**Міністерство освіти і науки**

**України Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»**

**Факультет інформатики та обчислювальної техніки Кафедра інформатики та програмної інженерії**

**ЗВІТ**

лабораторної роботи №8

з курсу «Програмні засоби проєктування і реалізації неромережевих систем»

Тема: «Рекурентні нейронні мережі LSTM»

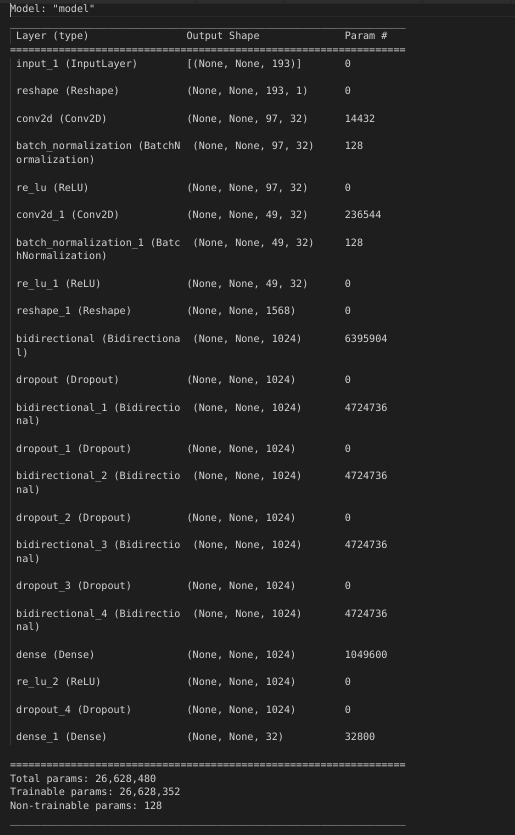
| Перевірив:  Шимкович В. М. | Виконав:  Студент Гр. ІП-01 Шпилька В.С. |
| --- | --- |

Київ 2023

**Завдання:** Написати програму, що реалізує нейронну мережу типу СNN-bi-LSTM для розпізнавання мови в текст.

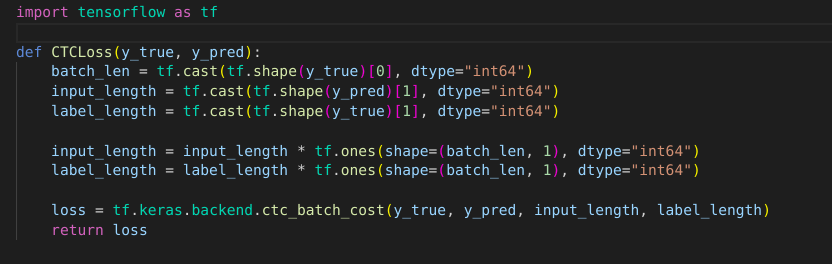
1. Реалізація нейронної мережі:





Нейрона мережа складається з 2 згорткових шарів та 5 рекурентних. Це дозволяє на першому етапі виділити корисні риси зі спектрограми, а потім за допомогою рекурентних шарів розпізнати звук на конкретному кроці.

Як функція втрат було реалізовано CTCLoss:



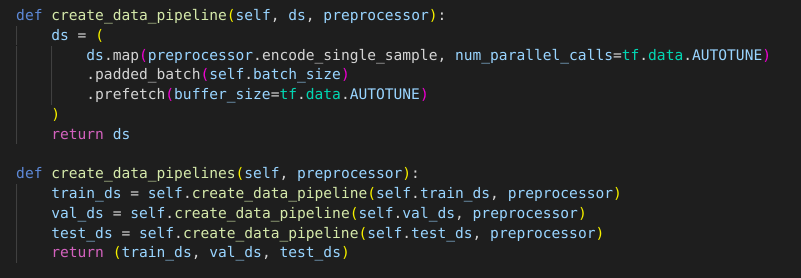
1. Створення пайплану для завантаження даних.

Для тренування було використано LJSpeech dataset. Для цього було створено tf dataset з шляхів до файлів та їх транскрипцій.

Далі використовується наступна функція препроцесінгу:

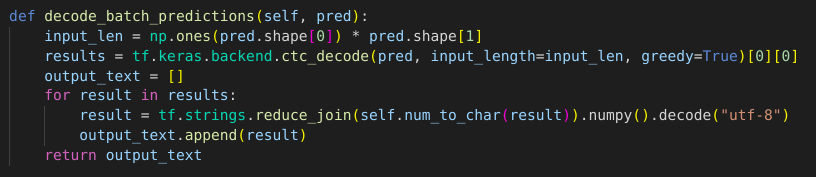
а.) Читається файл  
б.) Трансформація до потрібного формату  
в.) Створення спектрограми  
г.) Кодування символів в числа





