Assesment_2_Report.docx

by Hamid Raza

Submission date: 07-May-2024 12:26PM (UTC+0100)
Submission ID: 232504425
File name: Assesment_2_Report.docx (630.96K)
Word count: 276
Character count: 1664

Intelligent Systems and Machine learning

Assessment 2

Student Name: Hamid Raza

Student ID: C7473166

Task 1: Problem overview and Data Visualization

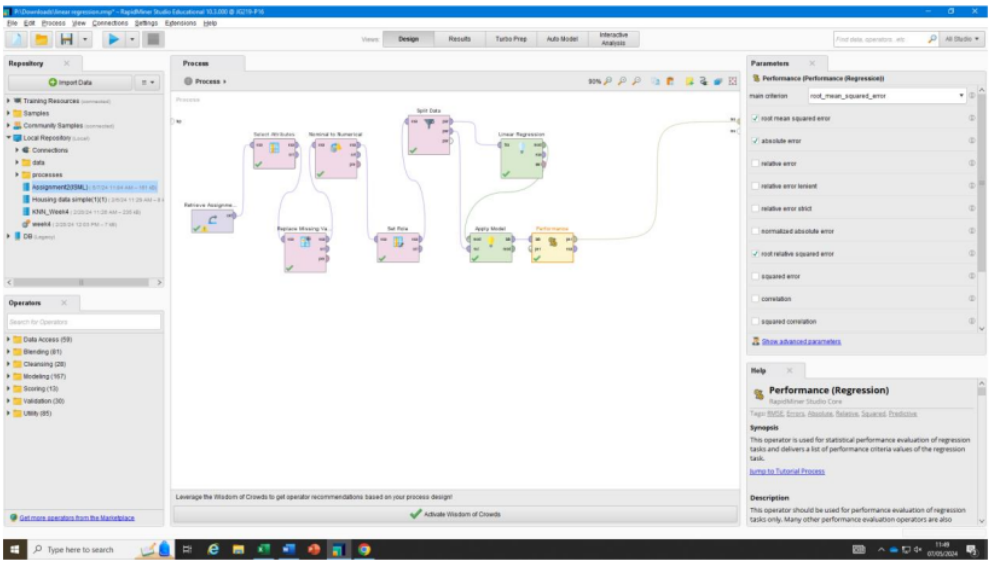
Objective

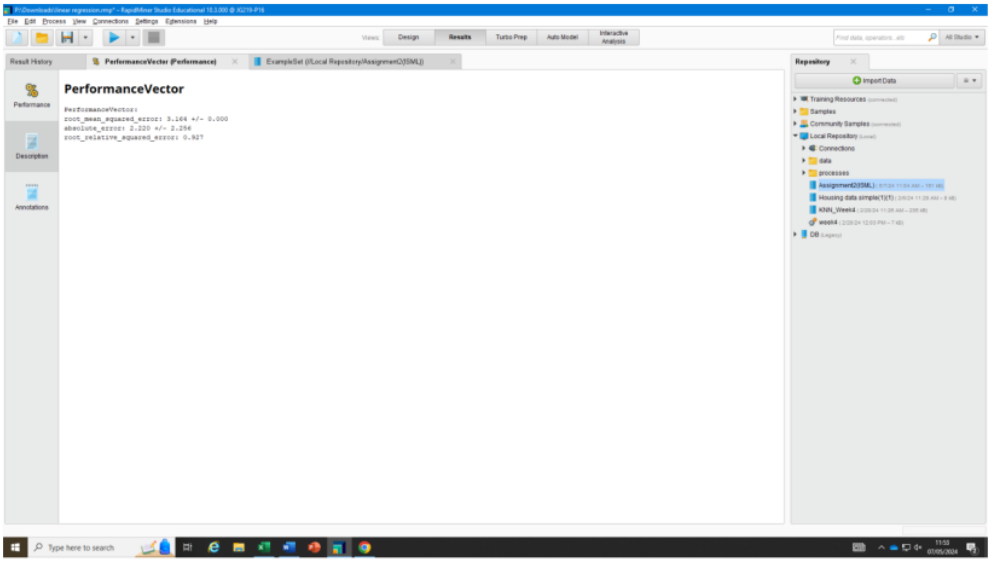
- The purpose of this study is to predict students' final grades (G3) using a range of factors. The goal is to determine the best effective prediction model by comparing Linear Regression with Decision Tree Regression.

TASK 2: Data Pre-processing Techniques Used for Handling Missing Values

- Missing values were resolved to ensure that the prediction models are robust. To ensure data integrity, techniques such as imputation were used when needed.

