**Assignment 10- Output**

Task 1 –

**a. Dataset Selection:**

1. Dataset: "Titanic Dataset"

Link: <https://github.com/awesomedata/awesome-public-datasets/tree/master/Titanic>

2. Dataset: "Bank Credit Scoring Dataset"

Link: <https://www.kaggle.com/datasets/kapturovalexander/bank-credit-scoring>

**b. Justification**

1. Titanic Dataset: -

The Titanic dataset is a well-known and durable dataset that includes details on passengers on the Titanic, including whether or not they survived. For classification tasks, it makes a great dataset. Since the target variable is binary (survived or not), binary classification methods are perfect for it. The analysis potential of this dataset resides in figuring out the variables that affected the survival of passengers. Based on factors like age, gender, class, and other characteristics, machine learning models can be created to forecast whether a passenger would survive. The 'Name' column can be used to extract data, such as titles, and to create new features by combining already existing ones, such as family size, to produce new ones. The dataset is popular in education and useful for practicing different classification techniques.

2. Bank Credit Scoring Dataset: -

The "Bank Credit Scoring" dataset is very pertinent to classification and credit risk analysis tasks in machine learning. It includes details about bank customers, such as their credit-related characteristics and whether credit was extended to them (i.e., binary classification problem). Credit risk assessment, a critical responsibility for banks and financial institutions to analyze customers' creditworthiness and make educated lending decisions, offers the most opportunity for analysis of this dataset. Overall, the "Bank Credit Scoring" dataset is useful for both academic research and real-world financial applications. Financial institutions can learn from machine learning models that have been trained on this dataset about determining a customer's creditworthiness, lowering credit risks, and making wise lending decisions.

Exploration, Processing, Implementation of the Machine Learnings, and Visualisations for the “**Titanic**” dataset.

Task 2 –

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

Task 3 –

A screen shot of a computer

Description automatically generated

A diagram of a box plot

Description automatically generated

Task 4 –

A screenshot of a computer

Description automatically generated

Based on the above evaluation “Random Model Evaluation” model is having high accuracy as accuracy is more for that model. Additionally, Precision, recall and f1- score is also more for model 2. So, based on the above factors we can say that the “Random Model Evaluation” model is having high accuracy.

Task 5 –

Scatter Plot : -

A screen shot of a computer screen

Description automatically generated

HeatMap: -

A graph with red and blue squares

Description automatically generated

Bar Charts: -

A graph of a passenger class distribution

Description automatically generated

Histogram: -

A graph of age distribution

Description automatically generated

Boxplots: -

A graph of a box plot

Description automatically generated

Exploration, Processing, Implementation of the Machine Learnings, and Visualisations for the “**Bank Credit Scoring**” dataset.

Task 2: -

A screenshot of a computer program

Description automatically generated

Task 3: -

A screenshot of a computer

Description automatically generated

A blue line graph with black lines

Description automatically generated

Task 4: -

A screenshot of a computer program

Description automatically generated

Based on the above evaluation “SVM Model Evaluation” model is having high accuracy as accuracy is more for that model.

Task 5: -

A graph of different colored columns

Description automatically generated

A graph of age distribution

Description automatically generated

A graph of a number of people

Description automatically generated

A group of blue and white data

Description automatically generated

A screenshot of a graph

Description automatically generated