

**Assignment Number 2**

**CS-4063 Natural Language Processing**

**Fall 2024**

Name : Waqas Naveed

RollNo#:21F-9111

Name :Sohaib Shoukat

RollNo#:21F-9181

**National University of Computer and Emerging Sciences**

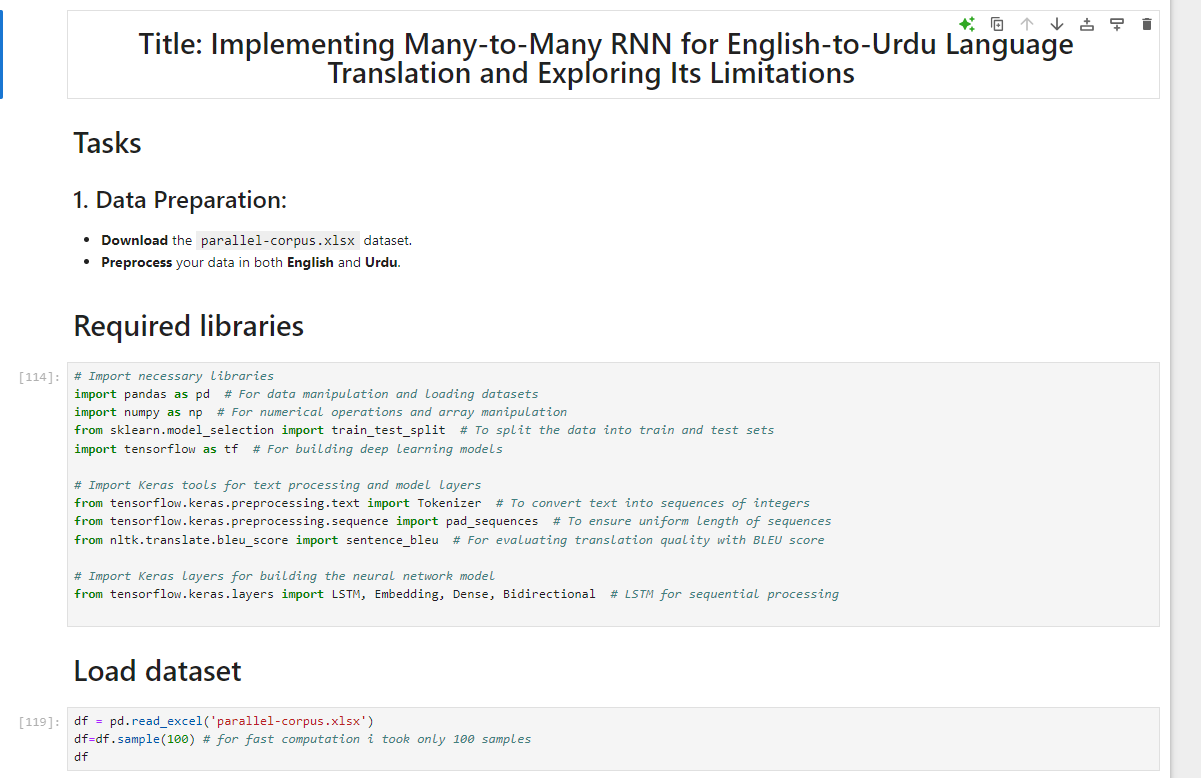
