

Multiple Linear Regression Analysis

Repository Overview

This repository contains the analysis of a dataset using Multiple Linear Regression. The notebook demonstrates the process of data preparation, exploratory data analysis (EDA), model training, and evaluation.

Introduction

In this project, we explore a dataset to predict income based on features such as age and experience. Multiple Linear Regression is employed to understand how these features collectively influence the target variable, income. The objective is to build a predictive model and evaluate its performance.

Dataset

The dataset consists of three columns:

- **age**: The age of individuals.
- **experience**: The years of experience.
- **income**: The income of individuals.

Data Source

The dataset is sourced from Kaggle: [Multiple Linear Regression Dataset](#).

Data Exploration

- **Initial Analysis:**
 - Explored dataset structure and data types.
 - Checked for missing values and outliers.
- **Summary Statistics:**
 - Provided descriptive statistics of the features and target variable.

Exploratory Data Analysis (EDA)

- **Visualization:**
 - **Scatter Plots**: Visualized relationships between features and income.

- **Distribution Plots:** Examined distributions of age, experience, and income.
- **Correlation Analysis:** Analyzed correlations between features to understand their relationships.

Data Preprocessing

- **Outlier Detection and Handling:**
 - Used statistical methods to identify and manage outliers.
- **Feature Selection:**
 - Selected relevant features (age and experience) for the model.

Model Selection and Training

- **Model Used:** Multiple Linear Regression
 - Trained the model using the training dataset.
 - Evaluated the model on the test dataset.

Model Evaluation

- **Performance Metrics:**
 - **Mean Squared Error (MSE):** 367,468.68
 - **R-squared (R^2):** 0.99
 - The model demonstrated high accuracy with a very low MSE and a high R^2 value, indicating a strong fit to the data.

Visualizations

- **Actual vs Predicted Income:**
 - Plotted to compare predicted values against actual values for both age and experience.

Conclusion

The Multiple Linear Regression model effectively captures the relationship between age, experience, and income. With an R^2 of 0.99, the model explains nearly all the variance in the income variable, and the low MSE indicates accurate predictions. This model is well-suited for predicting income based on the provided features.