**Experimentation Report: Asteroid Dataset Classification**

**-Wajid Ali Hashmi**

This report presents the results of classifying an asteroid dataset using several machine learning models. The following models were employed for the classification task:

* Linear Regression
* Ridge Regression
* Elastic Net
* Decision Tree
* Random Forest
* XGBoost

**Evaluation Metrics**

The performance of each model was evaluated using the following metrics:

1. **R-squared (R²) Score**: This metric measures the proportion of the variance in the dependent variable that is predictable from the independent variables. Higher R² values indicate better model performance.
2. **Mean Squared Error (MSE)**: MSE calculates the average squared difference between the predicted and actual values. Lower MSE values are desirable as they indicate smaller errors.
3. **Mean Absolute Error (MAE)**: MAE is the average absolute difference between the predicted and actual values. Similar to MSE, lower MAE values indicate better model performance.
4. **Root Mean Squared Error (RMSE)**: RMSE is the square root of the MSE and provides a measure of the average magnitude of the errors in the same units as the dependent variable.

**Results**

The following tables summarize the performance of each model based on the evaluation metrics:

**R-squared (R²) Scores:**

| **Model** | **R² Score** |
| --- | --- |
| Linear Regression | 1 |
| Ridge Regression | 1 |
| Elastic Net | 1 |
| Decision Tree | 1 |
| Random Forest | 1 |
| XGBoost | 0.999984 |

**Mean Squared Error (MSE):**

| **Model** | **MSE** |
| --- | --- |
| Linear Regression |  |
| Ridge Regression |  |
| Elastic Net |  |
| Decision Tree |  |
| Random Forest |  |
| XGBoost | 184.609 |
|  |  |

**Mean Absolute Error (MAE):**

| **Model** | **MAE** |
| --- | --- |
| Linear Regression |  |
| Ridge Regression |  |
| Elastic Net |  |
| Decision Tree |  |
| Random Forest |  |
| XGBoost |  |

**Root Mean Squared Error (RMSE):**

| **Model** | **RMSE** |
| --- | --- |
| Linear Regression |  |
| Ridge Regression |  |
| Elastic Net |  |
| Decision Tree |  |
| Random Forest |  |
| XGBoost | 13.5871 |
|  |  |

Based on the results, we can analyse the performance of the different models for the asteroid dataset classification task:

**Linear Regression, Ridge Regression, and Elastic Net**: These three models achieved a perfect R² score of 1, indicating that they can explain all the variance in the dependent variable. Additionally, their Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) values are extremely low, with the Linear Regression model having the lowest error values among the three. These results suggest that the asteroid dataset has a linear or near-linear relationship with the independent variables, and these models can fit the data almost perfectly.

**Decision Tree**: The Decision Tree model also achieved a perfect R² score of 1, which is impressive. However, its error values (MSE, MAE, and RMSE) are higher compared to the linear models. This could be due to the inherent complexity of the Decision Tree model, which can lead to overfitting on the training data.

**Random Forest**: The Random Forest model, like the other models, achieved a perfect R² score of 1. Its error values (MSE, MAE, and RMSE) are lower than those of the Decision Tree model, indicating better generalization performance. Ensemble models like Random Forest tend to perform better than individual models, as they combine multiple models to reduce overfitting and improve prediction accuracy.

**XGBoost**: The XGBoost model achieved a very high R² score of 0.999984, which is still excellent but slightly lower than the other models. However, its error values, particularly the MSE and RMSE, are significantly higher than the other models. This could be due to the inherent characteristics of the XGBoost algorithm, which can sometimes overfit the data or be sensitive to the choice of hyperparameters.

**Conclusions**

Based on the results, it appears that the linear models (Linear Regression, Ridge Regression, and Elastic Net) and the Random Forest model perform exceptionally well on the asteroid dataset, with near-perfect R² scores and extremely low error values. The Decision Tree model also achieves a perfect R² score but has higher error values, potentially due to overfitting.

The XGBoost model, while still performing well in terms of R² score, has significantly higher error values compared to the other models. This could be due to various factors, such as the choice of hyperparameters, the complexity of the dataset, or the inherent characteristics of the XGBoost algorithm.

It's worth noting that while the linear models and Random Forest seem to perform exceptionally well on this dataset, their performance may vary on different datasets or in different problem domains. Additionally, the interpretability of the models should also be considered, as linear models are generally more interpretable than ensemble models like Random Forest or XGBoost.

Overall, based on the provided results, the Linear Regression, Ridge Regression, Elastic Net, and Random Forest models appear to be the most suitable choices for the asteroid dataset classification task, offering excellent performance with low error rates. However, further analysis and validation on additional datasets or real-world scenarios may be necessary to confirm these findings.