**Department of Computer Science**

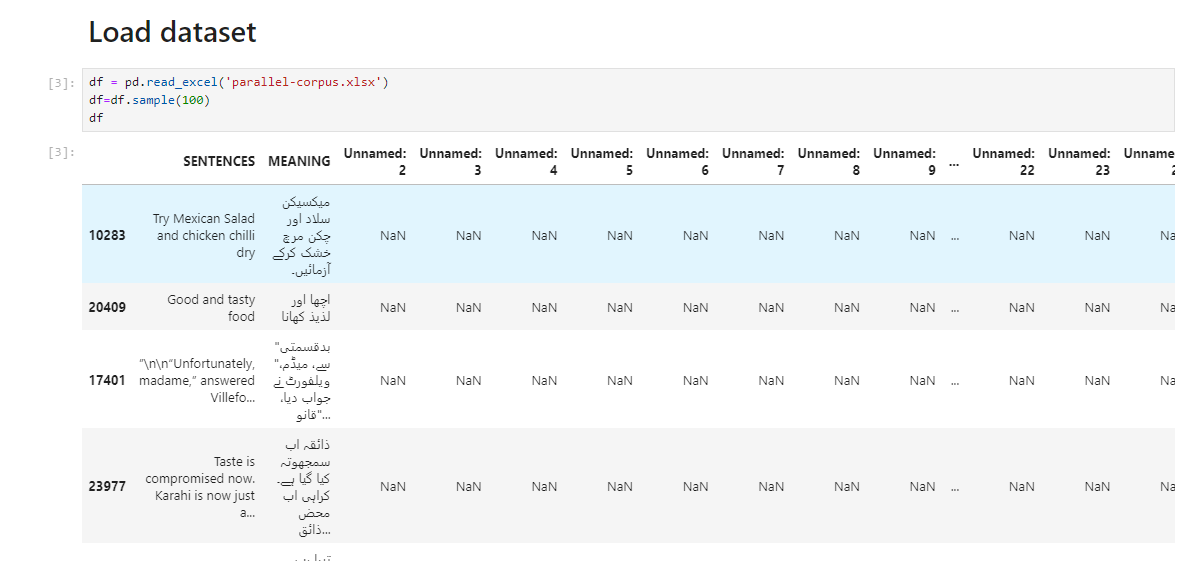
**Faisalabad, Pakistan**

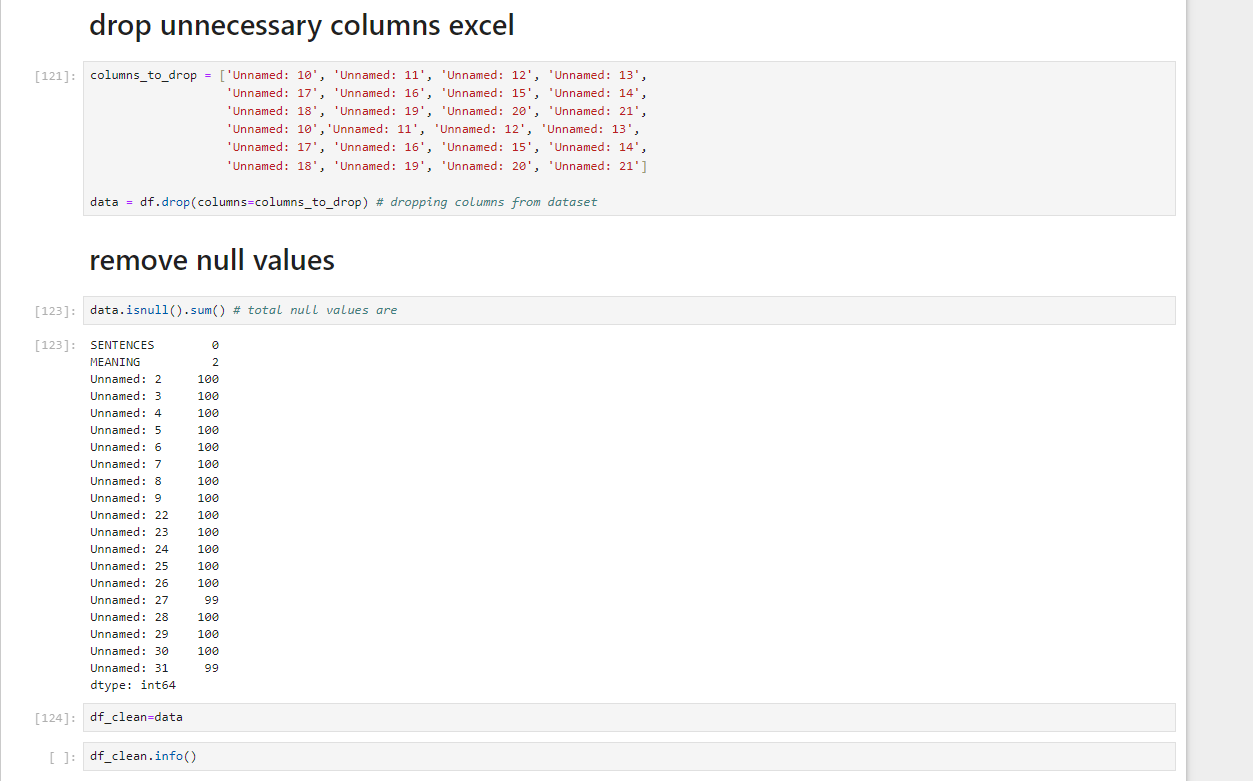
**Title: Implementing Many-to-Many RNN for English-to-Urdu Language Translation and Exploring Its Limitations**

