■ House Rent Prediction Using Multivariate Linear Regression

1. Introduction

In the modern housing market, rental prices are influenced by numerous factors such as city, number of rooms, furniture status, and insurance costs. With the availability of large datasets and advancements in machine learning, it has become possible to model and predict these prices using statistical techniques. This project applies Multivariate Linear Regression to predict house rent amounts based on several independent variables using the Brazilian Houses Dataset. By training the model on historical housing data, we aim to estimate rent prices more accurately and identify which features contribute most to rental cost variation.

2. Problem Statement

The task is to predict the monthly rent amount (y) for a given house using various measurable attributes such as city, number of rooms, bathrooms, parking spaces, fire insurance, and furniture status (independent variables x1, x2, x3, …). Mathematically, the model can be expressed as: $y = \beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \ldots + \beta nxn + \varepsilon$ where y is the Rent amount, x1, x2, …, xn are the features such as rooms, bathrooms, etc., $\beta0$, $\beta1$, …, βn are model coefficients, and ε represents random error. The goal is to build a regression model that minimizes prediction error and provides insight into how each variable influences rent.

3. Aims and Objectives

The key aims and objectives of this project are: 1. To preprocess the housing dataset — clean data, handle missing values, and encode categorical features. 2. To build a multivariate linear regression model that predicts rent amount using features such as city, rooms, bathrooms, parking spaces, fire insurance, and furniture type. 3. To evaluate model performance using metrics such as R^2 (accuracy) and Mean Squared Error (MSE). 4. To visualize the relationship between predicted and actual rent values through scatter plots and regression lines. 5. To analyze results and discuss the impact of various factors on rent prediction accuracy.

4. Results and Discussion

After training and evaluation, the model produced the following key outputs: - Model Coefficients: Represent how each feature impacts rent amount. - Intercept: The baseline rent when all other variables are zero. - R^2 Accuracy: Indicates how well the model explains rent variation. - Mean Squared Error (MSE): Represents the average squared difference between actual and predicted rent values. The scatter plot of actual vs predicted rent values showed that most points lie close to the ideal red line, indicating that the model’s predictions are fairly accurate. Some deviations exist, which may be due to unaccounted factors such as neighborhood quality, property condition, or seasonal variations. This confirms that multivariate linear regression is effective for identifying linear relationships between housing features and rent prices.

5. Conclusion

This project successfully demonstrated the use of Multivariate Linear Regression to predict house rent prices using real-world data. The model achieved a strong correlation between predicted and actual rent values, confirming the influence of factors such as city, number of rooms, parking spaces, and furniture status. Key takeaways include: - The linear regression model can effectively estimate rental prices using multiple input features. - Data preprocessing (especially cleaning and encoding) is essential for accurate predictions. - Visualization tools such as scatter plots help validate model accuracy visually. Future improvements could involve adding more features (e.g., property area, neighborhood quality), trying non-linear models like Random Forest or Gradient Boosting, and incorporating time-based or economic factors to reflect market trends.