**Netaji Subhas University of Technology**



# REPORT FILE

COMPUTER HARDWARE AND SOFTWARE

**GROUP MEMBERS:-**

Aryan Gangwar (2021UCS1606)

Devashis Soren (2021UCS1595)

Raveesh Gupta (2021UCS1581)

## **Project Title: Salary Prediction**

## **INTRODUCTION**

## The Salary Prediction project utilizes Decision Trees for accurate salary estimation. Employing the R programming language, the model considers factors like experience and education. Through data exploration and preprocessing, Decision Trees facilitate a straightforward yet effective approach to predict salaries. The project aims to provide organizations with a reliable tool for quick and informed decisions on compensation, leveraging the simplicity and interpretability of Decision Trees for efficient salary predictions.

## **OBJECTIVE**

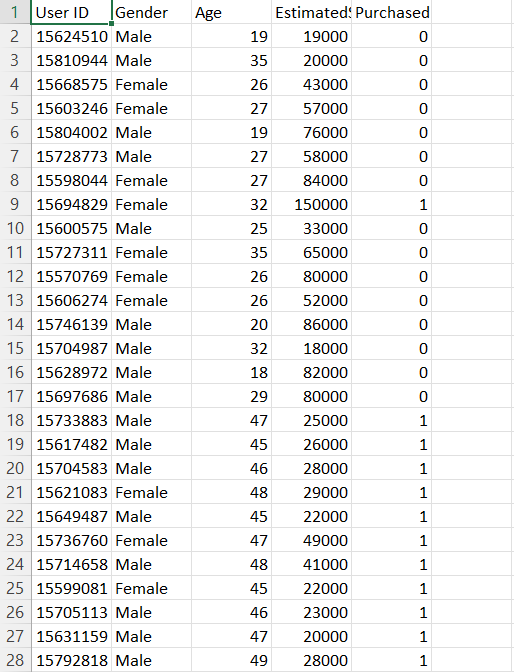
The project's objective is to employ Decision Trees for precise salary prediction. Utilizing R, the model considers key factors like experience and education. Through data preprocessing, the project aims to develop a streamlined and effective tool for organizations to make informed decisions on employee compensation, enhancing accuracy and efficiency in salary estimation.

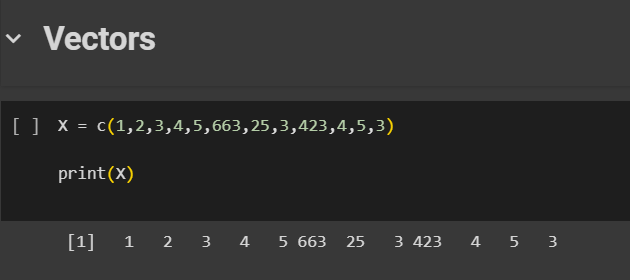
## **R**

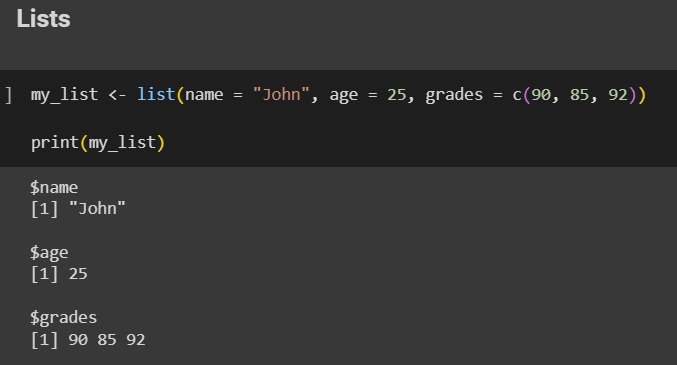
R is a powerful, open-source programming language and environment designed for statistical computing and data analysis. Widely used in academia and industry, R provides a vast array of statistical and graphical techniques. Its extensive package ecosystem and robust visualization capabilities make it a preferred tool for data scientists and statisticians. With a syntax conducive to data manipulation and analysis, R supports a variety of data formats and offers flexibility in handling diverse statistical models. The language's active community contributes to its continual development and ensures its relevance in the evolving field of data science and statistical analysis.

