Final Project: Titanic Survival Prediction

Moeen Khan

2024-11-17

## Introduction

This document provides an overview of the main functions being developed for the Titanic survival prediction project. We aim to predict the likelihood of passenger survival using machine learning models, including GLM, Decision Tree, and KNN.

## Main Functions

### 1. clean\_data()

This function preprocesses the Titanic dataset by: - Handling missing values. - Encoding categorical variables. - Normalizing numerical features.

clean\_data <- function(data) {  
 data <- na.omit(data)  
 data$Sex <- as.numeric(factor(data$Sex))  
 data$Class <- as.numeric(factor(data$Class))  
 return(data)  
}

## fit\_glm\_model()

This function builds a generalized linear model (GLM) for predicting survival.

fit\_glm\_model <- function(data) {  
 model <- glm(Survived ~ Class + Sex + Age, data = data, family = binomial)  
 return(model)  
}

## predict\_survival()

This function generates survival predictions for new data using the trained GLM.

predict\_survival <- function(model, new\_data) {  
 predictions <- predict(model, new\_data, type = "response")  
 return(ifelse(predictions > 0.5, 1, 0))  
}