Chapter 5: Forecasting

5.1. Introduction to Forecasting

Forecasting refers to the process of using historical data, identified patterns, and analytical or predictive models to estimate future outcomes. It enables organizations to anticipate trends, plan resources, and make informed decisions under uncertainty. Effective forecasting combines past performance with assumptions about future conditions and applies suitable models to generate reliable projections. The process often involves configuring forecast horizons, incorporating external influencing factors, and evaluating results for accuracy and consistency. Overall, forecasting serves as a critical bridge between data-driven models and real-world planning and decision-making.

The NPCYF Forecasting module is designed to generate forecasts based on previously built models while incorporating key assumptions and exogenous factors. It allows users to estimate exogenous variables through configurable collections and apply different estimation methods to specific dataset columns. The module then links forecast configurations, including forecast years, with selected trained model artifacts. Using batch-based execution, forecasts can be run across multiple models and configurations simultaneously. Detailed run statuses, execution details, and downloadable forecast results—such as state- and crop-wise projections—ensure transparency, traceability, and effective analysis of forecast outputs.

5.2. Purpose and Scopes

5.2.1. Purpose of Forecasting

- **Future Projection:** Estimate future outcomes based on historical data and trained predictive models.
- **Trend Anticipation:** Identify upcoming trends and patterns to support proactive planning.
- **Decision Enablement:** Support strategic, operational, and policy decisions with forecasted insights.
- **Scenario Analysis:** Evaluate the impact of different assumptions and external factors on future results.
  Planning and Allocation: Aid in effective resource planning, budgeting, and target setting.

5.2.2. Scope of Forecasting

- **Forecast Configuration:** Defining forecast horizons, periods, and related settings.
- **Model Integration:** Connecting trained model artifacts to forecasting workflows.
- **Exogenous Variable Estimation:** Estimating external variables that influence forecast outcomes.
- **Batch Forecast Execution:** Running forecasts across multiple models and configurations efficiently.
- **Result Analysis and Export:** Viewing, analyzing, and downloading forecast results for further use.
  -

5.3. NPCYF Forecasting Steps

5.3.1. Exogenous Estimation Config

In the Exogenous Estimation Config section, the *Config Collection* acts as the foundational layer for estimating exogenous variables required during forecasting. This subsection is focused on defining and organizing the set of configurations that will govern how external (exogenous) variables are projected into the future. These configurations are not the estimates themselves, but structured containers that hold rules, associations, and metadata needed for estimation. By centralizing these definitions, the system ensures consistency, reusability, and traceability across multiple forecasting exercises that depend on the same external variables.

The *Config Collection Items* subsection operationalizes the previously defined configuration collections by assigning concrete elements to them. Here, specific dataset collections and individual datasets—on which predictive models were originally built—are linked to the exogenous configuration. Each variable (column) in the dataset can then be associated with a distinct estimation method. This allows granular control, enabling different estimation techniques to be applied to different variables based on their nature or behavior. The workflow described—adding estimation methods, selecting configuration collections, mapping methods to columns, and saving mappings—ensures a systematic and auditable approach to exogenous variable estimation.

5.3.2. Forecast Config

Within the Forecast Config section, the *Config Collection* defines the core forecasting setup by specifying forecast configurations alongside the relevant forecast years. This subsection establishes the temporal scope and contextual parameters for forecasting, effectively bridging the gap between historical model outputs and future projections. By explicitly defining forecast years within a configuration, the system ensures that forecasts are aligned with the intended time horizon and that all downstream processes operate with a shared understanding of the forecast period.

The *Config Collection Items* subsection connects forecast configurations to the actual predictive models built earlier in the pipeline. This is where the forecast setup is tied to a specific model batch, typically represented through a serialized model artifact such as a pickle file. Selecting the appropriate model ensures that the forecast uses the correct trained logic and parameters. This linkage is critical because it determines which learned relationships are used to generate forecast values, thereby directly impacting forecast accuracy and interpretability.

5.3.3. Forecast Building

In the Forecast Building section, the *Forecast Batch Definition* subsection is responsible for assembling all required components into an executable batch. Here, the Forecast Config is linked with the corresponding Exogenous Config Collection, forming a complete and coherent forecasting unit. This batch-based approach allows multiple forecasts—potentially across different models or configurations—to be defined and managed together. It supports scalability and repeatability, making it easier to run complex forecasting scenarios without manually configuring each run.

