356 - val_loss: 1.1698 - val_acc: 0.0755
Epoch 98/100
2/2 [==============================] - 1s 299ms/step - loss: 1.8522 - acc: 0.2296 - val_loss: 1.2441 - val_acc: 0.0650
Epoch 99/100
2/2 [==============================] - 1s 301ms/step - loss: 1.9215 - acc: 0.2190 - val_loss: 1.1695 - val_acc: 0.0770
Epoch 100/100
2/2 [==============================] - 1s 301ms/step - loss: 1.9117 - acc: 0.2024 - val_loss: 1.1811 - val_acc: 0.0650

Input sentence: Şirketlerin EFF'den ggeçer not alması için devletlerin kulanıcılar üzerinde bulunduğu taleplerin açıklanması,
                kullanıcı gizliliği, devltlrn içerik kalrırma taleplerinin paylaşımı, verlrn saklanması konusunda aşık ilkeler
                ve sektörün kabul ettiği en iyi uyglm kurallarını uym katfgorilerimden geçer not alması gerekiiyyor.
Decoded sentence: Şirreeeere n FFFF'de e e     a a   e e e e e ooooooo  oo    e  er  e e  e  e erin  e e e e e e i e e e e e e
                  in  e e erin  e e e e e e in  e e erin  e e e e e e in  e e erin  e e e e e i e e e e e e e e e  e e e e i e e e
                  e e e e e  e e e e i e e e e e e e e  e e e e i e e e e e e e e  e e e e i e e e e e e e e e  e e e e
-
Input sentence: Buna göre Apple başta olmak üzege Adobe, Dropbx, Yahoo ve Wordpress 5 yıldız alırkenn, Google ve Micrsft ise sadece
                üç kategprüde geçer not alabildi.
Decoded sentence: Bunaa örer ppFFF'de e e     a  a   e e e e e ooooooo oo   o   e  er  e e  e  e erin  e e e e e e in  e e erin
                  e e e e e e in  e e erin  e e e e e i e e e e e e e e  e e e e i e e e e e e e  e e e e i e e e e e e e e
                  e e e e i e e e e e e e e  e e e e i e e e e e e e e  e e e e i e e e e e e e e  e e e e i e e e e
```

Karşılaşılan Problemler - Overfitting

```
Train on 2 samples, validate on 2 samples
Epoch 1/400
2/2 [==============================] - 9s 4s/step - loss: 2.7867 - acc: 0.1813 - val_loss: 1.4701 - val_acc: 0.0514
Epoch 2/400
2/2 [==============================] - 2s 971ms/step - loss: 2.7455 - acc: 0.1118 - val_loss: 1.3238 - val_acc: 0.0529
Epoch 3/400
2/2 [==============================] - 2s 1s/step - loss: 2.4300 - acc: 0.0997 - val_loss: 1.3923 - val_acc: 0.0317
Epoch 4/400
2/2 [==============================] - 2s 1s/step - loss: 2.6371 - acc: 0.0770 - val_loss: 1.2453 - val_acc: 0.0287
                                                    .
                                                    .
                                                    .
Epoch 397/400
2/2 [==============================] - 2s 932ms/step - loss: 0.1964 - acc: 0.6858 - val_loss: 1.8495 - val_acc: 0.0816
Epoch 398/400
2/2 [==============================] - 2s 923ms/step - loss: 0.5333 - acc: 0.5650 - val_loss: 1.5137 - val_acc: 0.1042
Epoch 399/400
2/2 [==============================] - 2s 930ms/step - loss: 0.3724 - acc: 0.6254 - val_loss: 1.6294 - val_acc: 0.0831
Epoch 400/400
2/2 [==============================] - 2s 927ms/step - loss: 0.2072 - acc: 0.6903 - val_loss: 1.6051 - val_acc: 0.0876

Input sentence: Şirketlerin EFF'den ggeçer not alması için devletlerin kulanıcılar üzerinde bulunduğu taleplerin açıklanması,
                kullanıcı gizliliği, devltlrn içerik kalrırma taleplerinin paylaşımı, verlrn saklanması konusunda aşık ilkeler
                ve sektörün kabul ettiği en iyi uyglm kurallarını uym katfgorilerimden geçer not alması gerekiiyyor.
Decoded sentence: Şirketlerin EFF'den geçer not alması için devletlerin kullanıcılar üzerinde bulunduğu taleplerin açıklanması,
                kullanıcı gizliliği, devletlerin içerik kaldırma taleplerinin paylaşımı, verilerin saklanması konusunda açık ilkeler
                ve sektörün kabul ettiği en iyi uygulama kurallarını uyma kategorilerinden geçer not alması gerekiyor.

