Problem Statement 2

SOCIAL NETWORK AD ANALYSIS

PROBLEM STATEMENT:

Try to understand the dataset of Social_Network_Ads.csv and try to find the best suitable ML algorithm and write the code in python for algorithm from scratch and try to achieve the below output plot.

OBJECTIVE:

- Data Preparation
- Feature Selection
- Model Building
- Results Visualization

BACKGROUND:

The given dataset is based on classification where to predict the Ads Purchased for new data. The dataset consists of 400 rows and 5 columns.

PREDICTIVE ANALYSIS:

- **DATA PREPARATION**

In preparation for inputting the X and y to our model, we must first split it into train and test sets.

- **FEATURE SCALING**

The X_train and X_test are scaled using a Standard scaler.

- **MODEL BUILDING**

I have used Decision Tree for modeling, because only this model give the desired output for the given dataset. Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction of the individual trees.

CONFUSION MATRIX:

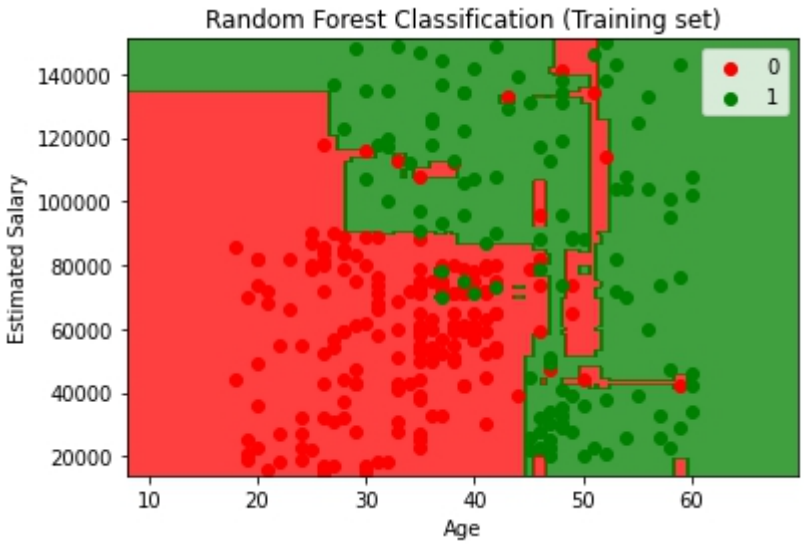
```
[[63  5]
 [ 4 28]]
```

MODEL ACCURACY:

```
0.91
```

- **VISUALIZING THE TRAIN AND TEST RESULTS:**

TRAIN SET:



TEST SET: