IntelliDecision.ai
User Manual

Table of Contents

1. Introduction

This User Manual (UM) provides the information necessary for analysts to effectively use the ID.ai platform for building and documenting the development of statistical models needed for data driven business Decision-making.

Corestrat's ID.ai offers model-building intelligence by removing the complexity of developing a predictive model without you having to write a single line of code. AI embedded Model.ai builds predictive models employing multiple ML techniques in a few seconds for the uploaded data and for your target goals

The no-code, enterprise-ready platform enables you to build and deploy classification and/or regression predictive models in a few clicks. ID.ai helps enterprise users to get closer to smarter business decisions by capturing actionable and hidden patterns in the data.

ID.ai automates a large part of repetitive machine learning steps to ease the tasks of data scientists & non data scientists, thereby enabling enterprises to swiftly adopt ML solutions while allowing them to focus on more complex issues.

2. Software Requirements Specifications

This section provides information about the minimum hardware and software configurations for installing the ID.ai software.

Operating System	Windows 10 & Higher
RAM	>=16 GB
Disk Space	>=200 GB
Software	MS Word 16 & Higher

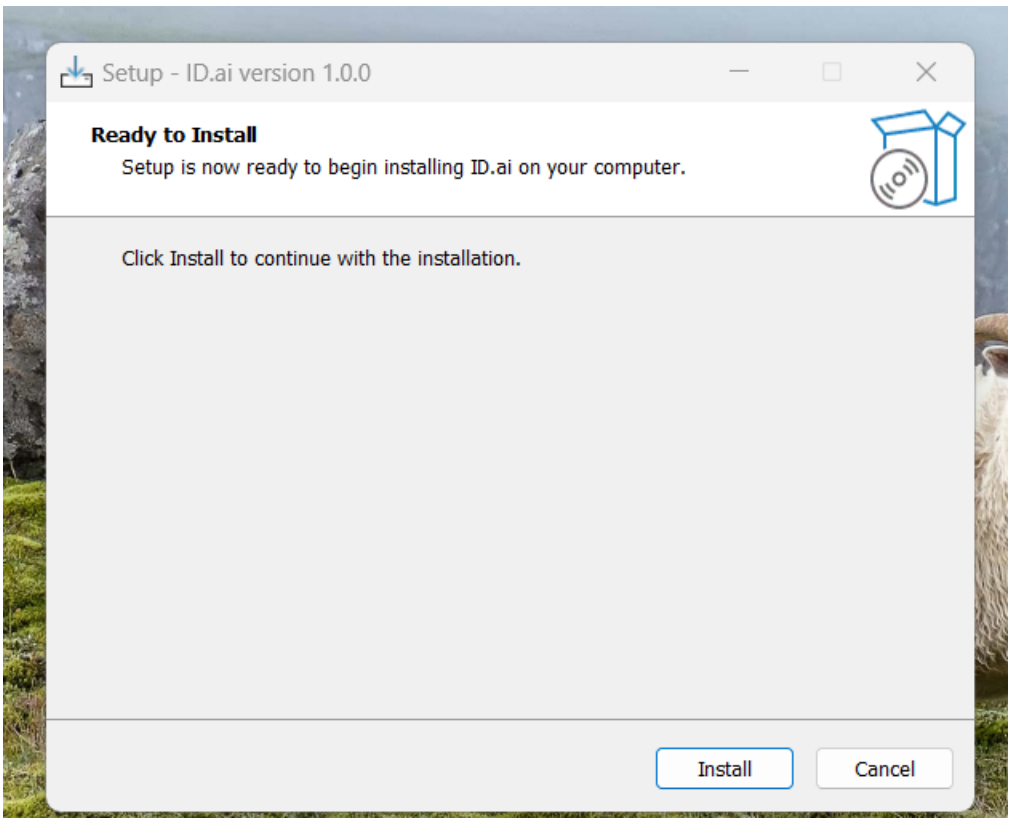
Any configuration below the acceptable versions/configurations described above will result in installation and/or performance issues of ID.ai.

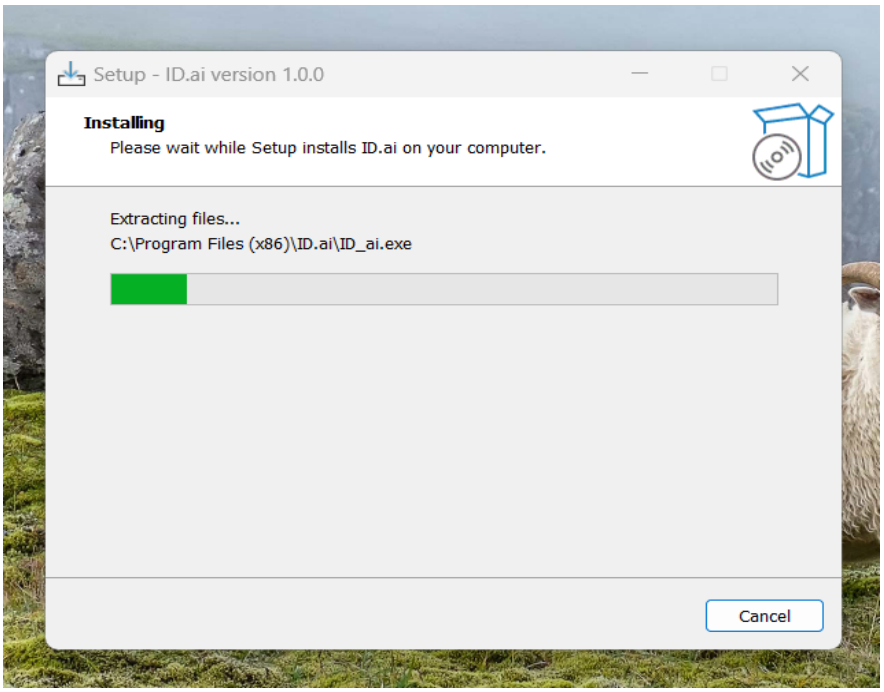
Installation of ID.ai

This section provides the steps to install the ID.ai application on the user's laptop.

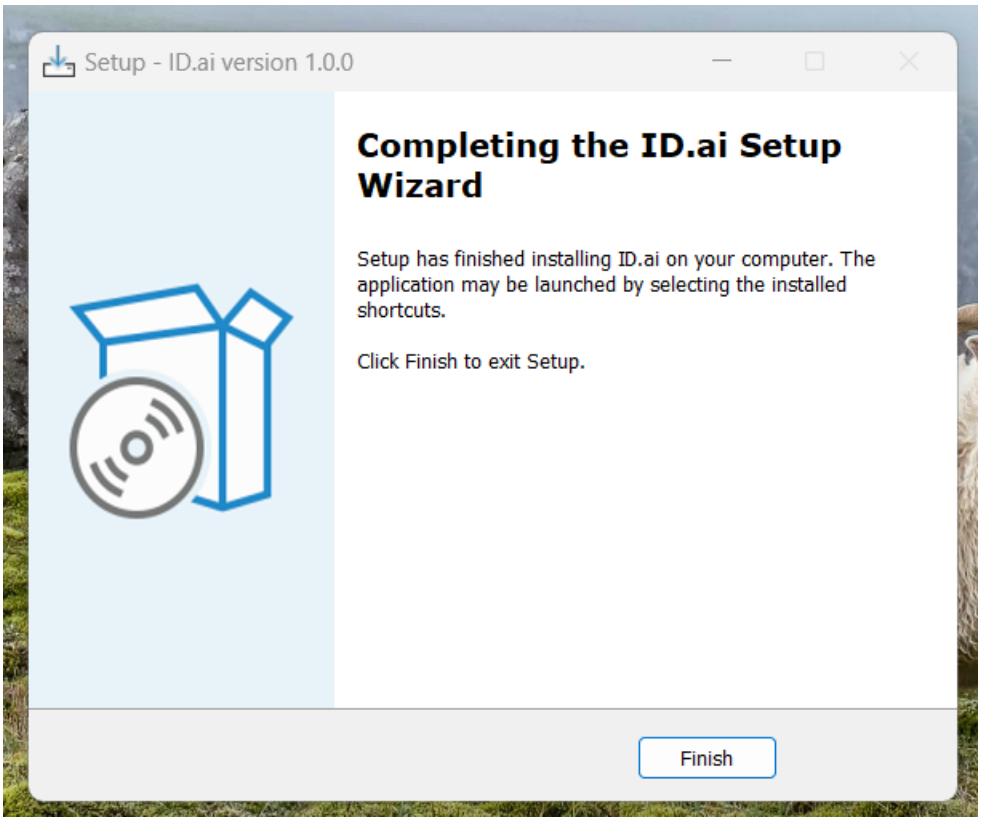
On purchase, Corestrat will provide an exe file named "ID.ai" which looks like the following image.

User should click on this exe file and will get the pop-up screen shown below requesting the user to proceed with the subsequent installation steps. Click on "Install". User should right click and "Run as Administrator" in case privileges are insufficient for direct installation.



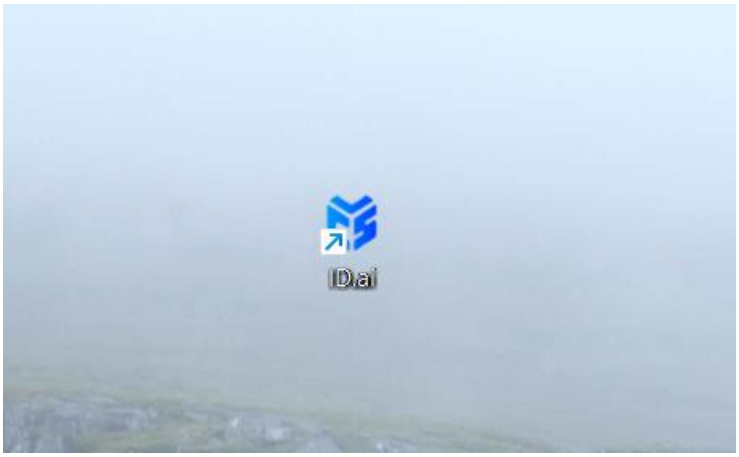


The user needs to click on “Finish” once installation is complete.



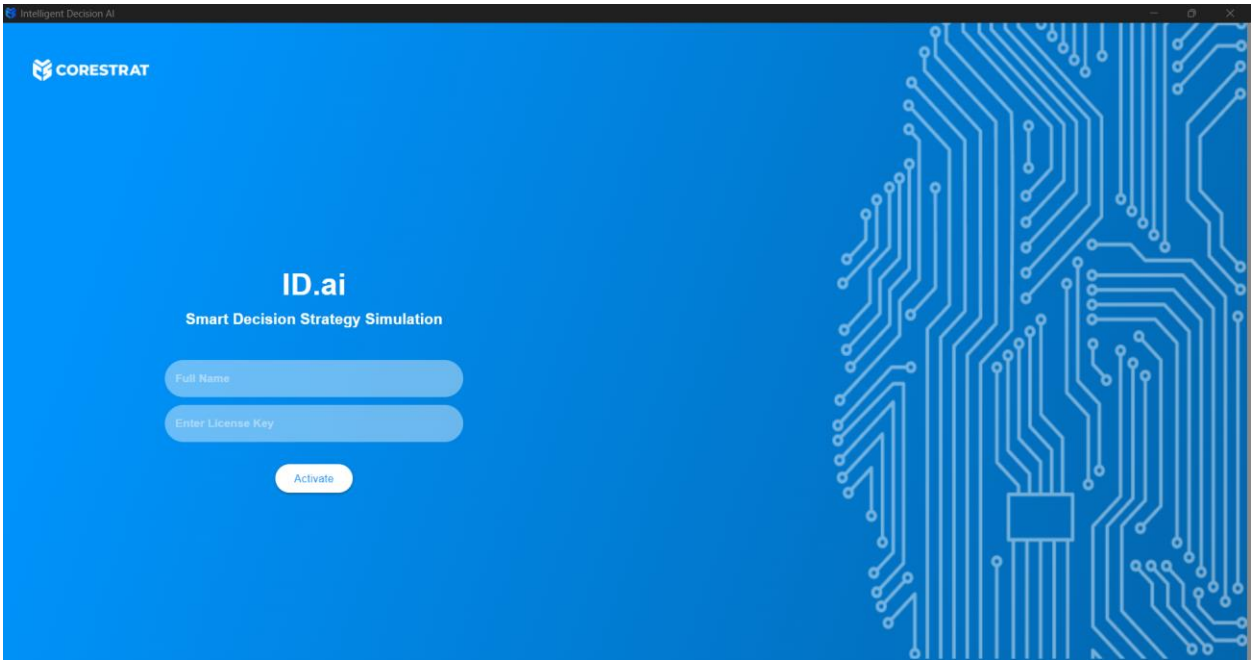
The user will get a shortcut to open the ID.ai application on the desktop.

Click on the ID.ai icon as shown below.

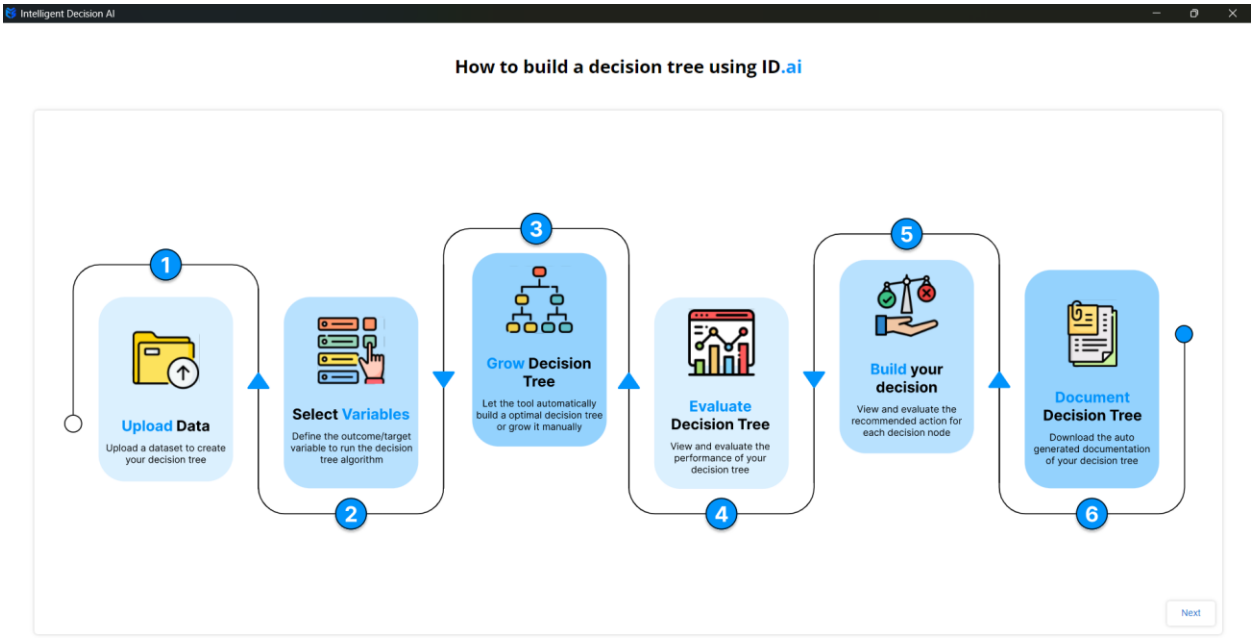


Getting Started

On opening the application user must input their username and license key which would have been provided at the time of purchase. Please enter these in relevant fields in the opening screen shown below to proceed further.



There are six stages in building the desired outcome to drive data driven decision making at scale; these stages are illustrated in sequence in the picture below.

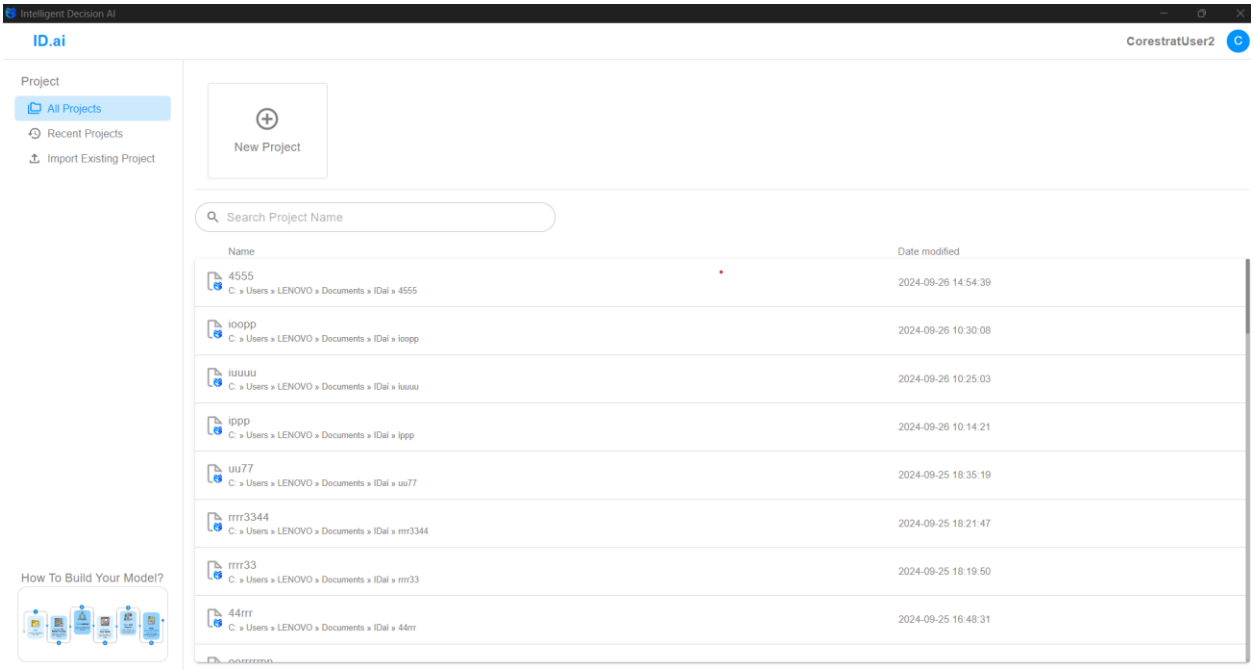


The steps within these six stages will be described in the subsequent sections.

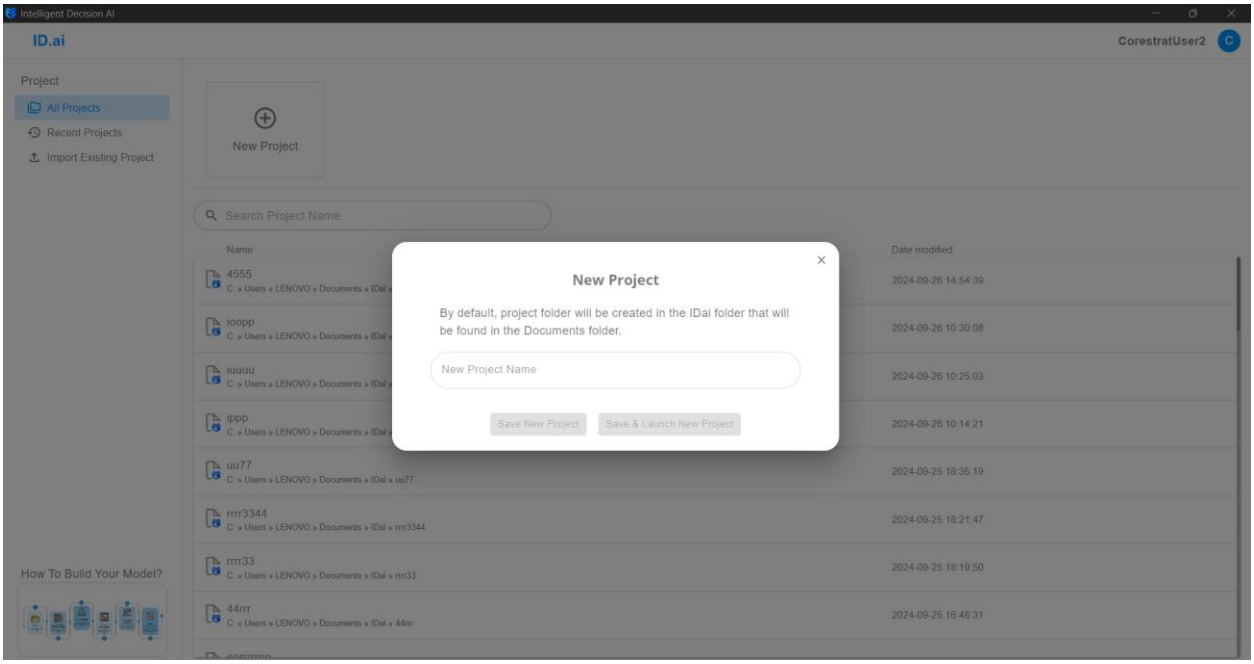
Upload Data Section

Project Creation & File Input

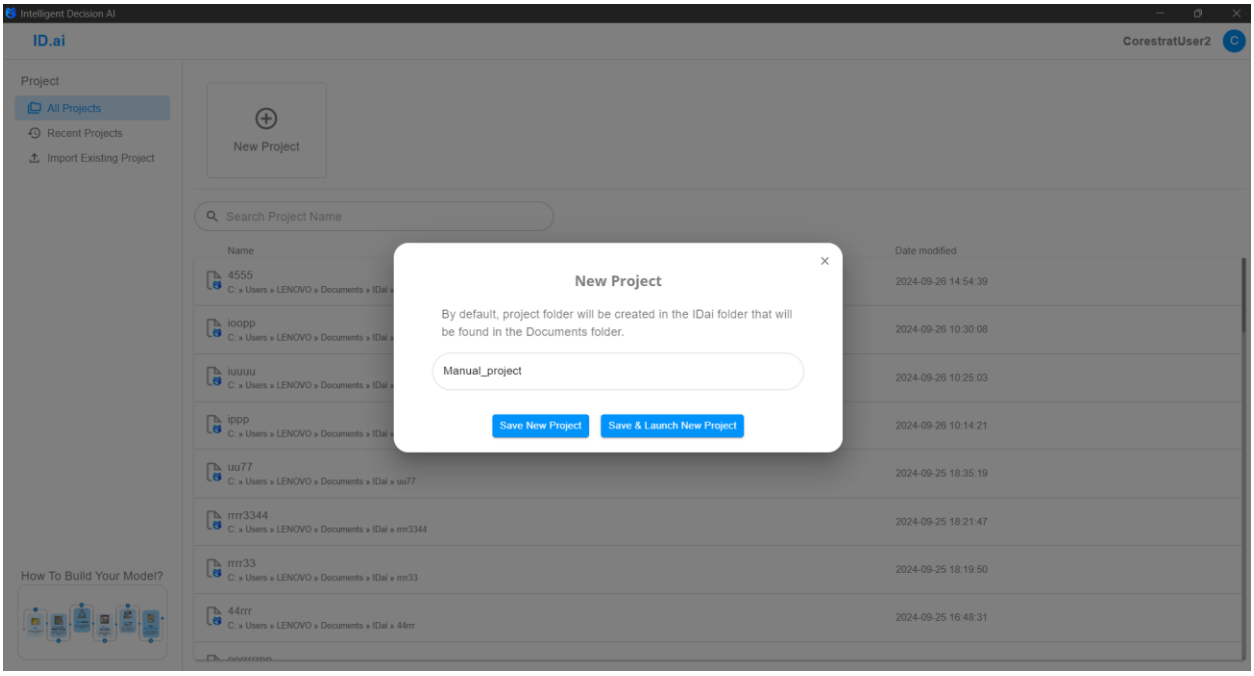
ID.ai can be used to build a new model from scratch or refine an existing model using new assumptions or additional data elements. Both options are illustrated in the screenshot below.



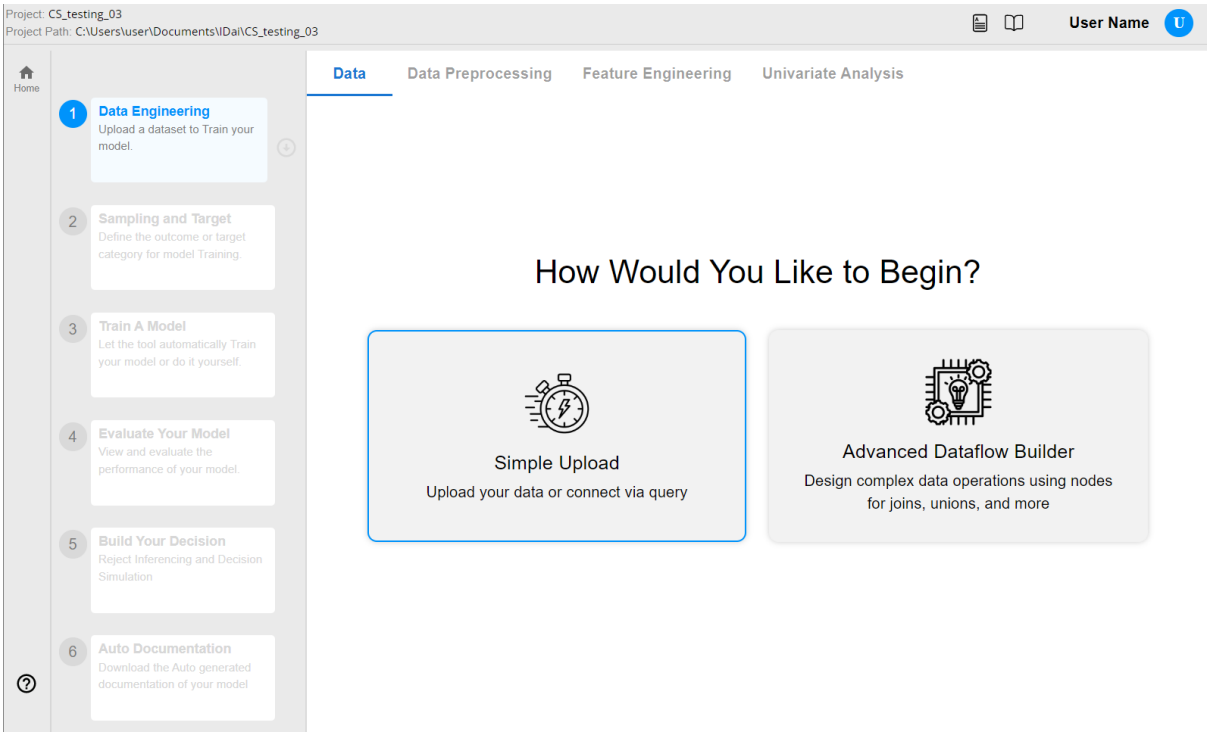
For building a new project, click on "**New Project**" and assign the project name in the pop-up screen as shown below.



Once user enters the name of the new project, they can either just save it and work on it later or start working on it immediately using the respective options in the pop-up as shown below.



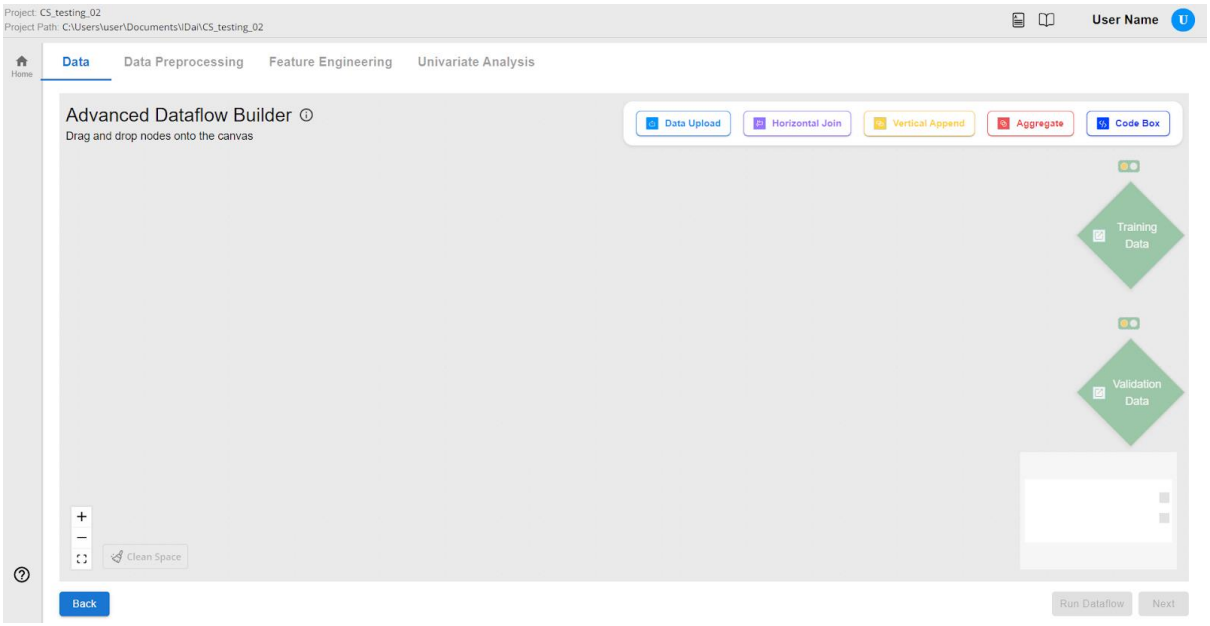
In the subsequent screen, the “project_name” and “project path” are displayed at top left side of the screen. For starting a new project, the user must point to the folder location and file name of the input data source.



When a user opens a project, this is the first screen they see. It lets them either upload data directly or use ID.ai's advanced Dataflow Builder for more control.

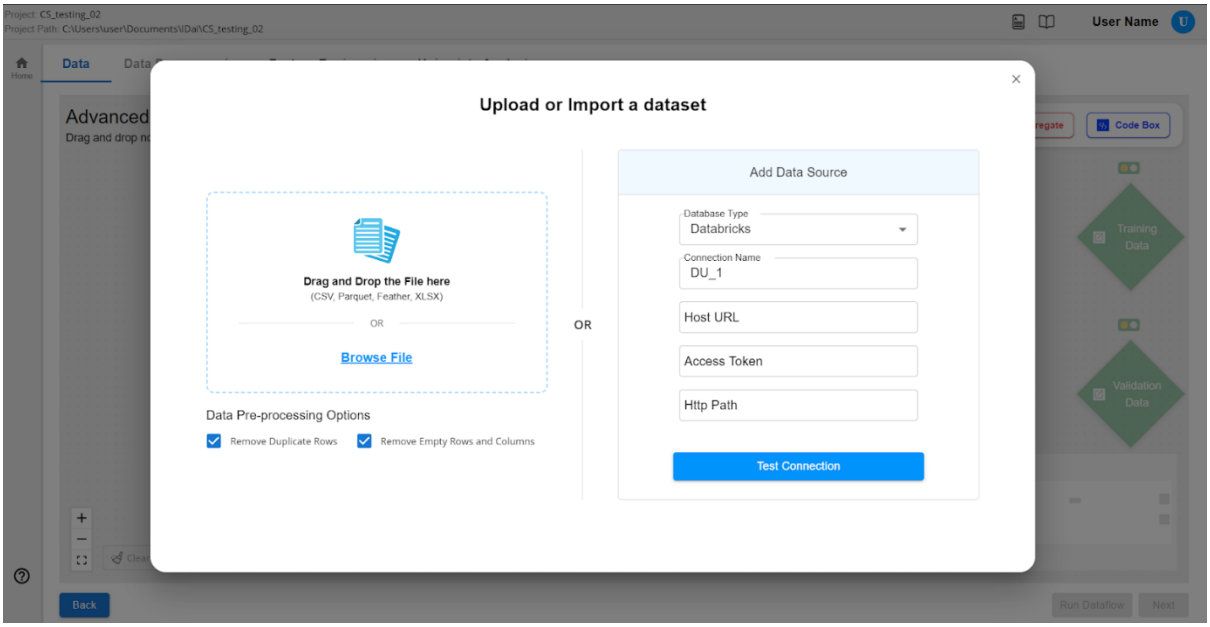
The **Advanced Dataflow Builder** enables users to create complex dataflows with ease by using drag-and-drop action nodes to perform powerful data transformations.

When the **Advanced Dataflow Builder** is selected by the user, a canvas is displayed as shown in the following image.



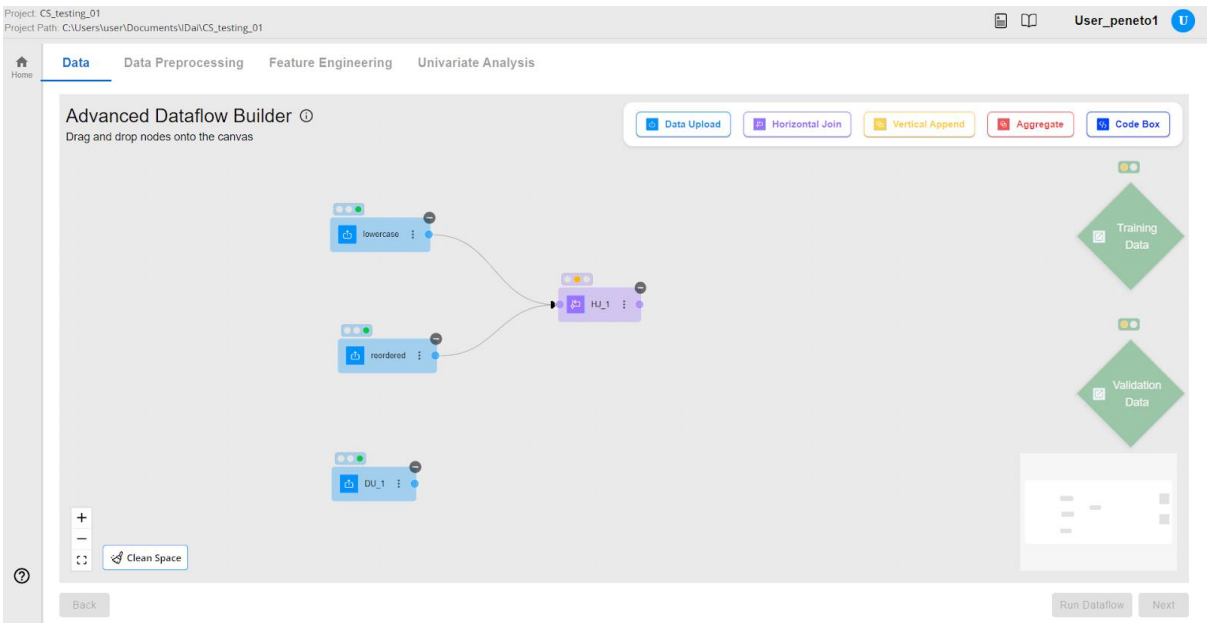
The first step is for the user to add **Data Upload** boxes to the whiteboard. Upon dragging the Data Upload component to the canvas, the screen shown below will appear.

Similar to **Simple Upload**, the user can upload datasets in any of the following formats: CSV, Excel, Feather, Parquet, and flat text files with delimiters. Data can also be imported from the available data sources: Databricks, MySQL, and Snowflake.

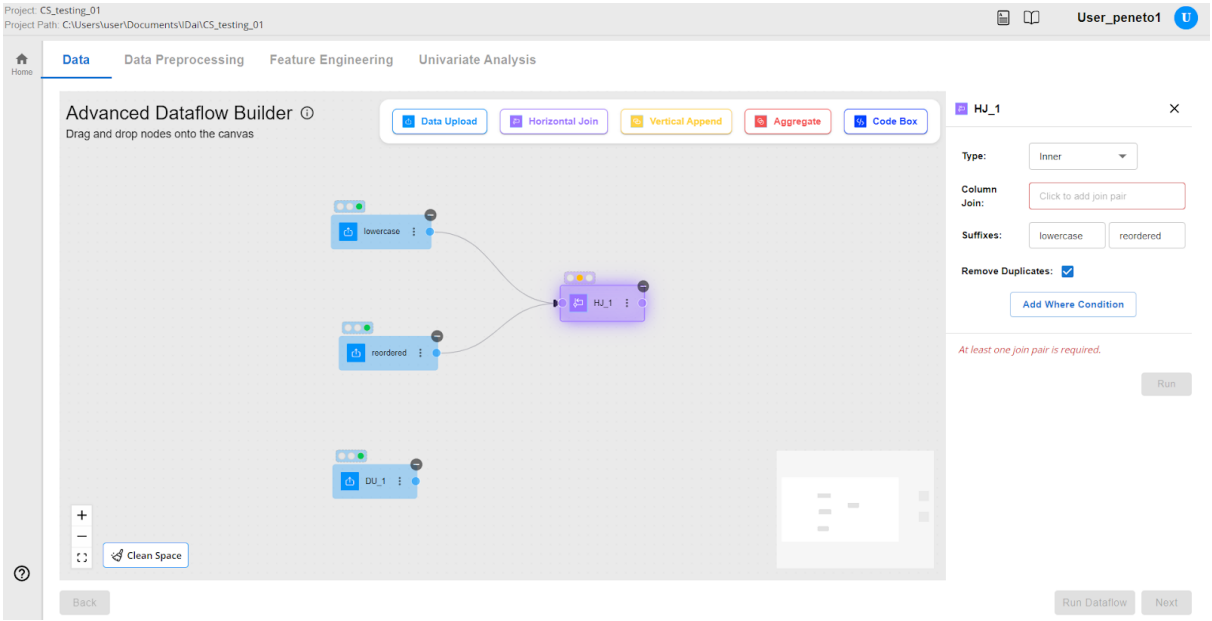


Once a Data Upload box has been successfully configured (indicated by the green light on top of each action box), the user can begin connecting it to action nodes.

The Dataflow Builder supports the following: horizontal and vertical joins of datasets, generating dataset statistics using the Aggregate box, and executing Python code using the Code box.

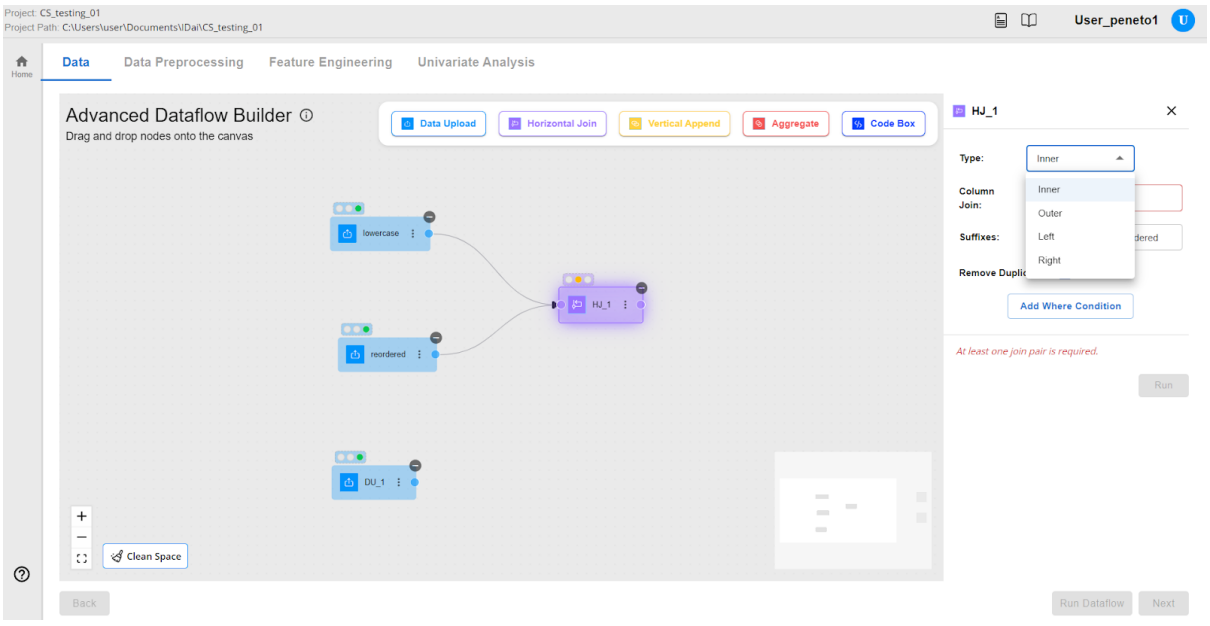


To use the **Horizontal Join** component, the user can drag and drop the box onto the canvas. Upon selecting the newly added box, a configuration sidebar will appear (as shown in the following image).



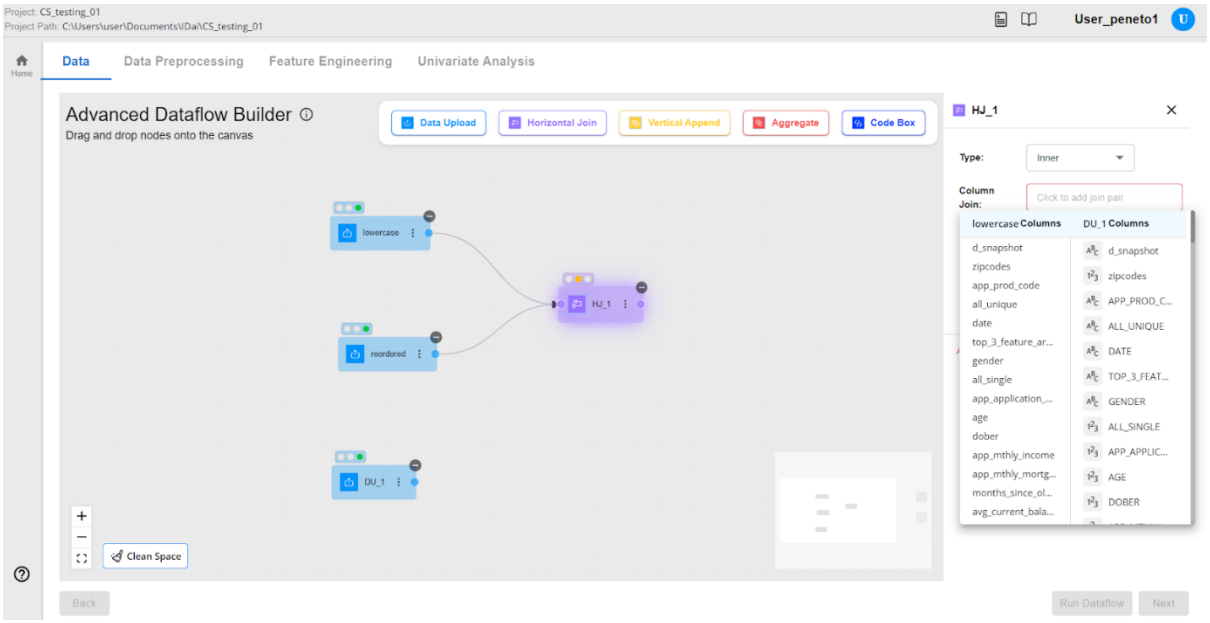
From this sidebar, the user can choose the type of join, select the columns for the join, customize table suffixes, and optionally add a custom WHERE condition to further filter the data.

ID.ai's Dataflow Builder supports Inner, Outer, Left, and Right joins.



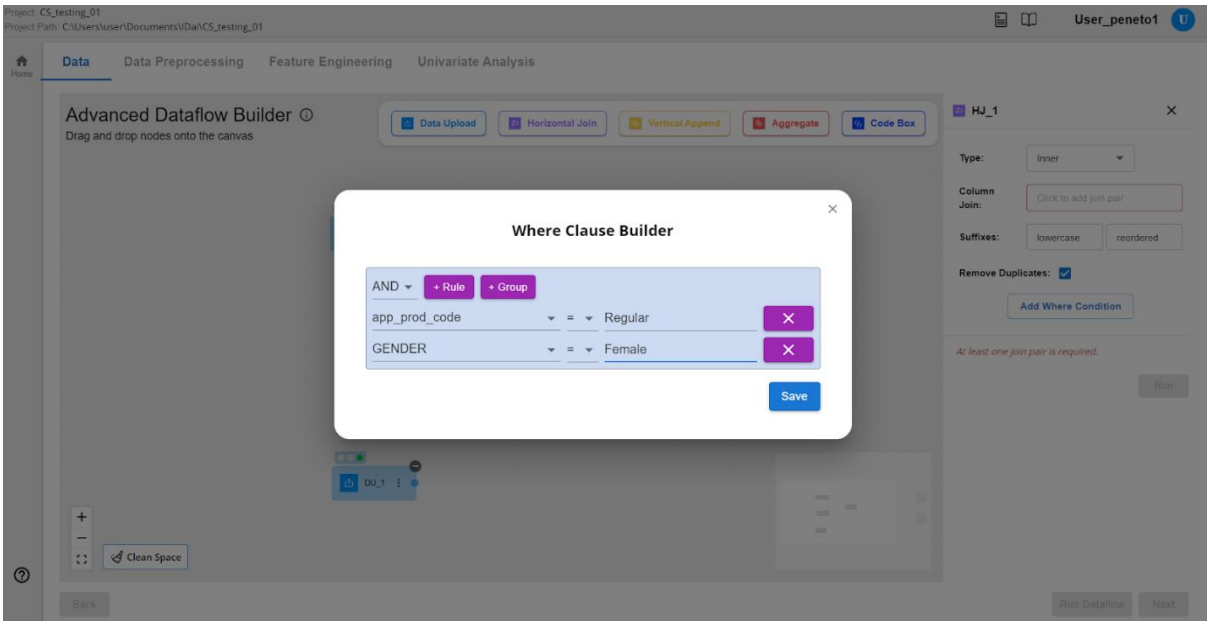
The column join section displays all columns from the datasets connected to the

horizontal join box, allowing the user to make selections accordingly. A custom WHERE condition can be added by clicking **Add Where Condition**.

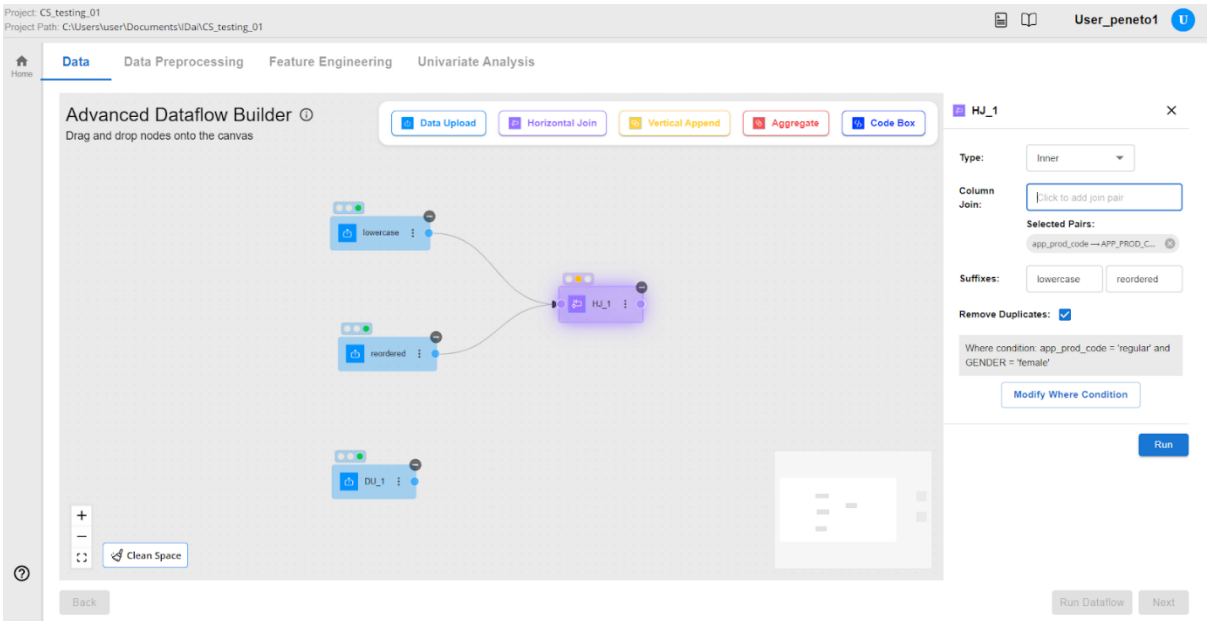


Clicking **Add Where Condition** opens a pop-up window (shown below) where the user can define the condition (rule) and group rules for better structure and optimization.

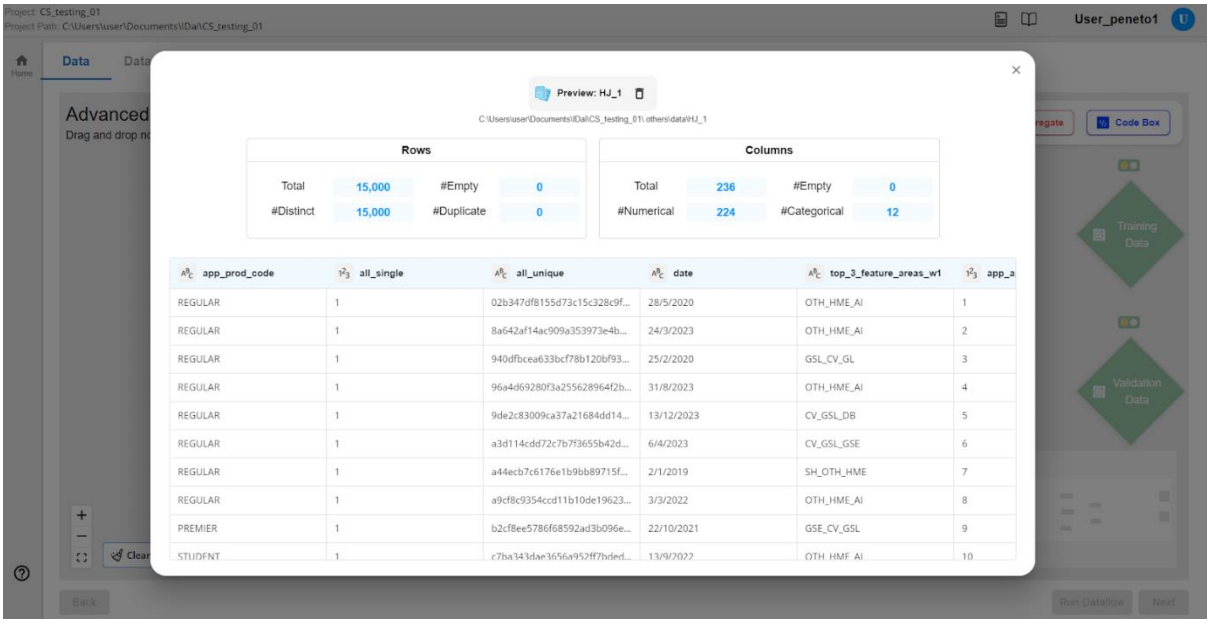
Once the condition is defined, the user can click **Save** to apply it.

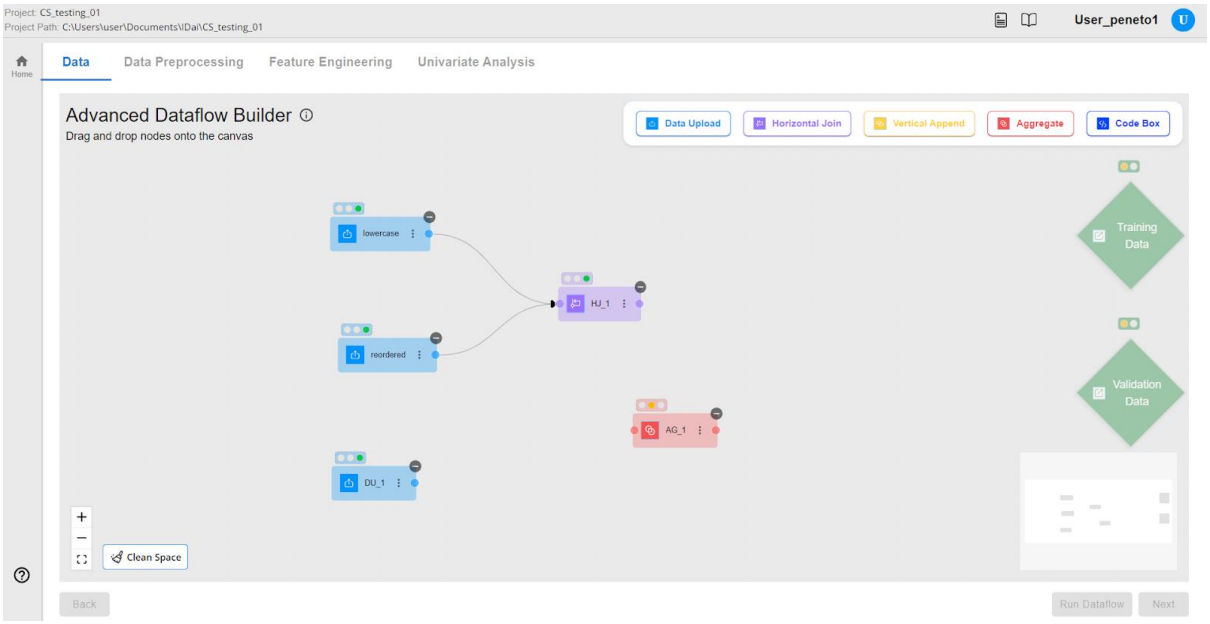


If changes are needed to an existing WHERE condition, the user can click **Modify Where Condition** to make edits.

