

# Phase 5: Report of the hyperparameter Tuning for XGBoost

**Andres Camilo Viloria Garcia**  
aviloriagarcia00@mylangara.ca  
*Data Analytics*  
*Langara College*  
Vancouver, BC

**Diana Sayuri Ortiz Montes**  
dortizmontes00@mylangara.ca  
*Data Analytics*  
*Langara College*  
Vancouver, BC

## Context

The scope of this step of the process is only focus on the best model now, which is Xgboost. As, we started evaluating what hyperparameter could have more impact in the stock price prediction, we created a table describing all hyperparameters including (see [5.2 XGBoost hyperparameters description.xlsx](#)):

- **Type of Hyperparameter:** This column categorizes the hyperparameters based on their type or category.
- **Hyperparameter Name:** This column specifies the name of the hyperparameter.
- **Default Value:** This column indicates the default value assigned to the hyperparameter if no specific value is provided.
- **Short Definition:** This column provides a brief description or explanation of the hyperparameter.
- **Range of Values and Meaning:** This column defines the valid range of values for the hyperparameter and provides additional context or meaning for those values.

## Clarification

In this phase of the project, it has been determined that the same hyperparameters that were previously evaluated will be considered for the hyperparameter tuning: `n_estimators` and `max_depth`. The previous table containing the hyperparameter information will prove valuable for future phases.

## Hyperparameter Tuning Process

The hyperparameter process was performed on just two hyperparameters with the following values:

- `n_estimators`: 100, 200, 400
- `max_depth`: 6, 50, 100

## Hyperparameter Tuning Results

The results of all tuning process can be see in the sheet named "1 Day Phase 4" in the file [5.3 Hyperparameter Tuning Comparison Table .xlsx](#) .

The XGBoost model, specifically the one with hyperparameters "Without Demographics; n\_estimator = 100, max depth = 100," stands out as the top performer among the options provided. It consistently achieves lower MAPE and MPE values compared to other hyperparameter combinations for both the 60-day and 240-day predictions. However, it is worth noting that the hyperparameters "Without Demographics; n\_estimator = 100, max depth = 50" yield results that are quite close, with the best MTT (training time). This suggests that the choice between the two options depends on the specific focus or priority. Both models have their strengths, and the decision should be based on whether achieving the lowest prediction errors (MAPE and MPE) or optimizing training time (MTT) is more important.

It's also important to mention that changing the "n\_estimator" parameter didn't have a big impact on the model's performance. However, adjusting the "max\_depth" parameter made a noticeable difference. Different values for "max\_depth" affected how accurate the predictions were, as shown by the MAPE and MPE values. This means that choosing the right "max\_depth" value is more important for improving the model's performance, while changing "n\_estimator" had less of an effect.

## Suggestions

In order to enhance future phases, it is recommended to assess the hyperparameters highlighted in green in the file "[5.2 XGBoost hyperparameters description .xlsx](#)" These highlighted hyperparameters are as follows:

- eta
- max\_depth
- subsample
- sampling\_method
- colsample\_bytree,colsample\_bylevel,colsample\_bynode
- lambda
- alpha
- Objective
- eval\_metric
- n\_estimators

If you need more details what are these hyperparameters see the file "[5.2 XGBoost hyperparameters description .xlsx](#)" and see [the official documentation](#).