PPT For Salary Prediction Project

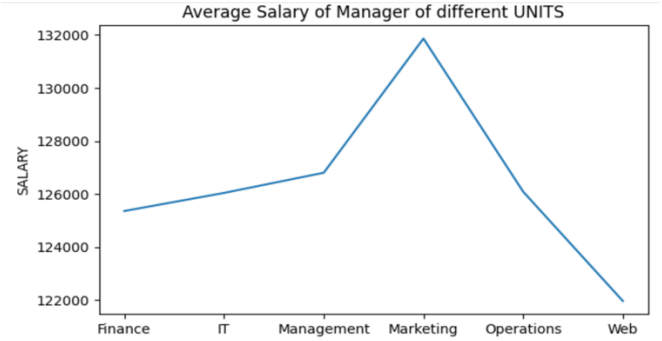
Problem: I am provided with a dataset consisting of some information about employees , and our task is to predict their Salary according to the data provided .

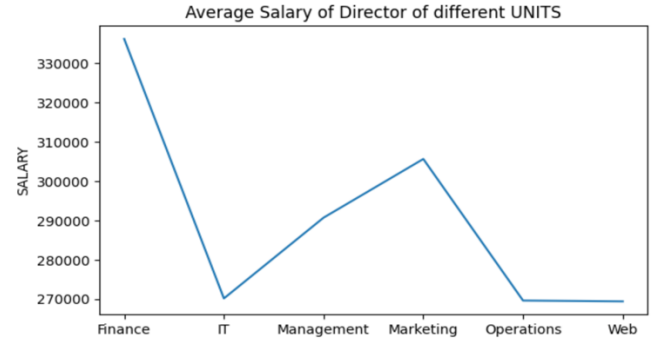
Libraries Used: I have used NumPy, Pandas, Matplotlib and Seaborn for this task .

DATA ANALYSIS

- ➔ There are total six designation of employees in this dataset , that are Analyst , Manager , Senior Manager , Director , Associate and Senior Analyst.
- ➔ Also there are six units in which employees works that are Finance, Marketing, Operations , Management , Web and IT.

Analyzing how salary varies of employers of different profession in different units





KEY FINDINGS :

1) We can see that the average salary of Analyst in Web UNIT is highest, and it is least in IT UNIT

2) Average salary of Associate is maximum in Operations UNIT and least in IT UNIT .

3) Average salary of Senior Analyst is maximum in IT UNIT and least in Marketing UNIT .

4) Average salary of Senior Manager is maximum in Finance UNIT and least in Marketing UNIT .

5) Average salary of Manager is maximum in Marketing UNIT and least in Web UNIT .

6) Average salary of Director is maximum in Finance UNIT and least in Web UNIT .

Analyzing, is there any gender discrimination in SALARY ?

KEY FINDINGS :

1) There is not much difference in salaries of MALE and FEMALE Analyst in different units.
2) For Associates , there is discrimination between salaries of MALE and FEMALE employees.
-> MALE Associates are paid less in Finance, Management and IT UNIT.
-> FEMALE Associates are paid less in Operations UNIT.
3) For Senior Analyst , there is gender discrimination in Salaries of MALE and FEMALE employees in Finance UNIT.
4) For Senior Manager , there is gender discrimination in Salaries of MALE and FEMALE employees in Finance , Web , Operations and Management UNIT.
5) For Manager , there is gender discrimination in Salaries of MALE and FEMALE employees in Finance , IT and Management UNIT.
6) For Director , there is gender discrimination in Salaries of MALE and FEMALE employees in Web , IT , Operations , Marketing and Management UNIT.

Analysis of change in Salary with Age of Employee

KEY FINDINGS:

There is not a gradual change in Salaries of employees with change in their Age .

Analyzing change in salaries of employees with change in years of experience they have

KEY FINDINGS:

Here also there is not a gradual increase in salaries of employees with increase in PAST EXPERIANCE they have.

-> But in some departments like Director in Operations UNIT etc. there is a linear increase in salary with increase in Past Experience.

DATA PREPROCESSING

APPROACH :

- ➔ I have imputed the blank space in columns like AGE , RATINGS etc. with zero .
- ➔ It seems that features like FIRST NAME , LAST NAME , DOJ , CURRENT DATE doesn't helps in learning , and can increase the complexity of model , hence I have dropped these columns .
- ➔ I have encoded the columns which have categorical values like UNIT etc. using One Hot Encoder .

MODEL DEVELOPMENT

KEY FINDINGS

LINEAR REGRESSION:

➔ R2 Square score on Test Set is : 0.962
➔ MSE on Test Set is : 0.038
➔ R2 Square score on Train Set is : 0.952
➔ MSE on Train Set is : 0.047

DECISION TREE REGRESSOR:

➔ R2 Square score on Test Set is : 0.922
➔ MSE on Test Set is : 0.079
➔ R2 Square score on Train Set is : 0.999
➔ MSE on Train Set is : 0.0007

RANDOM FOREST REGRESSOR:

➔ R2 Square score on Test Set is : 0.958
➔ MSE on Test Set is : 0.042
➔ R2 Square score on Train Set is : 0.990
➔ MSE on Train Set is : 0.009

XGBOOST REGRESSOR:

➔ R2 Square score on Test Set is : 0.952
➔ MSE on Test Set is : 0.048
➔ R2 Square score on Train Set is : 0.981
➔ MSE on Train Set is : 0.018

➔ By looking at the stats of different algorithm above , I have concluded that Random Forest Regression is the best algorithm among all for this problem .

MODEL DEPLOYMENT

➔For Model Deployment , I have used Streamlit Framework.

➔ Here is a screenshot of the FrontEnd to predict salary.