Chapter 4: Model Building

4.1. Introduction to Model Building

Model building refers to the process of designing, configuring, and training machine learning models so they can learn patterns from data and generate accurate predictions or insights. It involves selecting appropriate algorithms, defining model parameters and hyperparameters, and systematically training models on prepared datasets. Model building also includes experimenting with different configurations, evaluating performance, and refining parameters to achieve optimal results.

The NPCYF Feature Engineering module handles these tasks in six steps, mentioned chronologically as follows:

1. Model Building Config
2. Model Building

4.2. Purpose and Scopes

4.2.1. Purpose of Model Building

- **Pattern Learning:** Enable models to learn meaningful relationships and patterns from historical data.
- **Prediction Generation:** Produce accurate predictions or forecasts for unseen or future data.
- **Decision Support:** Provide data-driven insights to support business and operational decisions.
- **Performance Optimization:** Identify the best model and parameter combinations for optimal results.
- **Automation Enablement:** Allow scalable and repeatable model training and deployment workflows.

4.2.2. Scope of Model Building

- **Algorithm Selection:** Choosing suitable machine learning algorithms based on the problem type.
- **Hyperparameter Configuration:** Defining and tuning parameters that control model behavior and learning.
- **Model Training:** Training models on prepared datasets using different configurations.
- **Evaluation and Comparison:** Assessing and comparing multiple models to identify the best-performing ones.
- **Artifact Generation:** Storing trained models and execution details for reuse, analysis, and deployment.

4.3. NPCYF Model Building Steps

4.3.1. Model Building Config

The Model Building Config section focuses on defining and organizing the parameters required to build machine learning models in a structured manner. This part enables systematic tuning of model configurations before execution. The process begins with the creation of a Config Collection, which acts as a logical grouping of related configurations for a specific algorithm. For example, multiple ARIMA parameter combinations can be grouped under a single ARIMA Config Collection with a clear name and description. This organization ensures clarity, reuse, and easy management of different experimental setups during the model-building process.

Within each Config Collection, Hyperparameters are defined to specify the algorithm name, required parameters, and whether each parameter is mandatory. These hyperparameters form the foundation for creating valid model configurations. The Config Collection Items subsection then defines actual configurations by assigning concrete values to the previously defined hyperparameters for a selected algorithm. Each item represents a unique model setup. Together, these subsections enable controlled experimentation by systematically generating multiple model configurations that can later be used in batch-based model building and execution.

> 🔎 ***Technical Insight***
>
> *Hyperparameters are configuration settings defined before training that control a machine learning model's learning process and structure. They influence model complexity, convergence behavior, and overall performance but are not learned directly from data. For further details, refer to Appendix III*

4.3.2. Model Building

The Model Building section is responsible for executing the actual creation of machine learning models using the configurations defined earlier. This process is driven through batches that combine dataset collections with config collections. In the Model Build Batch Definition subsection, a batch is created by assigning a name and selecting the appropriate dataset and configuration collections. This setup enables

multiple models to be trained simultaneously across different datasets and parameter combinations, ensuring consistency and scalability in model training workflows.

Once defined, the batch is executed through Model Build Batch Execution, where the selected batch is run and its execution status is monitored to verify successful model creation. The Model Build Batch Run Details subsection provides detailed information about each model built, including execution time, date, and associated batch. Finally, Model Build Artifacts stores the generated outputs, such as pickle files for each trained model, allowing models to be retrieved, analyzed, or deployed later with full traceability.

> 🔎 ***Technical Insight***
>
> *Evaluation metrics are quantitative measures used to assess how well a machine learning model's predictions match actual outcomes. They help compare models, detect overfitting, and determine how effectively a model generalizes to unseen data. For further details, refer to Appendix III*

4.4. Tutorial - Model Building

4.4.1. Tutorial Part 1 - Model Building Config

Go to the Model Building module on the dashboard, and then click on the “Model Building Config” card.

You’ll be taken to the corresponding section - Model Building Config.

