**Assignment-10**

Ravali Gandhavarapu

8886593

Big Data Solution Architecture, School of Applied Computer Science & Information Technology

Prog8420: Programming for Big Data

Prof. Jomis Varikayanickal

Assignment Due Date: 02 Aug, 2023

**Assignment 10**

**Task 1: Dataset Selection**

I have chosen the titanic data set. It has the information about passengers on the Titanic, like age, gender, class, and whether they survived or not. I thought this dataset would be suitable for classification tasks, where the objective could be to predict whether a passenger survived or not based on the provided data.

**Task 2: Data Exploration with Python**

**Dataset**

**A screenshot of a computer screen

Description automatically generated**

**Summary Data**

**A screenshot of a computer screen

Description automatically generated**

**Data types:**

**A screenshot of a computer

Description automatically generated**

**Visualization:**

**A group of blue and white graphs

Description automatically generated**

**Task 3: Data Preprocessing with Python**

**A screenshot of a computer

Description automatically generated**

**Task 4: Implement Machine Learning Models with Python**

**A black screen with white text

Description automatically generated**

From the comparison, we can see that the Random Forest model performs better than the SVM model for this dataset. It has higher accuracy, precision, recall, and F1 score. So this is the better model.

**Task 5: Visualization with Python**

**A graph showing the number of survival distribution

Description automatically generated with medium confidence**

**A graph with blue and orange dots

Description automatically generated**