Report (Assignment 7)

Roll No. : IIT2022004
Name : Vatsal Bhuva

Performance Metrics

- **Test Accuracy**: Achieved an accuracy of **82.12%** on the test set.
- **Precision**: **86.21%** – When the model predicts that a passenger survived, it is correct approximately 86% of the time.
- **Recall**: **67.57%** – The model successfully identifies around 67.5% of all actual survivors.
- **F1 Score**: **75.6%**, reflecting a good balance between precision and recall.

Key Observations

- The model shows rapid learning in the initial epochs, with both training and validation performance plateauing around epoch **30**.
- A mild case of **overfitting** is observed, where training accuracy slightly exceeds validation accuracy.

Challenges Encountered & Potential Improvements

- **Missing Data**: The **Age** feature contains many missing values. While imputation with the median provides a quick fix, more sophisticated techniques (e.g., regression-based imputation) may yield better results.
- **Feature Engineering**: Further improvements could be achieved by:
    - Extracting **titles** (e.g., Mr., Miss, Dr.) from passenger names.
    - Parsing **cabin information** to derive deck levels.
    - Grouping passengers by **family size** to capture social relationships.
- **Model Tuning**:
    - Exploring different **architectures** with added layers or **dropout** to reduce overfitting.
    - Adjusting **learning rates**, applying **regularization**, or using different **optimizers** for better generalization.

- **Class Imbalance**: While not heavily imbalanced, applying **class weights** or **oversampling** could improve recall by ensuring minority class instances (survivors) are better learned.