Project Summary: Predicting High-Risk Passengers for Maritime and Travel Insurance Pricing

Business Problem:

Maritime insurance companies need a reliable method to assess the risk of passengers during potential maritime disasters. By identifying high-risk passengers, insurance companies can adjust their pricing models and ensure fair, competitive, and financially sound premiums for travelers. Traditionally, premiums were based on broad categories, which led to inaccuracies in pricing. This project aims to leverage predictive models to address this issue.

Data Understanding:

The project uses the Titanic dataset as a representative case of maritime passenger data. The dataset contains demographic and travel-related features.

Data Exploration and Feature Engineering:

- **EDA**: Exploratory Data Analysis (EDA) was performed to visualize distributions (age, fare, gender) and survival rates across different groups.
- **Handling Missing Values and Duplicates**: Duplicates were removed, and missing values were treated.
- **Outliers**: Outliers in key numeric features (FamilySize, Age) were handled by treating extreme values.
- **Feature Engineering**: A new combined feature, Sex_Pclass, was created to represent the interaction between a passenger’s gender and socio-economic status.
  Another feature, FareClassRatio, was engineered by dividing the fare by the class category (higher classes with lower values).
- **Scaling**: Min-max scaling was applied to continuous features (Age, Fare, FamilySize) to normalize the data.

Models Used:

1. **Logistic Regression**:
   - Logistic Regression was used as a baseline model for binary classification (Survived or Not Survived).
   - Hyperparameters such as epochs (number of training iterations) and eta (learning rate) were tuned using multiple configurations.
   - Best model achieved **84% accuracy** with parameters: epochs=20000, eta=0.1.

2. **Shallow Neural Network (Shallow_ANN)**:
    - A shallow neural network was implemented with one hidden layer using tanh and sigmoid activations.
    - The neural network was trained with various numbers of neurons, learning rates, and epochs.
    - Key hyperparameters included:
        - Neurons: 3, 6, 10, 15, and 20.
        - Learning rate (eta): 1e-4, 1e-3, 1e-2, 5e-2, and 1e-1.
        - Epochs: 500, 1000, 2000, 3000, and 5000.
    - The model output probabilities through the sigmoid function, and binary cross-entropy was used as the loss function.
    - The best model achieved 62.5% accuracy with parameters: neurons=3, eta=0.0001, epochs=3000
3. **Feedforward Artificial Neural Network (ANN)**:
    - A more complex neural network with multiple hidden layers and various architectures.
    - Architecture configurations included layers like [6, 8, 4, 2] and [12, 6, 4, 2] for modeling.
    - Regularization strategies:
        - **L2 Regularization**: Values of 0, 1e-4, 1e-3 were tested to control model complexity and reduce overfitting.
        - **Dropout**: Dropout rates (0, 0.3, 0.5) were applied to reduce overfitting by randomly ignoring a percentage of neurons during training.
    - Learning rates tested were 1e-3, 1e-2, and 1e-1, with epochs ranging from 3000 to 7000.
    - The training curves were plotted to visualize loss minimization, and accuracy was reported after each run.
    - The best model achieved 83.8% accuracy with parameters: architecture=[6, 8, 4, 2], activations=[tanh*4], eta=0.01, epochs=7000, l2_lambda=0, dropout=0.3

Conclusion:

This project successfully demonstrates the use of predictive modeling to assess passenger risk in maritime disasters, using the Titanic dataset as a case study. By applying models like **Logistic Regression**, **Shallow Neural Networks**, and **Feedforward Artificial Neural Networks**, the study achieved strong predictive accuracy, with the **best Binary Logistic Regression 84% accuracy**.

Overall, the project highlights the benefits of using data-driven models to modernize the insurance industry. With these techniques, insurers can more accurately predict risk, price premiums fairly, and maintain profitability even in the face of unpredictable maritime disasters.