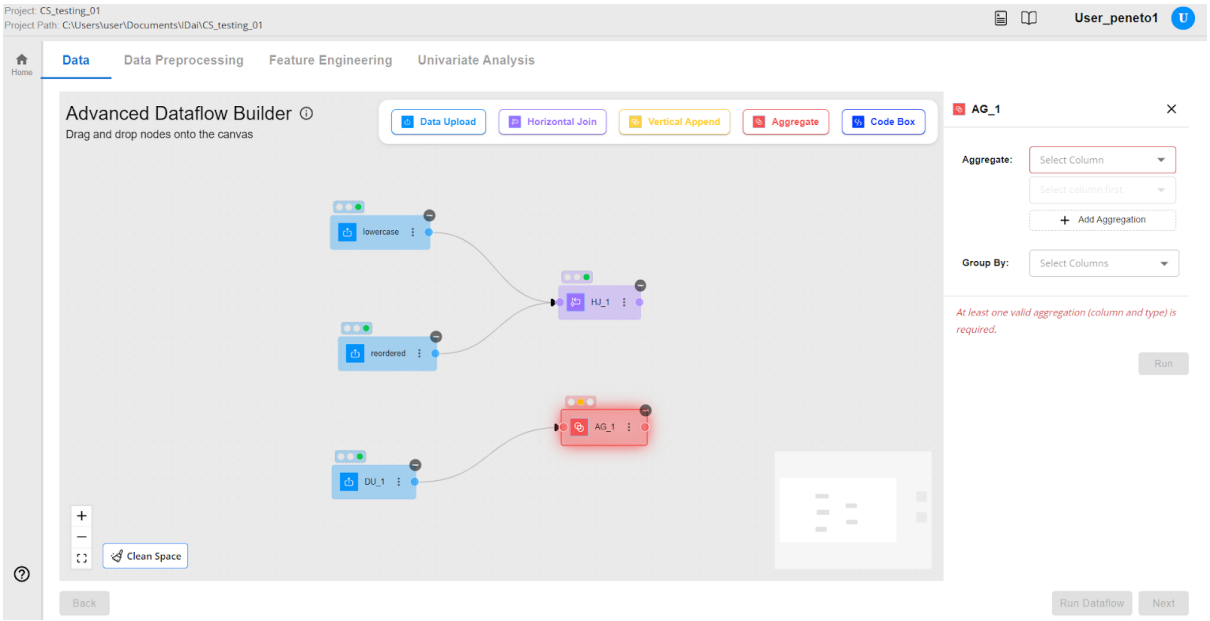


After configuring the necessary settings, the user can click **Run** to execute the join node, after which the output will be displayed as shown.



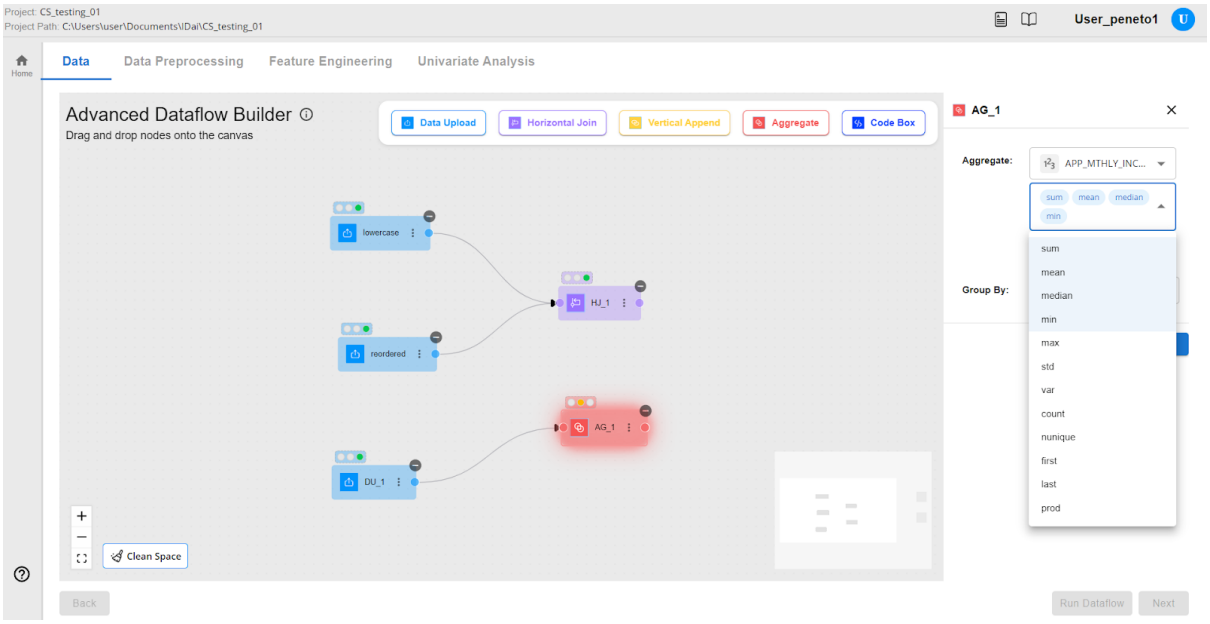


To view dataset statistics, the user can use the **Aggregate** box.

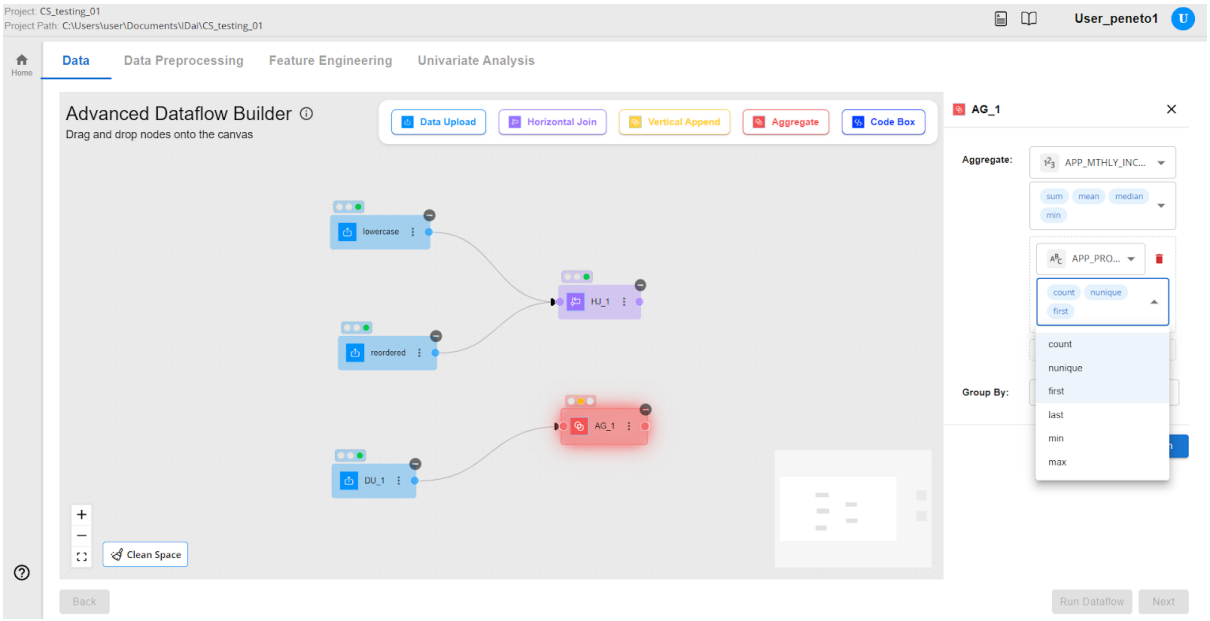


The user must select a column for aggregation.

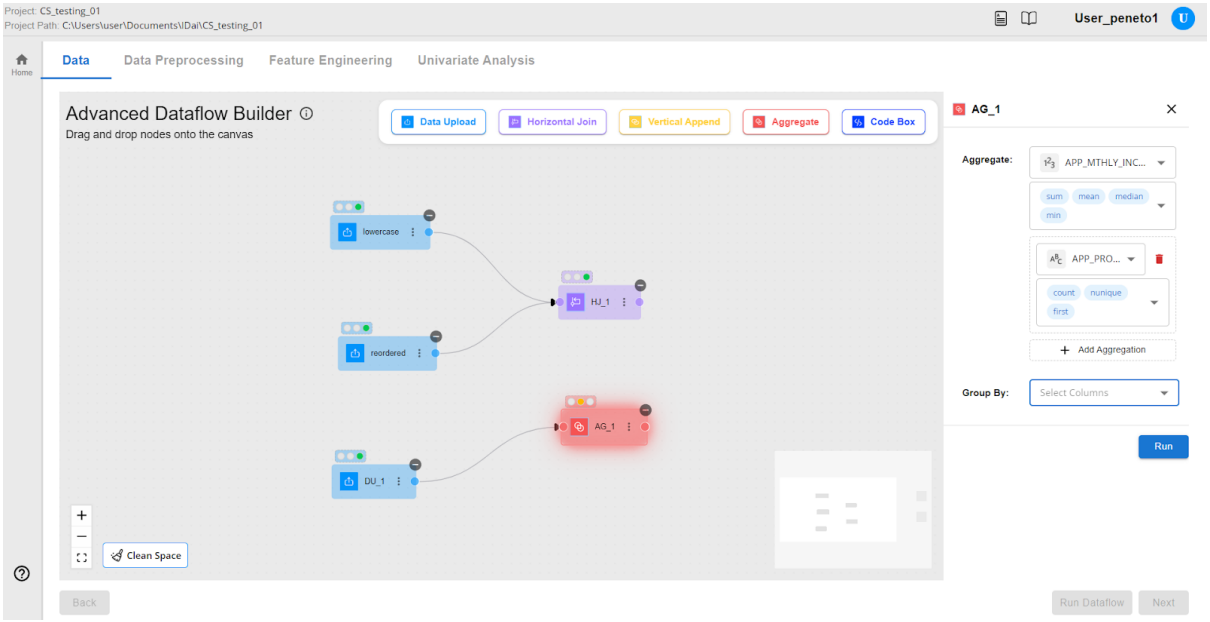
For numerical columns, the available aggregation types are shown in the following screen.



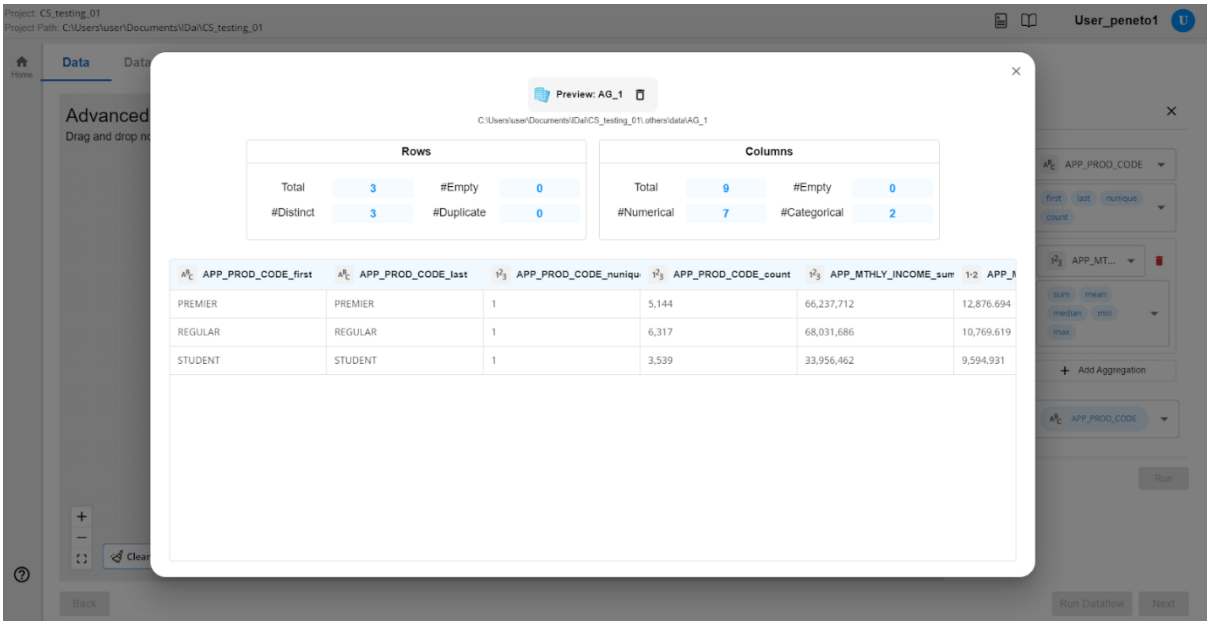
For categorical columns, the available aggregation types are shown in the following screen.



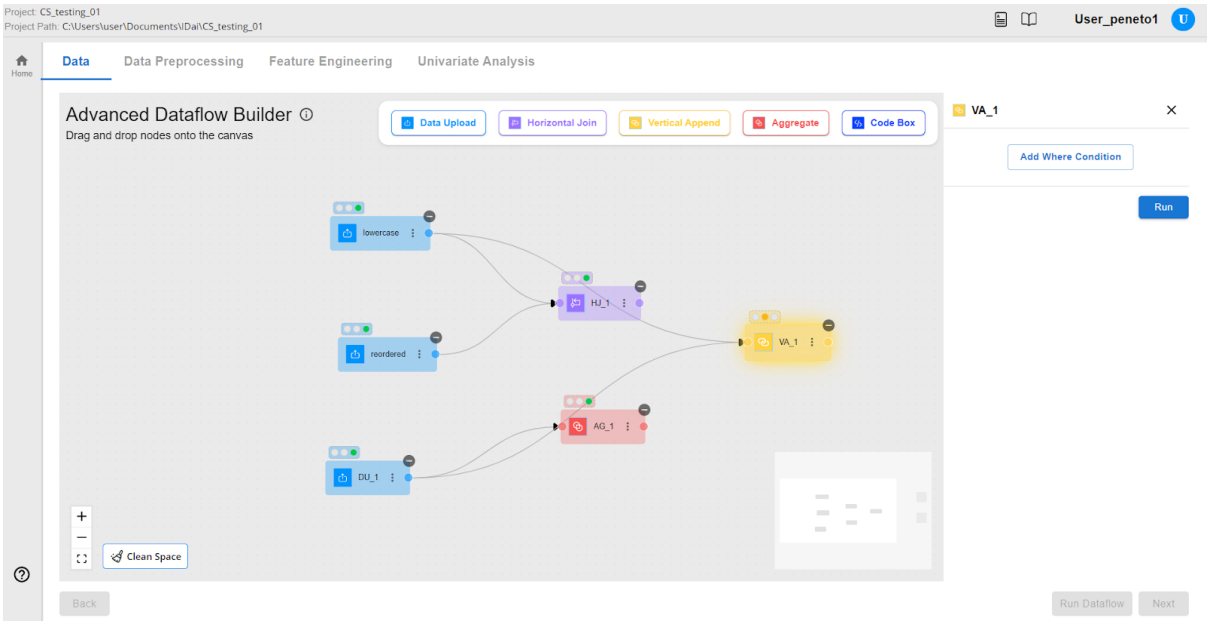
The user must also select a column for grouping the datasets.



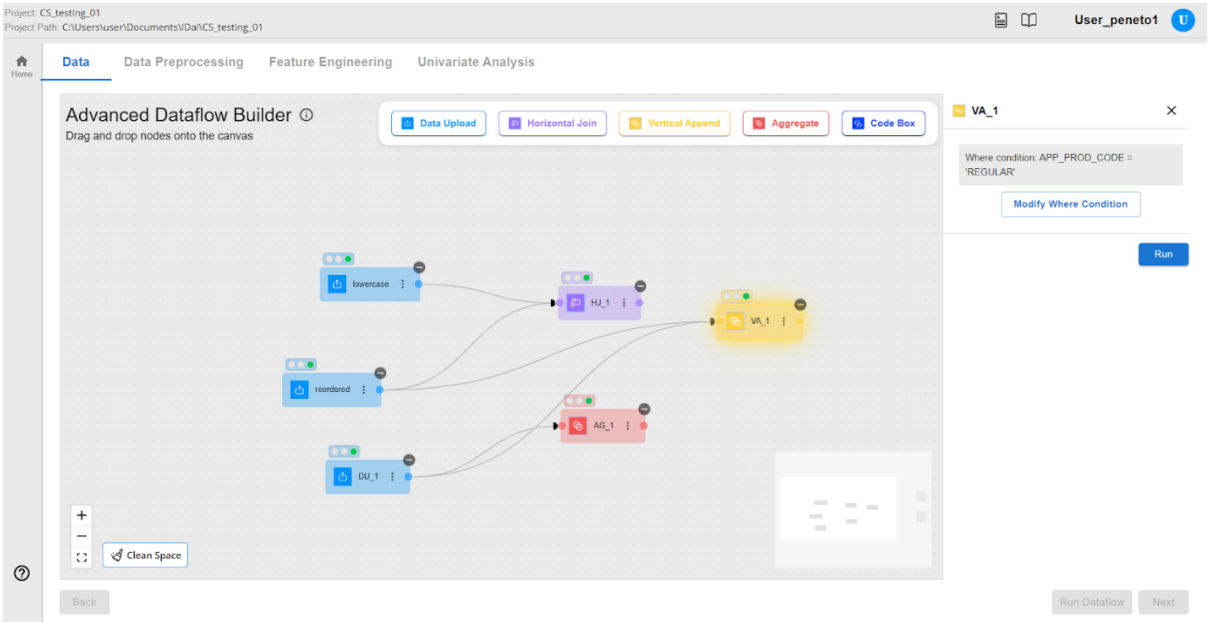
Once the settings are applied, clicking **Run** will execute the aggregate node, and the output will be displayed as shown.

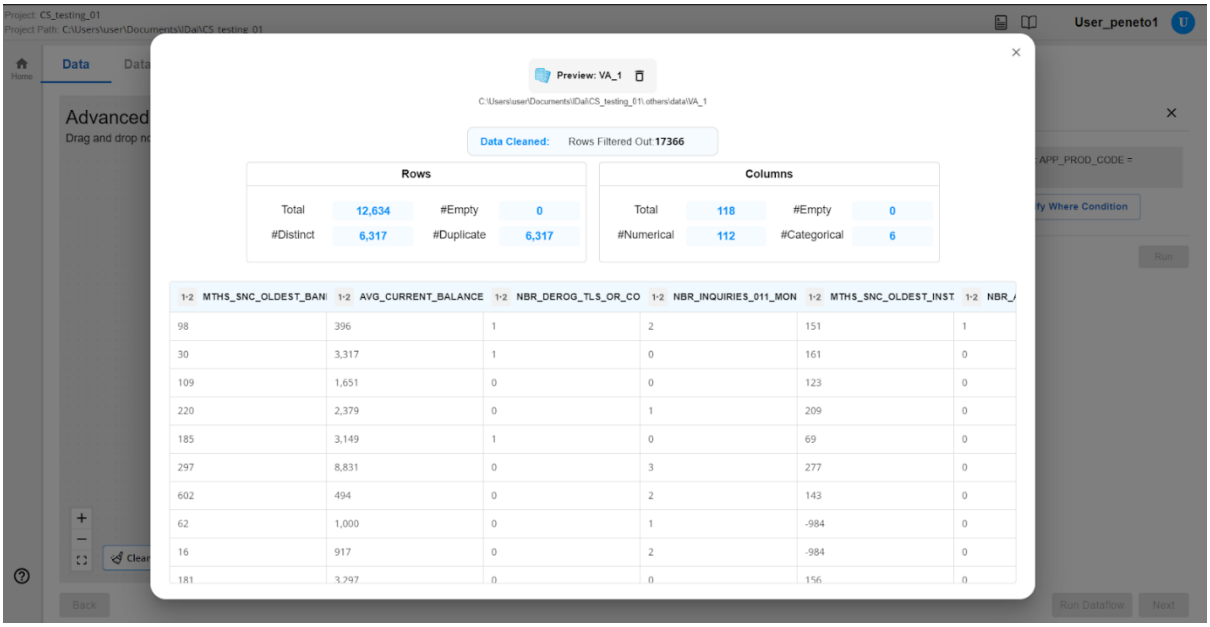


The **Vertical Join** component allows the user to stack one dataset on top of another. This is possible only when the datasets have identical column names (column order may differ).

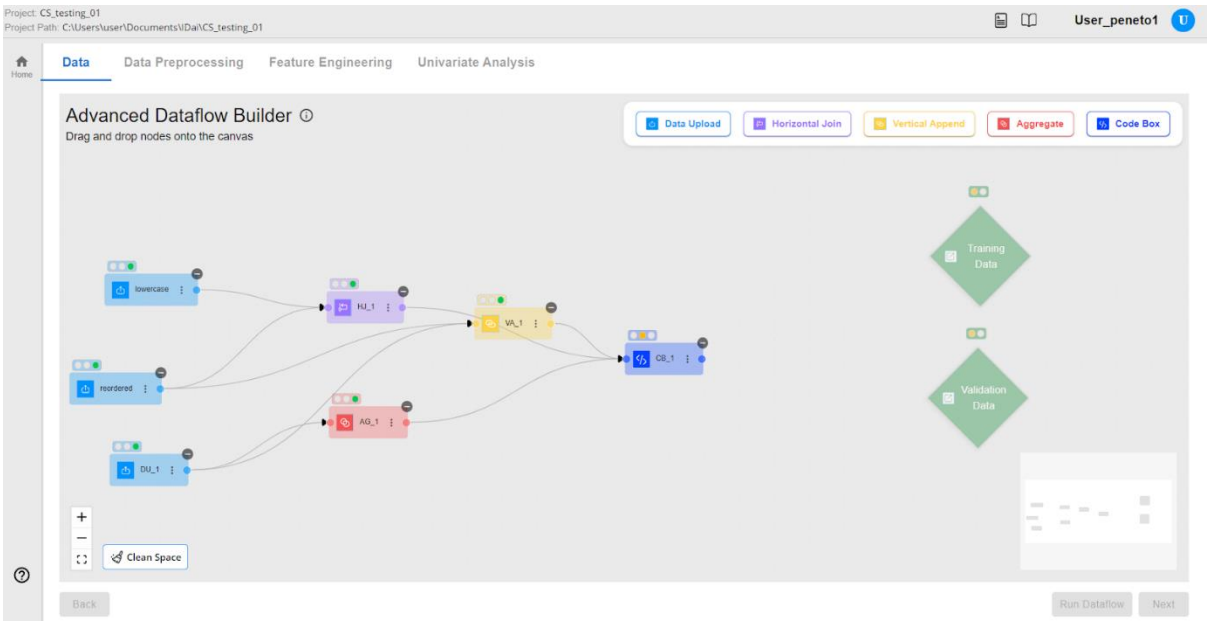


Similar to the Horizontal Join and Aggregate box, the user can add a custom WHERE condition to filter the data further.

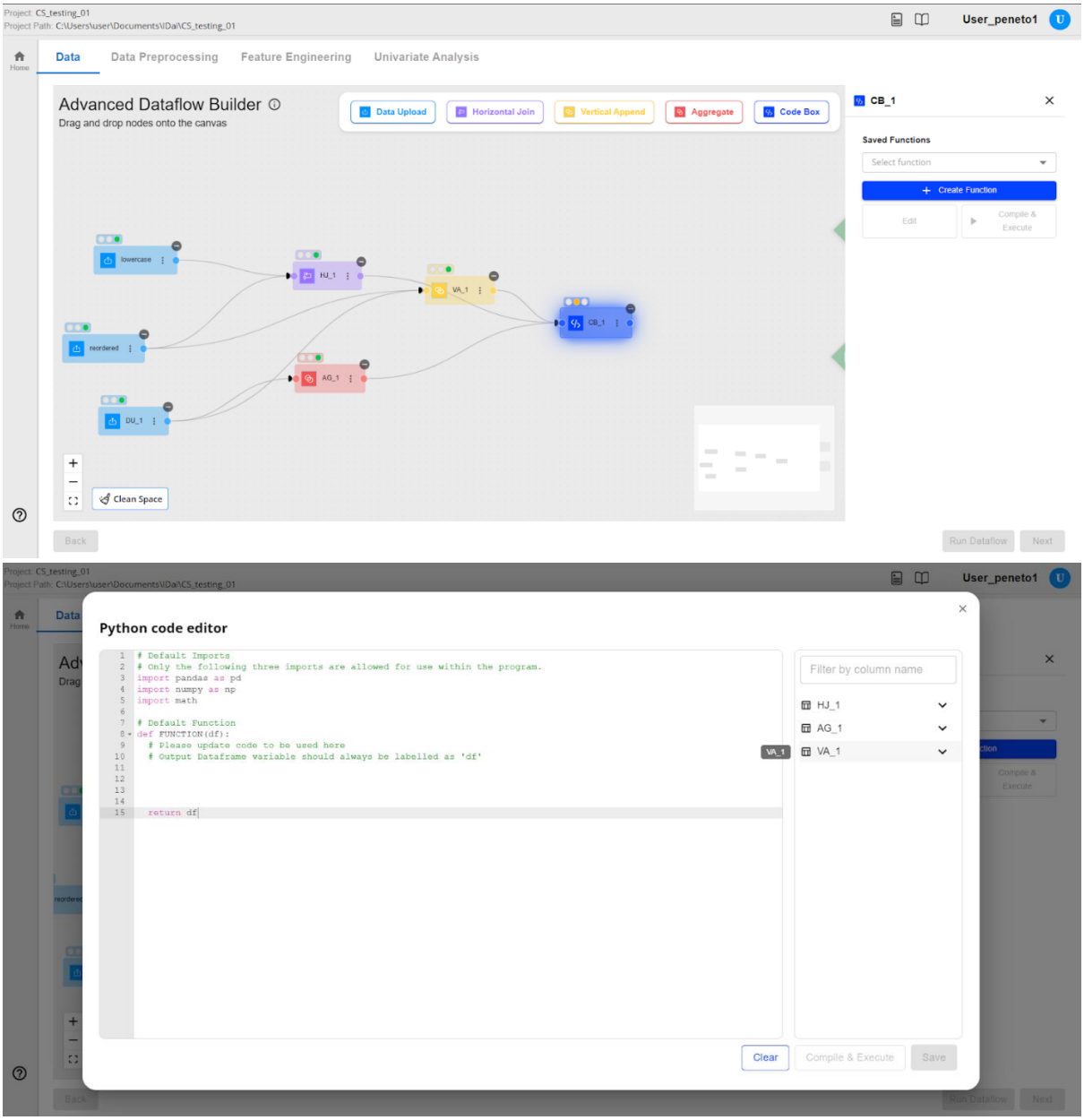




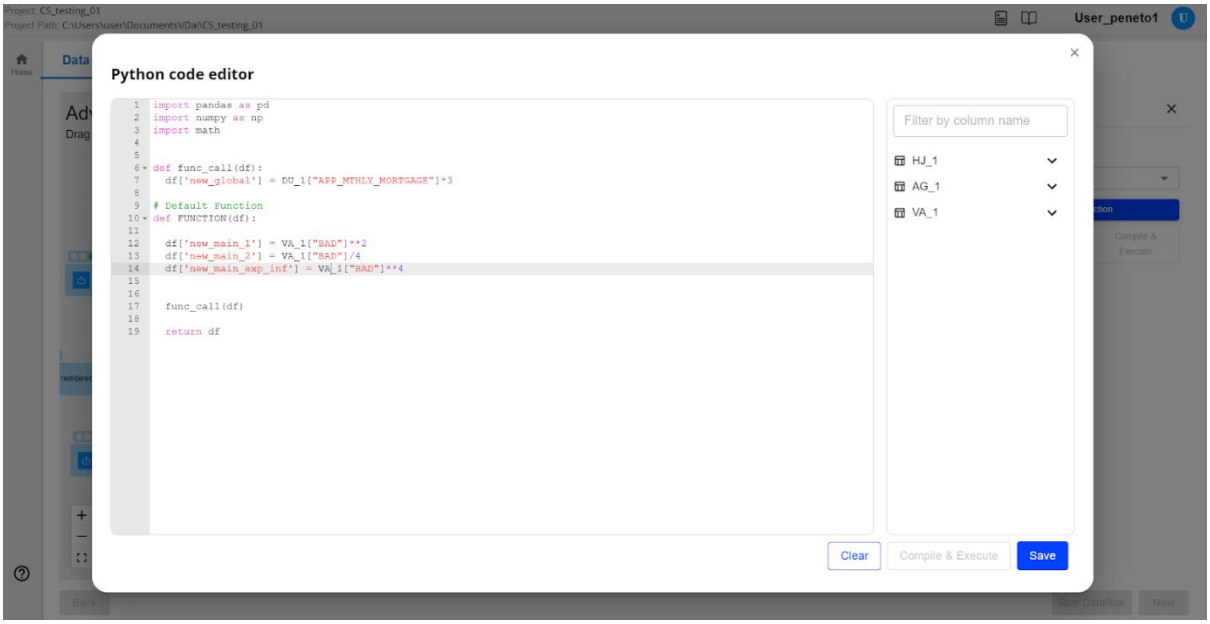
The **Code box** enables the user to apply Python code to the dataset.

