Lora Milam

Western Govenors University

D213 Advanced Data Analytics

23 August 2023

D213 Performance Assessment Task 2

**1 Introduction**

Lora Milam Masters Data Analytics (8/11/2023) Program Mentor:d213@wgu.edu

**1.1 Research Question**

**1. Summarize one research question that you will answer using neural network models and NLP techniques. Be sure the research question is relevant to a real-world organizational situation and sentiment analysis captured in your chosen dataset.**

**1.2 Objective and Goals**

**2. Define the objectives or goals of the data analysis. Be sure the objectives or goals are reasonable within the scope of the research question and are represented in the available data.**

**1.3 Prescribed Network**

**3. Identify a type of neural network capable of performing a text classification task that can be trained to produce useful predictions on text sequences on the selected data set.**

**2 Data Preparation**

B.  Summarize the data cleaning process by doing the following:

**2.1 Data Exploration**

 Perform exploratory data analysis on the chosen dataset, and include an explanation of each of the following elements:

•   presence of unusual characters (e.g., emojis, non-English characters, etc.)

•   vocabulary size

•   proposed word embedding length

•   statistical justification for the chosen maximum sequence length

**2.2 Tokenization**

2.  Describe the goals of the tokenization process, including any code generated and packages that are used to normalize text during the tokenization process.

**2.3 Padding Process**

  Explain the padding process used to standardize the length of sequences, including the following in your explanation:

•   if the padding occurs before or after the text sequence

•   a screenshot of a single padded sequence

**2.4 Categories of Senitment**

4.  Identify how many categories of sentiment will be used and an activation function for the final dense layer of the network.

**2.5 Steps to Prepare the Data**

5.  Explain the steps used to prepare the data for analysis, including the size of the training, validation, and test set split.

**2.6 Prepared Dataset**

6.  Provide a copy of the prepared dataset.

**3 Network Architecture**

**3.1 Model Summary**

**1. Provide a line graph visualizing the realization of the time series.**

**3.2 Network Architecture**

**2. Describe the time step formatting of the realization, including any gaps in measurement and the length of the sequence.**

**3.3 Hyperparameters**

**3. Evaluate the stationarity of the time series.**

**4 Model Evaluation**

D.  Evaluate the model training process and its relevant outcomes by doing the following:

**4.1 Stopping Criteria**

1.  Discuss the impact of using stopping criteria instead of defining the number of epochs, including a screenshot showing the final training epoch.

**4.2 Training Process**

2.  Provide visualizations of the model’s training process, including a line graph of the loss and chosen evaluation metric.

**4.3 Fit**

3.  Assess the fitness of the model and any measures taken to address overfitting.

**4.4 Predictive Accuracy**

4.  Discuss the predictive accuracy of the trained network.

**5 Summary and Recommendations**

**5.1 Code**

E.  Provide the code used to save the trained network within the neural network.

**5.2 Functionality**

F.  Discuss the functionality of your neural network, including the impact of the network architecture.

**5.3 Recommendations**

G.  Recommend a course of action based on your results.

**6 Reporting**

**6.1 Reporting**

H.  Create your neural network using an industry-relevant interactive development environment (e.g., a Jupyter Notebook). Include a PDF or HTML document of your executed notebook presentation.

**7 Supporting Documentation**

**7.1 Sources**

Western Governors University. (n.d.). D213 Advanced Data Analytics. Salt Lake City.