Task 4 Began with the retrieval of the preprocessed dataset, the type of models used were mainly supervised, because they perform more efficiently with Time-Series data.

The first model used was Linear Regression, a set role as label was applied to the (btc\_market\_price) attribute to identify it as the target of prediction, allowing rapid miner to distinguish between it and the other attributes. a split of 70:30 was Selected to make sure the model had a sufficient portion of the data to learn from while keeping enough data for testing.

For the performance, the Performance (regression) operator was used because it is specifically designed to evaluate regression models by providing essential metrics such as RMSE, Absolute Error, Correlation and **R²**.

Linear Regression showed strong predictive performance, proving to be an appropriate choice for predicting patterns in Time-Series data.

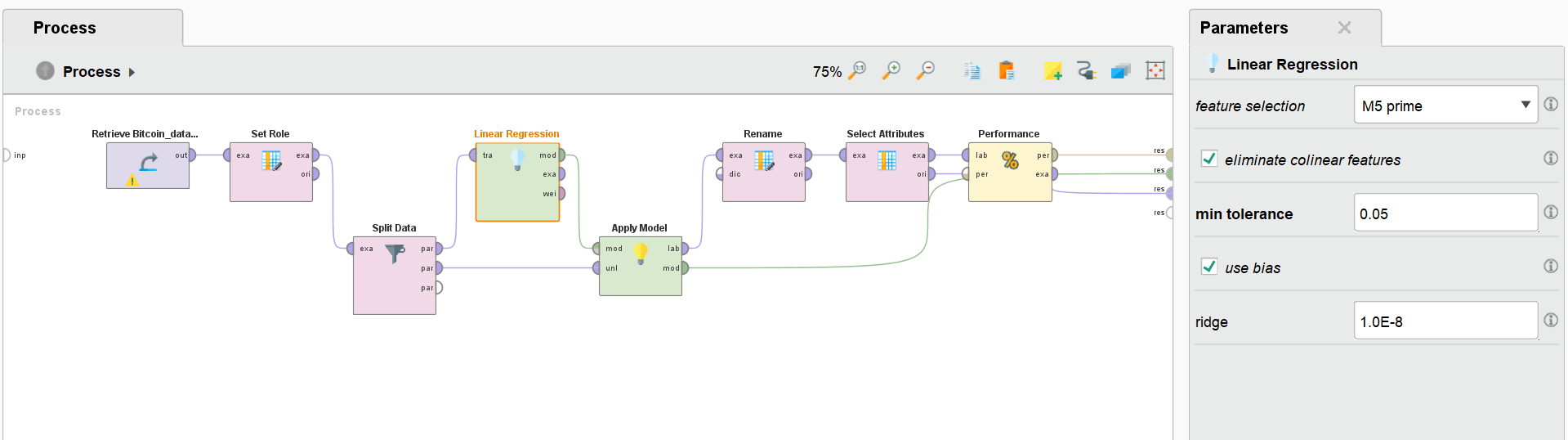


Figure - Process Flow of the Linear Regression Model Including Parameter Configuration.

The next model used was Random Forest. Initially, its results were comparatively weaker, this happened because the algorithm isn’t well suited with time series data and is better at handling regular datasets with features that don’t depend on each other. The model treats each row separately and doesn't remember what came before or after.

A screenshot of a computer

AI-generated content may be incorrect.

Figure - Process Flow of the Random Forest Model Including Parameter Configuration.

To mitigate this issue, multiple tuning strategies were implemented. These methods include adjusting the number of decision trees and enabling the random split option. This was done to increase the models’ diversity and make training quicker.

A graph of a person and person

AI-generated content may be incorrect.A graph of a person and person

AI-generated content may be incorrect.Before tuning, the model showed poor alignment with the target’s trend. However, after adjustments, the performance improved across all evaluation metrics. While it did not match or even get close to the performance results of the Linear Regression model. Tuning the Random Forest model has shown enough improvement to be considered for Ensemble model testing.

Figure - Random Forest Prediction Plot (After tuning)

Figure - Random Forest Prediction Plot (Before tuning)

The original untuned model gives a flat prediction line that does not match the real price trend, which is a clear case of underfitting. After tuning the model tracks the overall trend somewhat better, especially in the early years. But it starts struggling in the following years because of the rapid jump in bitcoins price.