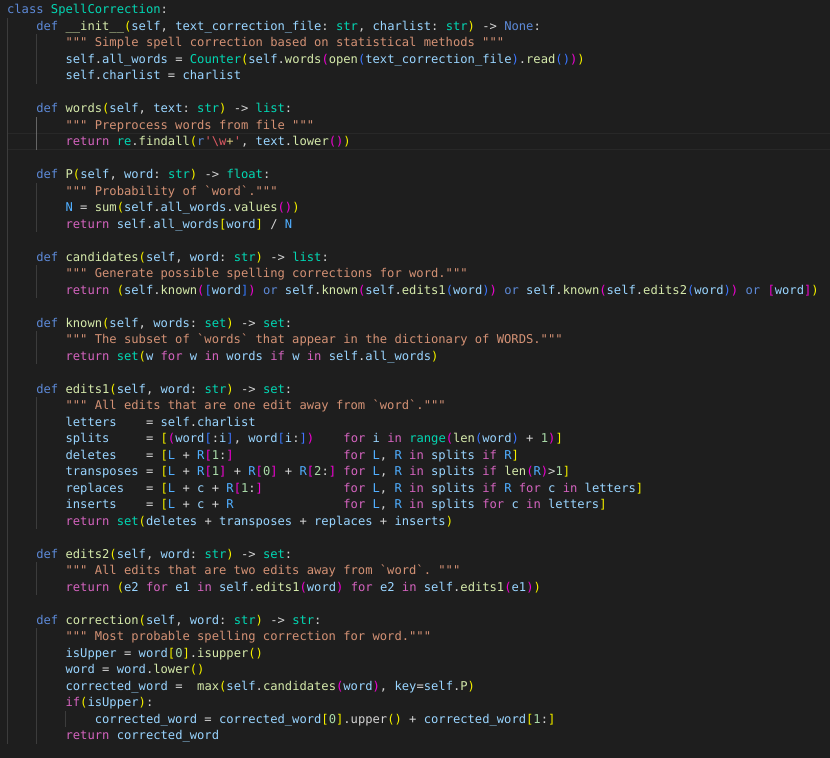
1. Постпроцесінг:

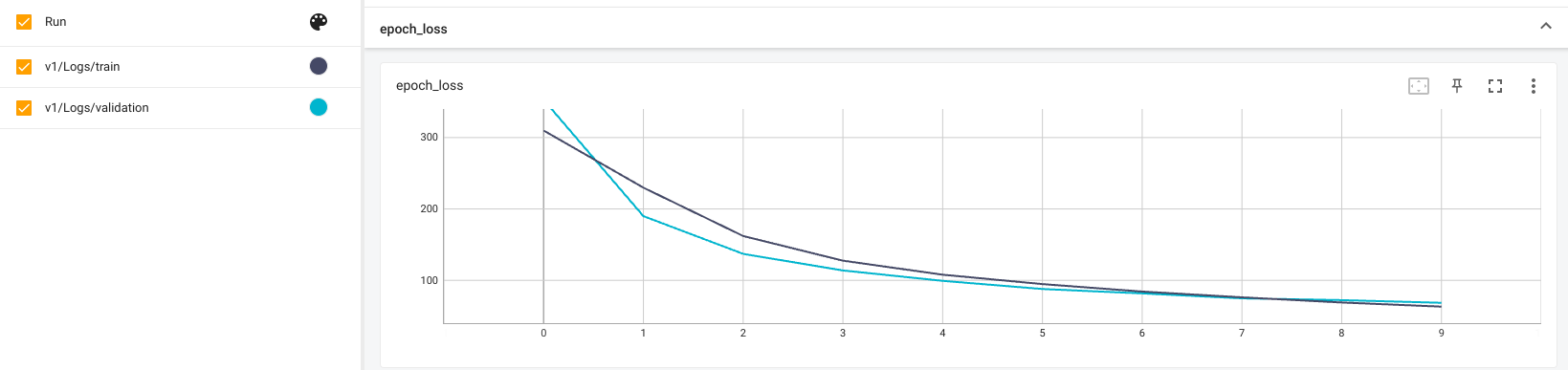
Оскільки в результаті виконання ми отримаємо массив ймовірностей послідовності, його треба перетворити назад до тексту. Для цього використовується ctc decoding і його варіант – жадібний.



Також було реалізовано простий алгоритм виправлення слів. Для цього парситься деякий великий файл у окремі слова і підраховується кількість кожного слова.

