**Neural Network Model Experimentation Report**

**Project Overview:**

This report details the experimentation and performance evaluation of various neural network architectures on the IMDB dataset. The goal of the project is to explore the effects of architectural changes, including varying hidden layers, activation functions, loss functions, and the use of regularization techniques, to optimize model performance.

**Dataset:**

The dataset used for training and validation is the IMDB movie reviews dataset, where the task is to classify reviews as either positive or negative. The data is split into training and testing sets.

**Experimentation Outline:**

The following variations were implemented in the neural network models:

1. **Hidden Layers:** Models with one, two, and three hidden layers were tested.
2. **Hidden Units:** Varying numbers of units (16, 32, 64) were used in the hidden layers.
3. **Loss Function:** The mse (Mean Squared Error) loss function was tested as an alternative to the commonly used binary\_crossentropy.
4. **Activation Function:** The tanh activation function was experimented with in place of the default relu.
5. **Regularization and Dropout:** L2 regularization and dropout were applied to reduce overfitting and improve the model's generalization ability.

**Model Configurations and Results:**

**Model 1: Single Hidden Layer**

* **Architecture:**
  + Input layer
  + 16 hidden units with relu activation
  + Output layer with sigmoid activation
* **Loss Function:** binary\_crossentropy
* **Accuracy:** 87.68%
* **Observation:** The simplest model with a single hidden layer performed decently with an accuracy close to 87.68%. This configuration is fast to train and shows reasonable performance, but it leaves room for improvement.

**Model 2: Two Hidden Layers with Tanh Activation**

* **Architecture:**
  + Input layer
  + 64 hidden units with tanh activation
  + Dropout (0.5)
  + 64 hidden units with tanh activation
  + Dropout (0.5)
  + Output layer with sigmoid activation
* **Loss Function:** mse
* **Accuracy:** 87.68%
* **Observation:** Adding a second hidden layer with more units and switching to the tanh activation function yielded a comparable accuracy. The use of the mse loss function, which is less common for classification tasks, did not significantly alter performance.

**Model 3: Three Hidden Layers with Regularization**

* **Architecture:**
  + Input layer
  + 32 hidden units with L2 regularization and tanh activation
  + Dropout (0.5)
  + 32 hidden units with tanh activation
  + Dropout (0.5)
  + 32 hidden units with tanh activation
  + Dropout (0.5)
  + Output layer with sigmoid activation
* **Loss Function:** mse
* **Accuracy:** 87.67%
* **Observation:** Introducing L2 regularization and three hidden layers did not yield an improvement in accuracy, but the model showed better generalization due to regularization and dropout, preventing overfitting during training.

**Model 4: Three Hidden Layers with Fewer Units**

* **Architecture:**
  + Input layer
  + 16 hidden units with L2 regularization and tanh activation
  + Dropout (0.5)
  + 16 hidden units with tanh activation
  + Dropout (0.5)
  + 16 hidden units with tanh activation
  + Dropout (0.5)
  + Output layer with sigmoid activation
* **Loss Function:** mse
* **Accuracy:** 87.68%
* **Observation:** Reducing the number of hidden units while maintaining three hidden layers and using dropout resulted in a model with similar accuracy (~87.68%). This suggests that adding more units or layers doesn't necessarily improve performance beyond a certain point.

**Key Insights:**

* **Hidden Layers and Units:** Increasing the number of hidden layers from one to three did not lead to a noticeable improvement in accuracy. Similarly, varying the number of hidden units had little effect on performance.
* **Activation Function:** The tanh activation function performed similarly to relu, making it a viable alternative, though no major advantage was observed.
* **Loss Function:** The use of mse instead of binary\_crossentropy did not improve performance and resulted in longer training times. For binary classification tasks, binary\_crossentropy is still the preferred choice.
* **Regularization and Dropout:** Adding L2 regularization and dropout successfully mitigated overfitting. This is particularly important for improving the model's generalization to unseen data.

**Conclusion:**

Across the different models, the neural network's performance remained stable at around 87.67% to 87.68% accuracy. While increasing the complexity of the model with additional hidden layers and units did not significantly improve accuracy, regularization techniques like dropout and L2 regularization helped in maintaining robust performance on the validation set, indicating that the model was not overfitting to the training data.

Future improvements could include further tuning of the regularization parameters, experimenting with different optimizers, or increasing the number of training epochs to enhance model performance.