**1. Problem Statement:** The problem involves predicting a continuous target variable based on several input features. This type of problem is common in fields like finance (predicting stock prices), real estate (predicting house values), or health (predicting disease progression or patient outcomes). The core goal is to build a model that can accurately predict future outcomes using historical data.

**2. About the Dataset:** The dataset consists of multiple features (independent variables) and one continuous target (dependent variable). Before applying machine learning models, the dataset underwent preprocessing steps such as handling missing values, feature scaling (if needed), and splitting into training and testing sets. The quality, size, and relevance of the dataset are crucial for developing a robust regression model.

**3. Application of Regression:** Regression analysis is suitable for solving problems where the target variable is continuous. It helps in understanding relationships between features and predicting outcomes. By fitting a mathematical model to the data, we can estimate how changes in inputs affect the output. This can support data-driven decision-making.

**4. Algorithms Used:**

* **Linear Regression:**
  + Linear Regression assumes a linear relationship between input features and the target.
  + It is simple, interpretable, and computationally efficient.
  + It works well when the data shows linear trends but struggles with non-linear patterns.
* **Random Forest Regressor:**
  + An ensemble learning method that builds multiple decision trees and averages their predictions.
  + It handles non-linear relationships, interactions between features, and is robust to overfitting.
  + It can model complex patterns in data without the need for feature scaling or assumptions about data distribution.

**5. Handling the Problem Individually:**

* **Linear Regression:**
  + Tries to find the best-fitting straight line through the data points.
  + Sensitive to outliers and multicollinearity.
  + Provides clear insights into feature importance via coefficients.
* **Random Forest:**
  + Build multiple decision trees on random subsets of data.
  + Aggregates predictions, which reduces variance and improves generalization.
  + Naturally captures feature interactions and handles high-dimensional datasets better.

**6. Comparison of Results:**

|  |  |  |
| --- | --- | --- |
| **Metric** | **Linear Regression** | **Random Forest Regressor** |
| RMSE | Higher | Lower |
| R² Score | Moderate | Higher |
| Residual Plot | More spread | More centered |

* **Interpretation:**
  + Random Forest outperforms Linear Regression in prediction accuracy and error distribution.
  + While Linear Regression is easier to interpret, it lacks flexibility.
  + Random Forest, though less interpretable, provides better results due to its ability to model complex relationships.

**7. Conclusion:** For this dataset and problem, Random Forest is the preferred choice due to its superior performance in capturing non-linear patterns and providing more accurate predictions. However, Linear Regression remains valuable for its simplicity and transparency, especially when model interpretability is a priority.