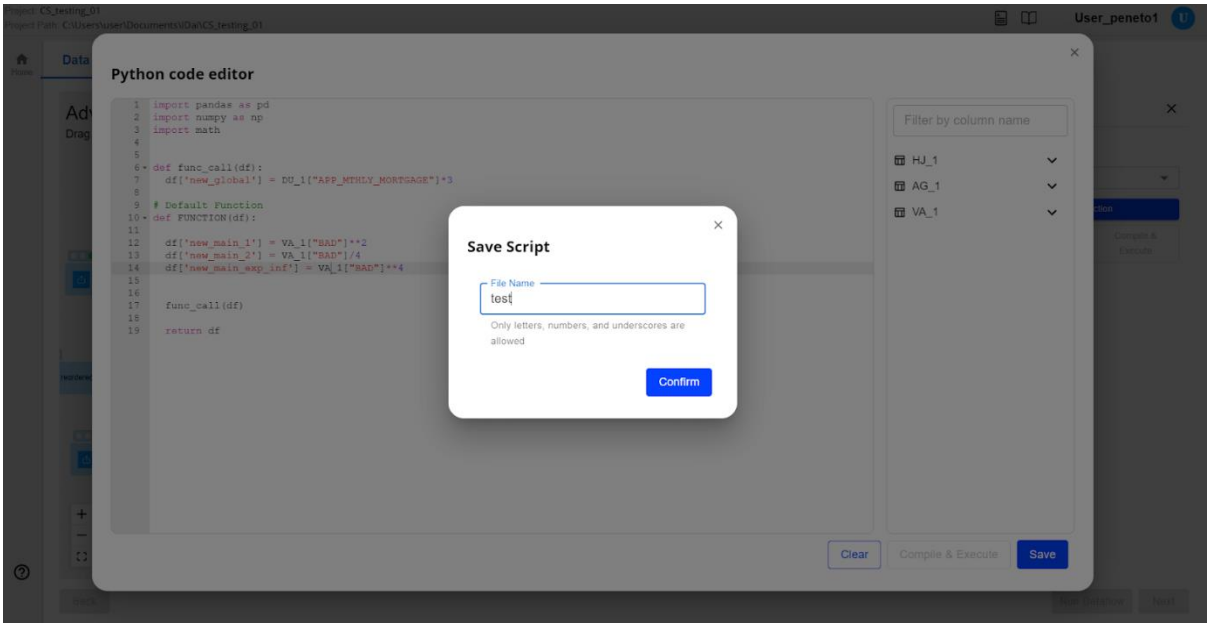


To create a new function, the user clicks **Create Function**, which opens a coding interface.

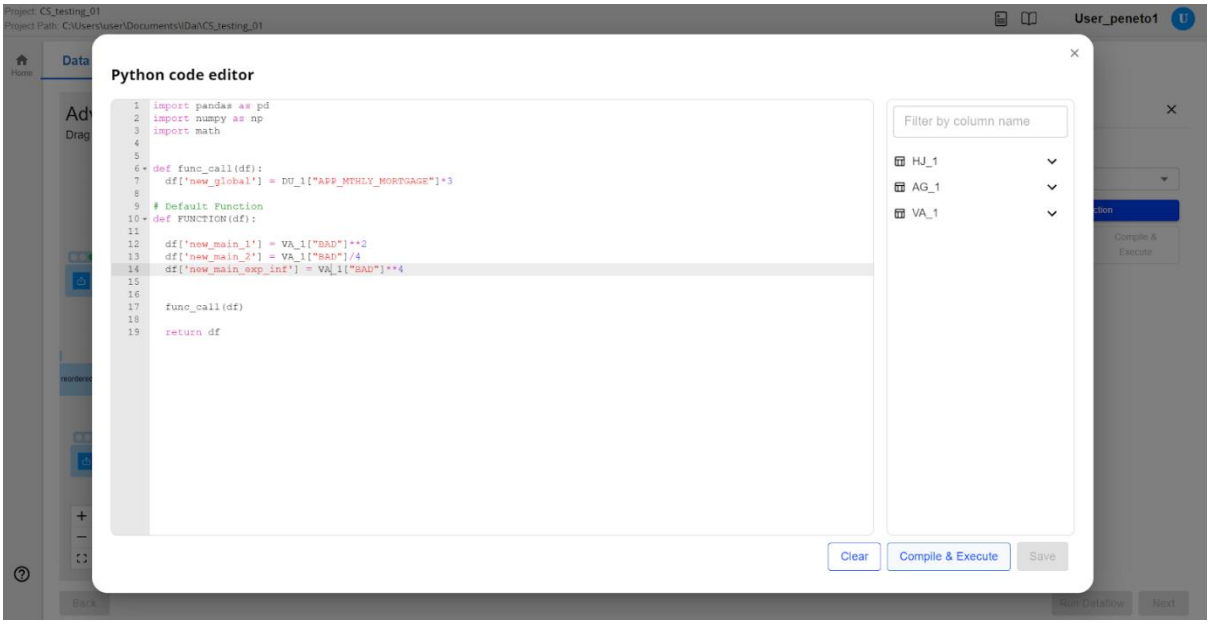


Once the desired code is entered, the function can be saved with an appropriate name for reuse within the workflow.

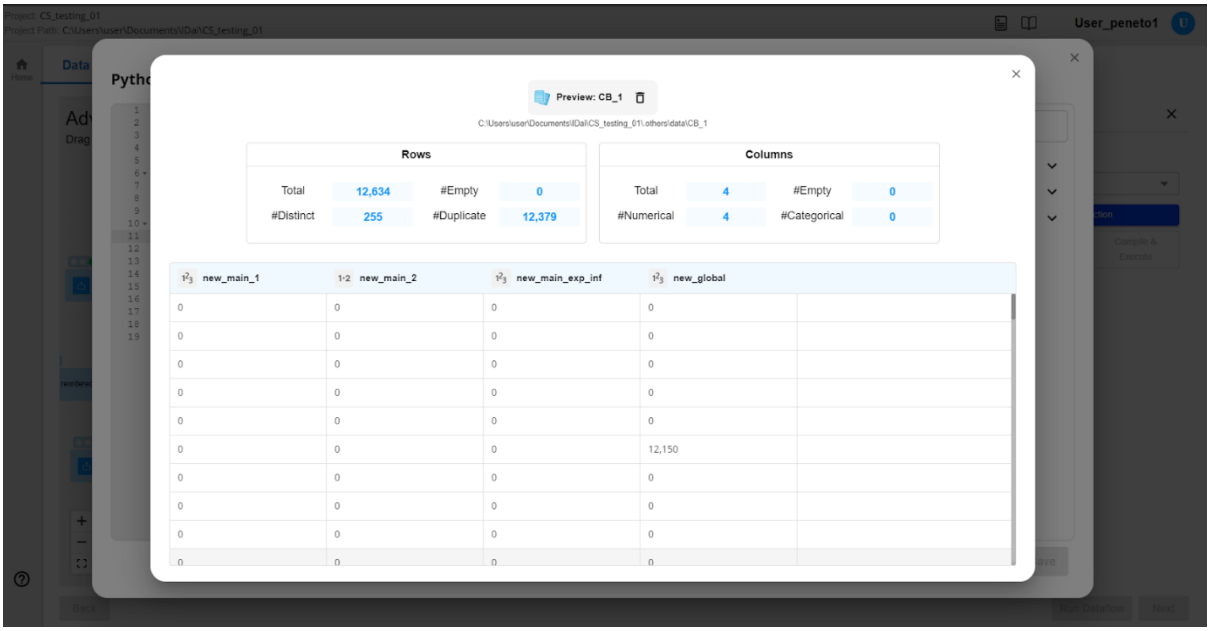




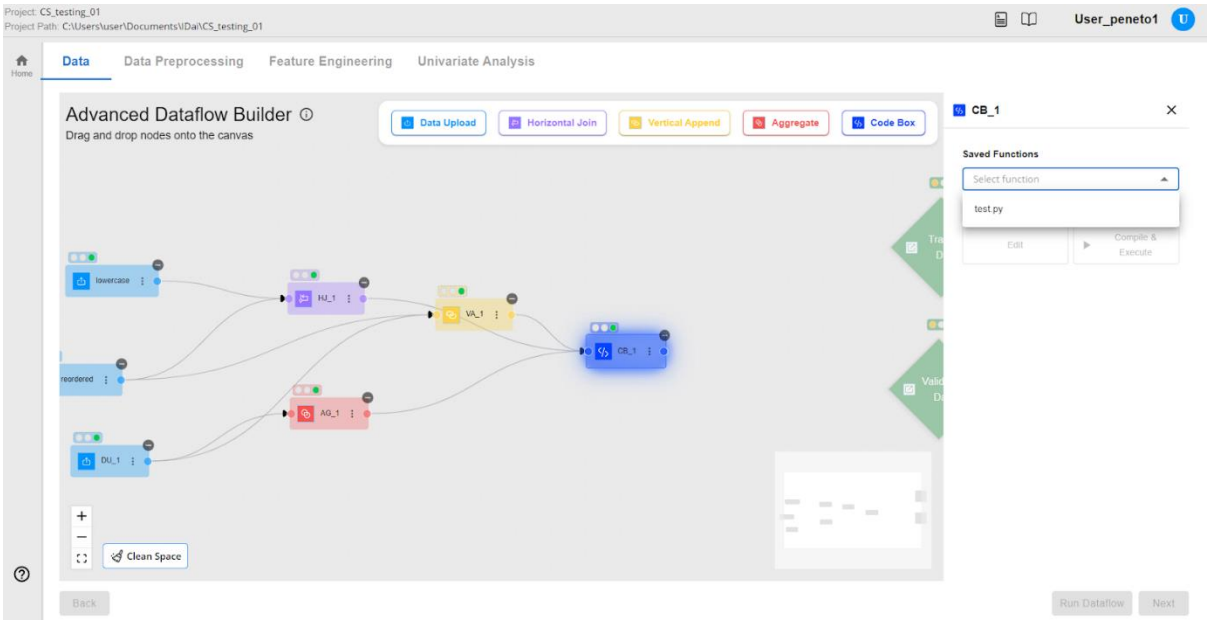
After saving the function, the user can compile and execute the code.



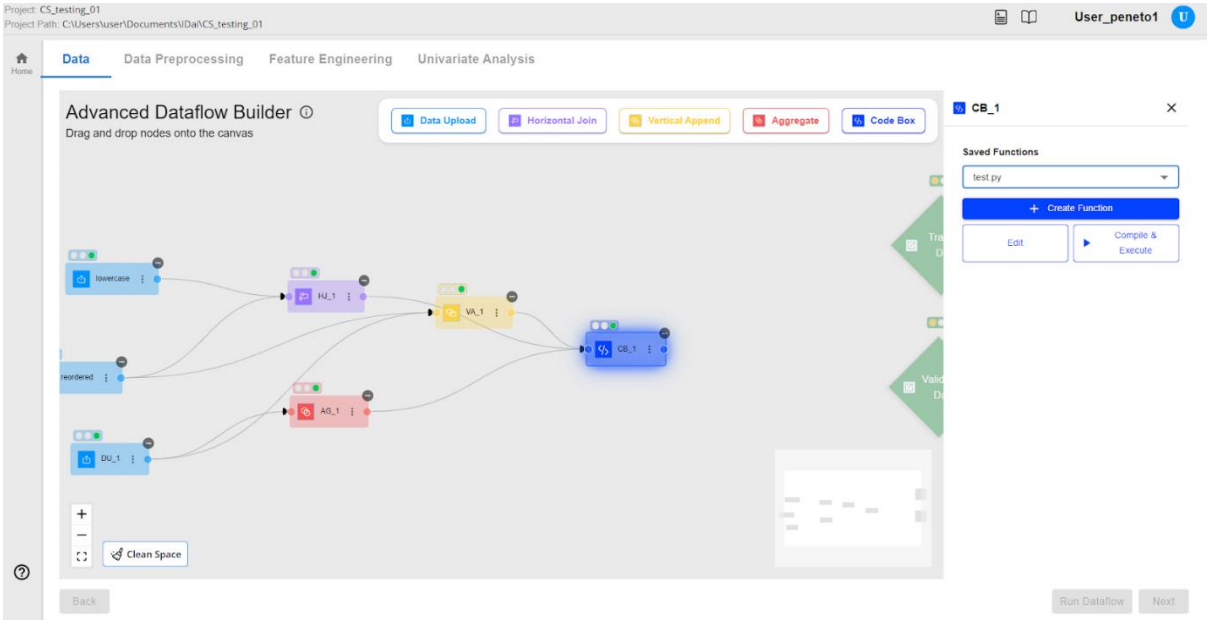
If the code compiles and executes successfully (without errors), an output screen will display the changes made to the dataset.



To use an existing function, the user can select it from the **Saved Functions** dropdown menu.

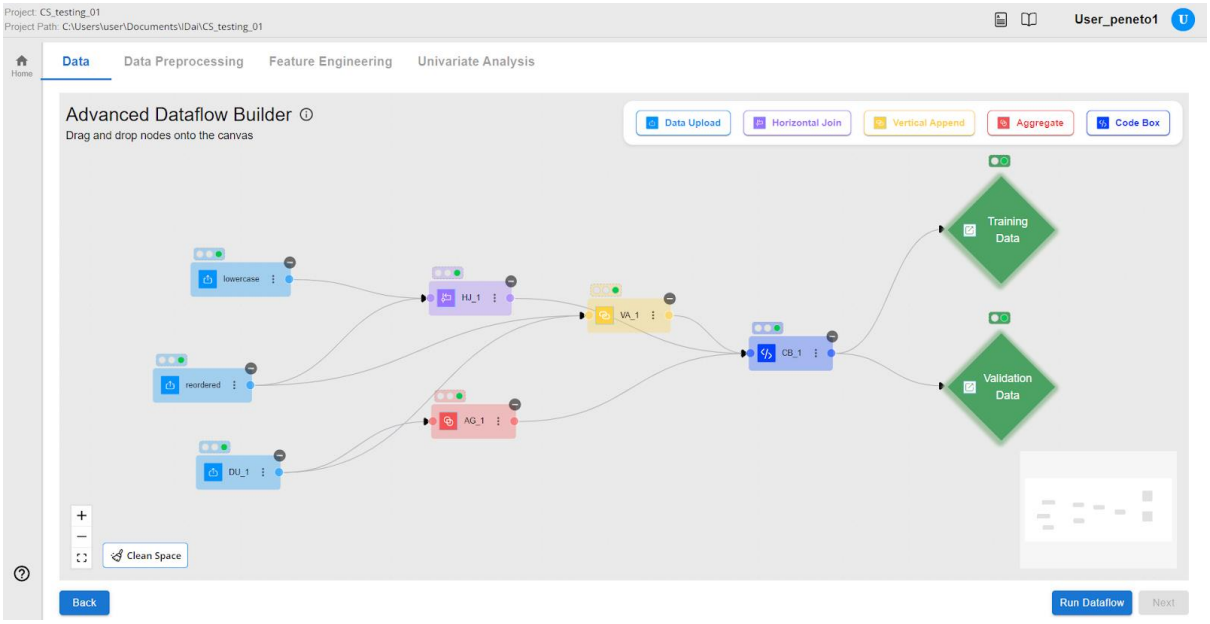


After selecting the desired function, the user can either edit it or click **Compile and Execute** to run it.

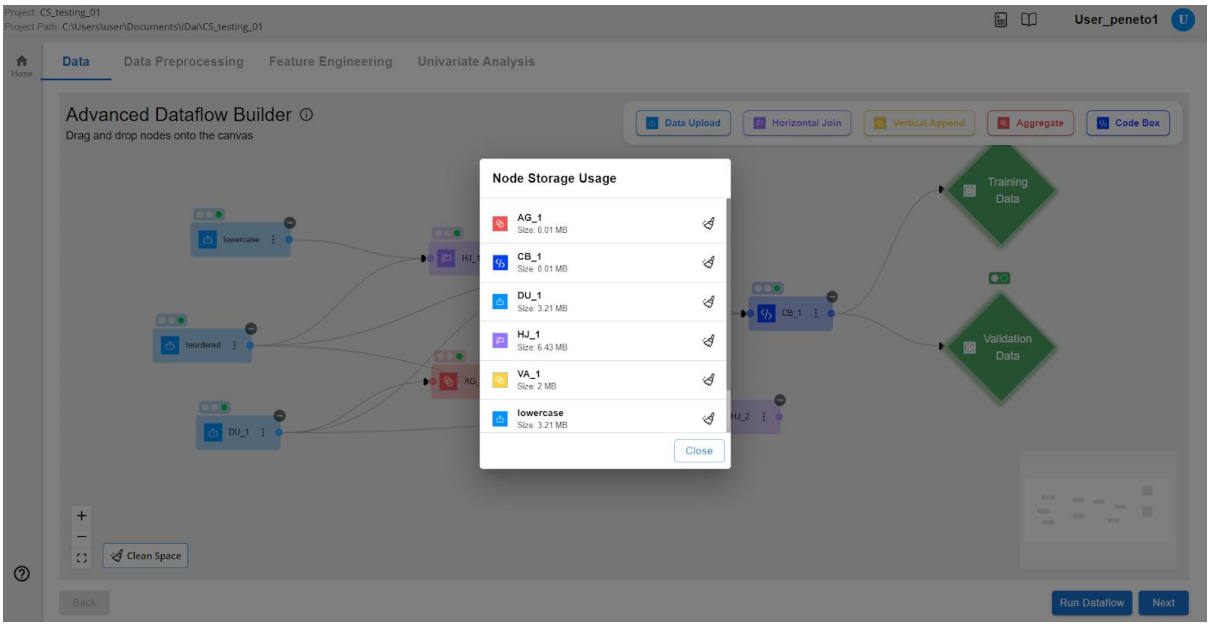


Once the desired workflow is complete, the user can connect its final node to the **Training Data** and **Validation Data** components.
This creates two dataframes: one for training and one for validation. Later in the process (during BYD), the validation dataset can be used directly without re-

uploading or creating a new dataset.

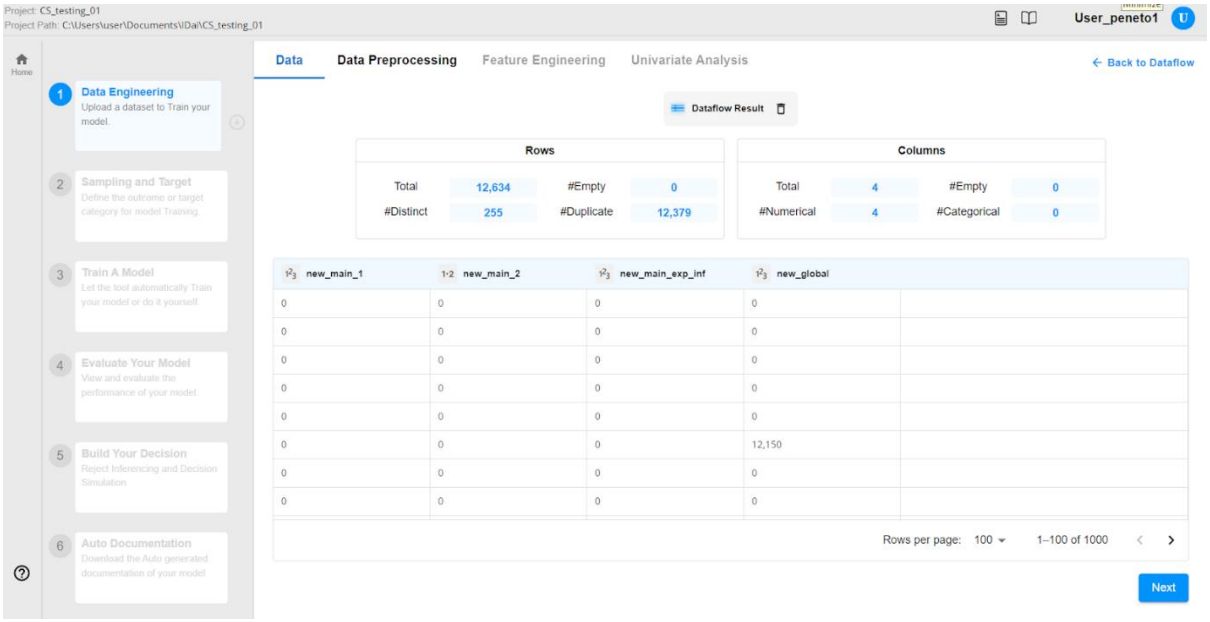


To clear the memory used by each node, the user can select **Clean Space** to free the memory occupied by that node.



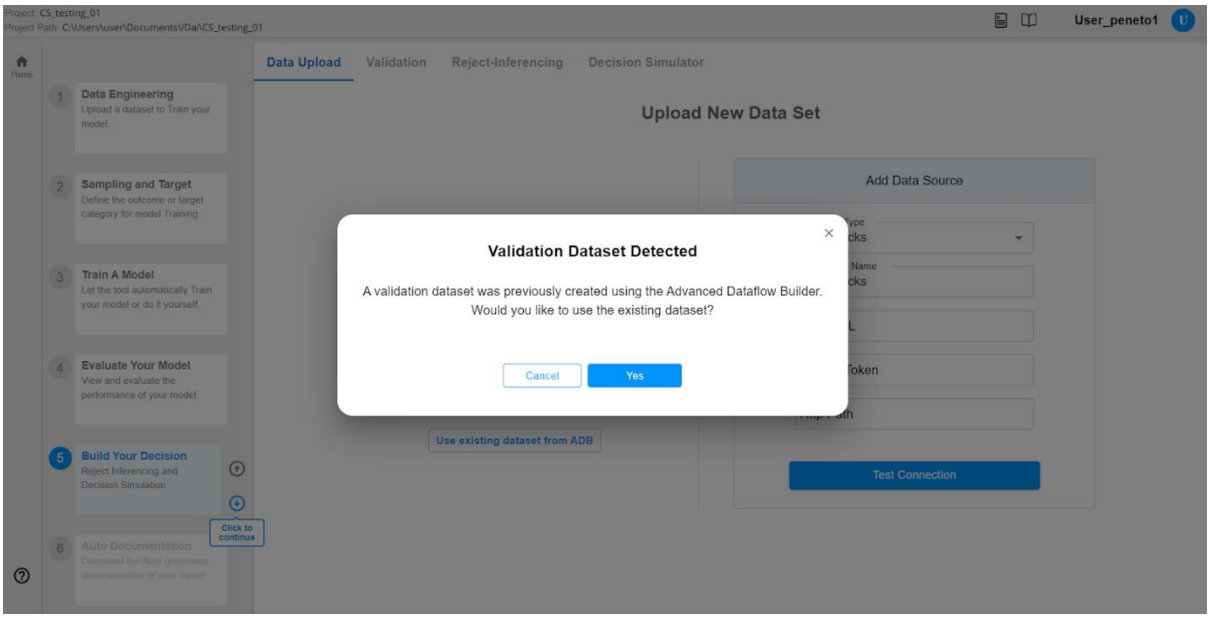
Once the user clicks ‘Next’ and the input file has been uploaded, a summary record count will be displayed along with the first 1000 rows’ actual data. Do

check that the row and column count match with the actual input. The screen also provides information on empty and duplicate rows & columns.

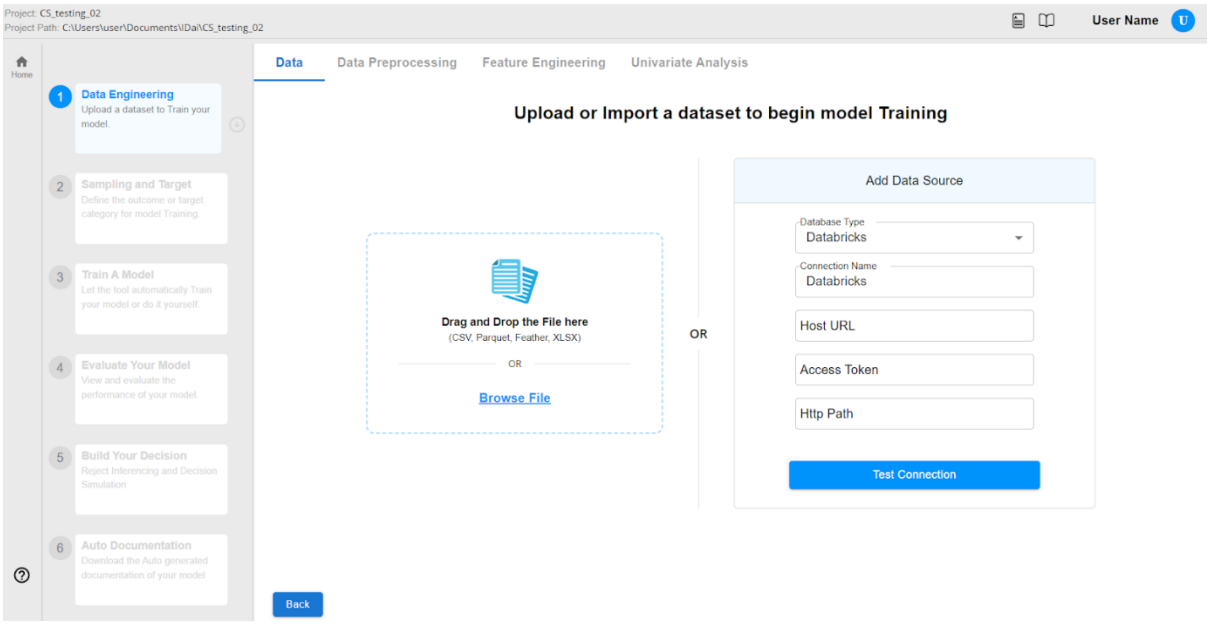


The validation dataset can be used during BYD. A prompt will appear, allowing the user to select the validation dataset from the Dataflow Builder.

