Salary Prediction Using Linear Regression

A comprehensive machine learning project that demonstrates how supervised learning algorithms can predict employee compensation based on experience and demographic factors.
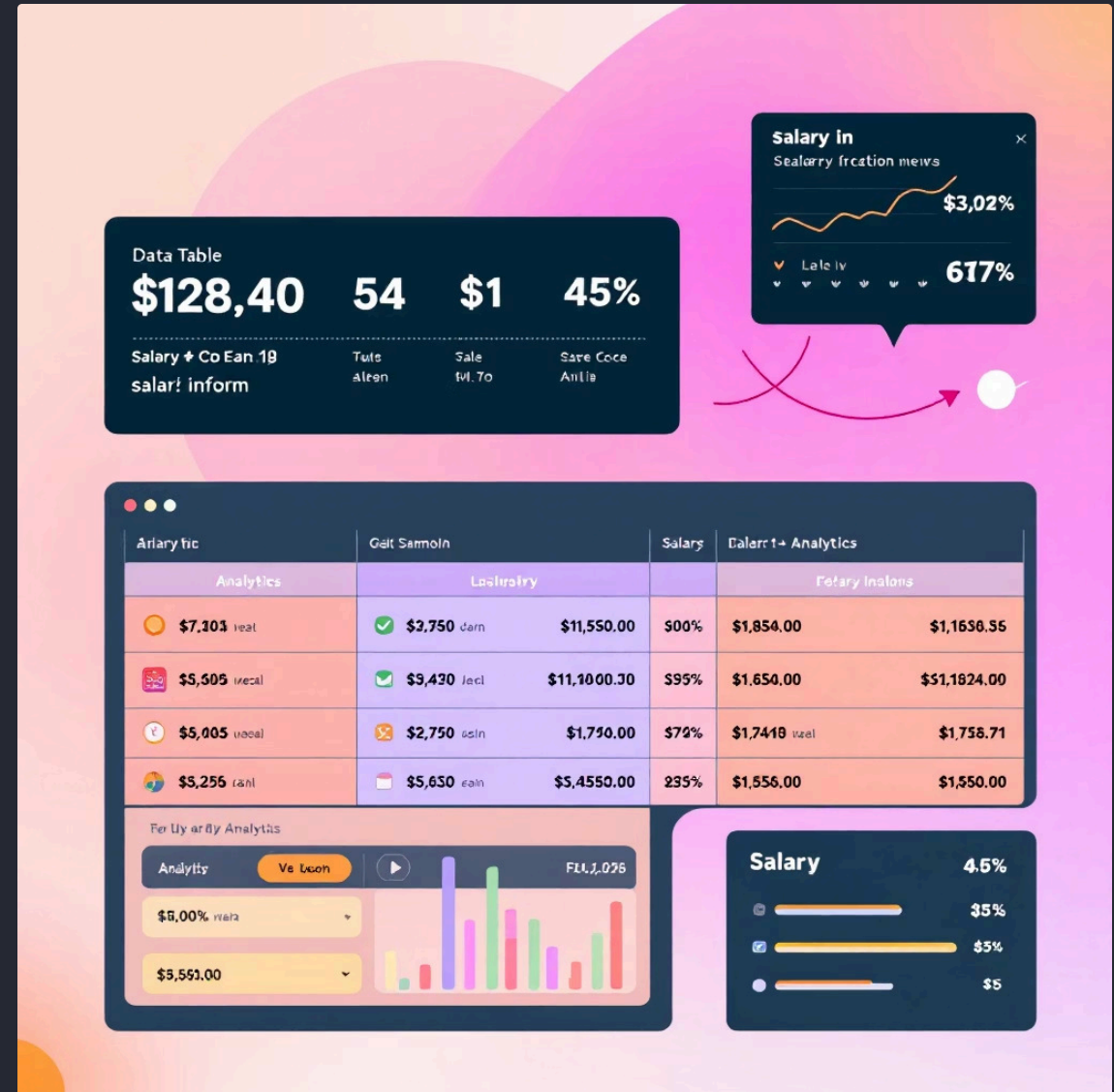
Project Overview and Dataset

Dataset Characteristics

We created a realistic dataset comprising 50 employee records with three key variables:

- Years of Experience

- Age

- Salary (target variable)

The data reflects real-world patterns where compensation typically increases with professional experience and employee maturity.

Implementation Workflow

01
Load Dataset

Import CSV file using pandas library for data manipulation

02
Feature Selection

Separate input features (YearsExperience, Age) from target variable (Salary)

03
Train-Test Split

Divide data into 80% training set and 20% testing set for model validation

04
Model Training

Train Linear Regression model using scikit-learn framework

05
Evaluation

Measure model accuracy using multiple metrics (R^2, MAE, MSE, RMSE)

06
Prediction & Visualisation

Generate predictions and create scatter plots with regression line

Model Evaluation Metrics

R² Score

Coefficient of determination measures how well the model explains variance in salary data. Values closer to 1 indicate better fit.

