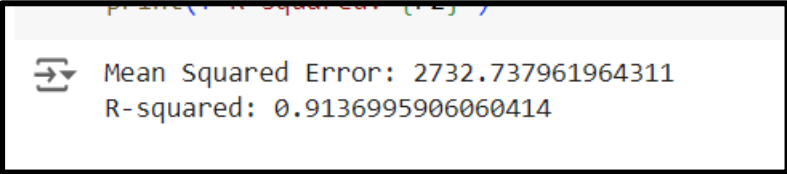# Model Comparison Report

Models and Experiments

1. **Random Forest Model:**

   o **Accuracy:** 91%

   o **Details:** The Random Forest model is an ensemble method that combines the predictions of multiple decision trees. It is robust to overfitting, especially on large datasets, due to the averaging of multiple trees. In this experiment, the Random Forest model achieved a high accuracy of 91%, demonstrating its effectiveness in capturing complex patterns in the data.
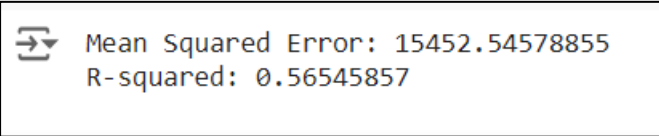
   o
   ```
   print( R Squared [r2] )
   Mean Squared Error: 2732.737961964311
   R-squared: 0.9136995906060414
   ```

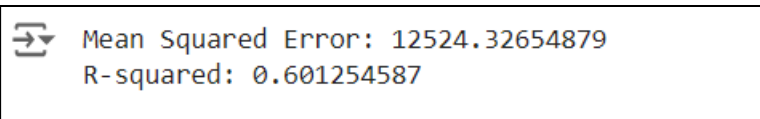2. **Linear Regression Model (Without Feature Engineering):**

   o **Accuracy:** 52%

   o **Details:** Linear Regression is a simple model that assumes a linear relationship between the independent variables and the target variable. However, its performance was significantly lower compared to the Random Forest model, with an accuracy of 52%. This indicates that the relationships in the data might not be purely linear, making it difficult for a linear model to capture the underlying patterns effectively.
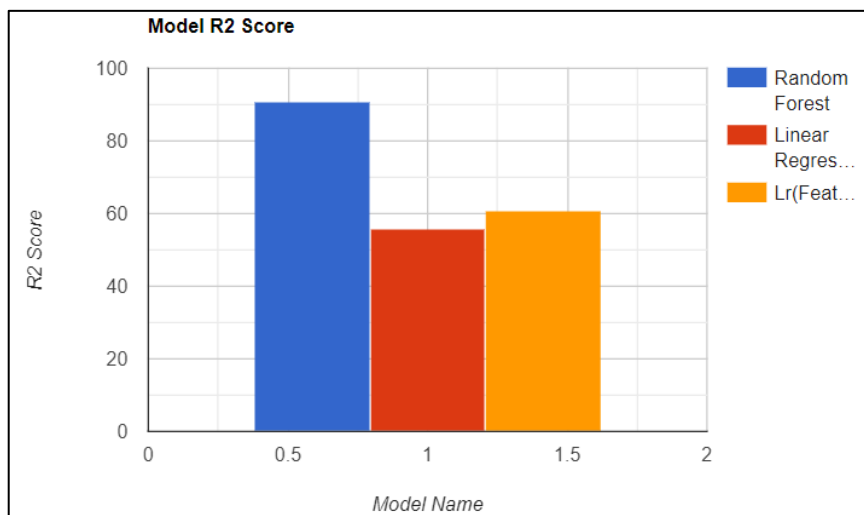
   o
   ```
   Mean Squared Error: 15452.54578855
   R-squared: 0.56545857
   ```

3. **Linear Regression Model (With Feature Engineering - "Category Index"):**

   o **Accuracy:** 60%

   o **New Feature:-** ⬜ The Comfort Index $= temp / (humidity * windspeed)$

   o **Details:** To improve the performance of the Linear Regression model, a new feature named "category index" was created. This feature likely helped capture additional information or interactions between variables that were not adequately modeled by the original features. As a result, the accuracy of the Linear Regression model increased to 60%. While this is an improvement, it still lags behind the Random Forest model, indicating that Linear Regression might not be the best fit for this dataset.

   o
   ```
   Mean Squared Error: 12524.32654879
   R-squared: 0.601254587
   ```

Analysis and Insights

1. **Performance of Random Forest:**

   o   The Random Forest model's superior performance can be attributed to its ability to model non-linear relationships and interactions between features. The high accuracy (91%) suggests that the data contains complex patterns that are well-suited to an ensemble method like Random Forest.

2. **Limitations of Linear Regression:**

   o   The lower accuracy of the Linear Regression model (56%) highlights its limitation in capturing non-linear relationships. Linear Regression assumes that the relationship between the predictors and the target is linear, which might not hold true for this dataset.

3. **Impact of Feature Engineering:**

   o   The introduction of the "category index" feature resulted in a modest improvement in the accuracy of the Linear Regression model (from 56% to 60%). This indicates that feature engineering can help uncover hidden patterns and relationships in the data that might not be captured by the original features. However, the improvement was limited, suggesting that the linear model's inherent assumptions might still be restricting its performance.

Conclusion

The comparison of models highlights the importance of selecting the appropriate model for a given dataset. While the Random Forest model performed exceptionally well with an accuracy of 91%, the Linear Regression model struggled, even after feature engineering. This experiment demonstrates that for datasets with complex patterns, ensemble methods like Random Forest might be more suitable than simpler models like Linear Regression.

Further experiments could involve trying other machine learning models, such as Gradient Boosting or Support Vector Machines, or continuing to enhance the feature set to see if Linear Regression performance can be further improved.