**Neural Network Analysis of IMDB**

**Assignment - 2**

**Introduction**

This report presents an effort of developing a neural network model for the movie review dataset from IMDB for classifying the reviews as positive or negative. The project proposes data loading and pre-processing, as well as modeling data using TensorFlow and Keras with emphasis on high accuracy of classification available at learning algorithms.

If data is simple, that is, there are equations between sorted values and meanings, the process is called Data Loading and Preprocessing.

**Loading the Dataset**

The IMDB dataset with both positive and negative reviews of 50000 samples was loaded using Keras. The data set was split on their corresponding training and testing set automatically, while the reviews were preprocessed in which each review was converted into seque(nces of integers where each integer denotes a word).

**Decoding Reviews**

As a further preprocessing step and to have a better understanding of the content of the reviews and to also correctly manage the data, the first review in the training set was translated from the .txt back to English words using the dictionary provided at the IMDB online website.

**Preparing the Data**

Vectorization

All the integer sequences extracted from the reviews were converted to binary vectors using the multi-hot encoding method. This conversion of the sequences brings those into the format which can be fed into the neural network, thus each review is given an ability of 10,000 dimensional vector, laying emphasis whether the presence of words or not.

They include Model Architecture, where an architectural prototype which is reusable and independent of specific technologies is created, and Compilation where the model is translated to an implementation language without affecting its reusability.

Binary classification based neural network model is consist of several dense layers. The added layers with activation functions and steps of regularization provide a capacity to learn more complicated patterns of the data.

**Layers and Activation**

**The network includes:**

**Input layer:** Accepts preprocessed data.

**Dense layers:** Use Tanh for the hidden layers and Sigmoid for the output layer to activate the data.

**Output layer:** Uses a sigmoid activation function to produce the probability of the inputs of a review being positive.

Other factors such as dropout to help hinder overfitting were also incorporated in the network.

**Training and Evaluation**

The training data was vectorized and trained on the model iterative epochs and training accuracy together with cross-validation accuracy was computed to check the performances against over-fitting.o Training accuracy was 90.72 percent at the fourth epoch.o At the end of the displayed training, validation accuracy hit 87.17%.o This retrained model was tested and their accuracy was determined to be 86.51 percent with loss of 0.1331, proving the model is stable.metrics recorded for both training and validation datasets to monitor progress and prevent overfitting.

**Results**

* Initial Model Accuracy:
* Training accuracy reached 90.72% by the fourth epoch.
* Validation accuracy reached 87.17% at the end of the training.
* Retrained Model:
* The retrained model achieved a test accuracy of 86.51% with a loss of 0.1331, demonstrating stable performance.

**Conclusion**

Thus, the analysis of the resulting model based on a neural network confirms good classification results on the IMDB movie review dataset and high stability of the performance indicators. The proposed model’s test accuracy of 87.45% at one time showcases how deep neural networks could be applied to natural language processing tasks in the future. Future work could expand through broader architectures or different datasets, to ascertain the ability of the model.