Mean Absolute Error (MAE)

Average absolute difference between predicted and actual salaries, providing intuitive error measurement.
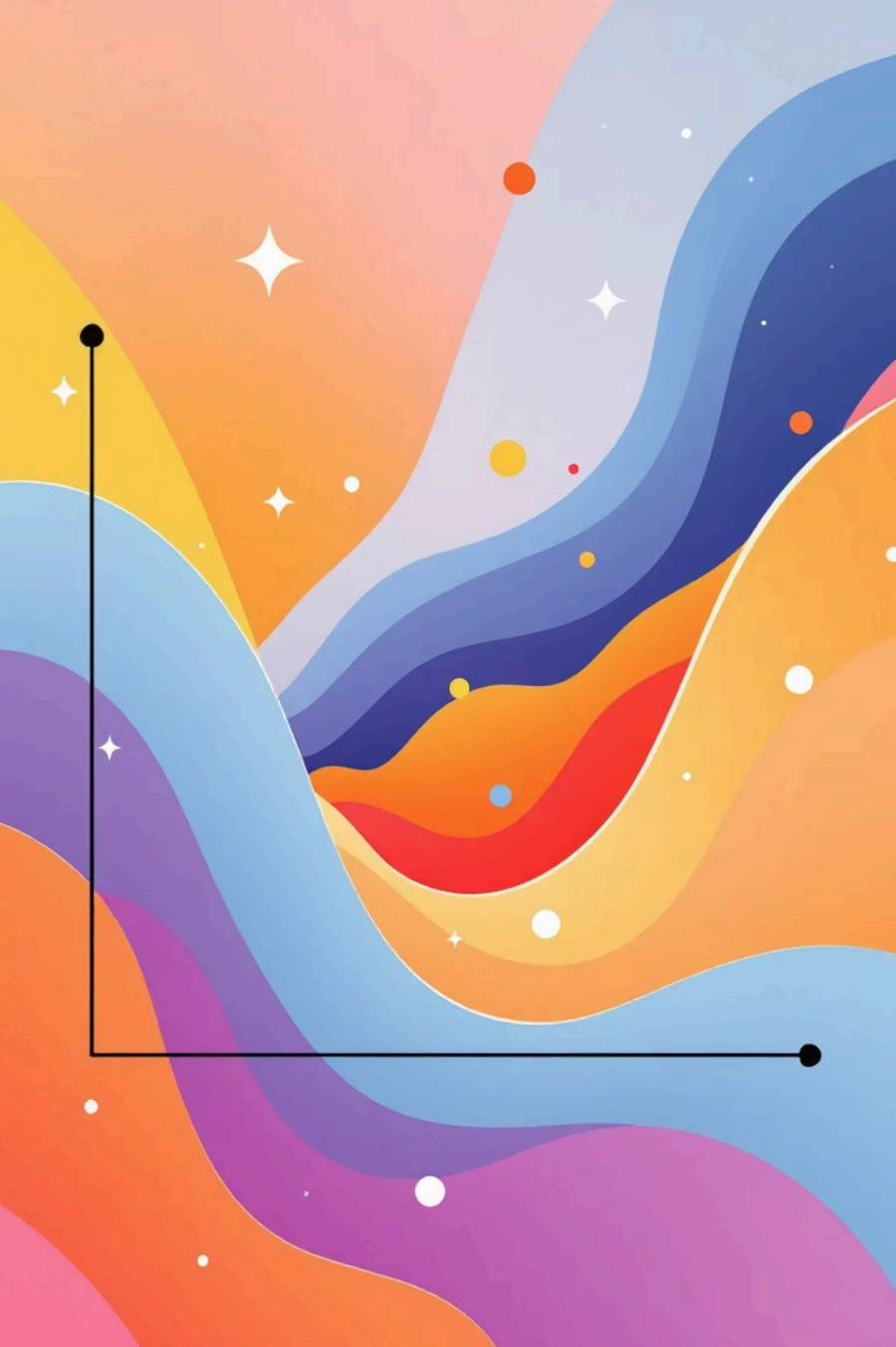
Mean Squared Error (MSE)

Average of squared differences, emphasising larger prediction errors more heavily.

Root Mean Squared Error (RMSE)

Square root of MSE, expressed in same units as salary for easier interpretation.

Visualising the Results

The scatter plot visualisation displays actual salary data points alongside the fitted regression line, clearly demonstrating the linear relationship between years of experience and compensation. This visual representation helps validate model performance and communicate findings effectively to stakeholders.

Key Takeaways and Future Directions

✓ Successful Implementation

Linear Regression effectively models the relationship between employee experience and salary, demonstrating core supervised learning concepts.

📊 Complete ML Workflow

The project showcases the entire machine learning pipeline from data loading through model evaluation and visualisation.

🚀 Expansion Opportunities

Consider enhancing the model with additional features, exploring advanced algorithms like Random Forest or Gradient Boosting, or implementing cross-validation techniques.