Part 1: Many-to-Many Recurrent Neural Network (RNN) Implementation

**Tasks:**

****

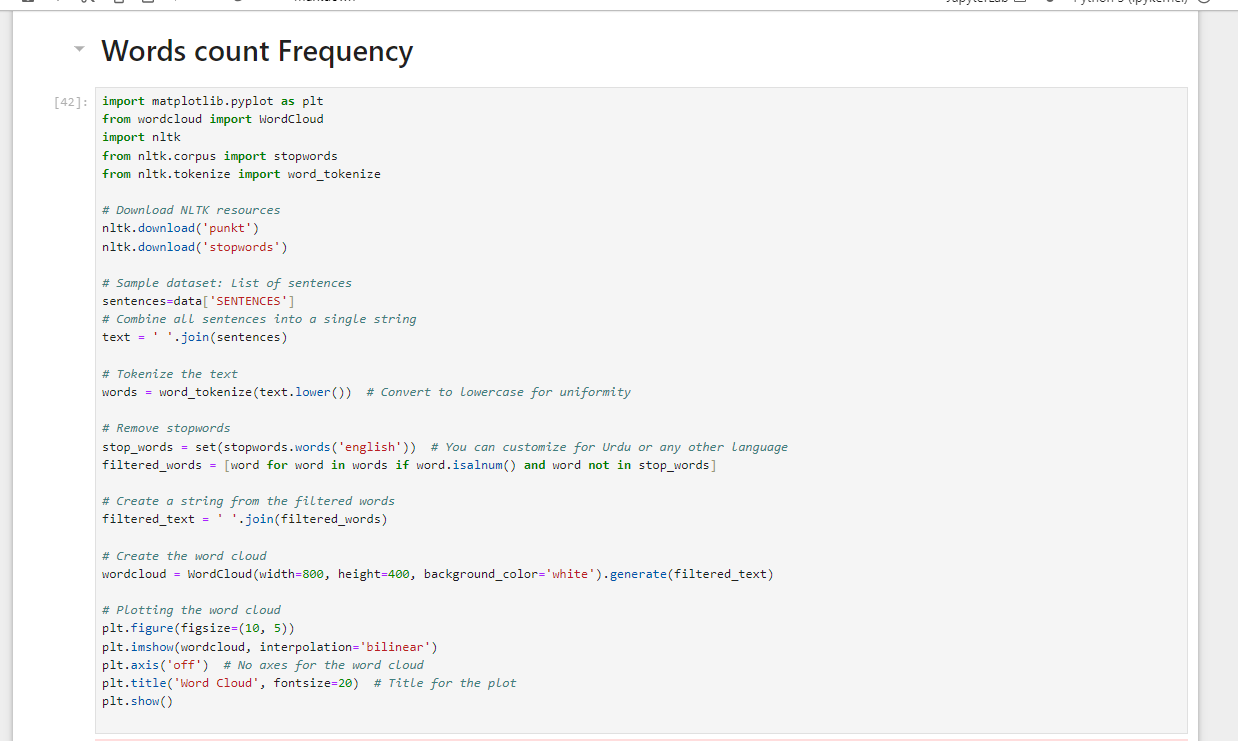
****

****

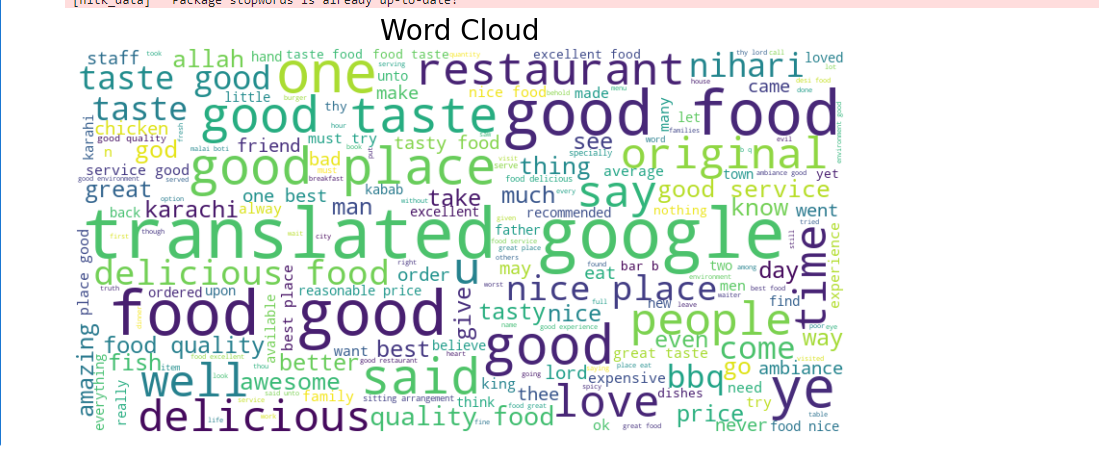
