**1. Root Mean Square Error (RMSE)**

**Definition:** RMSE is a measure of the differences between the predicted values generated by a model and the actual values observed. It is the square root of the average of the squared differences between predicted and actual values.

**Formula:**

RMSE  **=**

where is the actual value, is the predicted value, and n is the number of observations.

**Interpretation:**

RMSE gives a sense of how far the predictions are from the actual values in the same units as the target variable.

A lower RMSE indicates a better fit of the model to the data.

**2. Standard Deviation (SD)**

**Definition:** Standard Deviation measures the amount of variation or dispersion in a set of values. In machine learning, it can describe the spread of the data points around the mean, or the variation of the errors (residuals) in predictions.

**Formula:**

SD **=**

where represents individual data points and is the mean of the data points.

**Interpretation:**

A higher standard deviation means the data points are spread out over a wider range, indicating greater variability.

In the context of model residuals, a smaller SD indicates that the model's predictions are more consistent.

**3. R2 (R-squared)**

**Definition:** R2 is a statistical measure that indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It is also known as the coefficient of determination.

**Formula:**

R2 **=**

where SSres is the sum of squares of residuals, and SStot is the total sum of squares (proportional to the variance of the data).

**Interpretation:**

R2 values range from 0 to 1.

An R2 value of 0 indicates that the model does not explain any of the variability of the response data around its mean.

An R2 value of 1 indicates that the model explains all the variability of the response data around its mean.

A higher R2 value indicates a better fit of the model to the data.

**Usage in Machine Learning**

**RMSE** is commonly used in regression analysis to assess how well a model predicts a continuous outcome.

**Standard Deviation** is often used to understand the distribution of data or residuals, which can inform model performance and stability.

**R2** is used to measure the goodness-of-fit of a regression model, helping to understand how well the model explains the variance of the target variable.

These metrics together provide a comprehensive picture of model performance, allowing you to assess accuracy (RMSE), data spread (SD), and the model's explanatory power R2.

**Linear regression:** It is a type of machine-learning algorithm more specifically a supervised machine-learning algorithm that learns from the labelled datasets and maps the data points to the most optimized linear functions. which can be used for prediction on new datasets.

In our model the RMSE of linear regression is 13,795.35 that means the difference between predicted and observed values is 13,795.35. Also, this indicates that the model’s predictions deviate from the actual values by this amount.

The standard deviation of linear regression is 24,424.65 which represents the dispersion of the selling price values in our dataset. It's a measure of how spread out the values are around the mean. The standard deviation being higher than the RMSE suggests that the model is somewhat effective in reducing the prediction error compared to the variability in the data.