The user can also click **Use Existing Dataset from ADB** to import the validation dataset directly from the Dataflow Builder.



If user selects **'Simple Upload'** they are directed to this page

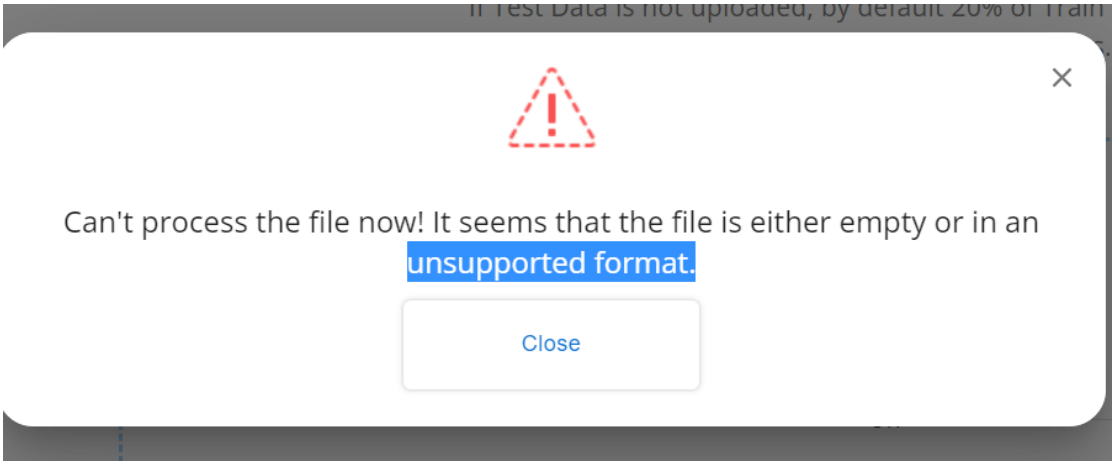


A) ID.ai supports data upload in any of the following formats:

1. CSV
2. Excel
3. Feather
4. Parquet
5. Flat text files with delimiters

The minimum number of records needed is 200 with at least 3 columns and a minimum target variable count of 10.

If an invalid file is used for input, the following error message pops up.

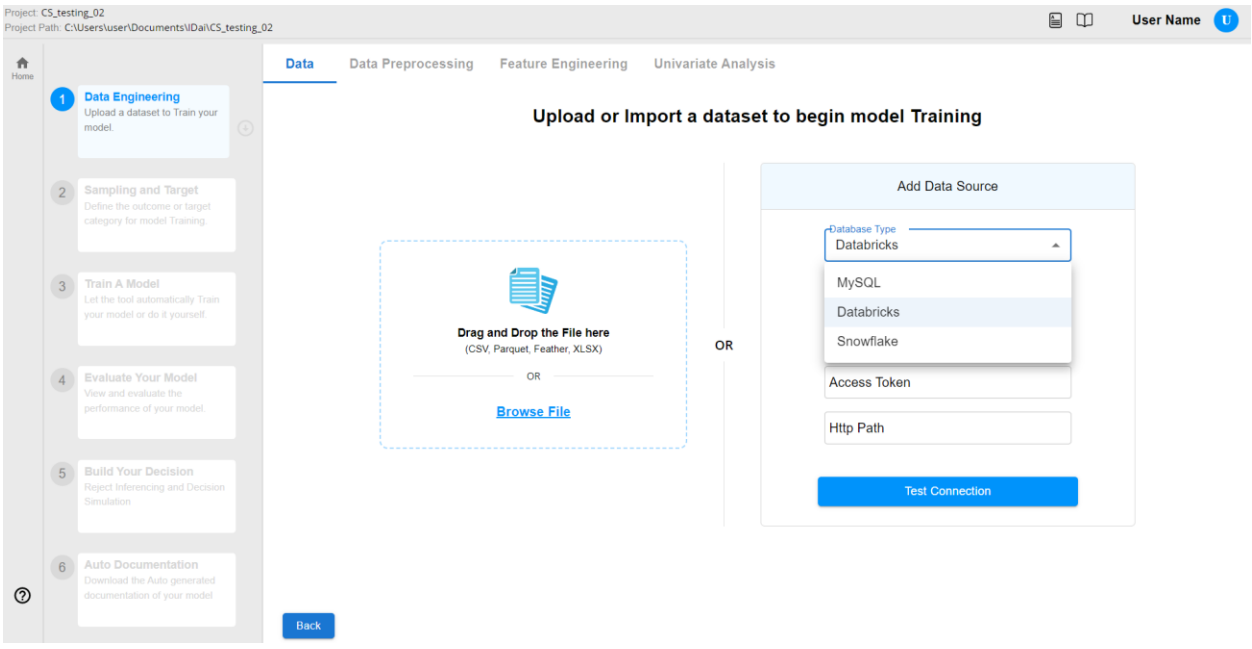


User needs to click on “Close” in the pop-up and go back and upload a dataset in one of the correct formats.

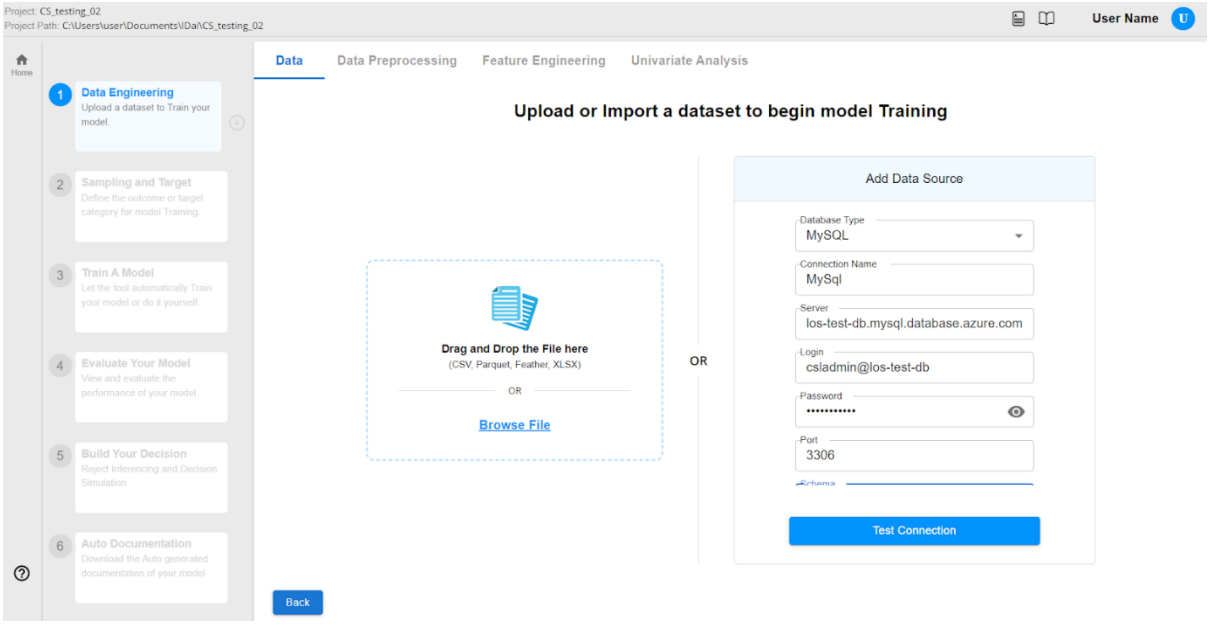
Import from a Data Source

B) ID.ai supports data imports from Databricks, MYSQL and Snowflakes

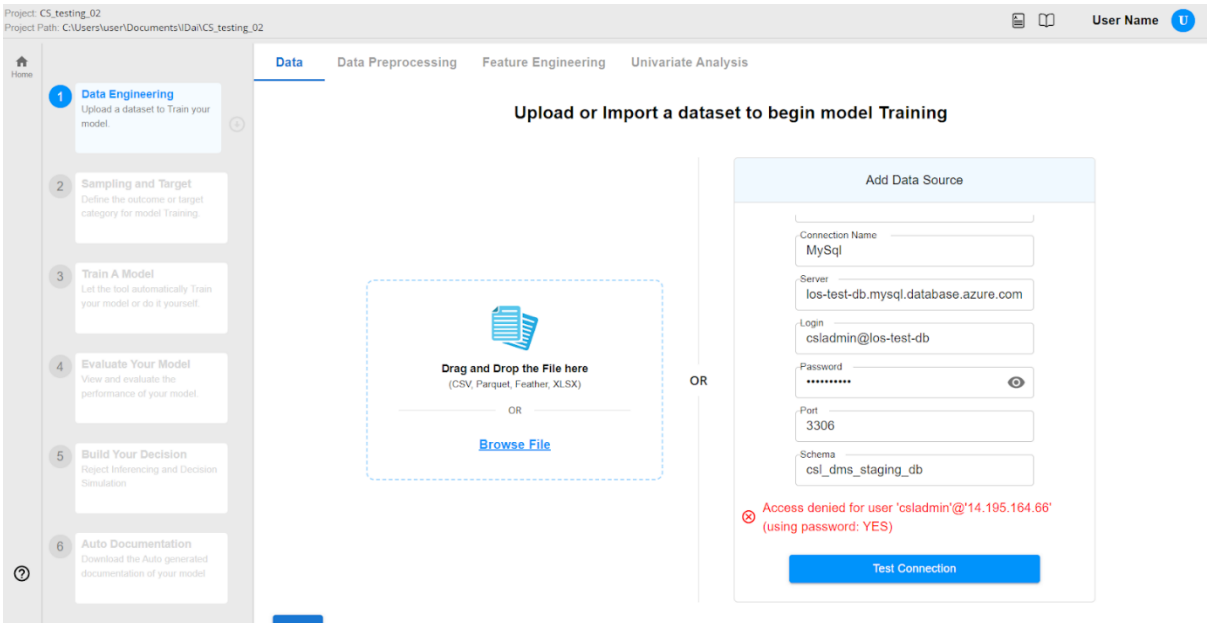
1. Select **Database Type** from the given options.



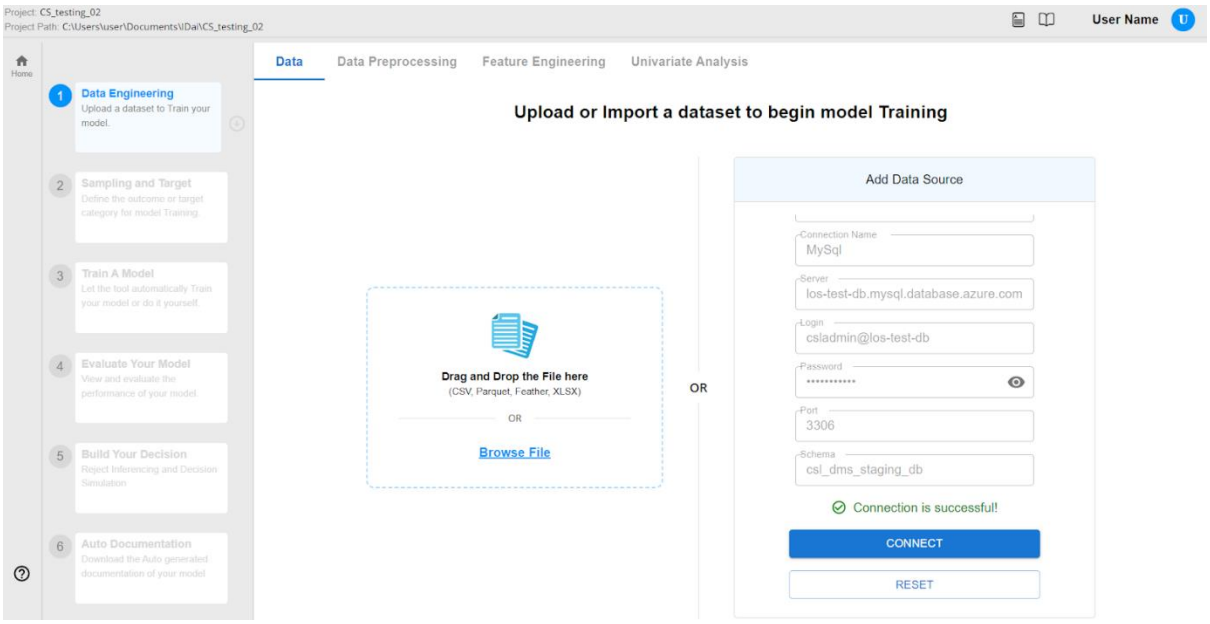
2. Add Credentials and Click on **Test Connection**.



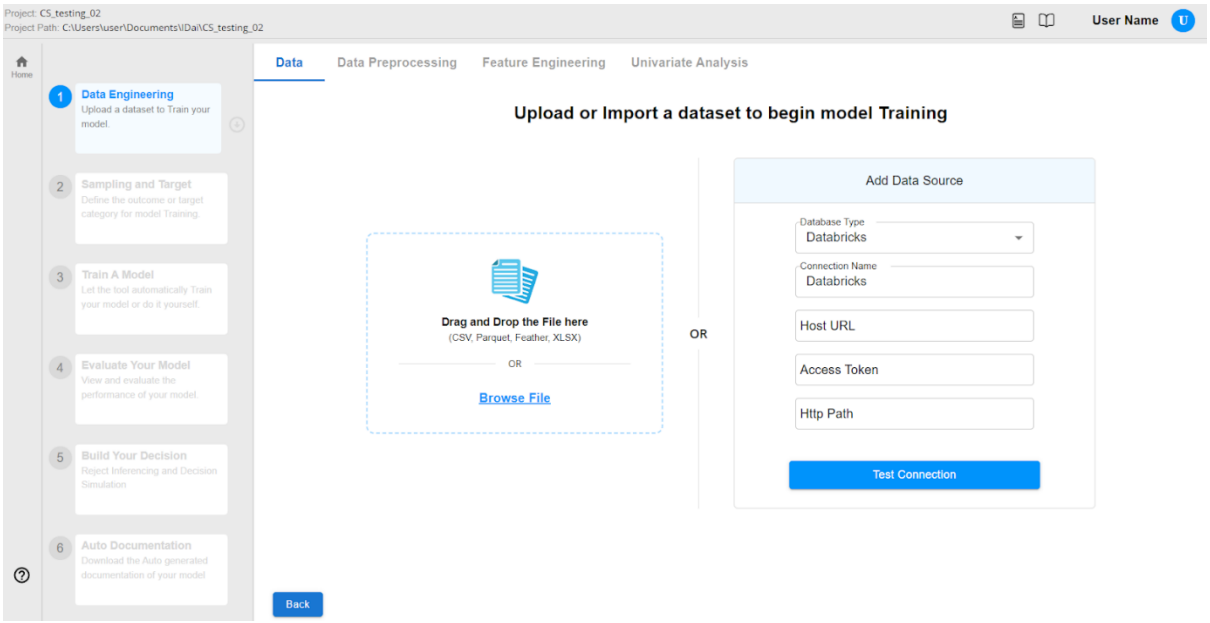
If any of the credentials used are invalid then user will be met with an appropriate error message as shown below upon clicking '**Test Connection**'



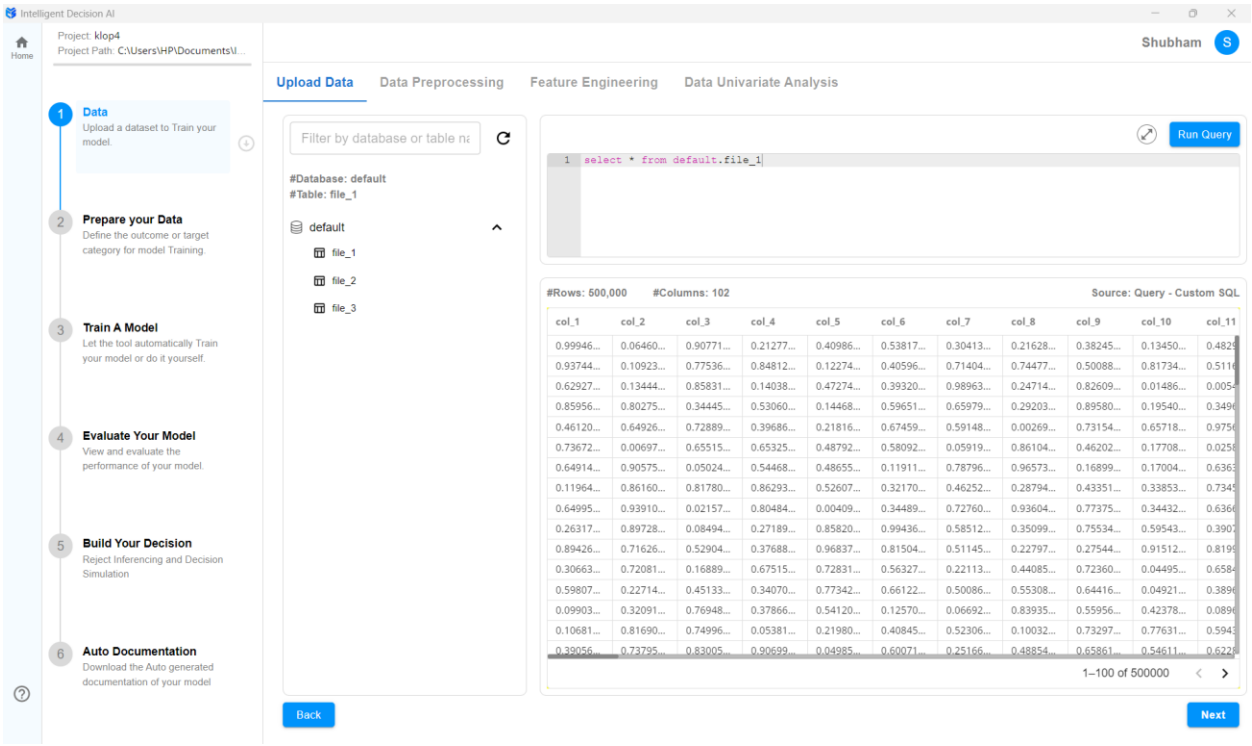
If the credentials are valid, the user will see the following screen:



Selecting '**Reset**' will revert the screen to its original state and clear the existing credentials.



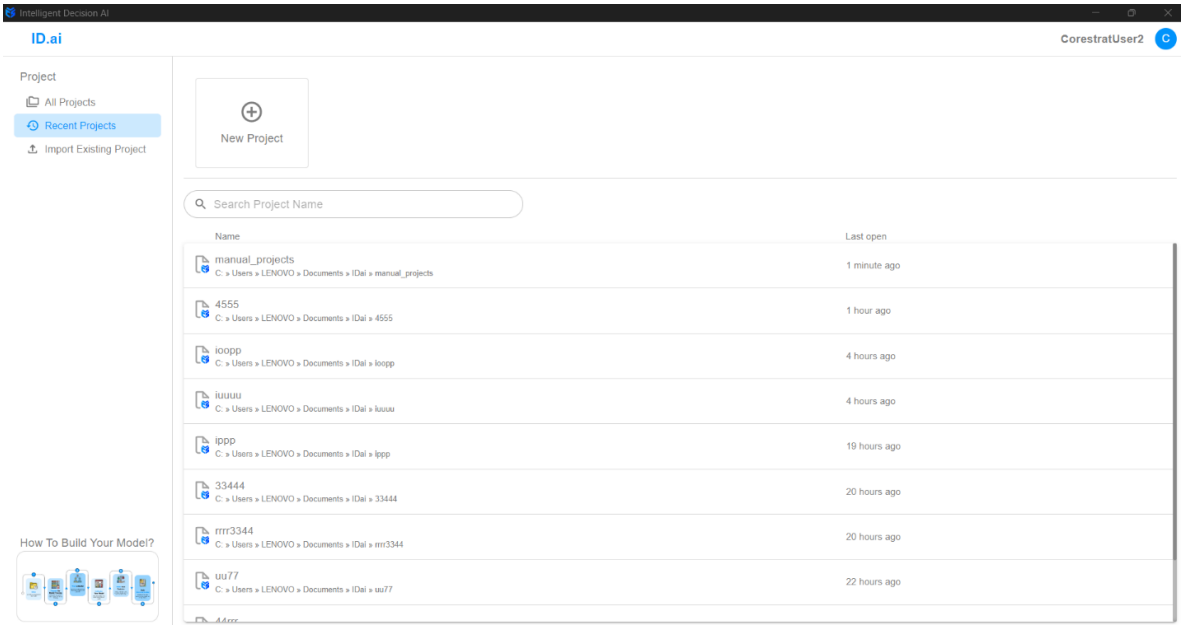
3. Write your SQL query to get and view the data. Click ***Next***, once you are content with dataset and want to use it for model building.



User needs to click on “**Close**” in the pop-up and go back and upload a dataset in one of the correct formats.

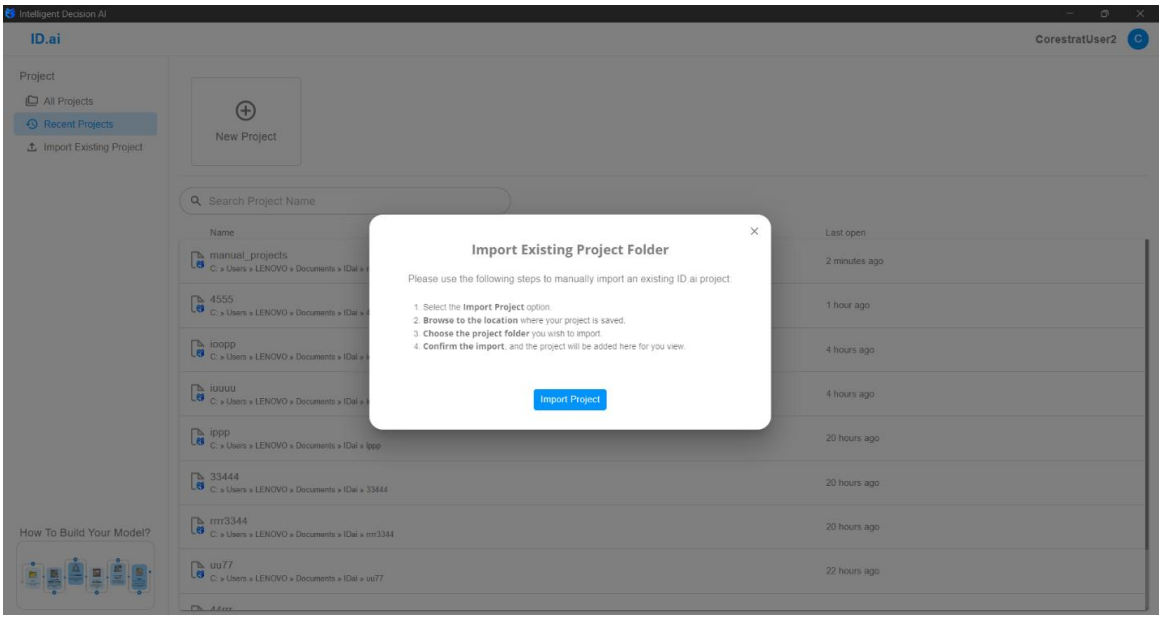
Existing Projects

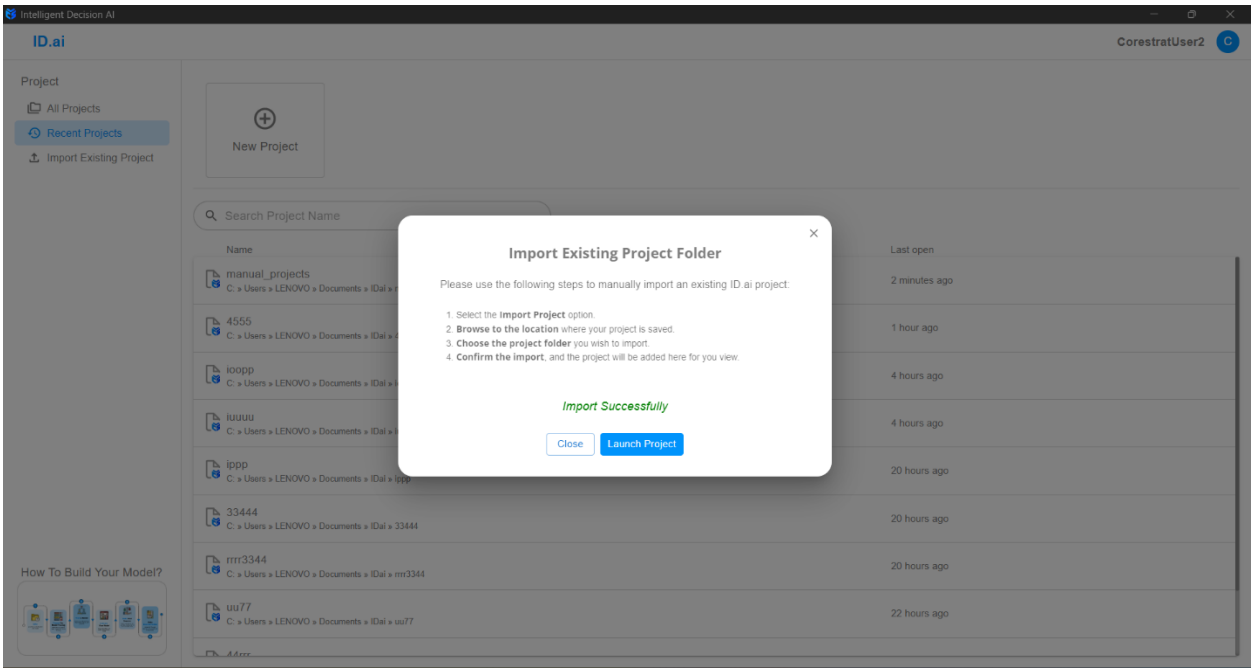
1. The list of recent projects worked on by the user earlier is provided; user also has the option to search the project names



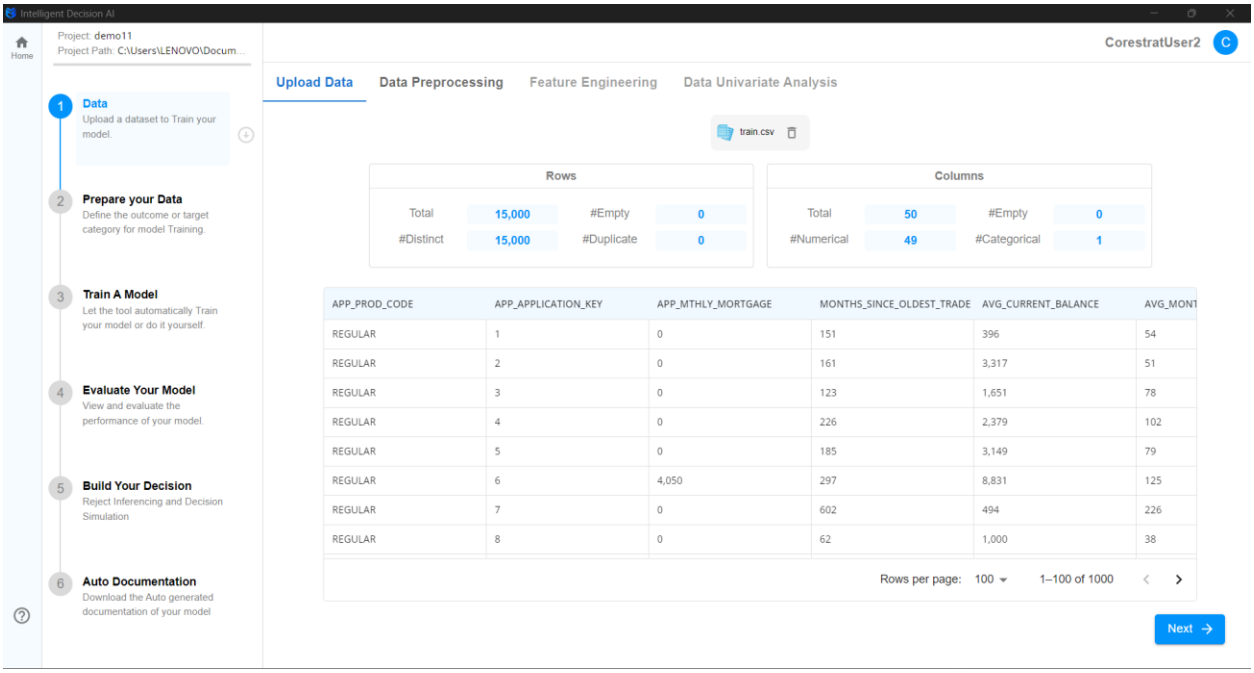
2. For importing a project from a different location user must

   a. Select the Import Existing Project option.
   b. Browse to the location where your project is saved.
   c. Choose the project folder you wish to import
   d. Confirm the import, and the project will be added here for you to view.