****

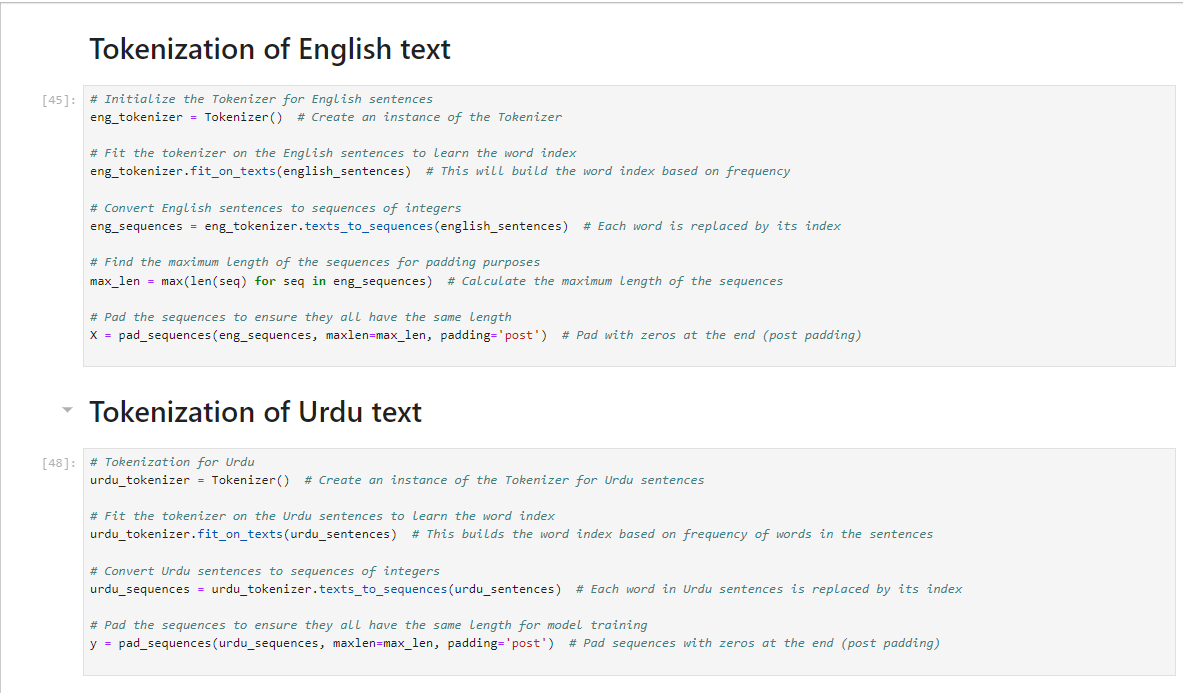
****

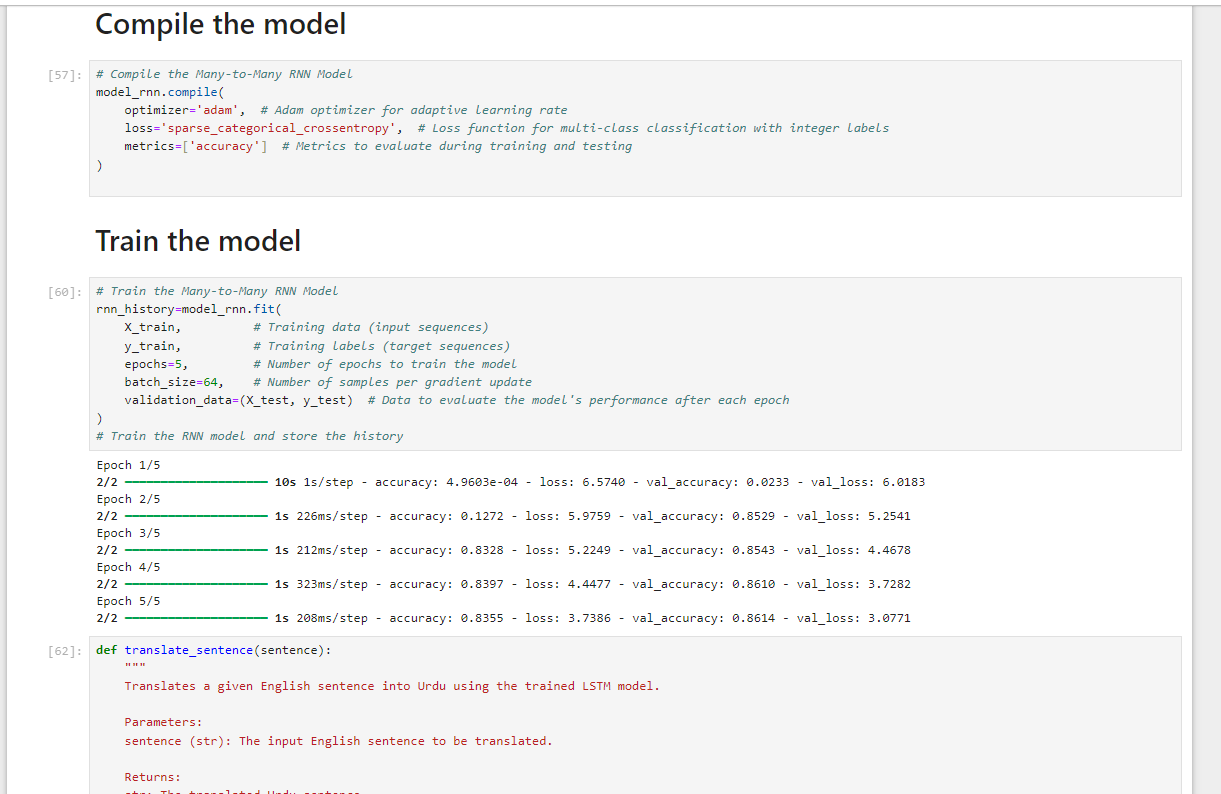
**And removed unnecassry words**

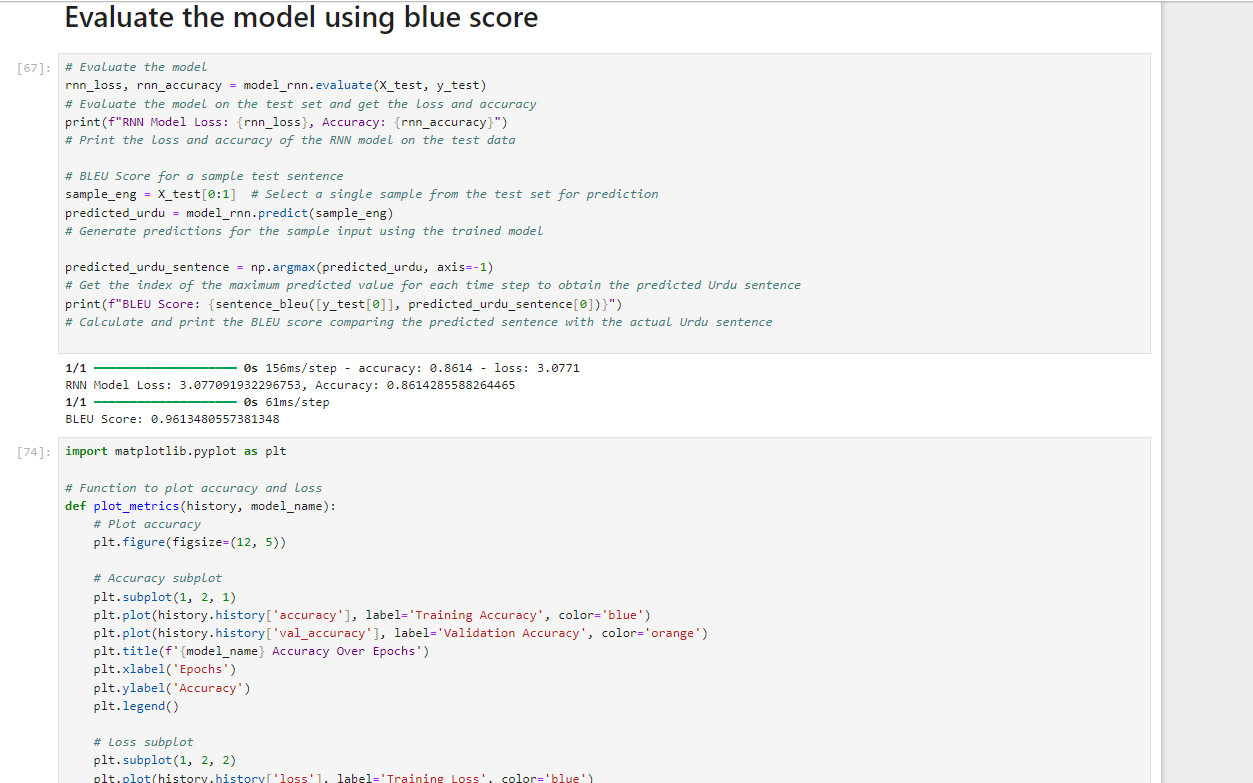
****

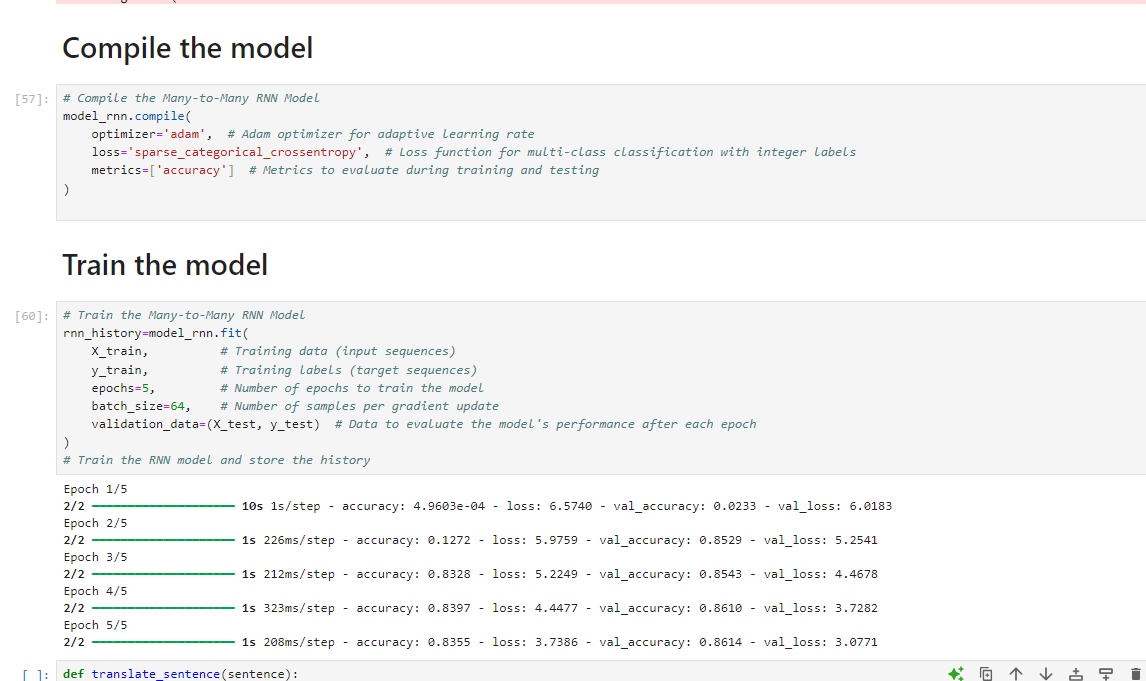
