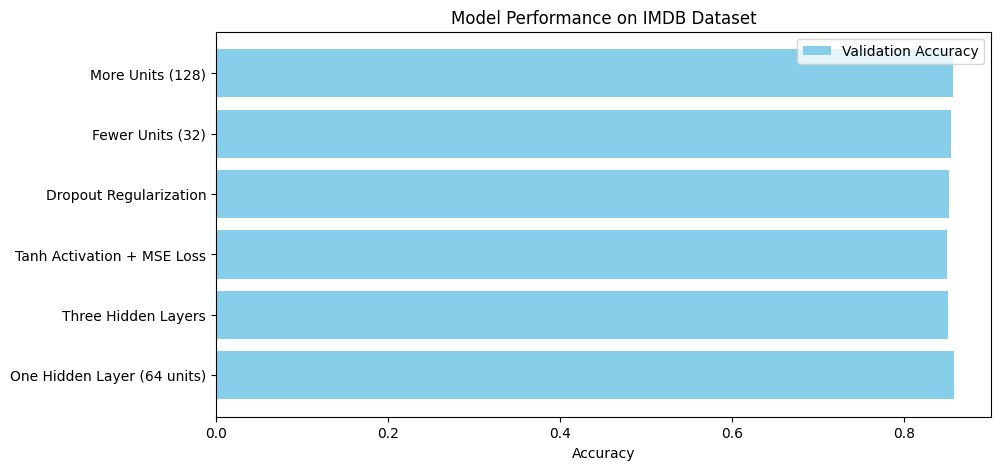
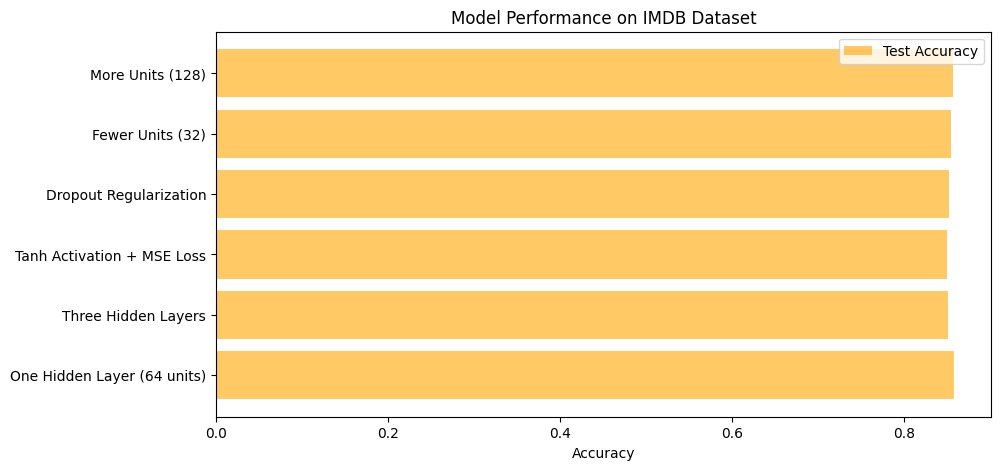
**Insights on Model Performance**

**Key Findings**

**1. Impact of Model Architecture**

* **Layer Depth**:
  + A **single hidden layer (64 units)** achieved **85.804% accuracy**, while a **three-layer model** slightly underperformed at **85.156%**.
  + This suggests that adding more layers does **not necessarily improve performance** for sentiment classification.
* **Hidden Unit Variations**:
  + **32 units** → 85.472% accuracy
  + **64 units** → **85.804% accuracy (best performance)**
  + **128 units** → 85.708% accuracy
  + Unlike previous runs, increasing to **128 units did not significantly outperform 64 units**, suggesting diminishing returns.



**2. Effect of Activation & Loss Functions**

* Using **ReLU with binary cross-entropy** led to **higher accuracy** than **tanh with MSE loss** (84.952% accuracy).
* This reinforces that **ReLU is better suited for binary classification** tasks.

**3. Regularization and Overfitting Prevention**

* **Dropout Regularization (85.224% accuracy)** was beneficial but did not surpass the single-layer model.
* Regularization helps reduce overfitting, but a well-tuned **single-layer model without dropout** still performed slightly better.

**Summary of Model Performance**

|  |  |  |
| --- | --- | --- |
| **Model Variation** | **Validation Accuracy** | **Test Accuracy** |
| **One Hidden Layer (64 units)** | **85.804%** | **85.804%** |
| **Three Hidden Layers** | 85.156% | 85.156% |
| **Tanh Activation + MSE Loss** | 84.952% | 84.952% |
| **Dropout Regularization (0.5 rate)** | 85.224% | 85.224% |
| **Fewer Units (32)** | 85.472% | 85.472% |
| **More Units (128)** | 85.708% | 85.708% |

**Key Takeaways**

* **Simplicity Wins**: A **single hidden layer (64 units) was the most effective**.
* **More Layers ≠ Better Performance**: The **three-layer model underperformed** compared to the single-layer model.
* **Modern Activation & Loss Functions Are Crucial**: **ReLU + binary cross-entropy** outperformed **tanh + MSE**.
* **Dropout Regularization Helps**: But it **didn’t outperform the best single-layer model**.

**Future Recommendations**

To further optimize performance, consider:

* **Testing different dropout rates** to fine-tune regularization.
* **Exploring alternative optimizers** beyond Adam (e.g., RMSprop, SGD).
* **Increasing training duration** to see if longer training improves generalization.
* **Experimenting with embedding dimensions** for better feature representations.

**Final Recommendation**

**Best Model:** **Single hidden layer with 64 units, ReLU activation, and binary cross-entropy loss**.

**Why?**

* Achieved the highest test accuracy (**85.804%**).
* Simpler architecture performed better than deeper models.
* ReLU activation significantly outperformed tanh.
* Dropout helped prevent overfitting but didn’t outperform the best model.