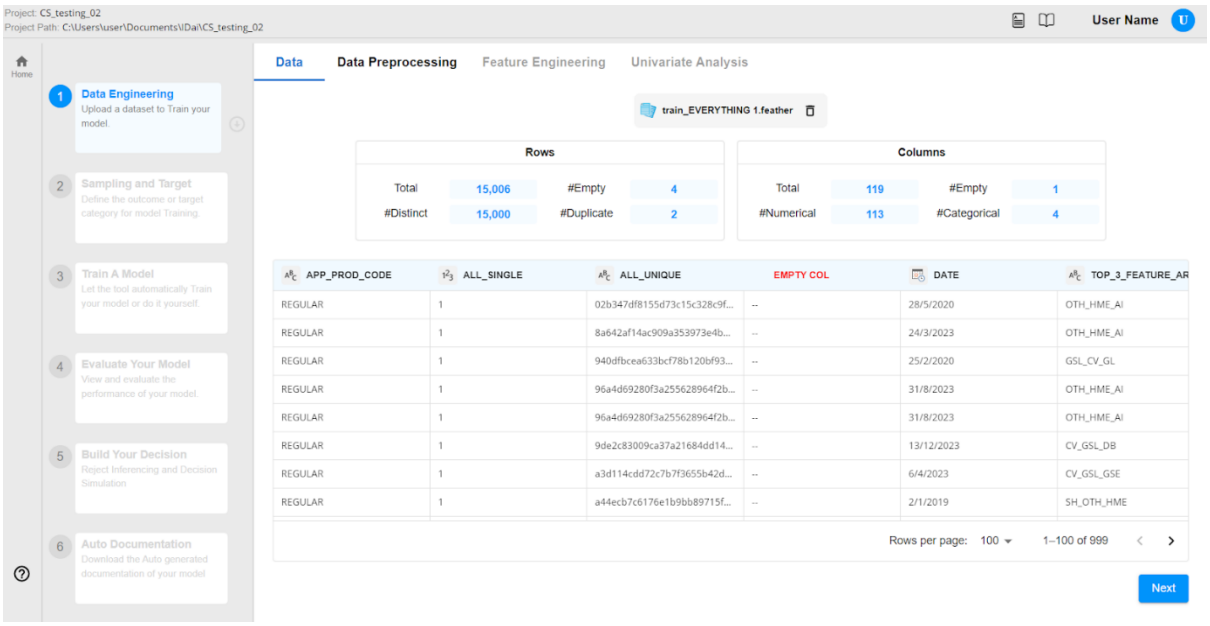


3. Once user clicks on "**Launch Project**" then imported project is successfully loaded

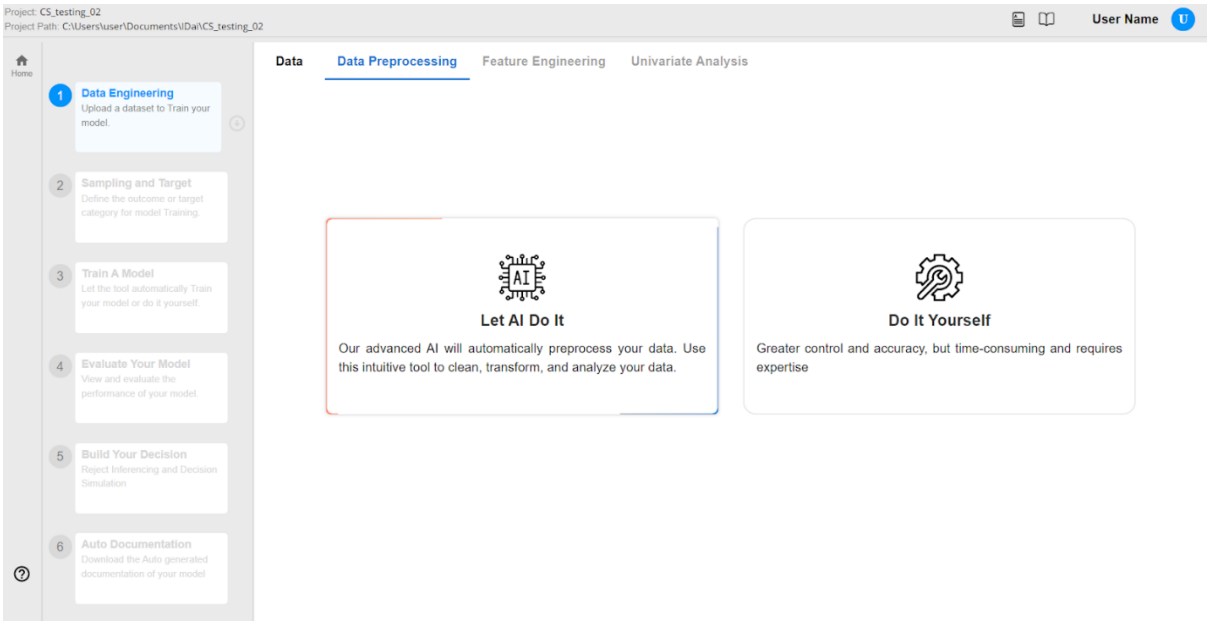


Input Data Management

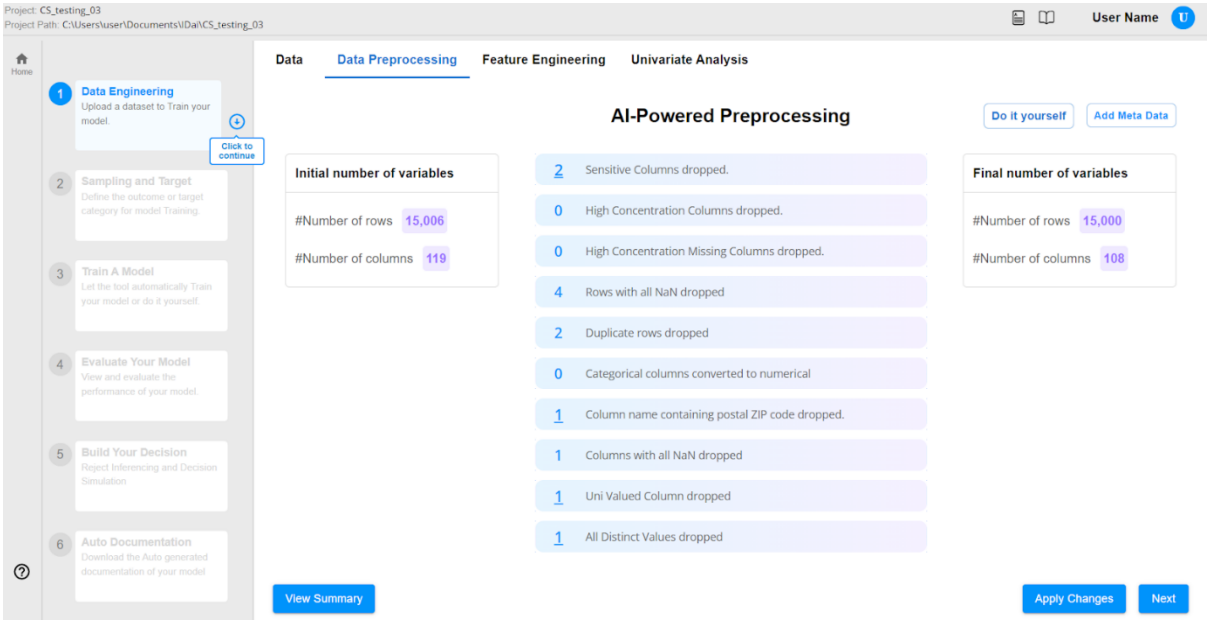
Once the input file has been uploaded, a summary record count will be displayed along with the first 1000 rows' actual data. Do check that the row and column count match with the actual input. The screen also provides information on empty and duplicate rows & columns.



Upon clicking the '**Next**' button the user is redirected to the Data Preprocessing screen, here the user can choose between '**Let AI Do It**' and '**Do It Yourself**'.



If '**Let AI Do It**' is selected, preprocessing is performed automatically using the default settings from the preprocessing screen.

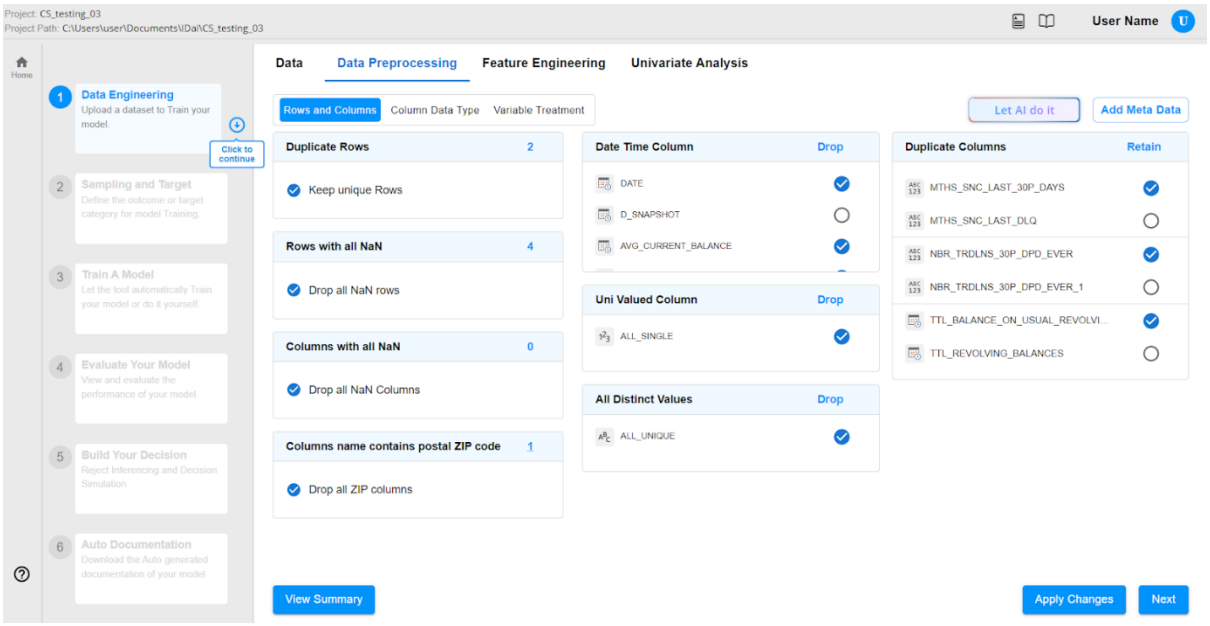


If '**Do It Yourself**' is selected, the following screen will be displayed.

ID.ai offers the following features on the Data Preprocessing screen:

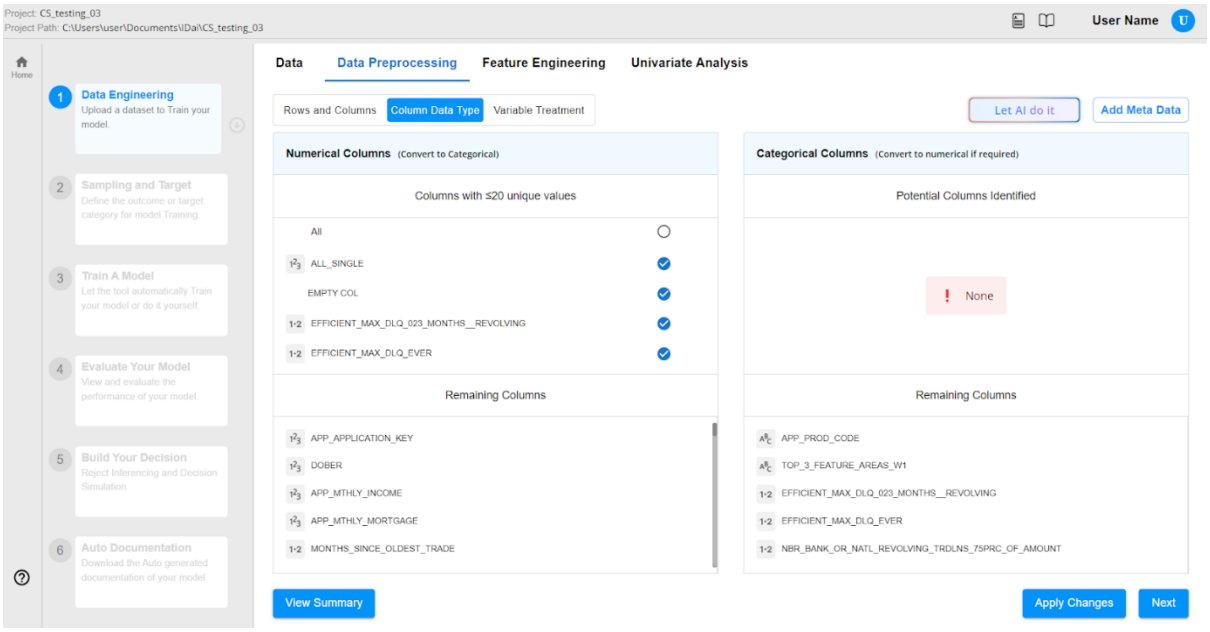
These features are grouped under the "**Rows and Columns**" tab, as shown on the screen below:

1. Remove duplicate rows, rows with all NaN values, columns with all NaN values, or columns with specific keywords (e.g., ZIP).
2. Identify and handle Date-Time columns, uni-valued columns, and columns with all distinct values, with options to drop them.
3. Detect duplicate columns and allow the user to retain one from each group.
4. Let AI handle preprocessing automatically or proceed with manual configuration using the available toggles and options.



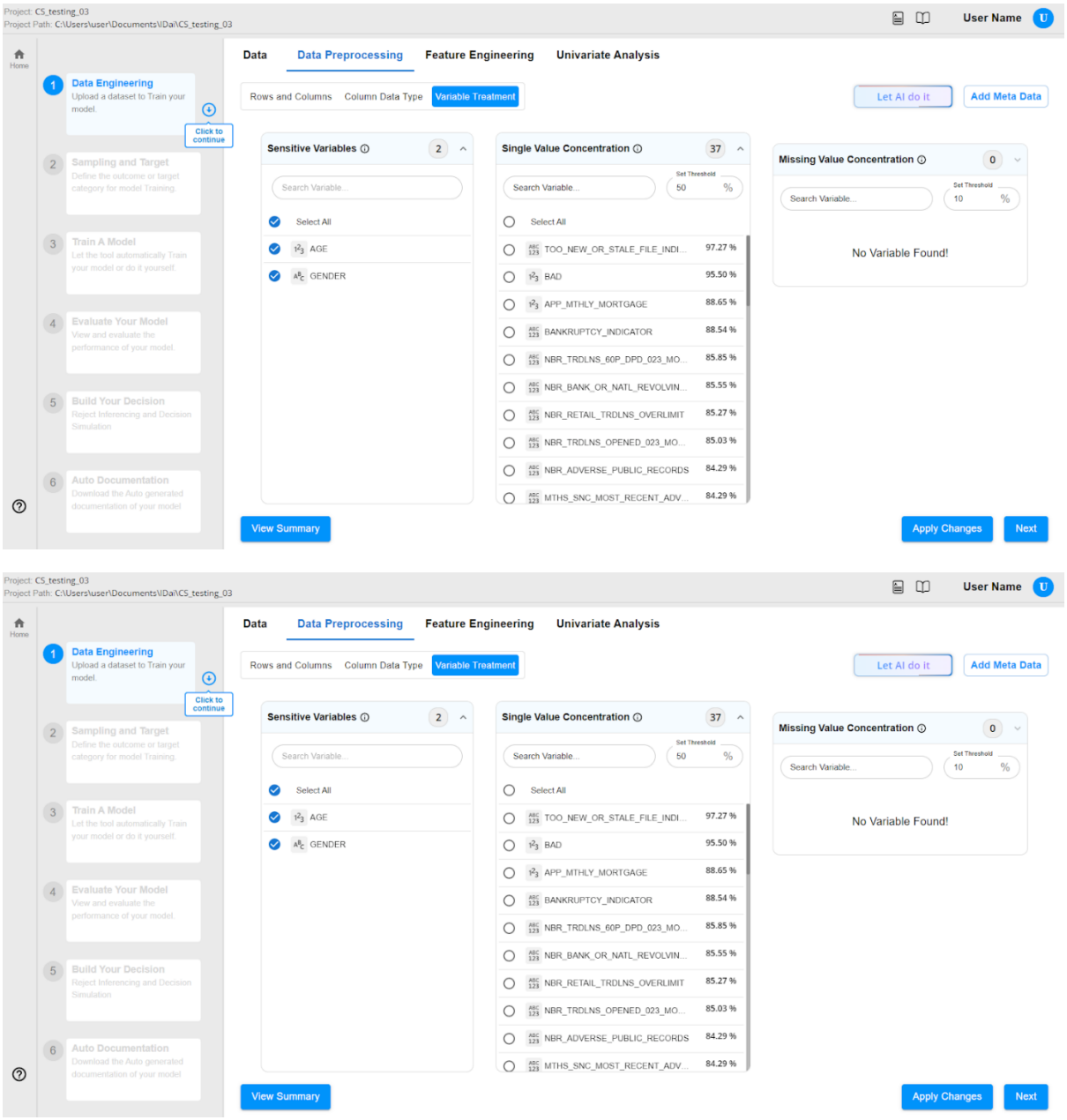
These features are grouped under the "**Column Data Type**" tab, as shown on the screen below:

1. Convert numerical columns to categorical columns and vice versa.
2. Provide a list of numerical columns with fewer than 20 distinct values as likely candidates for categorical treatment.
3. Highlight categorical (text/label-type) columns that may be suitable for numerical conversion.



These features are grouped under the "**Variable Treatment**" tab, as shown on the screen below:

1. Identifies sensitive variables such as age and gender that may require special handling for privacy or fairness. Users can select or deselect them as needed.
2. Highlights variables with a single value concentration (e.g., one value appearing in over 50% of rows), which may have low predictive value.
3. Displays variables with high missing value concentration, helping users review and decide on treatment.

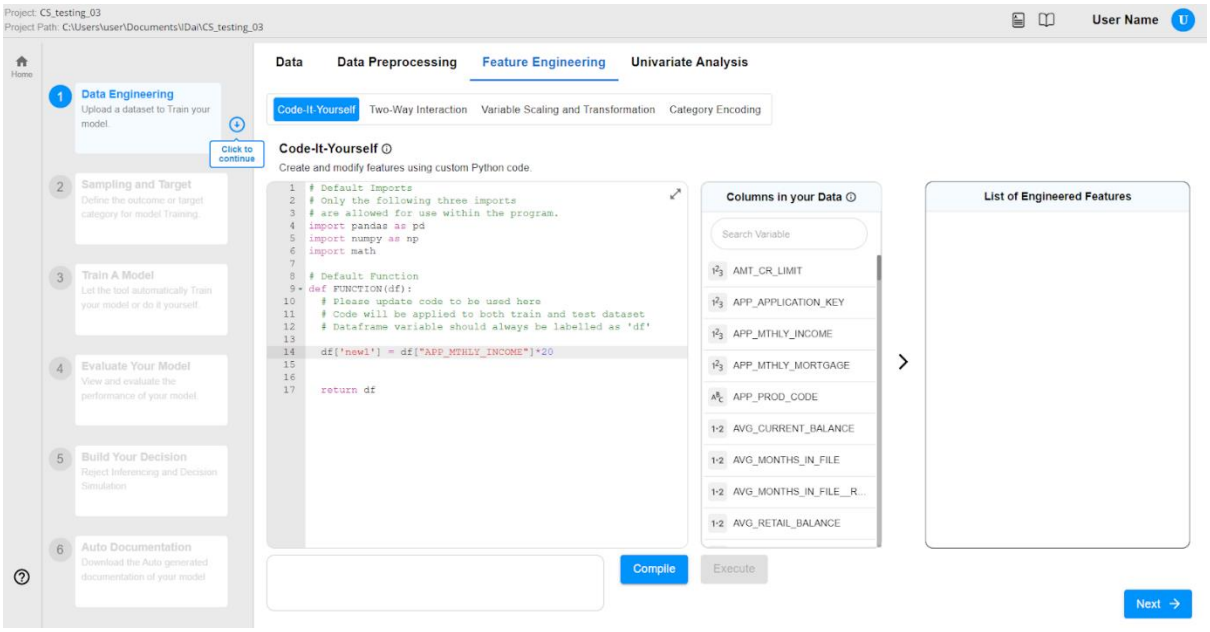


Feature Engineering:

Code-It-Yourself : Perform feature engineering by writing your own custom python code.

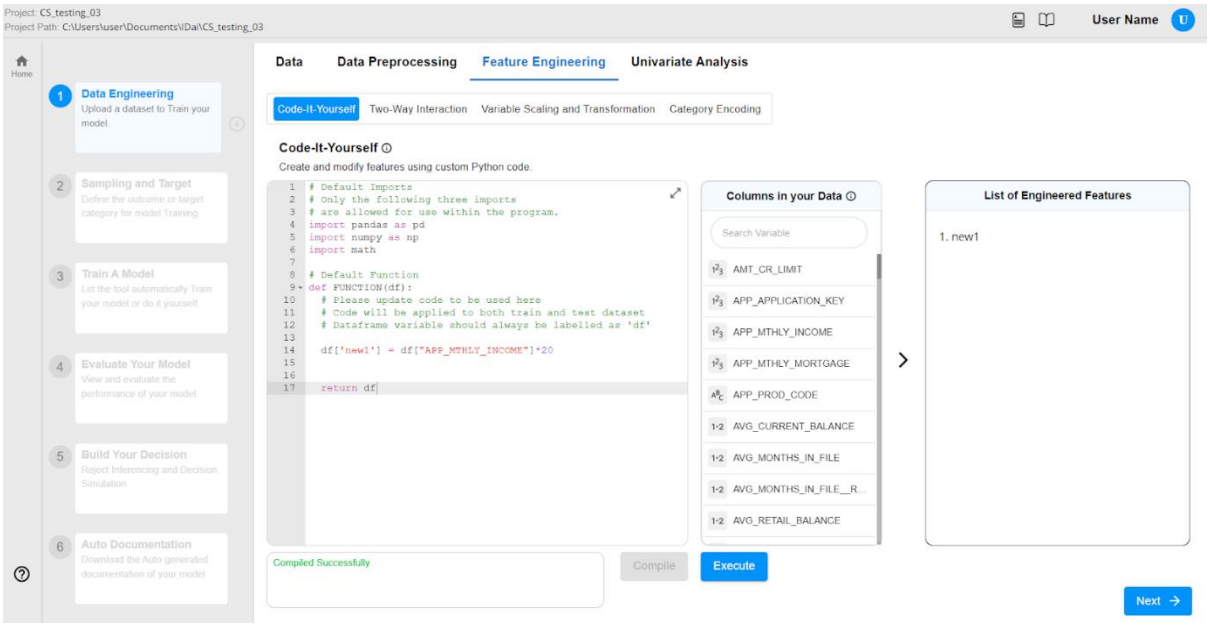
Simply input your code to create or modify features based on your data.

