Руководство по обучению модели нейросети

Обработка файлов(формируем выборку из html страниц):

1)формирование датафрейма(таблица,у которой индексированы столбики и строки)

2)Расстояние Левенштейна(метрика сходства между двумя последовательностями)

Токенизация(преобразование данных):

1)инициализация токенайзера для обработки текста

2)подготовка данных машинного обучения

3)Определяем макс. длину

4)Создаём модель RNN и компилируем её

5)обучаем модель на исходных и целевых последовательностях

6)преобразуем модель в последовательность чисел

7)получаем предсказание модели

Руководство по использованию модели: