Ayesha Fathima

Afathima20760@ucumberlands.edu

**LSTM USING EPICURIOUS RECIPES DATASET**

Professor. David Ostrowski

**Assignment: Text Generation using LSTM Network with Epicurious Recipes Dataset**

**Introduction**

This project involves using the Epicurious Recipes dataset to build and train a Long Short-Term Memory (LSTM) network that generates recipe text. The network is trained on a filtered text corpus containing recipe titles and instructions. Once trained, the model is used to generate new recipe text by predicting the next word in a sequence based on a given prompt, with output generated at two different temperature values.

**Dataset and Preprocessing**

The Epicurious Recipes dataset was obtained from Kaggle, containing columns such as title, directions, ingredients, and various nutritional information. The dataset was cleaned to form a text corpus, combining recipe titles and directions. The text was tokenized, and sequences of words were created.

The following code snippet was used to filter and tokenize the data:

A screen shot of a computer program

Description automatically generated

**Building the LSTM Model**

An LSTM model was built with an embedding layer, two LSTM layers, and a dense output layer. The model was trained using the categorical cross-entropy loss function and the Adam optimizer.

The code for building and training the model is as follows:

A screenshot of a computer program

Description automatically generated

**Text Generation**

To generate text, a seed prompt is given to the trained LSTM model, which predicts the next word in the sequence. The prediction is influenced by a parameter called **temperature**, which controls the randomness of predictions. Higher temperatures (closer to 1.0) produce more diverse and creative text, while lower temperatures produce more predictable and repetitive text.

The following function was used to generate text:

A computer screen shot of a program

Description automatically generated

A screenshot of a computer program

Description automatically generated

**Generated Text**

The model was tested with two temperature values: **0.5** (lower temperature) and **0.8** (higher temperature). The seed prompt used for testing was "Recipe for roasted vegetables | chop 1" and "Recipe for chocolate ice cream | mix".

**Temperature: 0.5**

A screenshot of a computer program

Description automatically generated

**Temperature: 0.8**

**A screenshot of a computer program

Description automatically generated**

**Conclusion**

The LSTM network successfully generates coherent recipe instructions by predicting the next word in a sequence. Lower temperature values (0.5) produce more predictable and straightforward text, while higher temperature values (0.8) generate more diverse and creative responses. The model could be further improved by training on more epochs or tuning hyperparameters.

**REFERENCES**

Kaggle. (n.d.). *Epicurious Recipes dataset*. Kaggle. https://www.kaggle.com/datasets/hugodarwood/epicurious-recipes-with-rating-and-nutrition

GitHub. (n.d.). *Long Short-Term Memory (LSTM) implementation for text generation*. GitHub. <https://github.com> (Note: Include the specific GitHub repository link you referred to, if applicable.)

Chollet, F. (2015). *Keras: Deep Learning for humans*. GitHub. <https://github.com/keras-team/keras>

TensorFlow. (n.d.). *Tokenizer API documentation*. TensorFlow. https://www.tensorflow.org/api\_docs/python/tf/keras/preprocessing/text/Tokenizer