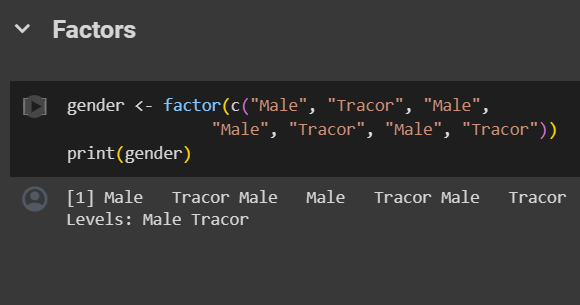
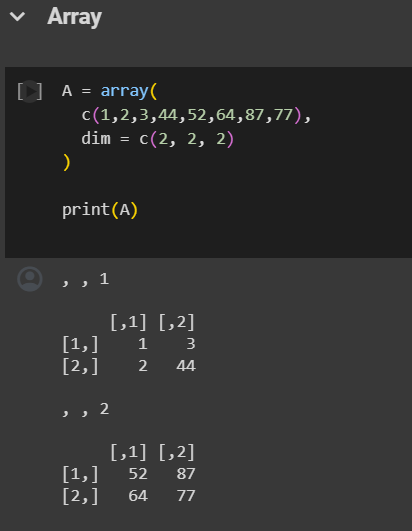
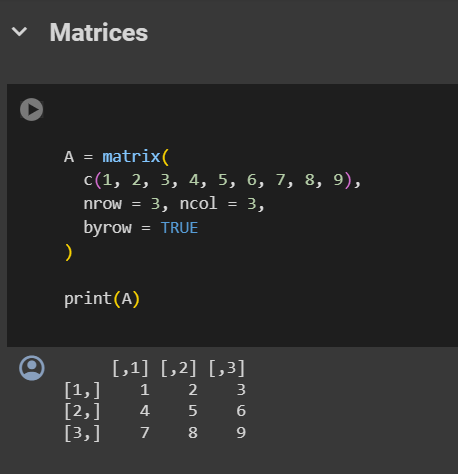
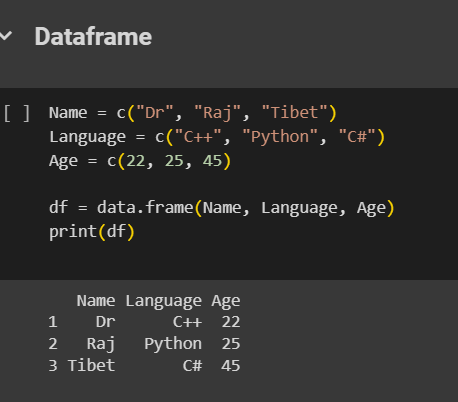
## **DATASET**

