

Salary Prediction using Linear & Polynomial Regression

1. Introduction

The objective of this project is to predict employee salaries based on multiple factors such as years of experience, age, gender, education level, and job title. Two regression models were implemented: Linear Regression and Polynomial Regression, to identify whether nonlinear relationships improve prediction accuracy.

2. Dataset Overview

The dataset includes the following features:

- Age
- Gender
- Education Level
- Job Title
- Years of Experience

Target variable: Salary

3. Data Cleaning and Preprocessing

Missing values were handled using the median for numerical columns. Categorical columns were encoded using label encoding to convert textual information into numeric format. The dataset was then split into training and testing subsets for model evaluation.

4. Model Training

Two regression models were trained:

1. Linear Regression – to model linear relationships between features and salary.
2. Polynomial Regression (degree=2) – to capture potential nonlinear relationships between features and salary.

5. Model Performance

☑ Linear Regression Results:

- R^2 Score: 0.6715
- MAE: 24,559.08

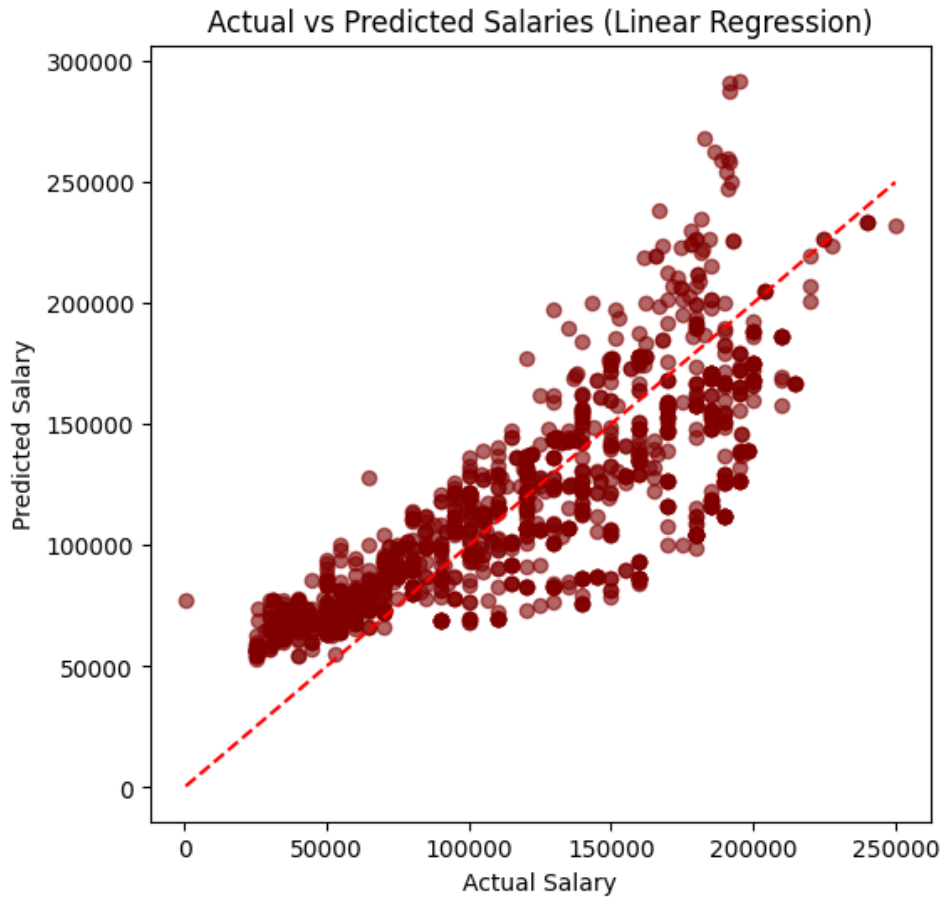
☑ Polynomial Regression Results:

- R^2 Score: 0.8626
- MAE: 14,957.27

The Polynomial Regression model achieved a significantly higher R^2 score and lower MAE, indicating that incorporating nonlinear relationships substantially improved prediction accuracy.

6. Visualization

The following figure illustrates the relationship between actual and predicted salaries for the Linear Regression model:



7. Conclusion

Linear regression provided a moderate prediction accuracy, explaining approximately 67% of the salary variation. However, the polynomial model (degree 2) captured more complex patterns in the data, improving the R^2 to 86%. This suggests that salary prediction benefits from modeling nonlinear interactions between experience, age, and education.