**Words frequency :**

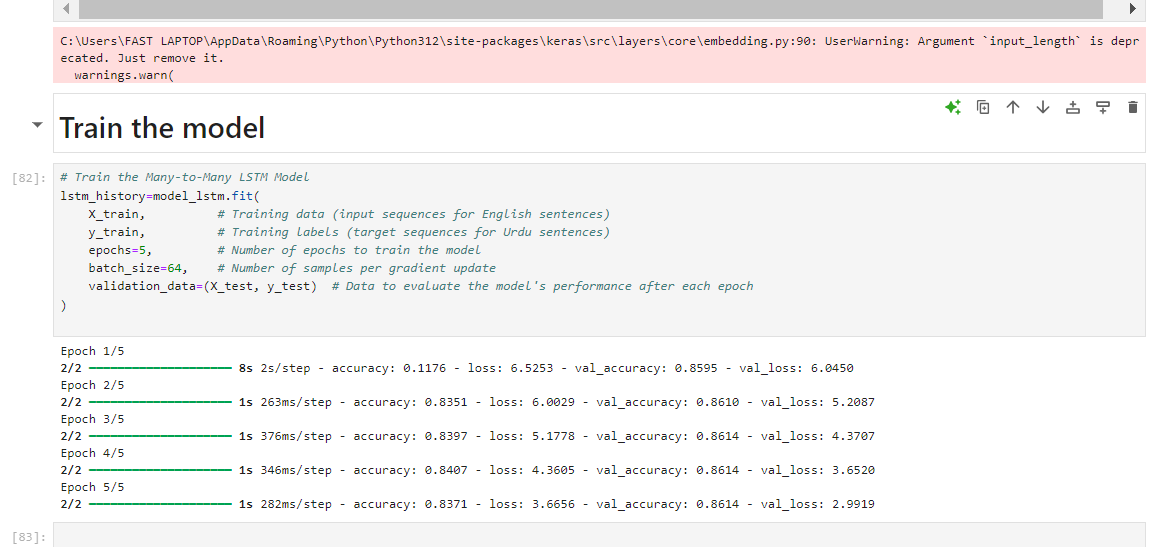
****

****

****

****

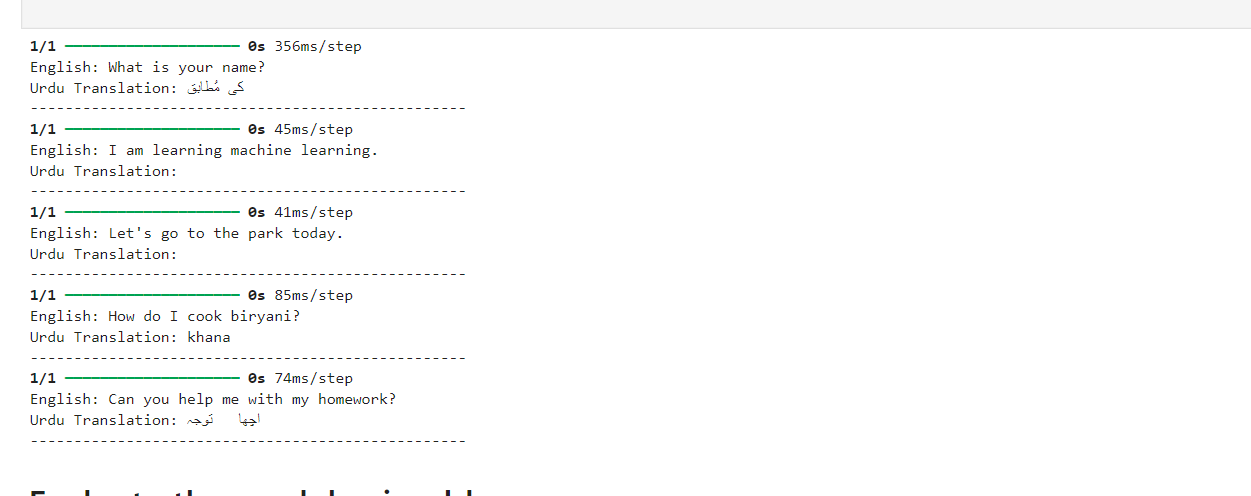
****

****

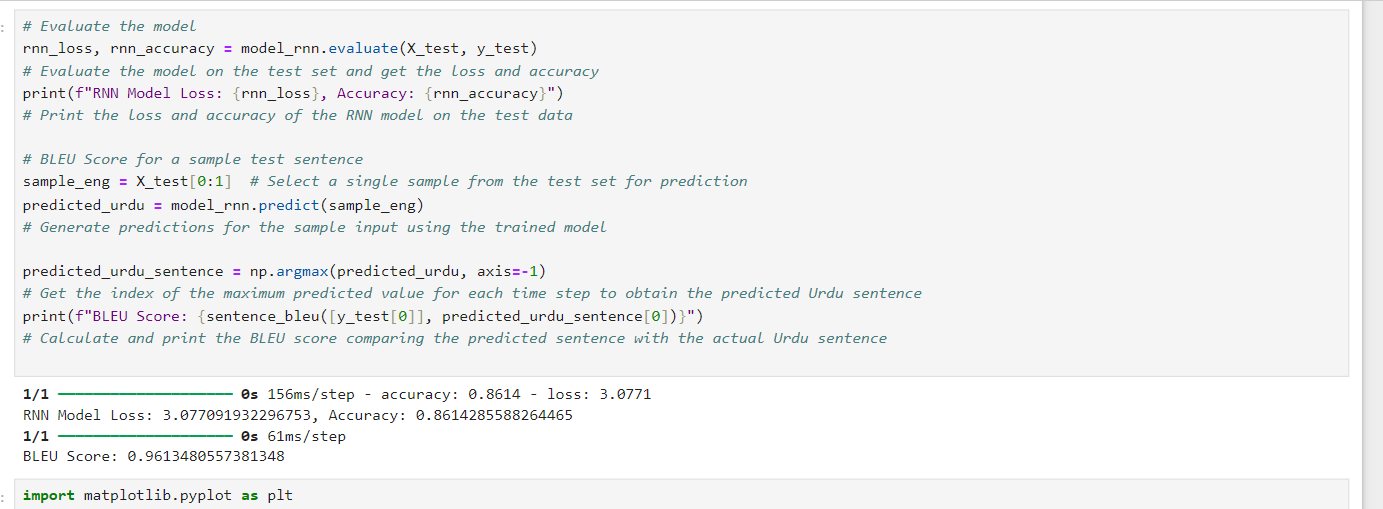
