

GEBZE TECHNICAL UNIVERSITY

CSE 484 HOMEWORK 3 REPORT

Yağız Hakkı AYDIN
1901042612

The Dataset

The dataset is a version of Turkish Wikipedia dump. It contains Turkish sentences for each line. It has 170458 sentences. Dataset is taken from [“https://www.kaggle.com/datasets/mahdinamidamirchi/turkish-sentences-dataset”](https://www.kaggle.com/datasets/mahdinamidamirchi/turkish-sentences-dataset).

As example, first 5 sentences from dataset is in following form:

Cengiz Han Çin'i kendi adamları ile idare etme yolunu ömrünün sonuna kadar sürdürdü.
Onlardan destek alarak muazzam bir imparatorluk inşa etti.
Yaptığı plana göre orduyu üçe ayırdı.
Ayrıca Android işletim sistemine ait uygulama mağazasında da geniş bir oyun seçeneği bulunmaktadır.
Bu konudaki ayrıntılı bilgilere Linux dağıtımları maddesinden ulaşılabilir.

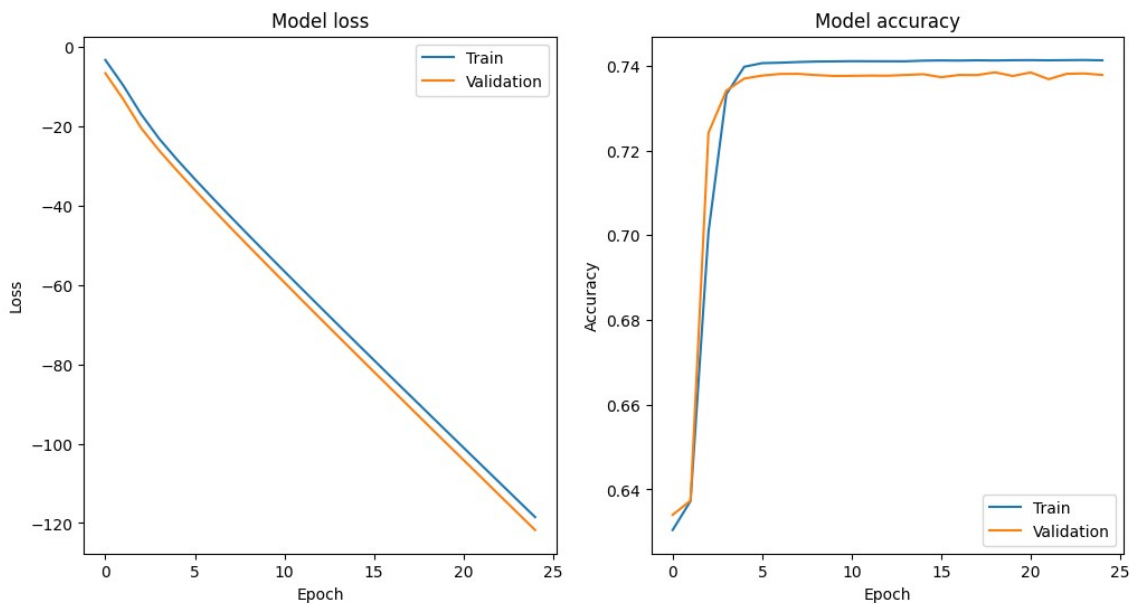
Sentences are written and stored in a list.

Then labeling applied on the sentences to check if sentence has separated suffix, has not separated suffix and has any suffix.

The Model

The model is an lstm structure with 128 units. With labeled data, model is trained in 25 epochs.

To measure model success, validation loss and validation accuracy metrics are used. Graphs for these metrics following.



The model is tested with 20 sentences and output of the script is in following:

1/1 [=====] - 1s 824ms/step

Test Sentence: Öğrenciler de geldi.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Öğrencilerde gelişme var.

Result: Suffixes are correctly written

1/1 [=====] - 0s 25ms/step

Test Sentence: Evdeyim.

Result: Suffixes are correctly written

1/1 [=====] - 0s 24ms/step

Test Sentence: Ev de oturuyoruz.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Arabada oturuyorum.

Result: Suffixes are correctly written

1/1 [=====] - 0s 25ms/step

Test Sentence: Araba da tamir ediliyor.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 24ms/step

Test Sentence: Kedi de uyuyor.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 25ms/step

Test Sentence: Kitap okuyorum.

Result: Suffixes are correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Kitapta bilgiler var.

Result: Suffixes are correctly written

1/1 [=====] - 0s 24ms/step

Test Sentence: Yarın geleceğim.

Result: Suffixes are correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Yarınki planımız nedir?

Result: Suffixes are correctly written

1/1 [=====] - 0s 24ms/step

Test Sentence: Ben de evdeyim.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 24ms/step

Test Sentence: Benim kitabım var.

Result: Suffixes are correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Kalem de düştü.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: O da katılıyor.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Bende bir şey yok.

Result: Suffixes are correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Bilgisayar da çalışıyor.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 22ms/step

Test Sentence: Kitap da okunabilir.

Result: Suffixes are not correctly written

1/1 [=====] - 0s 23ms/step

Test Sentence: Sınavda başarılı olacağım.

Result: Suffixes are correctly written

1/1 [=====] - 0s 26ms/step

Test Sentence: Sınav da zor olacak.

Result: Suffixes are not correctly written