Input sentence: Buna göre Apple başta olmak üzege Adobe, Dropbx, Yahoo ve Wordpress 5 yıldız alırkenn, Google ve Micrsft ise sadece
                üç kategprüde geçer not alabildi.
Decoded sentence: Buna göre Apple başta olmak üzere Adobe, Dropbox, Yahoo ve Wordpress 5 yıldız alırken, Google ve Microsoft ise sadece
                üç kategoride geçer not alabildi.
```

Modelin Özet Bilgileri

```
model.summary()
```

```
_________________________________________________________________________________________________
Layer (type)                 Output Shape              Param #     Connected to
=================================================================================================
input_7 (InputLayer)         (None, None, 45)          0
_________________________________________________________________________________________________
input_8 (InputLayer)         (None, None, 46)          0
_________________________________________________________________________________________________
lstm_6 (LSTM)                [(None, 256), (None,      309248      input_7[0][0]
_________________________________________________________________________________________________
lstm_7 (LSTM)                [(None, None, 256),       310272      input_8[0][0]
                                                                   lstm_6[0][1]
                                                                   lstm_6[0][2]
_________________________________________________________________________________________________
dense_3 (Dense)              (None, None, 46)          11822       lstm_7[0][0]
=================================================================================================
Total params: 631,342
Trainable params: 631,342
Non-trainable params: 0
_________________________________________________________________________________________________
```

2000 Samples – 200 Epoch – 100 Batch Size

Number of Wrong Samples:	2000
Number of Right Samples:	2000
Number of unique input tokens:	111
Number of unique output tokens:	113
Max sequence length for inputs:	852
Max sequence length for outputs:	856

Tabloda belirtilen değerler ile python programlamanın “numpy” kütüpanesinin “zeros” metodu ile veri kümesindeki toplam cümle adeti kadar en uzun cümle ve eşsiz (unique) karakterlerin uzunluğuna bağlı olarak, değerleri sıfır (“0”) olan vektörler oluşturulur.

Bidirectional LSTM İle Eğitme

```
Train on 1000 samples, validate on 1000 samples
Epoch 1/200
1000/1000 [==============================] - 83s 83ms/step - loss: 0.5056 - acc: 0.0942 - val_loss: 0.5468 - val_acc: 0.0152
Epoch 2/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.4486 - acc: 0.0158 - val_loss: 0.5399 - val_acc: 0.0126
Epoch 3/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.4461 - acc: 0.0155 - val_loss: 0.5357 - val_acc: 0.0252
Epoch 4/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.4440 - acc: 0.0171 - val_loss: 0.5344 - val_acc: 0.0200
Epoch 5/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.4411 - acc: 0.0182 - val_loss: 0.5311 - val_acc: 0.0270

.                                .                                .
.                                .                                .
.                                .                                .

poch 195/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0064 - acc: 0.1328 - val_loss: 0.6783 - val_acc: 0.0574
Epoch 196/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0043 - acc: 0.1331 - val_loss: 0.6872 - val_acc: 0.0569
Epoch 197/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0248 - acc: 0.1262 - val_loss: 0.6757 - val_acc: 0.0573
Epoch 198/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0064 - acc: 0.1328 - val_loss: 0.6817 - val_acc: 0.0575
Epoch 199/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0039 - acc: 0.1331 - val_loss: 0.6902 - val_acc: 0.0573
Epoch 200/200
1000/1000 [==============================] - 80s 80ms/step - loss: 0.0239 - acc: 0.1262 - val_loss: 0.6754 - val_acc: 0.0574
```

Bidirectional LSTM ile Modelin Özet Bilgisi

