

**GTU Department of Computer Engineering  
CSE 484/654 Natural Language Processing  
Fall 2021 - Homework 4 Report**

**Akif KARTAL  
171044098**

## 1) Problem Definition

The problem is to correct Turkish sentences written using English letters only by using a language model created with LSTM networks in Keras on Google Colabs.

## 2) Solution

The homework was finished as expected in homework pdf file. I didn't use n-grams I only use word tokenizer and LSTM. As a corpus I used very small portion of Turkish Wikipedia dump

(<https://www.kaggle.com/mustfkeskin/turkishwikipedia-dump>)

As you can see **chosen best sentences is not good**. Because my **dataset is very small** and I didn't use syllables.

### 2.1) Test Results with Sentences

Test Sentence: bilgisayar isletim sistemlerinin

Number of Possible Sentences: 4096

Best Sentence: bilgisayar işletim sistemlerinin

-----  
Test Sentence: ticari linux dagitimlarinin

Number of Possible Sentences: 1024

Best Sentence: ticari linux dağıtımlarının

-----  
Test Sentence: orta duzeyde zengin bir ailede

Number of Possible Sentences: 64

Best Sentence: orta düzeyde zengin bir ailede

-----  
Test Sentence: yeni adıyla turkiye

Number of Possible Sentences: 16

Best Sentence: yeni adıyla turkiye

-----  
Test Sentence: ilde bircok etkinlik

Number of Possible Sentences: 64

Best Sentence: ilde birçok etkinlik

-----

Test Sentence: çok sayıda koyun yetistirilir

Number of Possible Sentences: 2048

Best Sentence: çok sayıda köyün yetiştirilir

-----

Test Sentence: beklenen olumlu tepkiyi alamadı

Number of Possible Sentences: 64

Best Sentence: beklenen olumlu tepkiyi alamadı

-----

Test Sentence: güzel vakit geçiyor öneririm

Number of Possible Sentences: 1024

Best Sentence: güzel vakit geçiyor öneririm

-----

Test Sentence: konforlu ve temiz bir yer

Number of Possible Sentences: 32

Best Sentence: konforlu ve temiz bir yer

-----

Test Sentence: sıcak ve eğlenceli bir ortam

Number of Possible Sentences: 256

Best Sentence: sıcak ve eğlenceli bir ortam

-----

Test Sentence: bu yurt okula çok uzak

Number of Possible Sentences: 128

Best Sentence: bu yurt okula çok uzak

-----

Test Sentence: idare eder tavsiye ederim

Number of Possible Sentences: 16

Best Sentence: idare eder tavsiye ederim

-----

Test Sentence: fiyatlar çok pahali

Number of Possible Sentences: 16

Best Sentence: fiyatlar çok pahalı

-----

Test Sentence: oteller burdan ucuzdur

Number of Possible Sentences: 64

Best Sentence: oteller burdan ucuzdur

-----

Test Sentence: yurtda mutfak var is gorur

Number of Possible Sentences: 128

Best Sentence: yürtda mutfak var ısı görür

-----

Test Sentence: koskoca yurtda kutuphane yok

Number of Possible Sentences: 256

Best Sentence: koskoca yurtda kutuphane yok

-----

Test Sentence: nice dostluklarin kuruldugu yer

Number of Possible Sentences: 2048

Best Sentence: nice dostluklarin kurulduğu yer

-----

Test Sentence: cadirda kalirim daha iyi

Number of Possible Sentences: 64

Best Sentence: çadırdaki kalırım daha iyi

-----

Test Sentence: resimlere inanmayin bes para etmez

Number of Possible Sentences: 32

Best Sentence: resimlere inanmayın beş para etmez

-----

Test Sentence: çok bisey beklemeyin

Number of Possible Sentences: 32

Best Sentence: çok bısey beklemeyin

-----

Test Sentence: bu zamanda zor bulunacak bir yurt

Number of Possible Sentences: 128

Best Sentence: bü zamanda zör bulünaçak bır yürt

-----

Test Sentence: ogrenci dostu bir yurt

Number of Possible Sentences: 512

Best Sentence: öğrençi doştü bir yürt

-----

Test Sentence: calisanlarin davranislari cok iyi

Number of Possible Sentences: 2048

Best Sentence: çalışanların davranışları çok iyi

-----

Test Sentence: mac yayini yok bilginiz olsun

Number of Possible Sentences: 2048

Best Sentence: maç yayini yök bılgınız olşün

-----

## How to run

First add corpus file and test sentences file to your google drive. Then run the google colab file.