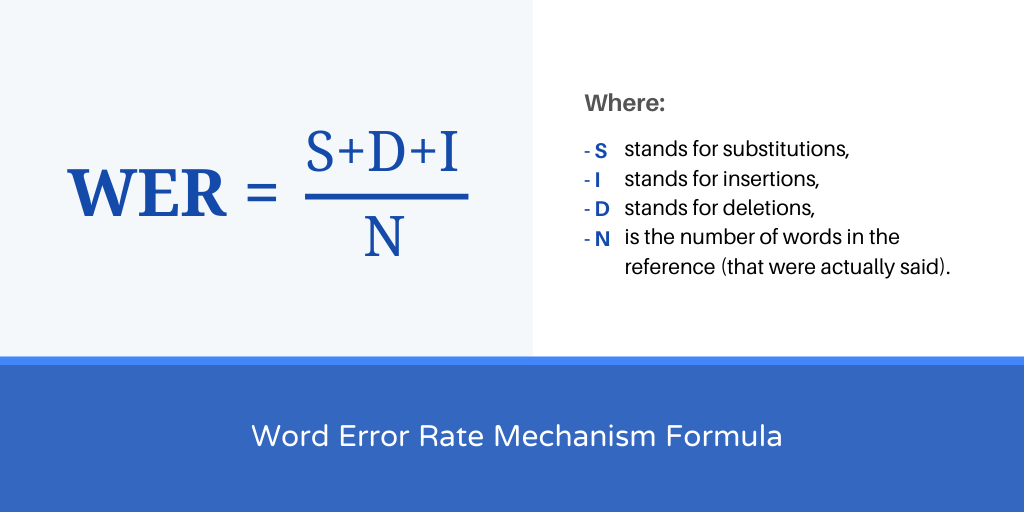
Далі функція приймає на вхід слово, якщо його немає в словнику, то вона генерує всі можливо варіанти даного слова з 2 змінами і шукає в словнику і повертає найімовірніше слово.



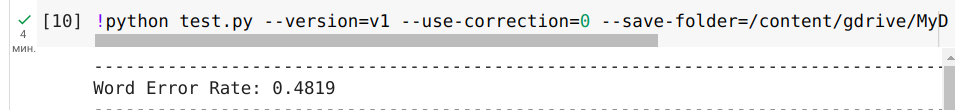
1. Тренування мережі:

Тренування проводилось на 10 епохах.

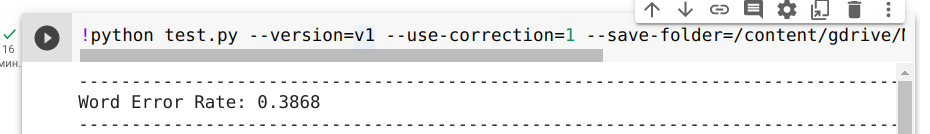
1. Тестування:

Для тестування було використано метрику wer. 

Було проведено тестування без використання виправлення слів:



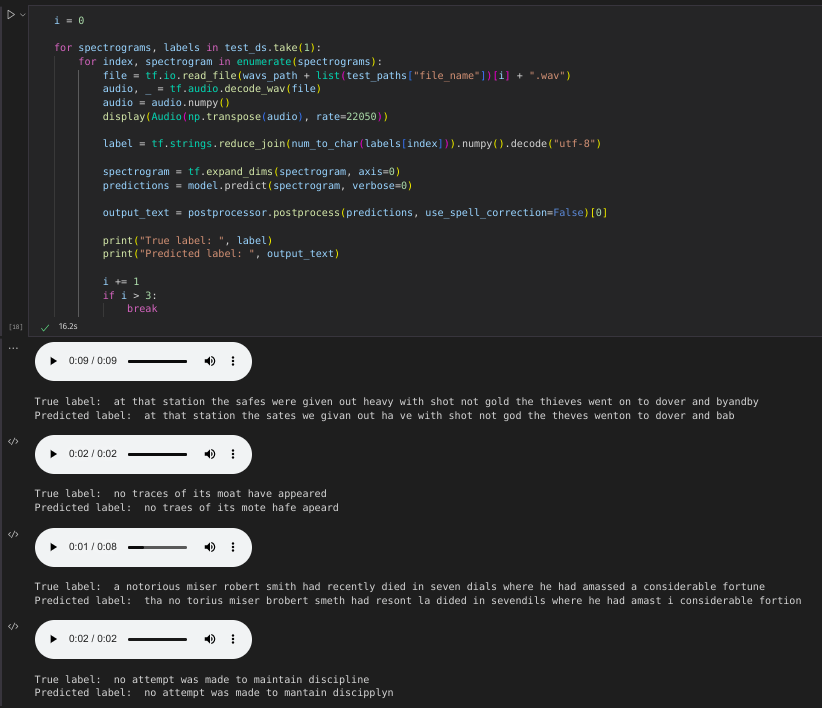
та за допомогою нього:



в результаті отримали покращення в 10 відсотків і фінальну помилку в 0.38. Тобто даний результат можна інтерпретувати як точність в 62 відсотки.

Варто тільки зауважити, що тепер працює тестування на 12 хвилин довше, але дану проблему можна покращити, якщо застосувати паралельний постпроцесінг для результату

Приклад використання:



**Висновок:** В результаті виконання лабораторної роботи було побудовано нейрону мережу СNN-bi-LSTM для розпізнавання мови в текст. Всього нейрона мережа має 26млн параметрів. Для LJSpeech датасету дана нейрона мережа показала гарні результати, а саме точність в 62 відсотки.