Social\_Network\_Ads.csv

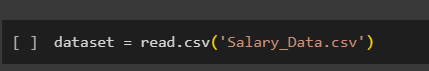


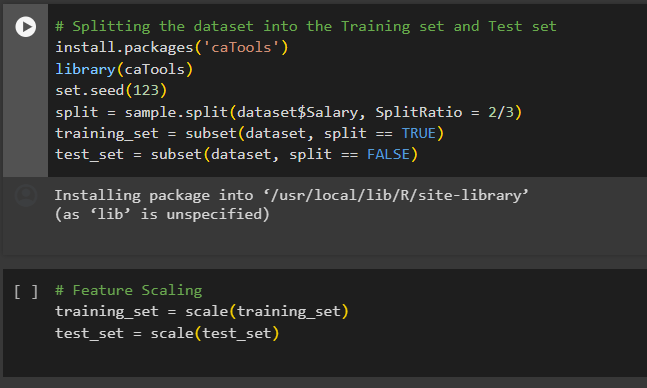
**DATA STRUCTURES IN R**

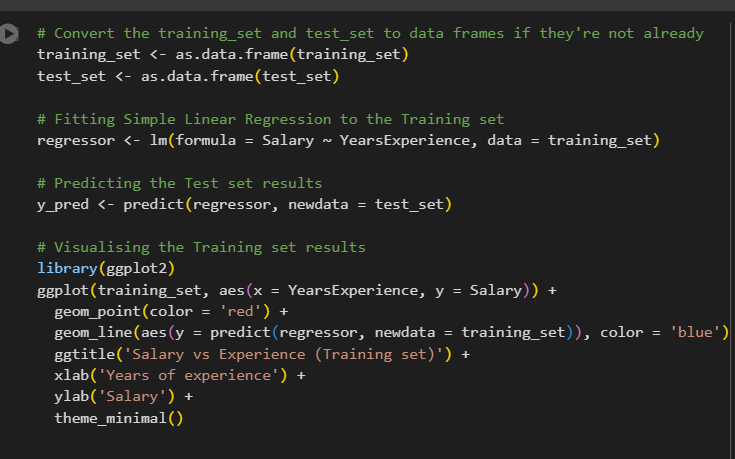
****

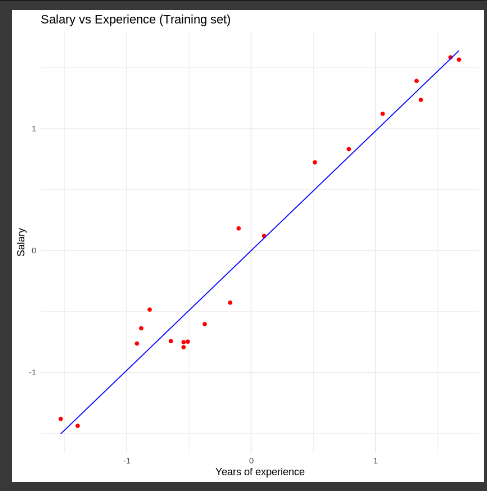
****

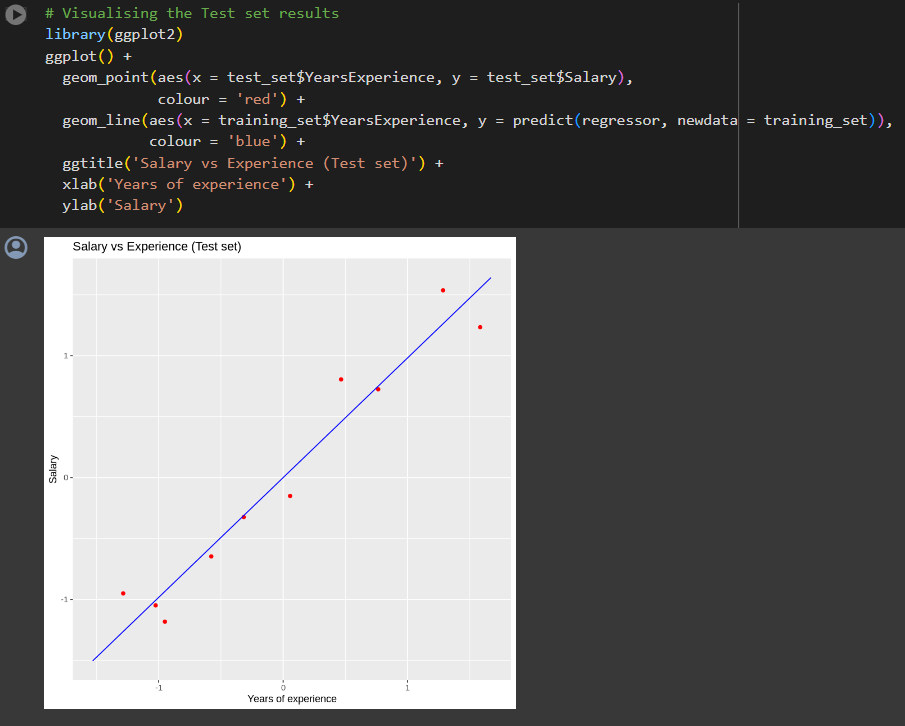
**LINEAR REGRESSION**

****

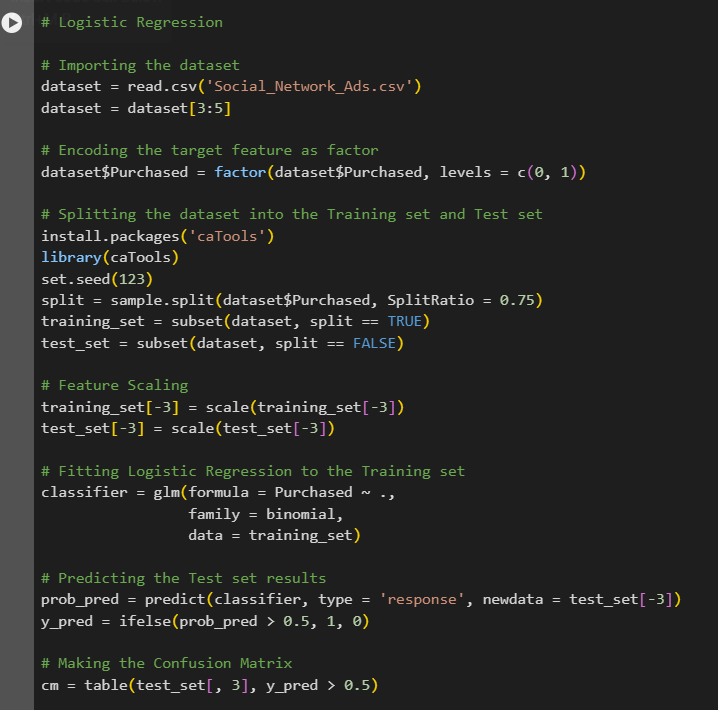
****

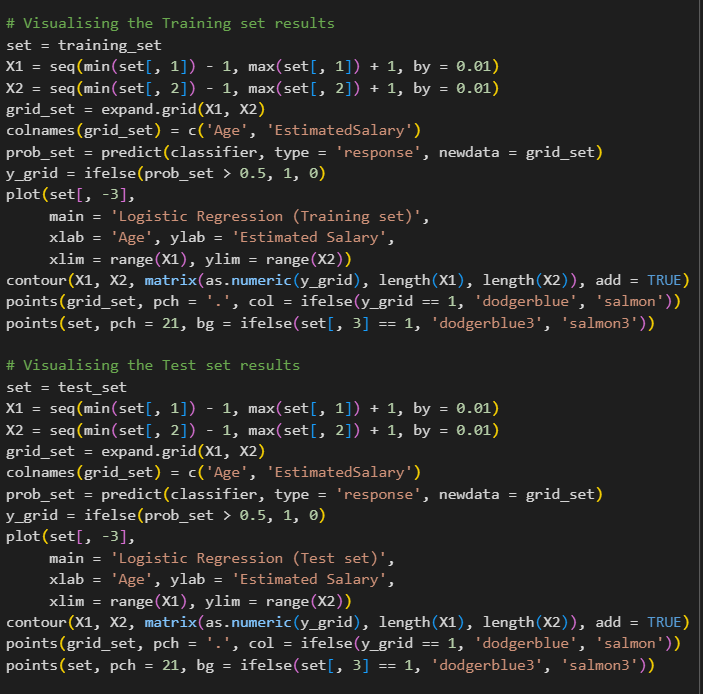
