Presented by Harsh Deep Pandey

SALARY PREDICTION

INTRODUCTION

- This project predicts **employee salary** based on features such as Age, Gender, Degree, Job Title, and Experience.
- The main objective is to **build a regression model** that can accurately estimate salaries using historical data.
- Helps companies automate salary estimation and create fair salary structures.

TOOLS & TECHNOLOGIES USED

- **Programming Language:** Python
- **Libraries:**
- pandas
- numpy
- matplotli
- scikit-learn
- **ML Algorithm:** Linear Regression
- **Environment:** Jupyter Notebook

DATASET DESCRIPTION

The notebook loads a file Dataset.csv having the columns:

- Age
- Gender
- Degree
- Job_Title
- Experience_years
- Salary

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0

Dataset operations performed:

- Viewing first rows
- Checking shape
- Renaming columns
- Checking and converting data types

	Age	Gender	Degree	Job_Title	Experience_years	Salary	Gender_Encoded	Degree_Encoded	Job_Title_Encoded	Age_scaled	Experience_years_scaled
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0	1	0	159	-0.750231	-0.761821
1	28.0	Female	Master's	Data Analyst	3.0	65000.0	0	1	17	-1.307742	-1.063017
2	45.0	Male	PhD	Senior Manager	15.0	150000.0	1	2	130	1.061680	0.744158
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0	0	0	101	-0.192720	-0.460625
4	52.0	Male	Master's	Director	20.0	200000.0	1	1	22	2.037324	1.497148
...	...	...	...	...	...	...	...	...	...	...	...
348	28.0	Female	Bachelor's	Junior Operations Manager	1.0	35000.0	0	0	68	-1.307742	-1.364212
349	36.0	Male	Bachelor's	Senior Business Development Manager	8.0	110000.0	1	0	111	-0.192720	-0.310027
350	44.0	Female	PhD	Senior Data Scientist	16.0	160000.0	0	2	115	0.922302	0.894756
351	31.0	Male	Bachelor's	Junior Marketing Coordinator	3.0	55000.0	1	0	63	-0.889609	-1.063017
371	43.0	Male	Master's	Director of Operations	19.0	170000.0	1	1	30	0.782924	1.346550

324 rows × 11 columns

MODEL BUILDING

✓ Linear Regression model

I have used:

```
from sklearn.linear_model import LinearRegression
```

Steps:

- Train model on X_train, y_train
- Predict salary on test data
- Compare predicted vs actual values

✓ Performance metrics

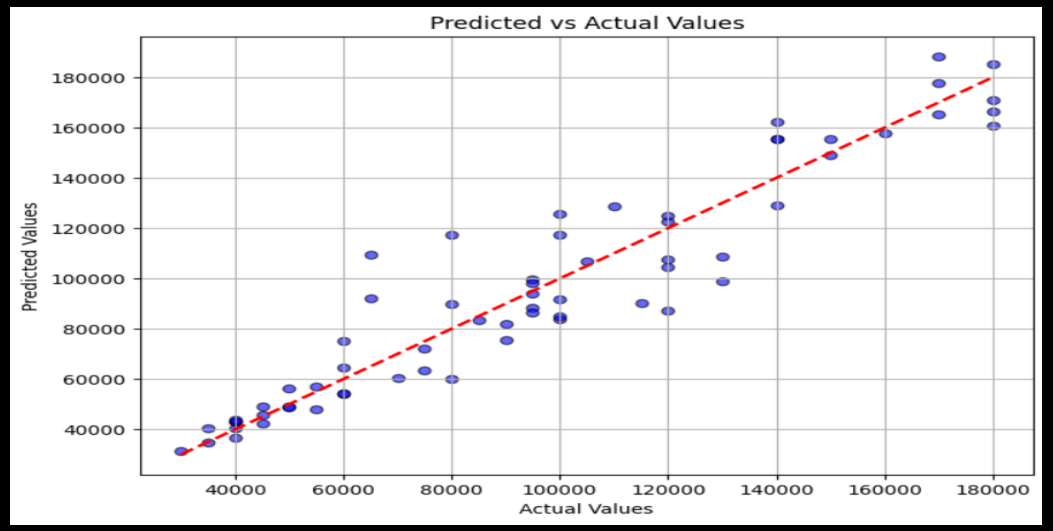
- Mean Absolute Error (MAE)
- R^2 Score

PROJECT WORKING — OUTPUT (SAMPLE SCREENSHOTS)

Dataset Head

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0

Model Accuracy



APPLICATIONS

- Useful for **HR departments** for salary benchmarking.
- Helps companies create **data-driven salary structures**.
- Can be used in **job portals** to estimate expected salary.
- Helps freshers and employees understand the **market salary trend**.

PROBLEMS FACED & SOLUTIONS

Problem Faced	Solution Implemented
Categorical data couldn’t be used directly	Applied Label Encoding (`cat.codes`)
Dataset imbalance for job titles	Ensured proper cleaning and preprocessing
Prediction errors initially high	Tuned model, cleaned data, removed noise
Plot visibility issues	Used matplotlib for clearer graphs

FUTURE SCOPE

- Add **multiple ML models** like Random Forest, XGBoost for better accuracy.
- Deploy as a **web application using Flask/Streamlit**.
- Include more salary-related features (City, Company Size, Skills).
- Build a **salary recommendation dashboard**.

PROJECT LINK

GitHub Repo: *https://github.com/Harsh-bot-1603/Salary_Prediction-Using-Linear-Regression.git*