The Superior University, Lahore

Assignment-I (Fall 2023)

Course Title:	Programming for AI			Course Code:	CAI601410	Credit Hours:	4
Instructor:	Prof. Rasikh Ali			Programme Name:	BSDS		
Semester:	4th	Batch: F23	Section:	BSDSM-4A	Date:	1st February, 2025	
Time Allowed:				Maximum Marks:			
Student's Name:	Asaad Iqbal			Reg. No.	SU92-BSDSM-F23-020		

Lab-Task 9
   1:   Sentiment Analysis using RNN

Task 1

Sentiment Analysis using RNN

1. Introduction

      This project focuses on analyzing customer reviews from Swiggy to determine the sentiment behind each review—whether it is positive or negative. Sentiment analysis is a key Natural Language Processing (NLP) task that helps businesses understand customer opinions and improve service delivery. In this case, an RNN-based model is trained on review data to predict sentiment effectively.

2. Dataset Overview

    The dataset used in this project is a CSV file named swiggy.csv, containing customer reviews and their associated average ratings. The relevant columns include:

- **Review**: Text review left by the customer.

- **Avg Rating**: Numerical rating given to the service/food.

3. Data Preprocessing

    To prepare the data for training:

- The reviews were converted to lowercase.

- Special characters and punctuation were removed using regular expressions.

- Sentiment labels were created from the Avg Rating column. Ratings above 3.5 were labeled as **positive (1)**, and others as **negative (0)**.

- Missing data entries were dropped to ensure clean inputs.

4. Tokenization and Padding

- A tokenizer was created to process the review texts, using a vocabulary limit of **5,000** most frequent words.

- The textual data was then converted into sequences of integers.

- All sequences were padded to a maximum length of **200 tokens** to maintain consistency for RNN input.

5. Data Splitting

The data was split into three sets:

- **Training Set (72%)**

- **Validation Set (8%)**

- **Testing Set (20%)**

Stratification was applied during splitting to maintain a balanced class distribution across sets.

6. Model Architecture

A Sequential model was developed with the following layers:

- **Embedding Layer**: Transforms each word index into a 64-dimensional vector.

- **SimpleRNN Layer**: Captures sequential dependencies with 64 units.

- **Dropout Layer**: Applies a dropout of 0.5 to reduce overfitting.

- **Dense Layer**: A single neuron with sigmoid activation to classify the sentiment.

7. Model Compilation and Training

- The model was compiled using **binary cross-entropy** as the loss function and **Adam optimizer**.

- Early stopping was applied to monitor validation loss, with patience set to 2 epochs.

- The model was trained for up to 10 epochs, using a batch size of 32.

8. Model Evaluation

After training, the model was evaluated on the test dataset to determine its accuracy. The trained model successfully generalized on unseen data, indicating good performance.

9. Sentiment Prediction

A prediction function was implemented to analyze new review texts and return the predicted sentiment along with its probability.

10. Sample Output

Input Review:
"The food was great."

Predicted Sentiment:

```
...    Review: The food was great.
       1/1 ──────────────── 1s 607ms/step
       Sentiment: Positive (Probability: 0.66)
```