Task 3. Technique 1: Linear Regression

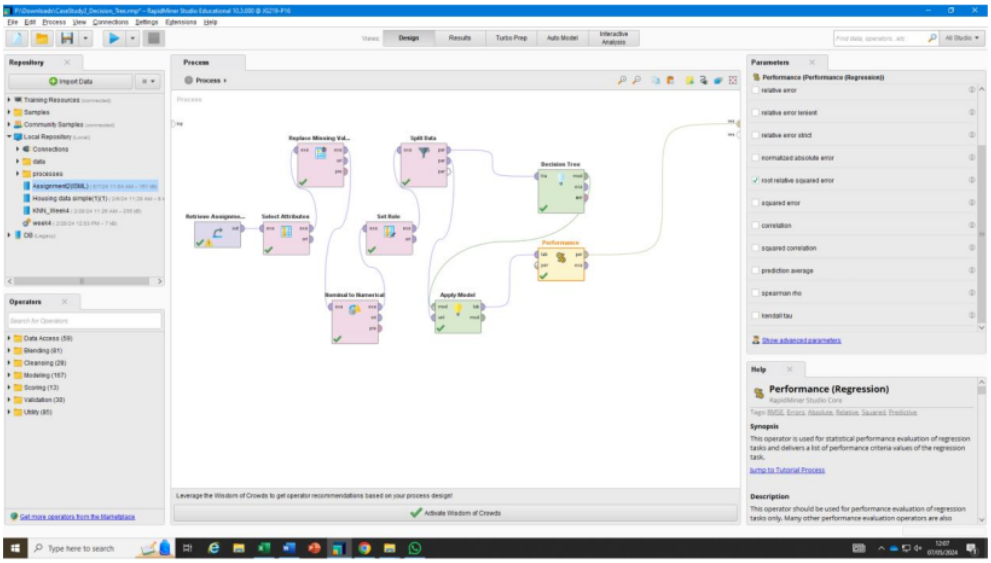


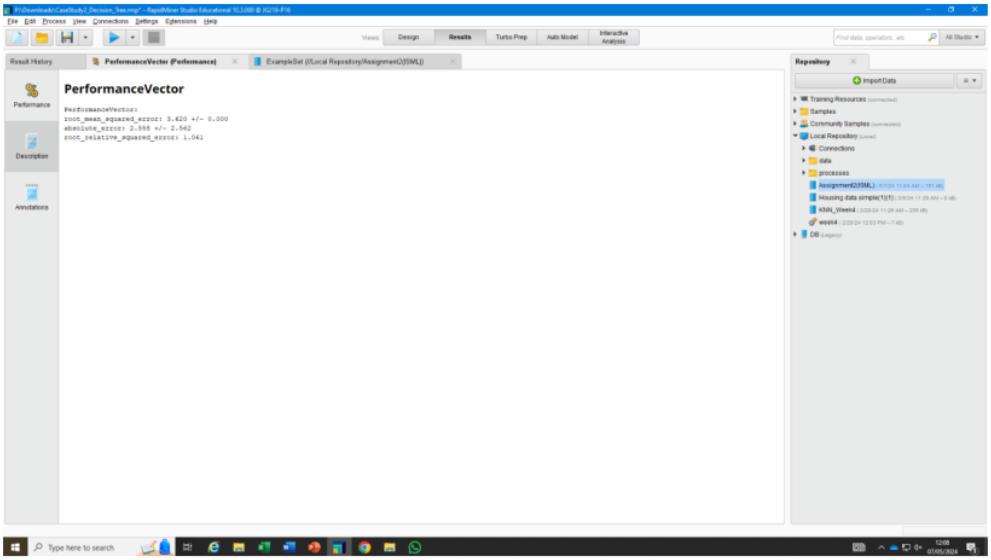


Motivation for Using Linear Regression

- Linear Regression was selected for its simplicity and interpretability, acting as a baseline for analysing linear relationships within the data.

Task 4. Technique 2: Decision Tree Regression.





Motivation for Using Decision Tree Regression

- Decision Tree Regression was chosen for its capacity to represent non-linear correlations and interactions between variables, providing a more flexible approach than linear techniques.

Task 5: Comparative Analysis and Summary. Conclusion: Performance Comparison.

Linear Regression performed marginally better than the Decision Tree model, with lower RMSE and Root Relative Squared Error values. This suggests a better match to the data, despite the simpler model structure.

The greater errors in Decision Tree Regression may indicate overfitting or that the model complexity did not convert into improved predictions for this dataset.

Conclusion

The results show that Linear Regression offered a more accurate and dependable prediction for the final grades (G3) under the conditions examined. While Decision Tree Regression provided insights into more complicated patterns, it underperformed on this particular dataset, emphasising the relevance of model selection depending on dataset features and prediction goal.