Name: Muhammad Sheharyar & Awais Ur Rehman
Roll no # 19122017 & 19122172
Semester: 6th sec B
To Miss Uzma Afzal

Report on Salary Base Data Analysis

OBJECTIVE:

The objective of this analysis is to gain comprehensive insights into the salary data of employees within a company. By exploring various factors such as age, gender, education level, job title, years of experience, and salary, the goal is to understand the dynamics influencing compensation. This analysis aims to provide organizations with valuable information to shape their compensation structures, ensuring fairness and competitiveness. Additionally, the report seeks to demonstrate the significance of utilizing decision trees and random forests algorithms for predictive modeling in this context.

INTRODUCTION AND BACKGROUND OF THE PROBLEM:

In the realm of Human Resources and Workforce Management, understanding the intricate relationships between employee demographics and compensation is crucial. The dataset under consideration contains essential details about employees, and the challenge lies in deciphering these details to optimize compensation strategies. The analysis will shed light on how factors such as age, education, and job title impact salary decisions, fostering transparency in organizational practices.

DATA COLLECTION:

The dataset utilized for this analysis was obtained from Kaggle at the following link: [Salary Base Data](https://www.kaggle.com/datasets/sinhasatwik/salary-base-data). Each row in the dataset represents a unique employee, and the columns provide information about age, gender, education level, job title, years of experience, and salary.

1. **Age:** Numeric representation of the age of each employee in years.

2. **Gender:** Categorical data indicating the gender of each employee (male or female).

3. **Education Level:** Categorical data specifying the educational level of each employee (high school, bachelor's degree, master's degree, or PhD).

4. **Job Title:** Categorical data denoting the job title of each employee, ranging from managerial roles to technical positions.

5. **Years of Experience:** Numeric representation of the number of years of work experience for each employee.

6. **Salary:** Numeric values representing the annual salary of each employee in US dollars.

DATA PREPROCESSING:

Prior to analysis, the dataset underwent preprocessing, including handling missing values and encoding categorical variables. Numerical scaling may have been applied for consistency, and any outliers were addressed to ensure the robustness of subsequent analyses.

MODELLING AND EVALUATION:

Decision Trees and Random Forests algorithms were chosen for predictive modeling. These algorithms are adept at predicting categorical variables like job title or gender based on other features. They excel at capturing complex relationships within the data, handling both numerical and categorical variables, and capturing non-linear patterns. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1 score.

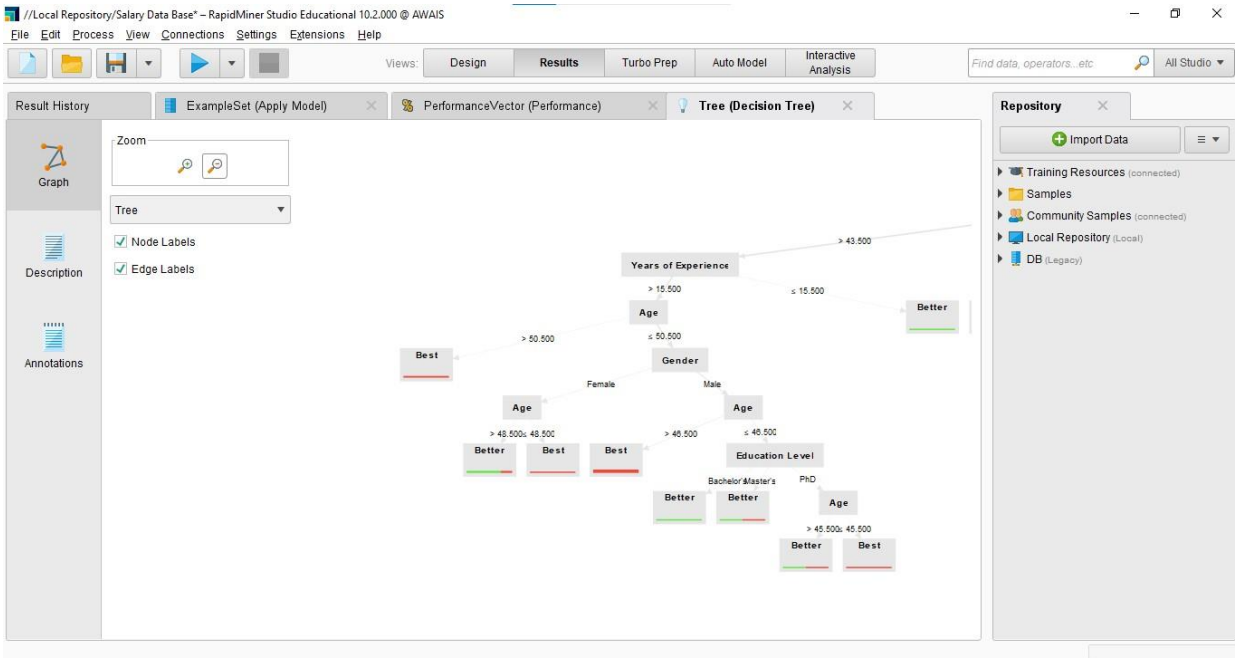
RESULTS:

The analysis provided valuable insights into the factors influencing employee salaries. Decision Trees and Random Forests effectively predicted categorical variables, contributing to a nuanced understanding of compensation dynamics.

CONCLUSIONS:

In conclusion, this analysis demonstrates the importance of leveraging salary data to inform compensation decisions. The utilization of Decision Trees and Random Forests allows organizations to navigate the complexities of employee demographics and tailor compensation practices for fairness and competitiveness. The transparency fostered by this approach supports informed decision-making in Human Resources and organizational strategies, ensuring compliance with employment regulations.

Result:



Accuracy: