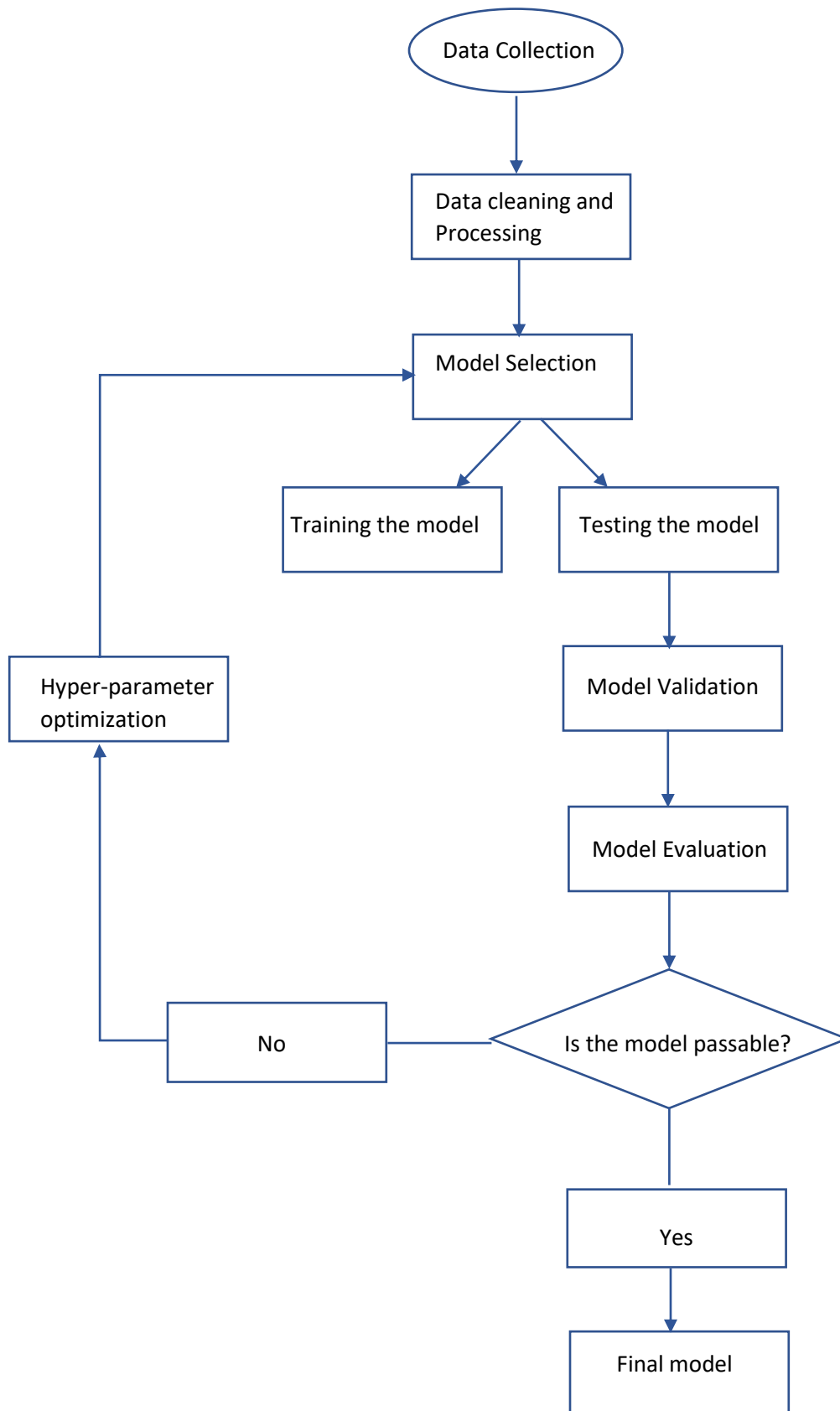


STAGES INVOLVED IN DEVELOPING AN ML MODEL:

Below are the key phases that are involved in the development of a robust and optimized machine learning model:

1. **Data Collection:** The data is received from concrete testing laboratories in PDF formats.
2. **Data cleaning and processing:** The required variables are scrapped, structured, normalized, and stored in the database.
3. **Model selection:** Effective machine learning models are selected that can produce high accuracy in predictions.
4. **Training and testing the model:** The selected models are trained with 70 percent of the data and tested with the remaining 30 percent of the data. The data is split into train and test sets to evaluate how well the model performs. The train set is used to fit the model, and the statistics of the train set are known. The test set is solely used for predictions. This could prevent the occurrence of 'overfitting', which is crucial in machine learning, as it describes an undesirable behavior of the model where it produces accurate predictions for the training data but fails to generalize to new, unseen data.
5. **Model validation:** To validate and test the performance of the model, resampling methods such as K-Fold cross validation, Leave-One-Out Cross Validation. These methods examine how the model performs on some testing data that was not used to fit or train the model.
6. **Model Evaluation:** In order to compare and analyze the performance of the machine learning models, various evaluation metrics are examined, that give prediction error rates and correlation between the variables. Part of them is Correlation coefficient (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).
7. **Hyper-parameter optimization:** The performance of the model is improved by tuning the parameters, specific to an algorithm.
8. **Final model determination:** The final ML model is determined by evaluating different models, optimizing the parameters, and assessing the performance metrics.

Stages involved in a developing an effective model.



In order to compare and analyze the performance of the machine learning models, various prediction error rates are examined. Some of them are:

- **Correlation coefficient (R^2)** characterizes the closeness between the true value of the data and the predicted value of the model. The closer the R^2 is to 1, the higher the prediction accuracy of the model.
- **Root mean square error (RMSE)** measures the accuracy of the model. The smaller the RMSE value, the higher is the prediction accuracy of the model.
- **Mean absolute error (MAE)** which refers to the magnitude of difference between the prediction of an observation and the true value of that observation. Smaller values of RMSE value indicate the higher prediction accuracy of the model.
- **Mean absolute percentage error (MAPE)**, a measure of prediction accuracy of a forecasting method. A robust model should have low MAPE.