This section is divided into three further subsections - “Config Collections”, “Config Collection Items” and Hyperparameters.

4.4.1.1. Step 1 - Config Collection

In this subsection, the collection is defined by a name and description. It is recommended to name it by the algorithm that’ll be used by the models, and describe it as such.

Fill in the "Title" and “Description” in the form below the table.

Then, click the “Add” button to add your newly defined collection to the table of existing collections. Upon successful completion of the process, a pop-up message will appear informing of the addition of the new config collection, and your newly defined collection would be visible in the table as well.

4.4.1.2. Step 2 - Hyperparameters

Click on the “Hyperparameters” button and you’ll be taken to the corresponding subsection.

Type in the “Algorithm Name” and suitable algorithm description in the form below the table.

Now click on the “Add Hyperparameters” button on the right side of the form to add your hyperparameters.

Add as many hyperparameters you want, and check the “mandatory” checkbox if any of your included hyperparameters is a must for training your model. After you’re done adding hyperparameters, click on the “save” button. It will save your list of hyperparameters under your config collection.

Finally, click on the “Add” button to add your algorithm, along with your specified hyperparameters to the table. A pop-up will appear after the successful addition, as well as your algorithm and hyperparameters appearing on the table.

4.4.1.3. Step 3 - Config Collection Items

Click on the “Config Collection Items” button to go to the relevant section.

From the drop down list, select your config collection that was defined in step 1. Then fill in the title and description in the form. Enter the value for test and train split and select the algorithm defined in step 2. After that, click on the “Hyperparameters” button to assign values to your hyperparameters.

Assign values to your hyperparameters, and then click on the save button.

Finally click on the “Add” button to add your config collection item to the table. A pop-up will appear upon successful completion of the process and your item will be displayed on the table as well.

Now that you’ve specified your model and hyperparameters, you’re ready to move on to the next part and actually build the model.

4.4.2. Tutorial Part 2 - Model Building

Click on the “Model Building” card on the “Model Building” module in the dashboard. It’ll take you to the corresponding section.

This section is further divided into four subsections - “Model Build Batch Definition”, “Model Build Batch Execution”, “Model Build Batch Run Details”, and “Model Build Artifacts”.

4.4.2.1. Step 1 - Model Build Batch Definition

Click on the “Model Build Batch Definition” button to go to the relevant subsection.

Fill in the “Title” and “Description” section of the form with the name and suitable description of your model building batch. Then from the “Select Dataset Collection” drop down list, choose a dataset collection. Then from the “Select Config Collection” drop down list, select a config collection created in part 1 of this tutorial. Add any comment if necessary.

Click on the add button to include your newly defined batch to the existing list. A pop-up will inform about successful completion of the process and your batch will be visible in the table.

4.4.2.2. Step 2 - Model Build Batch Execution

Click on the “Model Build Batch Execution” button to go to the relevant subsection.

From the “Select Model Build Batch” drop down list, select your batch.

Then click on the “Run” button to run your model on the selected dataset collection.

A popup message will appear upon successful completion of the process and your batch item will also be visible on the table.

4.4.2.3. Step 3 - Model Build Batch Run Details

Click on the “Model Build Batch Run Details” button to go to the corresponding subsection.

Select a model building batch from the drop down list and all the datasets trained under that batch will appear on the table above.

Choose whatever dataset you want to see the details of. To check logs, click on the “View Logs” button and a dialog box will appear displaying the relevant data.

To check what evaluation metrics are being used, click on the “Evaluation Metrics” button, and a dialog box will display all the relevant information.

4.4.2.4. Step 4 - Model Build Artifacts

Click on the “Model Build Artifacts” button to go to the relevant subsection.

From the drop down list, select your model building batch, and all the datasets trained under that batch will appear in the table above.

Pick any dataset of your choice, and then to download artifacts, click on the “Download Artifacts” button. A pop up will appear informing you that it has been downloaded in your local storage.