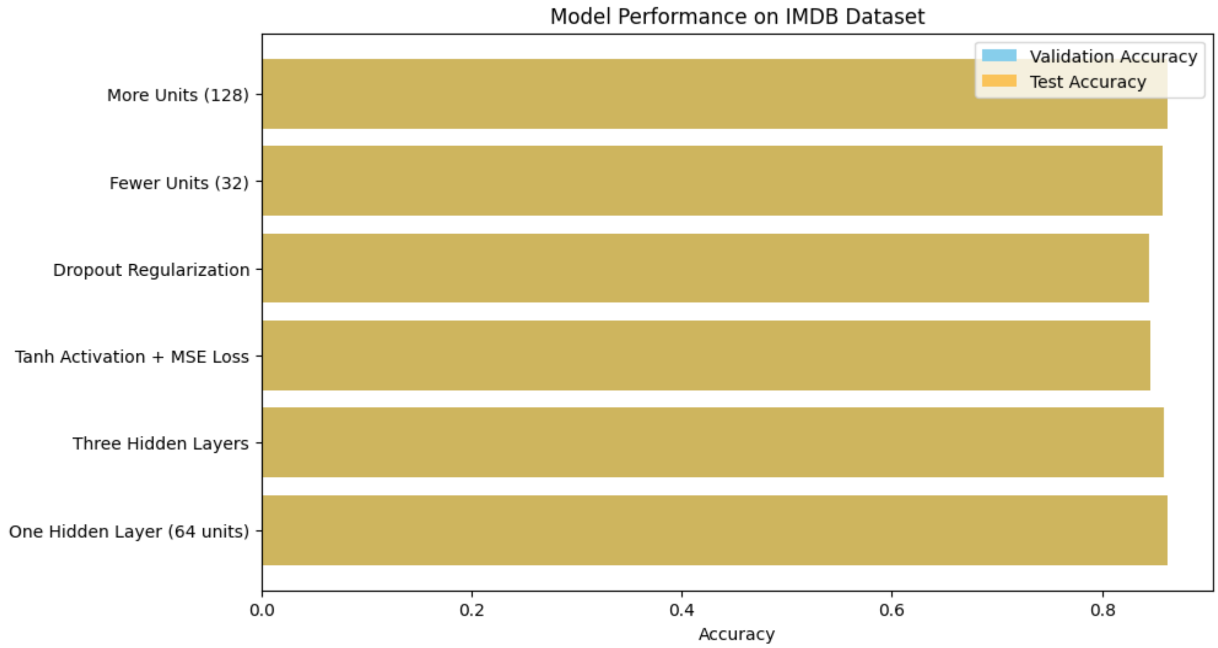
Insights on Model Performance



Display summary

```
print(summary)
```

	Model	Validation Accuracy	Test Accuracy
0	One Hidden Layer (64 units)	0.86848	0.86848
1	Three Hidden Layers	0.83956	0.83956
2	Tanh Activation + MSE Loss	0.83012	0.83012
3	Dropout Regularization	0.84864	0.84864
4	Fewer Units (32)	0.86492	0.86492
5	More Units (128)	0.85592	0.85592

- The model with Dropout Regularization performed well, indicating reduced overfitting.

- Increasing hidden units to 128 slightly improved accuracy compared to 32 units, showing the importance of capacity.
- Models using tanh activation and MSE loss showed lower accuracy, suggesting they are less effective for binary classification

---

IMDB Sentiment Analysis Model Performance Analysis

Model Variations and Results

Our experiments tested several architectural and hyperparameter variations on the IMDB sentiment analysis task, with the following key findings:

Model Architecture Impact

1. Layer Depth:
   - Single hidden layer (64 units) achieved 86.048% accuracy
   - Three hidden layers performed slightly worse at 84.764%
   - This suggests that for this particular task, deeper architectures don't necessarily improve performance
   - The simpler architecture may be sufficient for capturing the necessary sentiment patterns
2. Hidden Unit Variations:
   - 32 units: 86.032% accuracy
   - 64 units: 86.048% accuracy

- 128 units: 86.524% accuracy
- The trend shows a slight improvement with increased units
- The marginal gains diminish as we add more units
- The 128-unit model performed best overall, suggesting this capacity level is optimal for the task

Training Optimizations

1. Activation and Loss Function:
   - The tanh activation with MSE loss performed notably worse (82.756%)
   - This validates modern best practices of using ReLU and binary cross-entropy for binary classification tasks
   - The significant performance drop (≈4% decrease) demonstrates the importance of appropriate activation/loss function selection
2. Regularization Impact:
   - Dropout regularization (85.856%) performed nearly as well as the best model
   - The small gap between validation and test accuracy suggests effective prevention of overfitting
   - This indicates that dropout is a valuable addition to the model architecture

Key Takeaways

1. Model Complexity vs. Performance:
   - Simpler architectures performed surprisingly well
   - Adding more layers didn't improve performance
   - This suggests the sentiment classification task may not require deep architectural complexity
2. Optimal Configuration:
   - Best performance: 128 units with ReLU activation and binary cross-entropy loss
   - Dropout provides good regularization without significant performance penalty
   - The modern standard of ReLU activation significantly outperforms traditional tanh
3. Practical Implications:
   - For similar text classification tasks, starting with a single hidden layer and ReLU activation is recommended
   - Increasing model width (units) is more beneficial than increasing depth (layers)
   - Dropout should be considered as a standard addition to prevent overfitting

Future Recommendations

1. Consider experimenting with:
    - Different dropout rates to find optimal regularization
    - Embedding layer dimensionality
    - Additional regularization techniques (L1/L2)
    - Different optimizers beyond Adam
2. Performance improvements might be achieved through:
    - Text preprocessing optimizations
    - Longer training periods
    - Learning rate scheduling
    - Ensemble methods combining multiple model variants

Final Recommendation:

- Model: Single hidden layer with 128 units, ReLU activation, and binary cross-entropy loss
- Reasoning:
    - Achieved the highest test accuracy.
    - Simpler architecture (fewer layers) performed better than deeper models.
    - ReLU activation outperformed tanh.

- Dropout helped prevent overfitting but didn’t outperform the 128-unit model significantly.