**I just took 100 data point**

**Because model took long time for training that’s why its accuracy is not well**

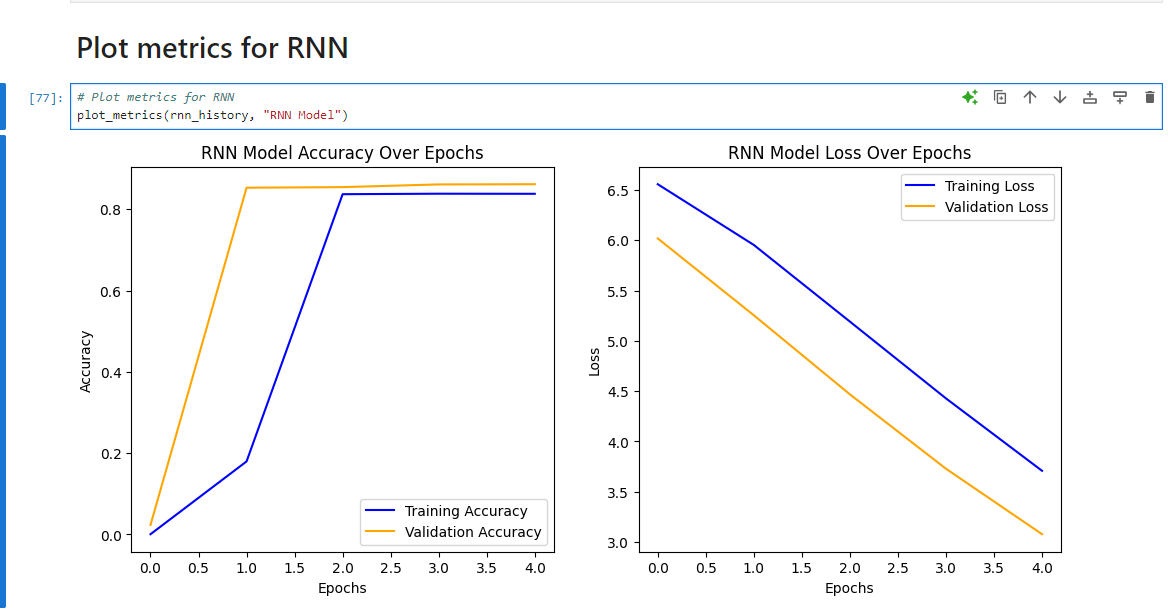
**It perform so well when I train on whole dataset**

