**Neural Network Analysis of IMDB**

**Assignment - 2**

**Introduction**

In order to categorize the IMDB movie review dataset as either positive or negative, this research attempts to create a neural network model. The project suggests pre-processing and data loading, together with data modeling with TensorFlow and Keras, with a focus on learning algorithms that can classify data with high accuracy.   
Preprocessing and data loading are the terms used when the data is simple, meaning that there are equations between sorted values and meanings.

**Loading the Dataset**

Using Keras, the IMDB dataset including reviews of 50,000 samples, both positive and negative, was loaded. The data set was automatically divided between training and testing sets, and prior to preprocessing, each review was transformed into an integer sequence, where each integer represented a word.

**Decoding Reviews**

The first review in the training set was converted from.txt to English words using the dictionary offered on the IMDB website as an additional preprocessing step, to better comprehend the content of the reviews, and to properly manage the data.

**Preparing the Data**

Vectorization

Using the multi-hot encoding technique, all of the integer sequences that were taken out of the reviews were transformed into binary vectors. The process of converting the sequences into a format that can be fed into a neural network gives each review a 10,000-dimensional vector, emphasizing whether or not words are there.   
They include Model Architecture, which creates a reusable architectural prototype that is not dependent on any particular technology, and Compilation, which converts the model into an implementation language without compromising its reusability.

A neural network model based on binary classification has multiple dense layers. The ability to learn more complex patterns in the data is provided by the additional layers with activation functions and regularization procedures.

**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**

Training accuracy and cross-validation accuracy were calculated to check for over-fitting after the training data was vectorized and trained on the model iteratively over epochs.o At the fourth epoch, training accuracy was 90.72 percent.o The displayed training ended with a validation accuracy of 87.17%.o The accuracy of this retrained model was assessed and found to be 86.51 percent with a loss of 0.1331, demonstrating the model's stability.In order to track development and avoid overfitting, metrics were recorded for both training and validation datasets.

**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**

The examination of the neural network-based model that was produced thus validates excellent stability of the performance indicators and good classification results on the IMDB movie review dataset. The test accuracy of 87.45% at one point for the suggested model demonstrates how deep neural networks may be used in the future for tasks involving natural language processing. To determine the model's capability, further research could use larger architectures or distinct datasets.