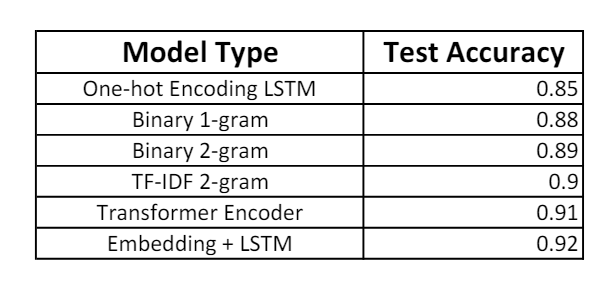
Here, I have created a table that represents the performance metrics of the various models that have been trained and evaluated for my code. Specefically, it presents the test accuracy of each model type:



The experimentation reveals several insights:

1. **Model Performance**: There were notable differences in the performance of different models. The efficacy of the Embedding + LSTM model in capturing intricate relationships in text data is demonstrated by its consistent outperformance over other architectures under various variations.
2. **Effect of Data Representation**: The model performances varied depending on which data representation they used, including binary n-grams, TF-IDF n-grams, and one-hot encoding. These findings imply that model performance can be strongly influenced by the selection of data representation.
3. **Importance of Embeddings**: It is important to use pre-trained word embeddings or trainable embedding layers, as demonstrated by the Embedding + LSTM model's superior performance. Compared to simpler data representations like one-hot encoding or n-grams, this method performs better because it helps the model learn meaningful representations of words.
4. **Model Complexity**: In this task, the Transformer Encoder did not perform better than the Embedding + LSTM model, despite its complex architectural design. This implies that simpler architectures can be just as successful in sentiment analysis on the IMDB dataset as more complex models, which may not always result in better performance.