```
_____
Layer (type)                 Output Shape        Param #     Connected to
==================================================================================================
input_1 (InputLayer)         (None, None, 111)   0
_____
bidirectional_1 (Bidirectional) [(None, 512), (None, 753664      input_1[0][0]
_____
input_2 (InputLayer)         (None, None, 113)   0
_____
concatenate_1 (Concatenate)  (None, 512)         0           bidirectional_1[0][1]     bidirectional_1[0][3]
_____
concatenate_2 (Concatenate)  (None, 512)         0           bidirectional_1[0][2]     bidirectional_1[0][4]
_____
lstm_2 (LSTM)                [(None, None, 512),  1282048     input_2[0][0]
                                                              concatenate_1[0][0]
                                                              concatenate_2[0][0]
_____
dense_1 (Dense)              (None, None, 113)   57969       lstm_2[0][0]
==================================================================================================
Total params: 2,093,681
Trainable params: 2,093,681
Non-trainable params: 0
_____
```

Eğitilmiş Verilerin Encode ve Decode İle Tahmini

-

Input sentence: Şirketlerin EFF'den ggeçer not alması için devletlerin kulanıcılar üzerinde bulunduğu taleplerin açıklanması, kullanıcı gizliliği, devltlrn içerik kalrırma taleplerinin paylaşımı, verlrn saklanması konusunda aşık ilkeler ve sektörün kabul ettiği en iyi uyglm kurallarını uym katfgorilerimden geçer not alması gerekiiyyor.

Decoded sentence: Şirketlerin EFF'den geler nok alması için devartelerine konulacaklara için tanatıla gerekildi.

-

Input sentence: Buna göre Apple başta olmak üzege Adobe, Dropbx, Yahoo ve Wordpress 5 yıldız alırkenn, Google ve Micrsft ise sadece üç kategprüde geçer not alabildi.

Decoded sentence: Buna göre Appler aht olası kamesi yaban etkiye bir ünümet olar kabul değildi.

Eğitilmiş Verilerin Encode ve Decode İle Tahmini

-

Input sentence: Googlle devletlerin içerik kaödırma ggibi taeplerini payylaşma konusunea pek de cömrt davrnmzkn, diğdr yandan da kullanıcıların ferlerini nasıl kendnd saakladdığını da tam olarak belirrtmmiyor.

Decoded sentence: Google devletlerin içerik kaldırma gibi taleplerini paylaşma konusunda pek de cömert davranmazken, diğer yandan da kullanıcıların verlerini nasıl kendinde sakladığını da tam olarak belirtmiyor.

-

Input sentence: Çidrosoft'ta da durum çok farklı değil.

Decoded sentence: Microsoft'ta da durum çok farklı değil.

-

Input sentence: Ancak raporda en kötü not aln şirkt Dhatsapp olluyor.

Decoded sentence: Ancak raporda en kötü notu alan şirket Whatsapp oluyor.

Kaynaklar

1. https://www.makaleler.com/ (**Dataset oluşturmak için makaleler**)
2. https://archive.ics.uci.edu/ml/datasets/TTC-3600%3A+Benchmark+dataset+for+Turkish+text+categorization (**TTC-3600: Benchmark dataset for Turkish text categorization Data Set**)
3. https://keras.io/ (**Keras Documentation: The Python Deep Learning library**)
4. https://medium.com/turkce/keras-ile-derin-%C3%B6%C4%9Frenmeye-giri%C5%9F-40e13c249ea8 **(Keras ile Derin Öğrenmeye Giriş)**
5. Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, Yoshua Bengio, 7 Oct 2014, "**On the Properties of Neural Machine Translation: Encoder-Decoder Approaches"**
6. Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, 19 May 2016, **"Neural Machine Translation by Jointly Learning to Align and Translate"**
7. Yann LeCun, Yoshua Bengio & Geoffrey Hinton, 28 May 2015, **"Deep learning"**
8. Tom Young, Devamanyu Hazarika, Soujanya Poria, Erik Cambria, Aug. 2018, **"Recent Trends in Deep Learning Based Natural Language Processing"**

TEŞEKKÜRLER