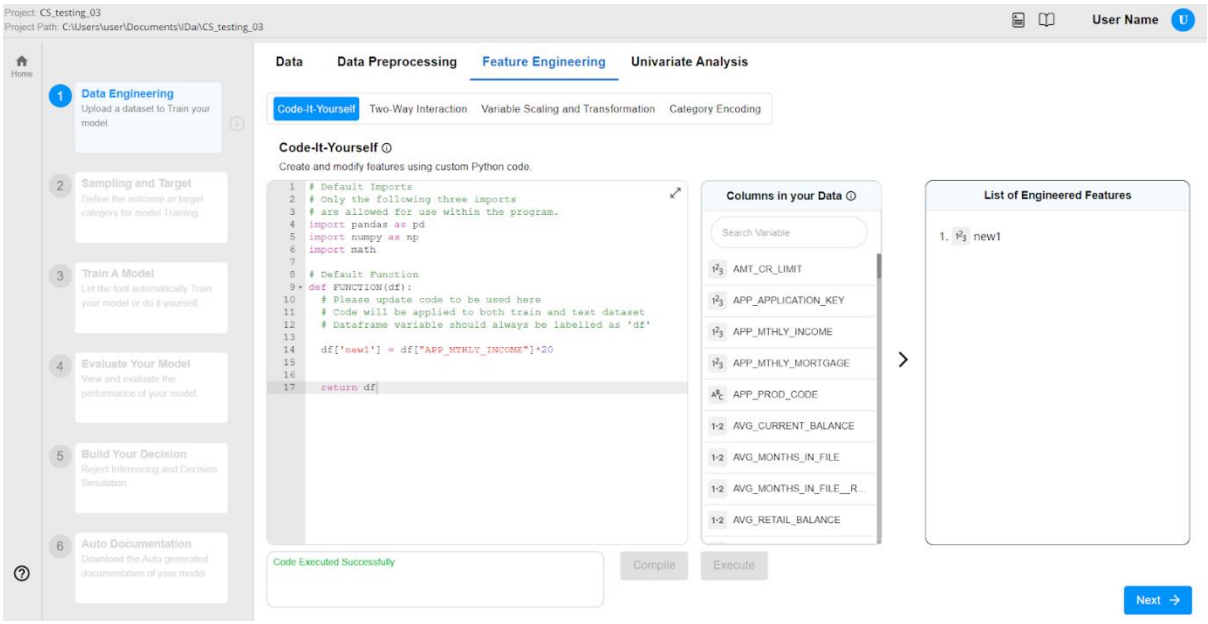
This gives you full flexibility to tailor features to your specific needs.



Variable List : Hover over each variable to check the variable type. Double click on the variable to populate it in the code box immediately.

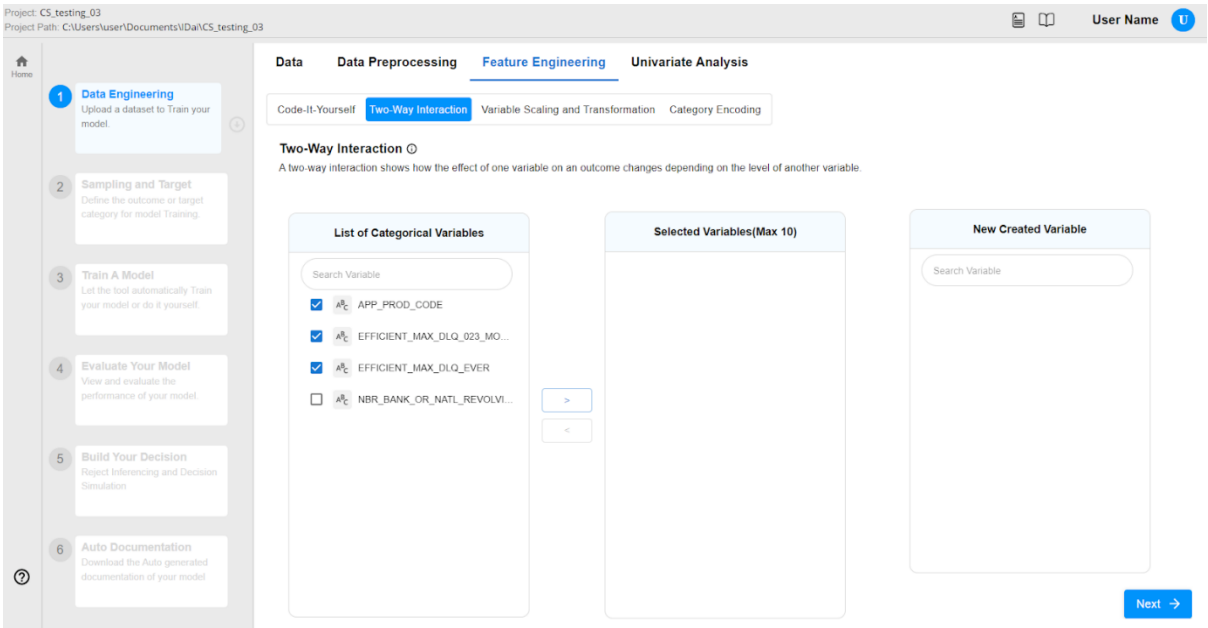
After click on “Compile” and “Execute” button then new variable name is populated in List of Engineering Features

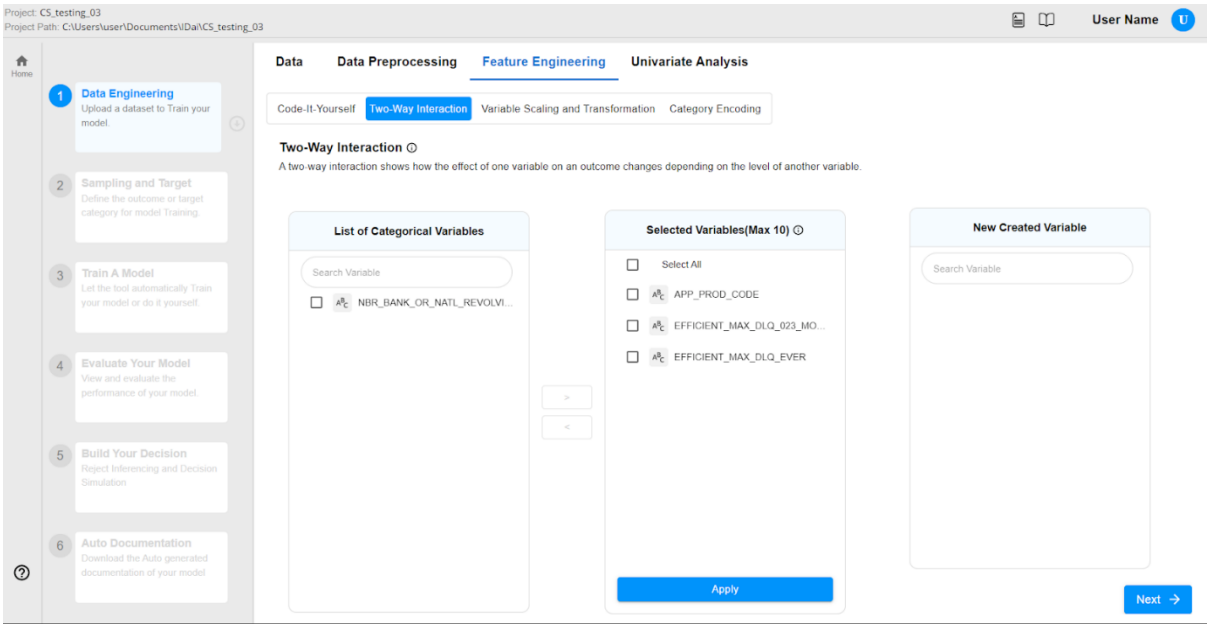




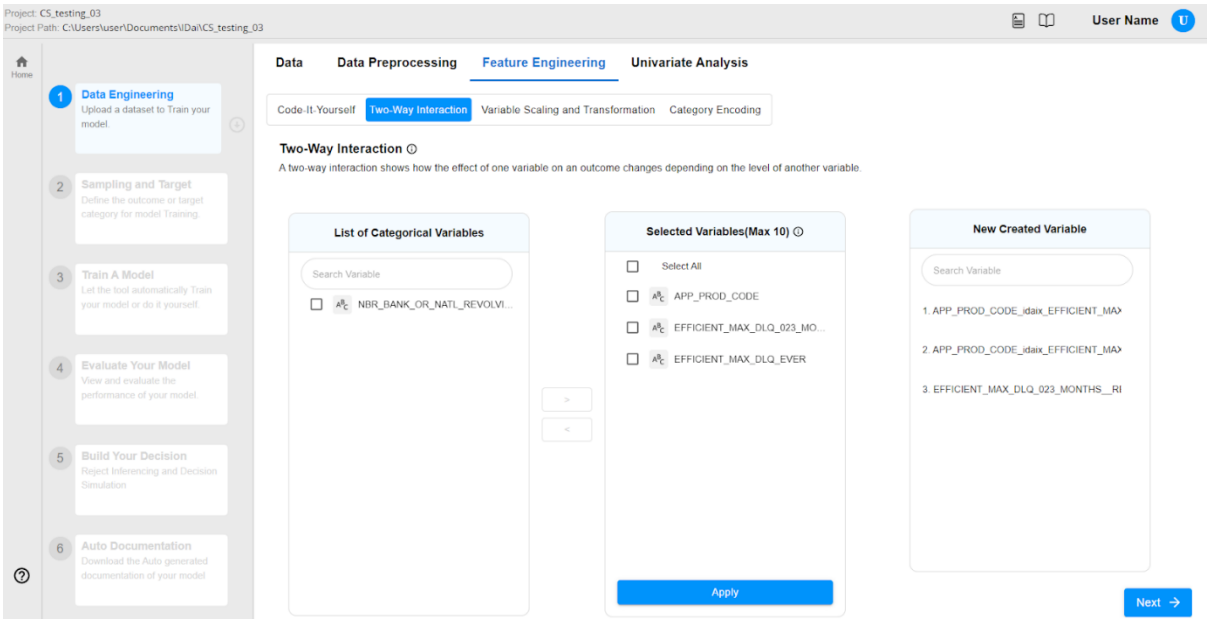
Two-Way Interaction: A two-way interaction shows how the effect of one variable on an outcome changes depending on the level of another variable.

The user selects variables from the ‘List of Categorical Variables’ and clicks the ‘>’ symbol to move them to the ‘Selected Variables’ list.



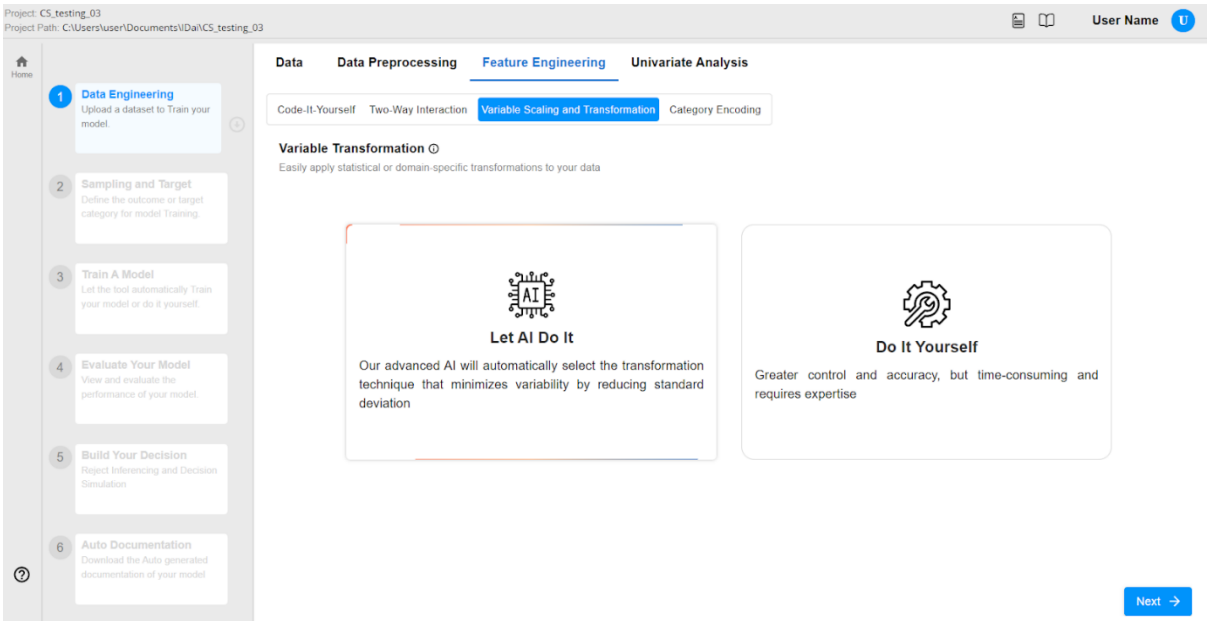


The user selects variables from the '**Selected Variables**' and clicks the '**Apply**' button to move them to the "**New Created Variable**" list.

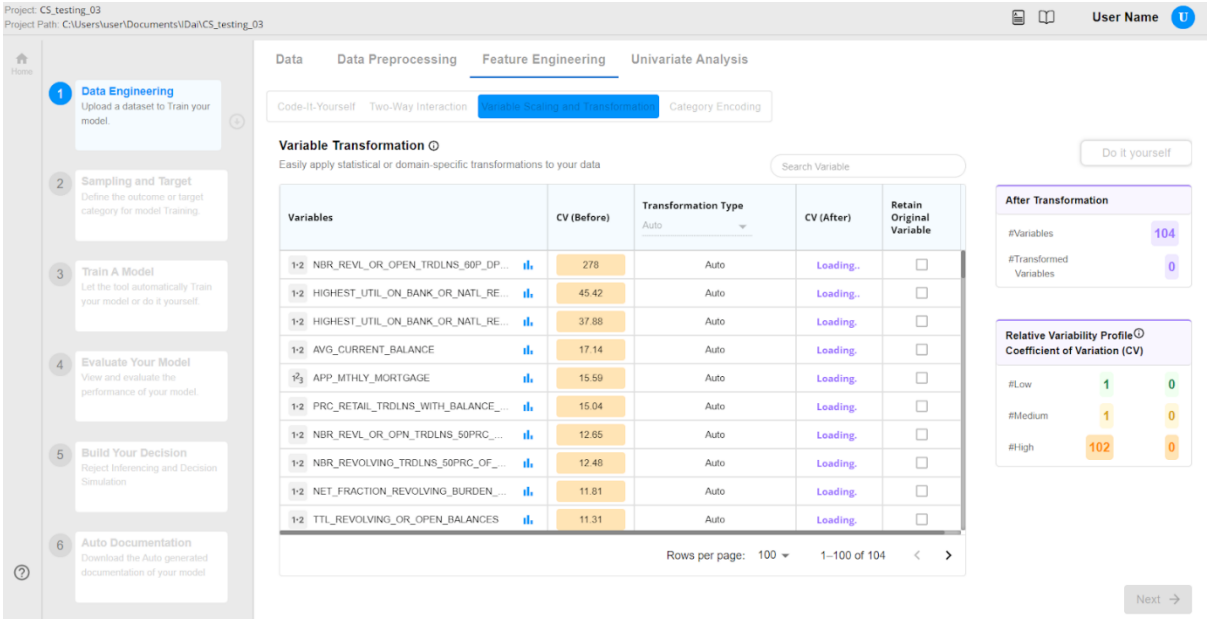


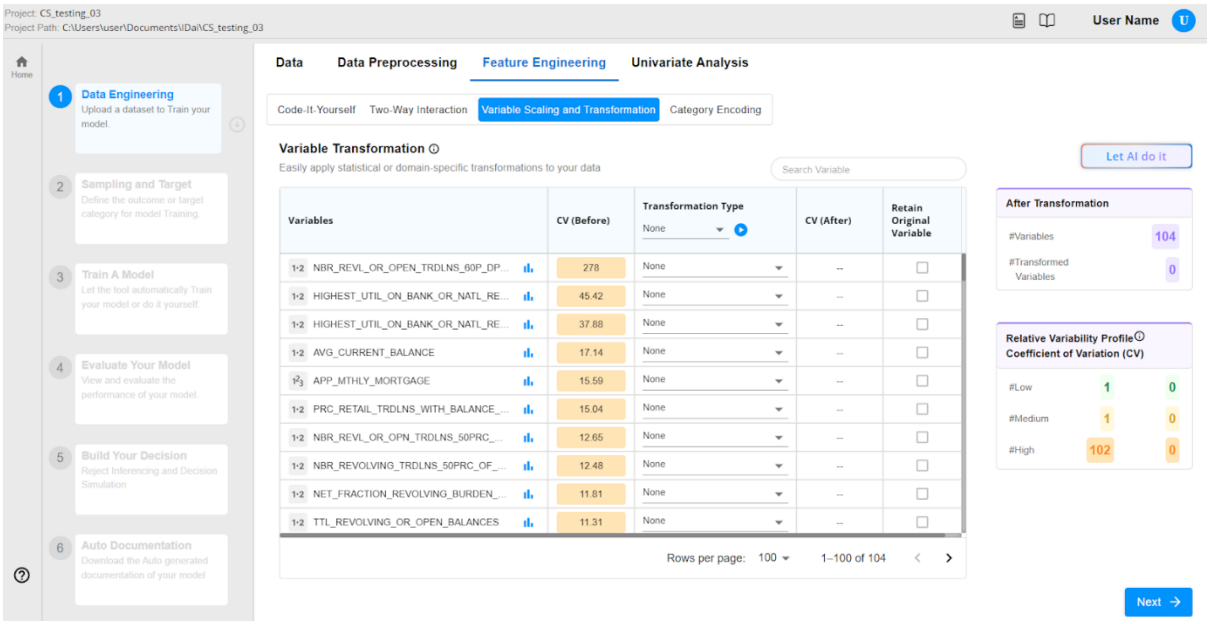
After clicking on "**Next**" button to direct the user to the variable transformation screen

Once user is moved to this section, they are prompted to select one of two options: '**Let AI Do It**' and '**Do it Yourself**'



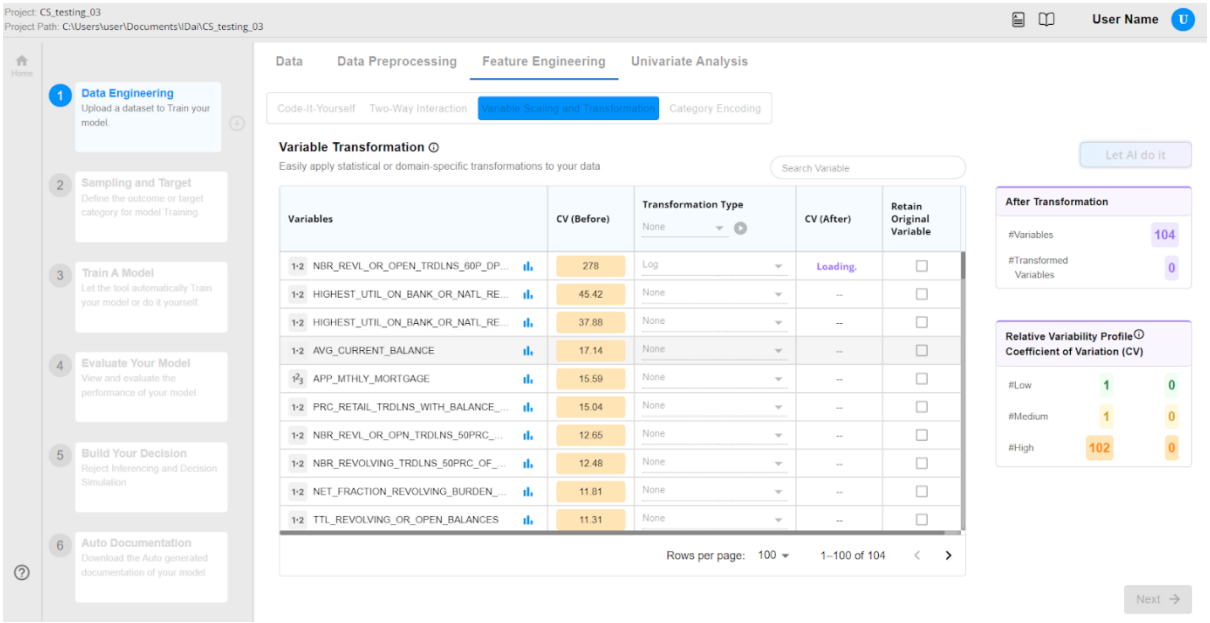
If the user selects '**Let AI Do It**', variables will be transformed using the most suitable transformation type from the available list.



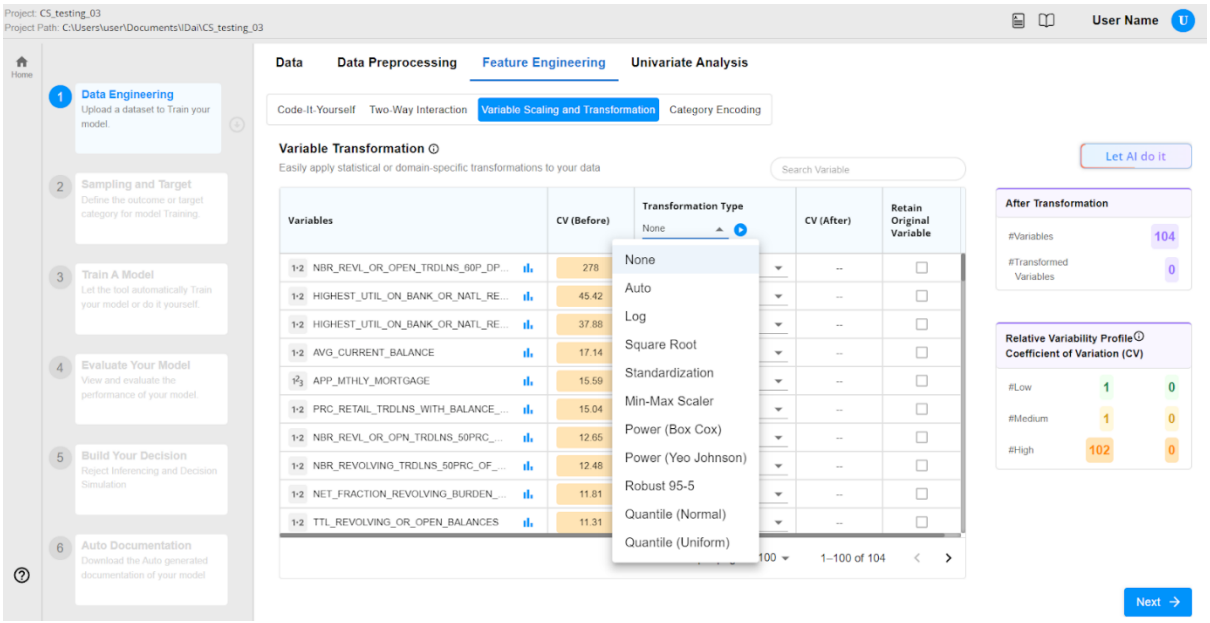


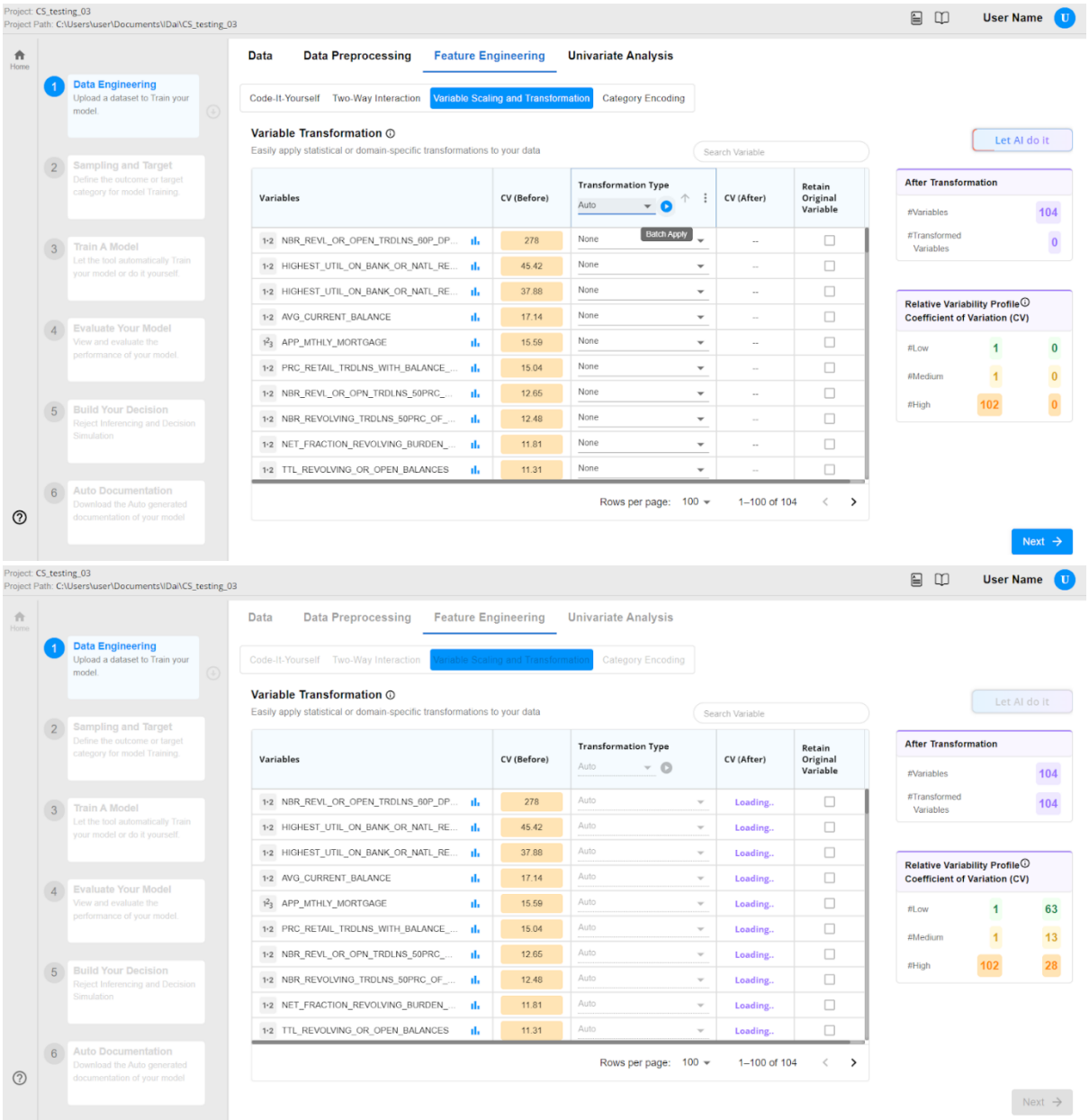
If the user selects '**Do It Yourself**', they can either apply a batch transformation or perform transformations individually.