****

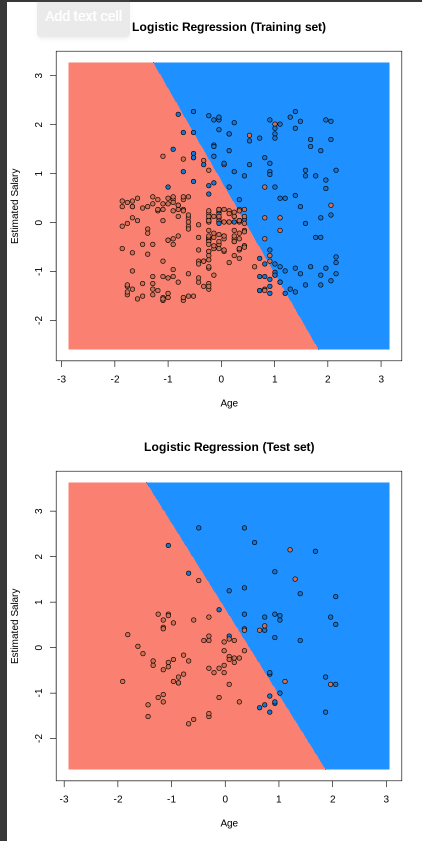
****

****

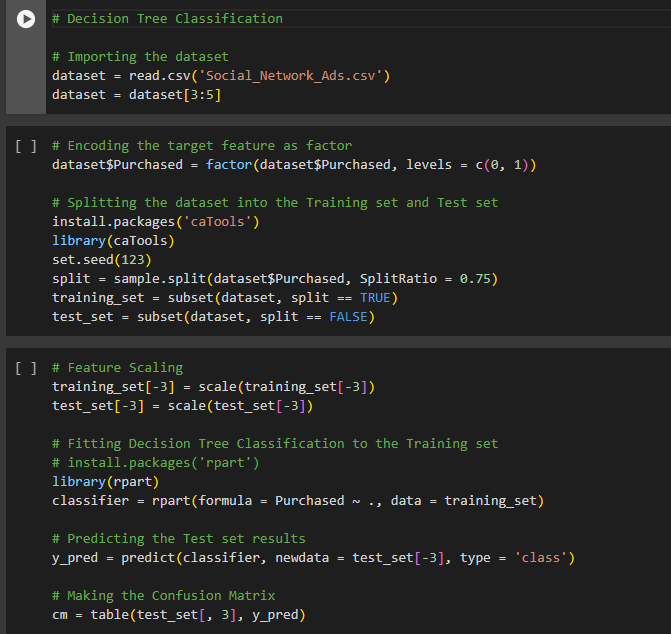
## **LOGISTIC REGRESSION**

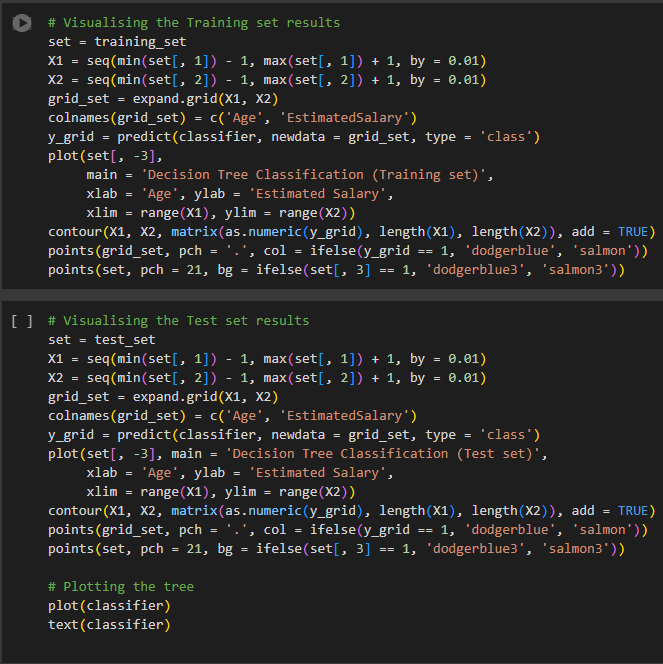


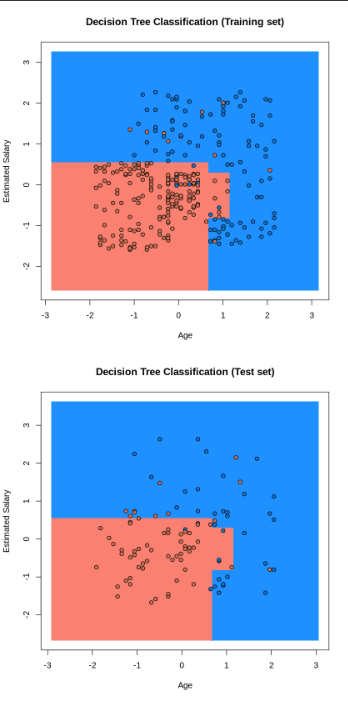


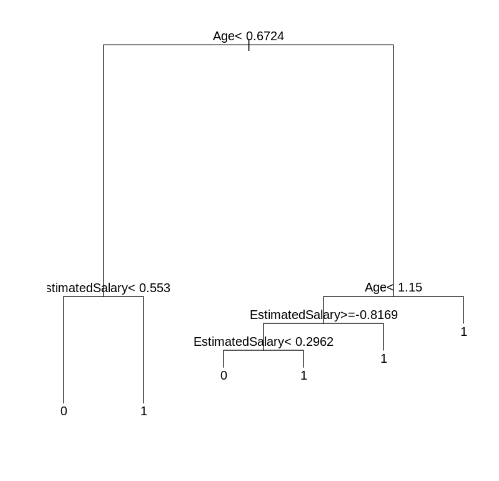


## **DECISION TREE ALGORITHM**









## **CONCLUSION**

In this data science project, we went through the customer segmentation model. We developed this using a class of machine learning known as unsupervised learning. Specifically, we made use of a clustering algorithm called K-means clustering. We analysed and visualized the data and then proceeded to implement our algorithm. Hope you enjoyed this customer segmentation project of machine learning using R