Logo

Regression Engine

Model Report

Overview

In this analysis, multiple linear regression was performed to examine the relationship between the dependent variable and several independent variables. This type of regression was chosen because it allows for the quantification of the individual impact of each predictor on the target variable while controlling for the effects of others. The model assumes a linear relationship between the predictors and the response, and it was trained on a preprocessed dataset with appropriate scaling and handling of missing values. The regression coefficients provide insights into the direction and magnitude of each feature's influence, while statistical metrics such as R^2, adjusted R^2, and p-values were used to assess the model's explanatory power and the significance of each predictor.

Target

- Price

Features

- X
- Y
- Z
- table
- carat
- clarity

Regression Equation

- 1.701
- 0.672(carat)
- 0.279(cut_Ideal)
- 0.323(cut_Premium)
- 0.347(cut_Good)
- 0.310(cut_Very Good)
- 0.443(cut_Fair)
- 0.091(color_E)
- 0.418(color_I)
- 0.556(color_J)
- 0.291(color_H)
- 0.124(color_F)
- 0.191(color_G)
- 0.031(color_D)
- 0.481(clarity_SI2)
- 0.313(clarity_SI1)
- 0.104(clarity_VS1)
- 0.169(clarity_VS2)
+ 0.019(clarity_VVS2)
+ 0.087(clarity_VVS1)
- 0.920(clarity_I1)
+ 0.180(clarity_IF)
+ 0.050(depth)
+ 0.009(table)
+ 1.159(x)
+ 0.024(y)
+ 0.108(z)

Assumption Check

1. Linearity

Test Result: failure

Feature(s) without a significant relationship:

const

clarity_SI1

clarity_VS1

clarity_VS2

clarity_VVS2

clarity_VVS1

clarity_I1

clarity_IF

cut_Ideal

cut_Premium

cut_Good

cut_Very Good

cut_Fair

2. Independence of Errors

Test Result: success
Durbin Watson Stat: 1.99

Threshold: 1.5 to 2.5

3. Normality of Errors

Result: failure
Jarque-Bera Stat: 0.0
Threshold p-value: 0.05

4. No Perfect Multicollinearity

Result: failure
High VIF Features: 13
Threshold VIF: between 5 to 10

5. Equal Variance of Errors (Homoscedasticity)

Result: failure
Breusch-Pagan Stat: 0.0
Threshold p-value: 0.05

Model Evaluation

- Mean Absolute Error (MAE): **0.1102** **(lower is better)**
- Mean Squared Error (MSE): **0.0372** **(lower is better)**
- Root Mean Squared Error (RMSE): **0.1929** **(lower is better)**
- R^2 Score: **0.9633** **(higher is better)**
- Adjusted R^2 Score: **0.9633** **(should be close to R^2)**

Conclusion

Based on the provided metrics, the model demonstrates **excellent** explanatory power, with an R^2 score of **0.9633**.

The prediction errors — Mean Absolute Error (MAE) = **0.1102** and Root Mean Squared Error (RMSE) = **0.1929** — are considered **minimal** based on common thresholds.

The adjusted R^2 score of **0.9633** being close to R^2 indicates that the model includes relevant features and is not overfitting.