**Problem Statement & Description**

The prediction of survival on the Titanic is a classic problem in data science. However, the process of building machine learning models for this task can be prone to biases, especially concerning demographic variables such as gender. The objective of this project is to investigate bias detection and mitigation in machine learning models predicting survival on the Titanic dataset. We aim to build and evaluate two machine learning models: decision forests and neural networks, to predict passenger survival based on various features available in the dataset. Therefore, the primary objective of this assignment is to build two machine learning models using decision forests and neural networks, respectively, to predict whether or not a passenger will survive. In this context, the predicted result of “1” represents male passengers, while “0” denotes female passengers.

**Brief Description of the 2 Methods Used**

1. **Neural Networks**: Neural networks consist of interconnected nodes organized in layers. In this assignment, a shallow neural network is utilized, which is composed of an input layer, one or more hidden layers, and an output layer. The network learns to map input features to output classes through a series of weighted connections and activation functions.
2. **Decision Forests**: Decision forests are an ensemble learning method consisting of multiple decision trees. Each tree in the forest independently learns a set of decision rules from the training data. During prediction, the final output is determined by aggregating the predictions of individual trees, typically through a voting mechanism.

**Experimental Results – Decision Forest**

* Model summary: the summary of the model shows the estimated coefficients for each feature.A close-up of a white background

  Description automatically generated
* Confusion matrix: 2 confusion matrices are generated. The first one corresponds to one subset of the test data, where 91 instances where the model predicted class 0 and the actual class was also 0, 21 instances where the model predicted class 1 but the actual class was 0, 4 instances where the model predicted class 0 but the actual class was 1, and 6 instances where the model predicted class 1 and the actual class was also 1.A computer screen shot of a computer code

  Description automatically generated
* Accuracy: the overall accuracy calculated for males was 0.7950820, meaning that around 80% of the predictions made by the model match the actual labels in the male group. Similarly, the accuracy for the females was 0.7894737, or 79%. A close up of a number

  Description automatically generated

**Experimental Results – Neural Network**

* Model training: a neural network model was trained using the nnet() function with one hidden layer of 5 neurons (size = 5). The training process converged successfully, as indicated by the absence of convergence warning messages. A screenshot of a computer

  Description automatically generated
* Model summary: the summary of the model shows the estimated coefficients for each feature along with their min, max, median, and mean values. A white background with blue text

  Description automatically generated
* Confusion matrix: the confusion matrices provide a detailed breakdown of the model’s predictions at different probability thresholds. Each row in the matrix represents a specific probability threshold, while each column indicates the number of instances predicted as either 0 or 1 at that threshold level. For example, at a threshold of 0.0569769589456554, the model predicted 58 instances as 0 (negative class) and 7 instances as 1 (positive class). Similarly, at a threshold of 0.0590997235580439, the model made one prediction for class 0 and zero predictions for class 1.  
   A number in a row

  Description automatically generated with medium confidence A number of numbers and digits

  Description automatically generated with medium confidence
* Accuracy: the overall accuracy calculated for males was 0.47540984, meaning that around 48% of the predictions made by the model match the actual labels in the male group. Similarly, the accuracy for the females was calculated to be 0.01754386, or 2%. These accuracies indicate that the neural network models did not perform very well in predicting the target variable for both male and female groups. A close-up of a network account

  Description automatically generated
* Bar graphs:

A comparison of different blue and red squares

Description automatically generated

**Conclusion**

Our analysis of bias detection and mitigation in machine learning models for predicting survival on the Titanic dataset yielded insightful results. While both decision forests and neural networks showed promising overall performance, there were variations in prediction accuracies across gender groups. These differences highlight the importance of considering demographic variables and conducting thorough evaluations to detect and mitigate biases in machine learning models. Further investigation and refinement of the models may be necessary to ensure fairness and equity in predictive modeling tasks.