**Problem Statement & Description**

The Titanic dataset project aims to explore and analyze passenger information from the RMS Titanic's maiden voyage using three distinct data analysis techniques: decision trees, logistic regression, and shallow neural networks. The dataset contains attributes such as passenger class, age, gender, fare, and survival status, with the objective of predicting passenger survival based on these attributes. This classification challenge serves as an opportunity to understand the strengths and weaknesses of different classification methods in predicting survival outcomes.

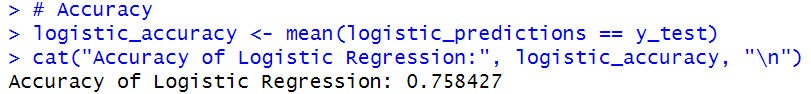
**Brief Description of the 3 Methods Used**

1. **Logistic Regression**: Logistic regression is a linear model used for binary classification tasks. It models the probability that a given input belongs to a particular class using the logistic function. It is a simple yet effective algorithm for classification tasks.
2. **Shallow Neural Networks**: Shallow neural networks, also known as feedforward neural networks, consist of an input layer, one or more hidden layers, and an output layer. In this assignment, we'll use a shallow neural network with a single hidden layer. The network will learn to map input features to output classes through a series of weighted connections and activation functions.
3. **Decision Trees**: Decision trees are a non-parametric supervised learning method used for classification and regression tasks. They learn simple decision rules inferred from the input features to predict the target variable. Decision trees partition the feature space into regions, assigning a class label to each region based on the majority class of training samples within that region.

**Experimental Results – Logical Regression**

* Model summary: the logistic regression model was trained using the glm() function with the binomial family on the training dataset (train\_data). The summary of the model shows the estimated coefficients for each feature along with their standard errors, z-values, and p-values. However, the model appears to have convergence issues, as indicated by the warning messages during the fitting, which may have impacted the reliability of estimated coefficients. A screenshot of a computer

  Description automatically generated
* Confusion matrix: evaluation on the testing dataset revealed that the logistic regression model correctly classified instances of survival and non-survival. However, it misclassified some instances, indicating potential limitations. A computer screen shot of a computer error

  Description automatically generated
* Accuracy: the overall accuracy of the logistic regression model on the testing dataset is approximately 75.84%. 

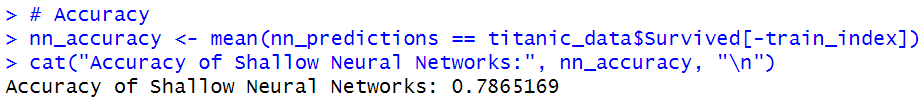
**Experimental Results – Shallow Neural Network**

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

A screenshot of a computer

Description automatically generated

* Confusion matrix: Evaluation on the testing dataset showed that the shallow neural network model correctly classified most instances of survival and non-survival. However, it misclassified some instances, suggesting areas for improvement. A screenshot of a computer error

  Description automatically generated
* Accuracy: the overall accuracy of the shallow neural network model on the testing dataset is approximately 78.65%. 

**Experimental Results – Decision Trees**

* Decision tree structure: the decision tree model was built using the rpart() function. The printed representation of the decision tree shows the hierarchical structure of the tree, with each node indicating a decision point based on a feature’s value. Terminal nodes, denoted by “\*”, represent the final classification decision. A screenshot of a computer code

  Description automatically generated
* Visualization: the decision tree was visualized using the prp() function from the rpart.plot package. A diagram of a number of numbers

  Description automatically generated with medium confidence
* Confusion matrix: Evaluation on the testing dataset showed that the decision tree model correctly classified most instances of survival and non-survival. However, it misclassified some instances, suggesting areas for improvement. A screenshot of a computer code

  Description automatically generated
* Accuracy: the overall accuracy of the decision tree model on the testing dataset is approximately 78.65%. A close-up of a number

  Description automatically generated

**Discussion**

The logistic regression model exhibited moderate performance in predicting passenger survival. Despite its simplicity, the model faced convergence issues during training, which raise concerns regarding the stability and reliability of the estimated coefficients. Further investigation is necessary to address these convergence issues and potentially improve the model's performance. Additionally, the misclassification of instances, particularly passengers classified as "survived" instead of "not survived", suggests potential overlap or ambiguity in the feature space. This highlights the need for more advanced modeling techniques to better capture the underlying patterns and nuances in the data.

On the other hand, the shallow neural network model demonstrated strong performance in predicting passenger survival, achieving a high accuracy rate. The successful convergence of the training process indicates the stability and reliability of the model for classification tasks. However, the misclassification of some instances suggests that further refinement of the model architecture and training parameters may be beneficial to improve performance and reduce errors. Overall, the shallow neural network model shows promise as a robust classifier for the Titanic dataset, effectively capturing the complexities of the data and providing accurate predictions.

Similarly, the decision tree model showed excellent performance in predicting passenger survival, achieving a high accuracy rate comparable to the shallow neural network model. The hierarchical structure of the decision tree allows for straightforward interpretation of the classification rules, making it a useful tool for understanding the data and making predictions. However, like the other models, the decision tree model also misclassified some instances, indicating potential areas for improvement. Nevertheless, the decision tree model remains promising for the Titanic dataset, offering high accuracy and interpretability, which are essential for practical applications and decision-making processes.