****

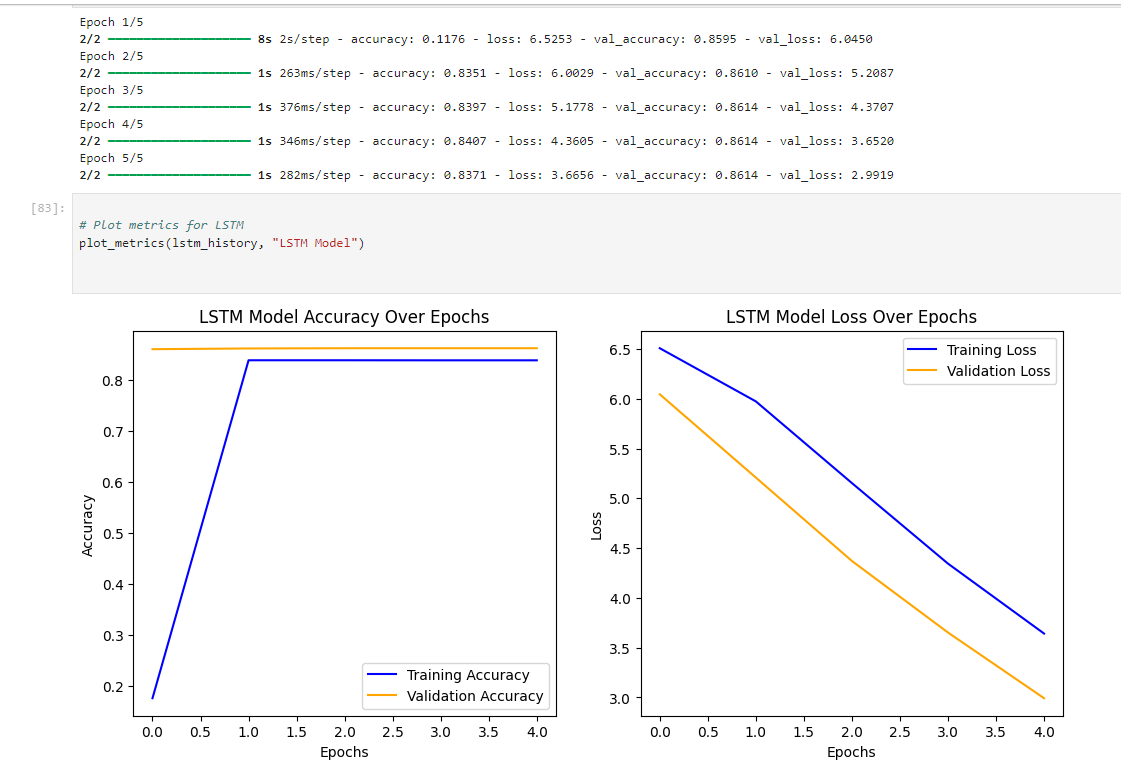
**Blue Score & Accuracy**

****

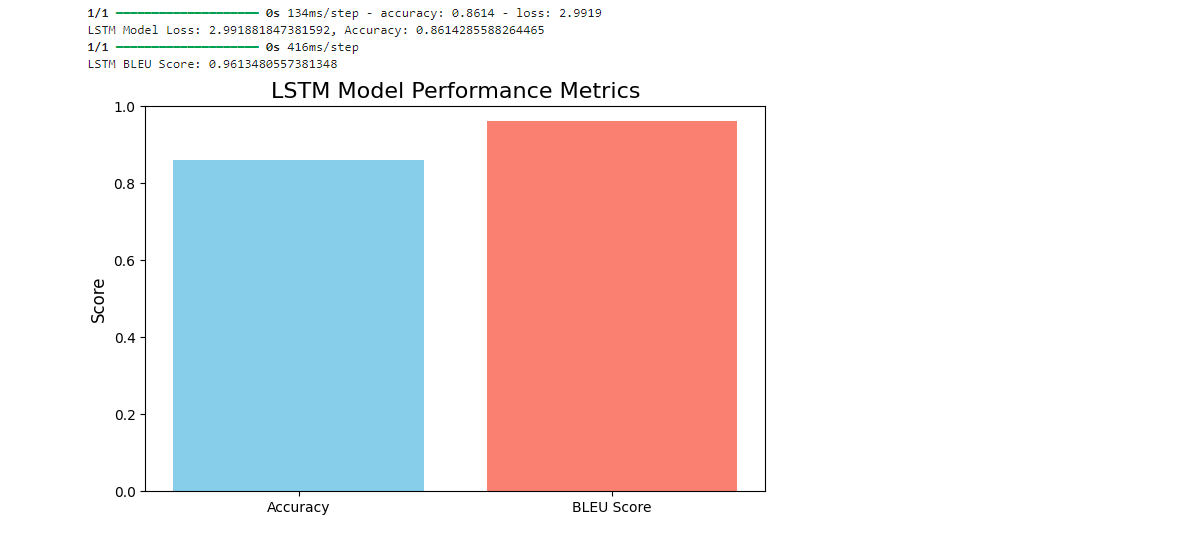
* **RNN Model Accuracy Over Epochs**

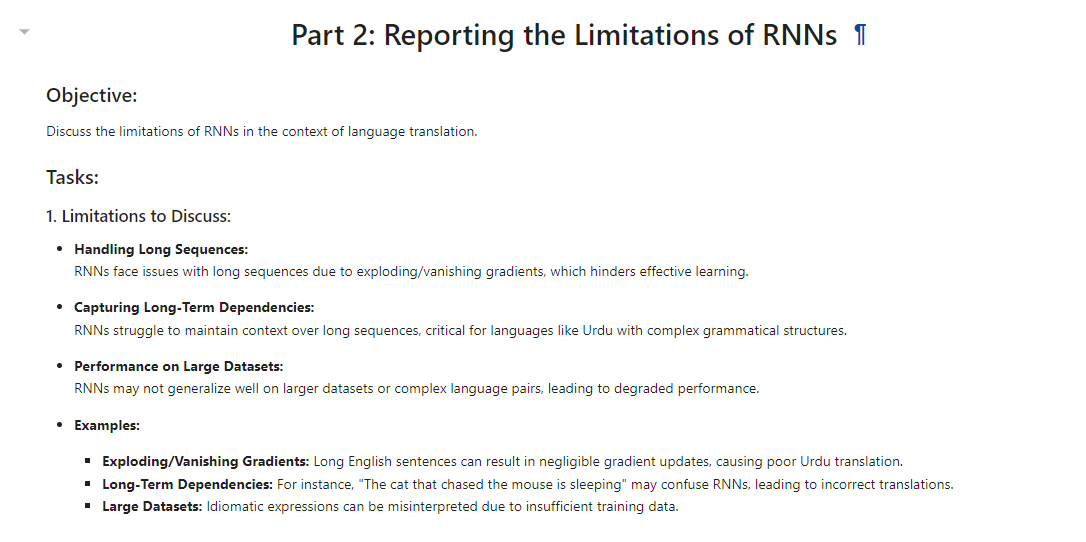
****

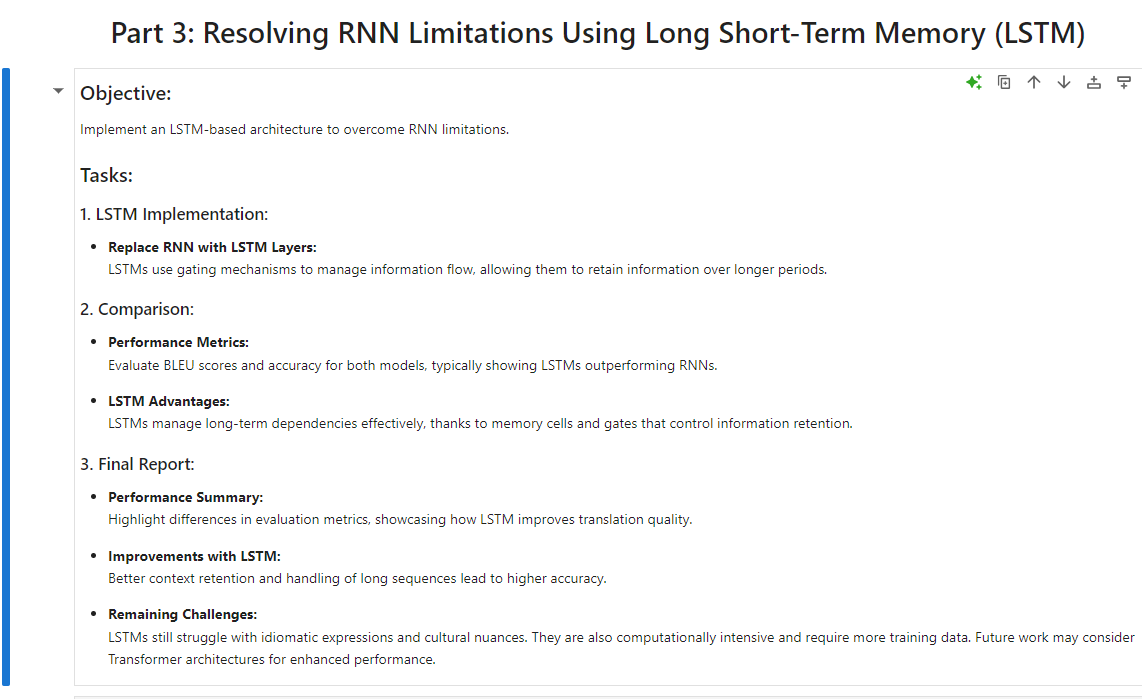
**LSTM Model Accuracy Over Epochs**

****

**Blue score**

****

****

****

**In conclusion:**

this assignment provided a comprehensive understanding of implementing and evaluating RNN and LSTM models for English-to-Urdu language translation. It highlighted the challenges RNNs face, such as handling long sequences and capturing dependencies, and demonstrated how LSTM can address these limitations. Through hands-on experience with data preprocessing, model training, and performance evaluation, we gained insights into the importance of sufficient data for effective learning. Although the limited dataset caused underfitting, the project showcased the potential of deep learning architectures for translation tasks and emphasized the need for further optimizations and data augmentation for real-world applications.