The *Forecast Batch Run* subsection handles the execution phase of the defined forecast batches. Once a batch is selected and triggered, the system runs the forecast and tracks its execution status. This includes indicating whether the run was successful or encountered issues. This step is crucial for operational monitoring, as it provides immediate feedback on the health and validity of the forecasting process, enabling users to quickly identify failures and take corrective action.

The *Forecast Batch Run Details* subsection provides deeper visibility into an executed forecast batch. Rather than just indicating success or failure, it exposes detailed information about the run, such as execution metadata and intermediate outcomes. This level of detail supports debugging, auditing, and performance analysis, allowing users to understand how a particular forecast was produced and to diagnose issues if results are unexpected.

Finally, the *Forecast Run Result* subsection presents the actual outputs of the forecasting process. It displays forecasted values segmented by dimensions such as state and crop, making the results directly usable for analysis and decision-making. Additionally, this section allows users to download the forecast outputs and view the estimated exogenous data, ensuring transparency between input assumptions and final predictions. This closes the forecasting loop by making results accessible, verifiable, and reusable.

5.4. Tutorial - Forecasting

5.4.1. Tutorial Part 1 - Exogenous Estimation Config

From the Forecasting module in the dashboard, click on the “Exogenous Estimation Config” card to go to the relevant section.

5.4.1.1. Step 1 - Config Collection

Under the “Config Collection” section, fill in “Title” and “Description” with the title and suitable descriptions of your configuration collection to estimate the exogenous variables.

Then click on the “Add” button to successfully include your newly defined collection to the list. Upon success, a pop-up will appear, and your new collection will be visible in the table as well.

5.4.1.2. Step 2 - Config Collection Items

Click on the “Config Collection Items” button to go to the corresponding subsection.

Fill in the form by selecting appropriate collection sets from the drop down lists and with proper name and suitable description.

Click on the “Assign Estimation Method” button to select appropriate estimation methods for your exogenous variables. A dialog box with a checklist will appear for you to select. Select your desired dataset column(s) as well as required feature filters from the drop down list.

Click on the + button to assign estimation methods. The drop down lists will appear on the task pane for your selection.

Select appropriate choices from the drop down lists and click on the “Save” button.

Check the box and click on the “Save Mapping” button. And the mapping will be saved.

5.4.2. Tutorial Part 2 - Forecast Config

Click on the “Forecast Config” card from the dashboard module, and it will take you to the corresponding page.

5.4.2.1. Step 1 - Config Collection

Fill the form with appropriate title, description and target year.

Click on the “Add” button. Your newly defined collection will be added to the table. A pop-up will appear informing of successful completion of the process.

5.4.2.2. Step 2 - Config Collection Items

Click on the “Config Collection Items” button to reach the corresponding subsection.

Fill up the form with appropriate values, and then click on “Artifacts” button - a checklist will appear for you to select appropriate artifacts.

Select as many artifacts as needed, and then click the “Save” button.

Now, click on the “Add” button to add your collection item to the table. Upon successful completion, a pop-up will appear with a success message.

5.4.3. Tutorial Part 3 - Forecast Building

Click on the “Forecast Building” card from the dashboard to go to the relevant page.

5.4.3.1. Step 1 - Forecast Batch Definition

Fill up the form with the appropriate title. Description and comment if necessary. Select a forecast config collection and a exogenous config collection from the respective drop down lists.

Click on the “Add” button. A pop-up will inform you of the successful addition of your definition, and the added definition will be visible in the table above.

5.4.3.2. Step 2 - Forecast Batch Run

Click on the “Forecast Batch Run” button to get to the relevant subsection.

Select a forecast batch from the drop down list and all the forecast configs under that batch will be visible in that table.

Select the batch you want to run and then click on the “Run Batch” Button. Your batch execution will start and a pop-up will inform you of the same.

5.4.3.3. Step 3 - Forecast Run Details

Click on the “Forecast Run Details” button to go to the corresponding subsection.

From the dropdown menu select a forecast batch to see all the executions under it in the table above.

Select a batch and click on the “View Config” button to see the configuration details.

5.4.3.4. Step 4 - Forecast Run Result

Click on the “Forecast Run Result” button to get to the corresponding subsection.

From the dropdown list select a forecast batch to see the results on the table above.

By clicking on the “Download” button, you can download the results if required.