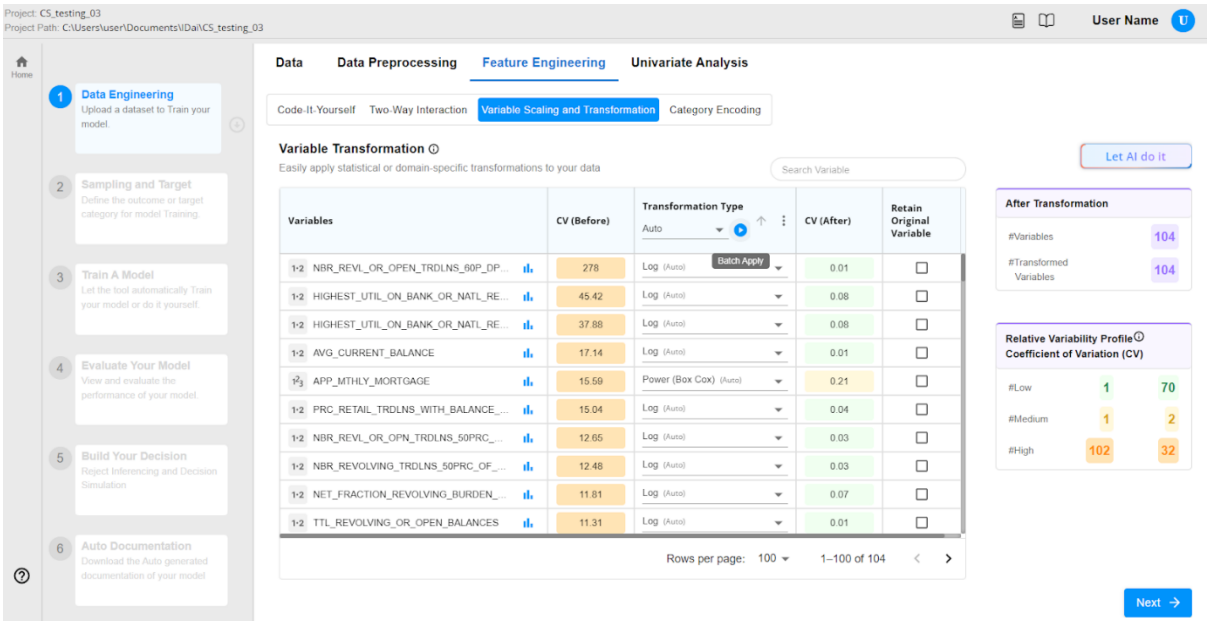
Individual transformations can be applied by choosing a transformation type within the row of the desired variable.



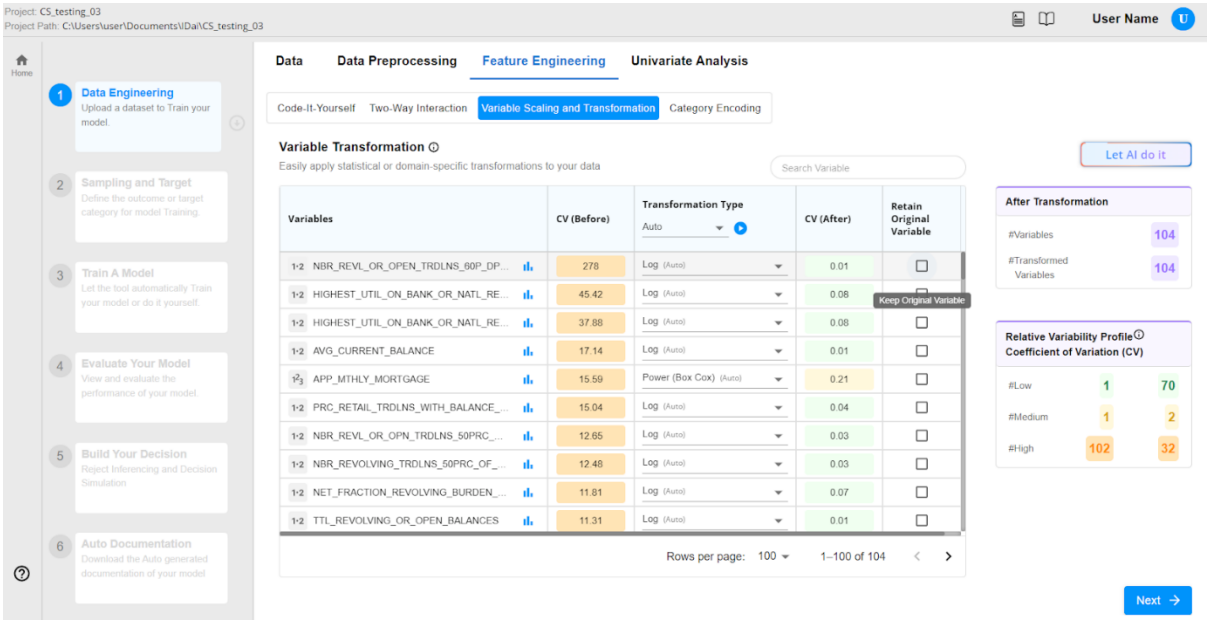
Batch transformations can be applied by selecting a transformation type and clicking the blue play button on the right.

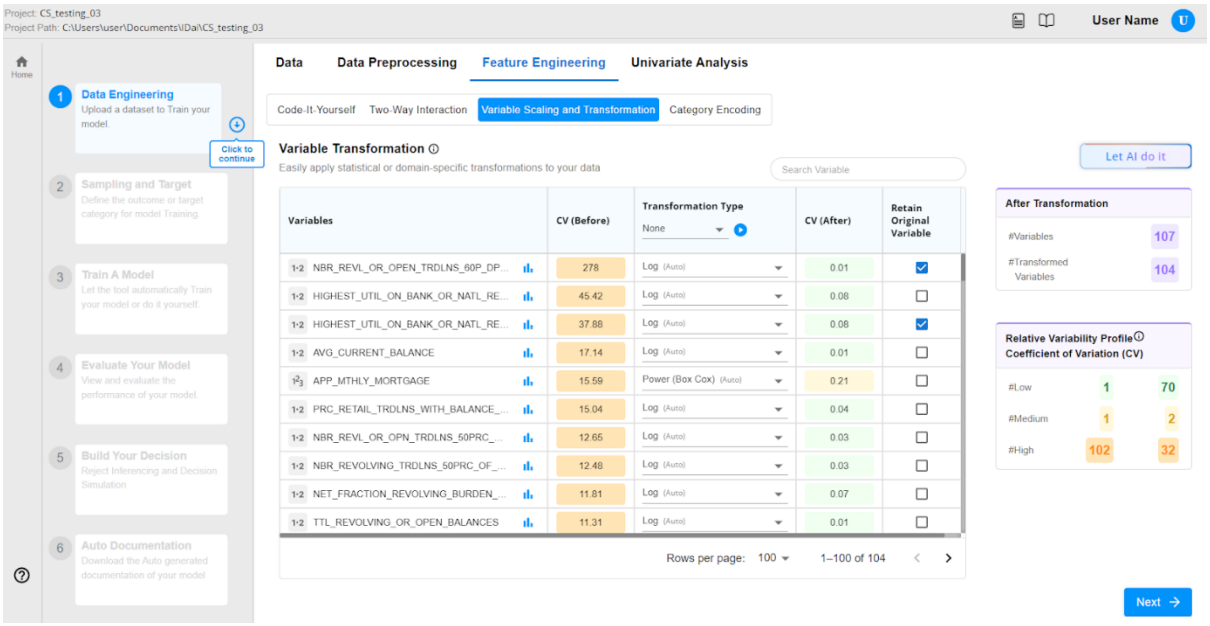



Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Data
Data Preprocessing
Feature Engineering
Univariate Analysis
Code-It-Yourself
Two-Way Interaction
Variable Scaling and Transformation
Category Encoding
Variable Transformation
Easily apply statistical or domain-specific transformations to your data
Search Variable
Let AI do it
Variables
CV (Before)
Transformation Type
Auto
Batch Apply
CV (After)
Retain Original Variable
NBR_REVL_OR_OPEN_TRDLNS_60P_DP... 278 None --
HIGHEST_UTIL_ON_BANK_OR_NATL_RE... 45.42 None --
HIGHEST_UTIL_ON_BANK_OR_NATL_RE... 37.88 None --
AVG_CURRENT_BALANCE 17.14 None --
APP_MTHLY_MORTGAGE 15.59 None --
PRC_RETAIL_TRDLNS_WITH_BALANCE_... 15.04 None --
NBR_REVL_OR_OPN_TRDLNS_50PRC_... 12.65 None --
NBR_REVOLVING_TRDLNS_50PRC_OF_... 12.48 None --
NET_FRACTION_REVOLVING_BURDEN_... 11.81 None --
TTL_REVOLVING_OR_OPEN_BALANCES 11.31 None --
Rows per page: 100
1–100 of 104
After Transformation
#Variables 104
#Transformed Variables 0
Relative Variability Profile
Coefficient of Variation (CV)
#Low 1 0
#Medium 1 0
#High 102 0
Next
Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Data
Data Preprocessing
Feature Engineering
Univariate Analysis
Code-It-Yourself
Two-Way Interaction
Variable Scaling and Transformation
Category Encoding
Variable Transformation
Easily apply statistical or domain-specific transformations to your data
Search Variable
Let AI do it
Variables
CV (Before)
Transformation Type
Auto
CV (After)
Retain Original Variable
NBR_REVL_OR_OPEN_TRDLNS_60P_DP... 278 Auto Loading..
HIGHEST_UTIL_ON_BANK_OR_NATL_RE... 45.42 Auto Loading..
HIGHEST_UTIL_ON_BANK_OR_NATL_RE... 37.88 Auto Loading..
AVG_CURRENT_BALANCE 17.14 Auto Loading..
APP_MTHLY_MORTGAGE 15.59 Auto Loading..
PRC_RETAIL_TRDLNS_WITH_BALANCE_... 15.04 Auto Loading..
NBR_REVL_OR_OPN_TRDLNS_50PRC_... 12.65 Auto Loading..
NBR_REVOLVING_TRDLNS_50PRC_OF_... 12.48 Auto Loading..
NET_FRACTION_REVOLVING_BURDEN_... 11.81 Auto Loading..
TTL_REVOLVING_OR_OPEN_BALANCES 11.31 Auto Loading..
Rows per page: 100
1–100 of 104
After Transformation
#Variables 104
#Transformed Variables 104
Relative Variability Profile
Coefficient of Variation (CV)
#Low 1 63
#Medium 1 13
#High 102 28
Next

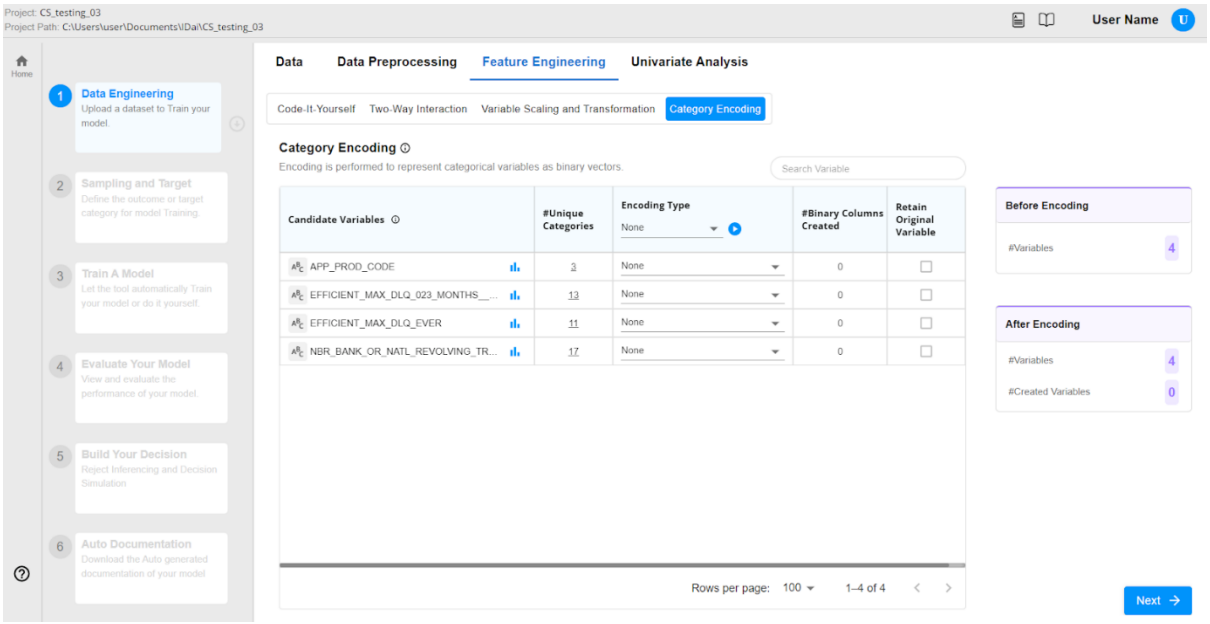


The user can choose to retain both the original and transformed versions of selected variables. If the original is not retained, only the transformed version will be used in subsequent steps.





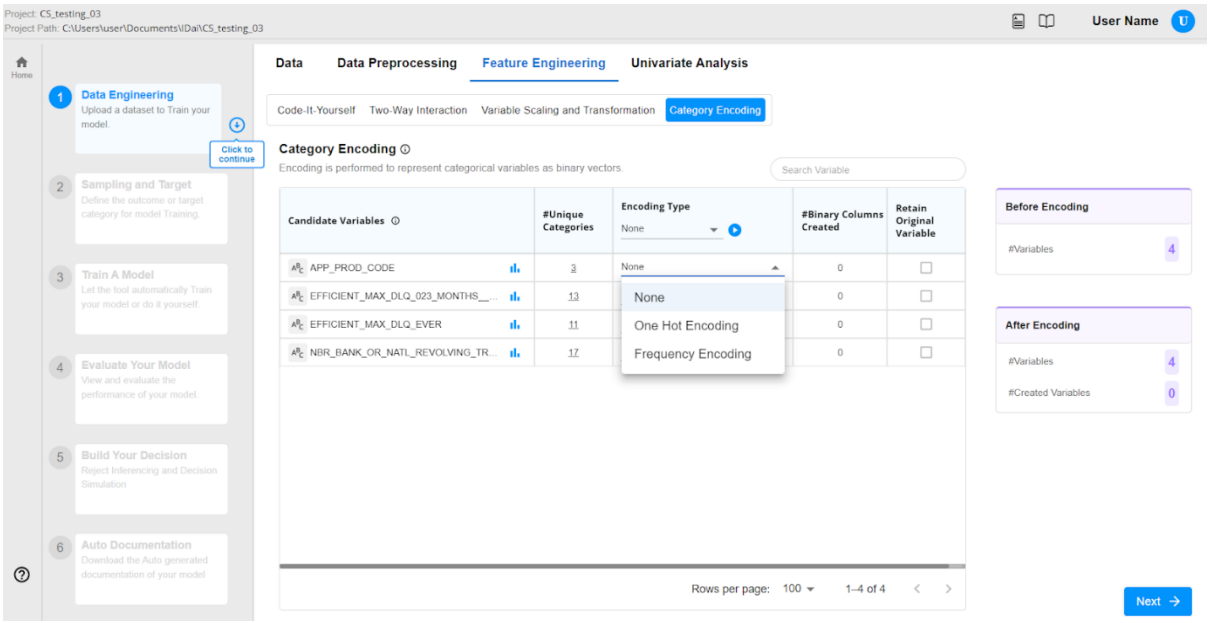
Upon clicking '**Next**', the user is taken to the Category Encoding screen.

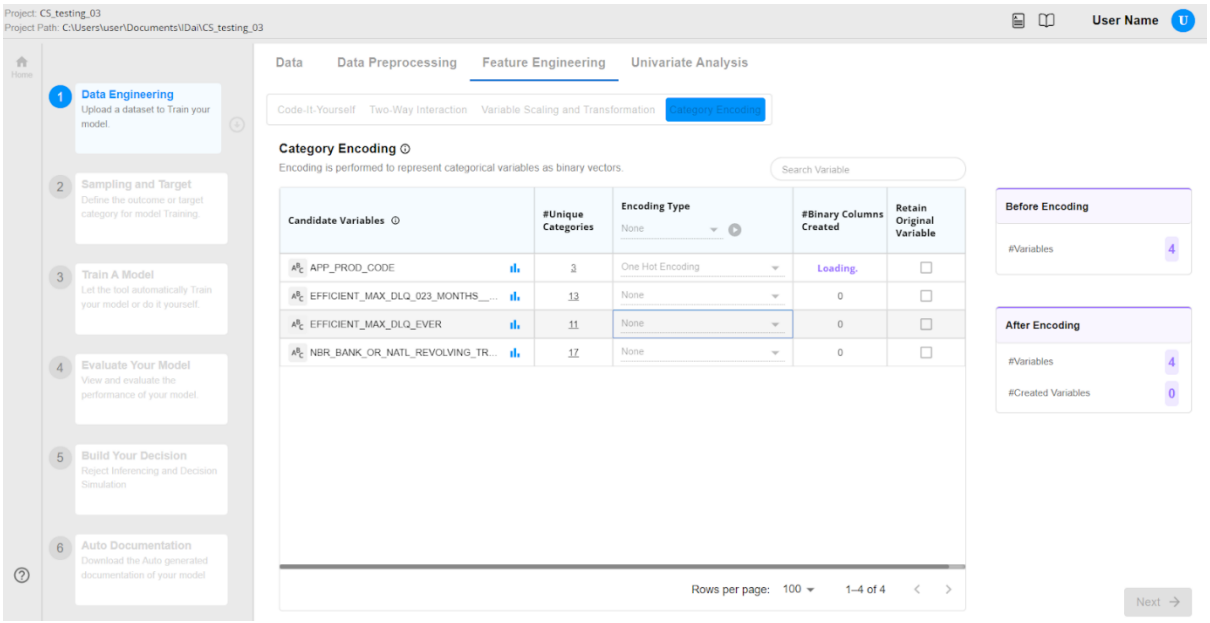


This screen enables the user to apply encoding techniques such as **One-Hot Encoding** and **Frequency Encoding** to categorical columns.

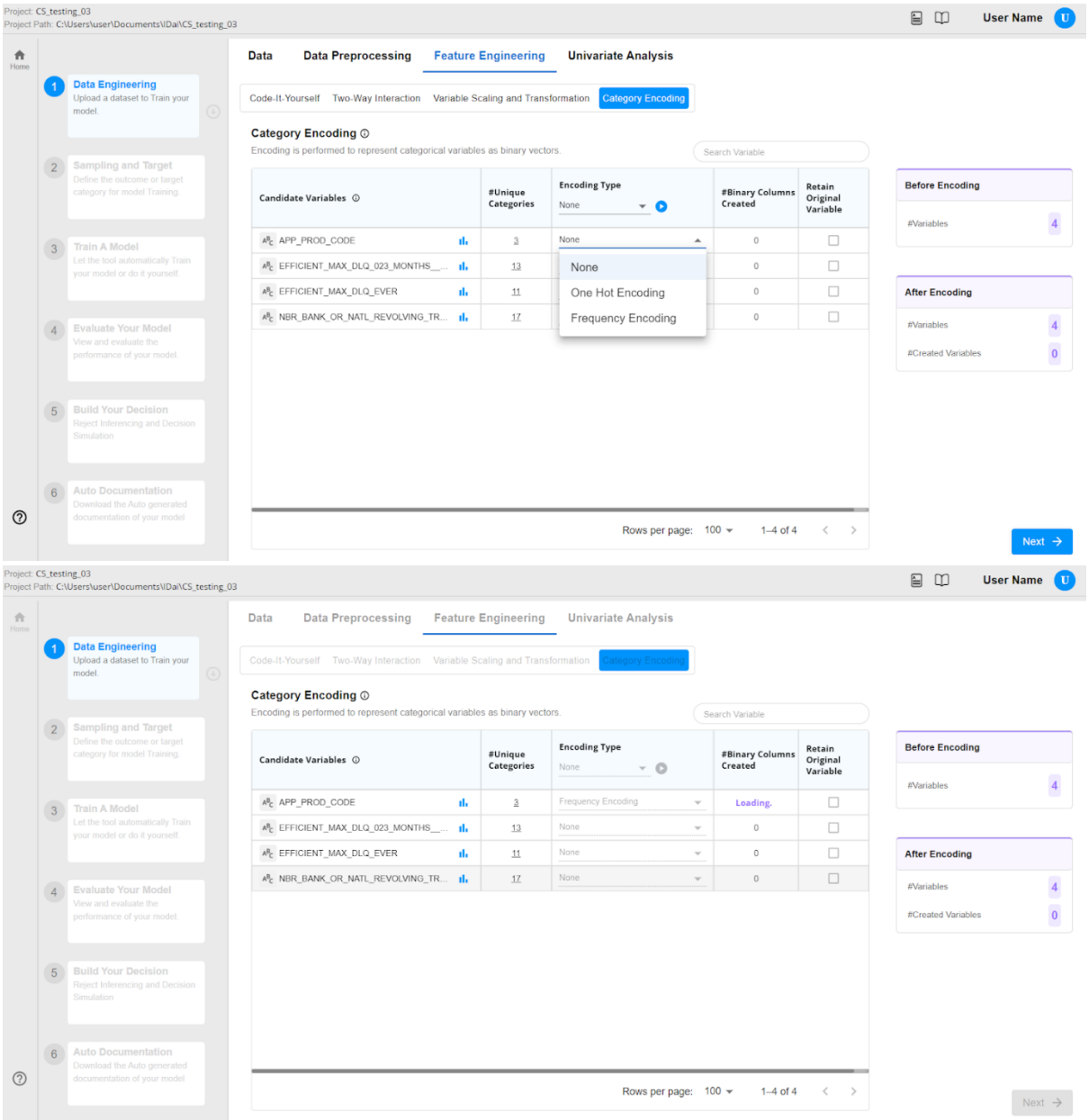
Batch and individual transformations are also supported, similar to the previous Variable Scaling and Transformation screen.

If the user selects **One-Hot Encoding**, each unique category within the chosen column/variable will be transformed into a separate column.



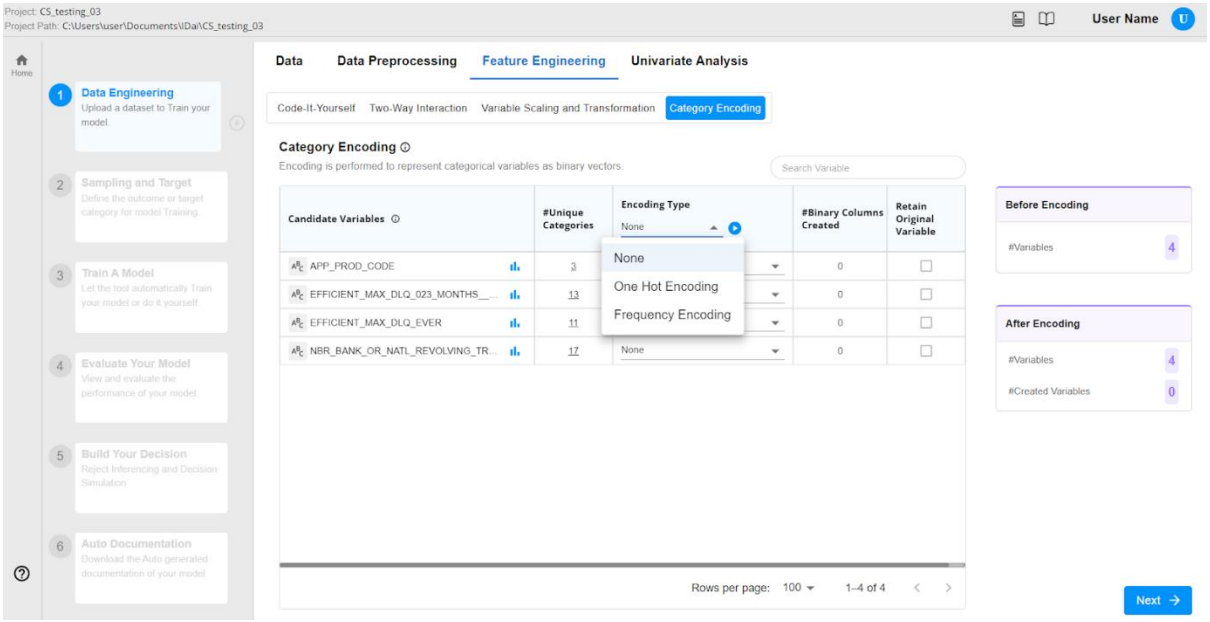


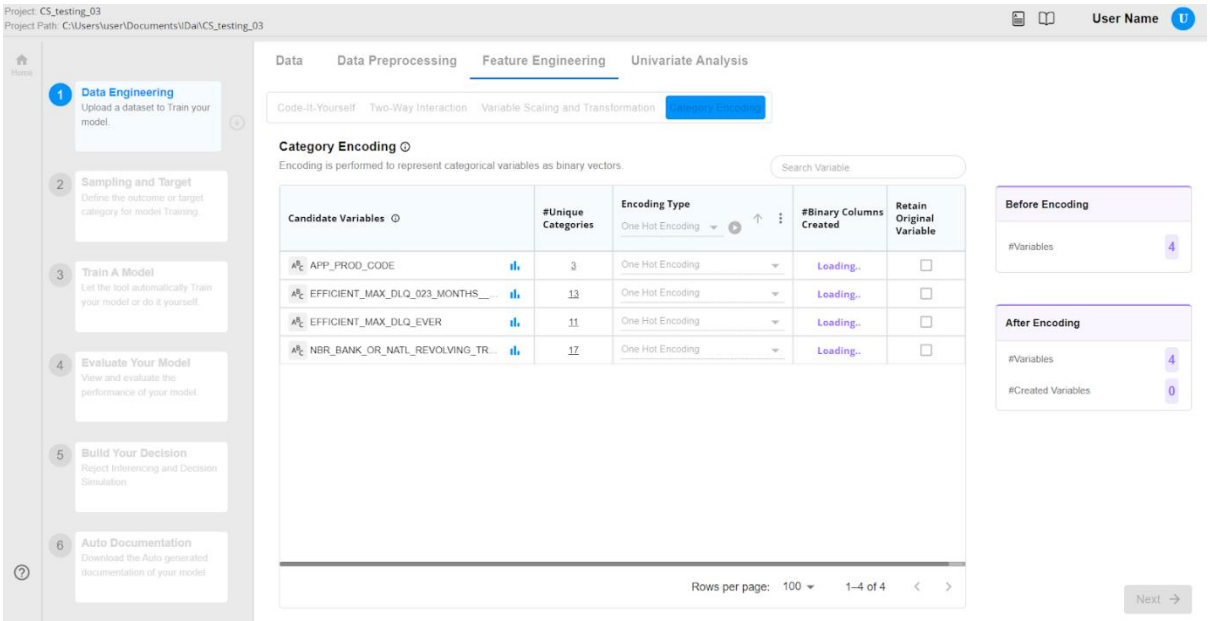
If the user selects **Frequency Encoding**, the system calculates the occurrence count of each unique value in the selected column/variable and replaces the original value with its corresponding count.

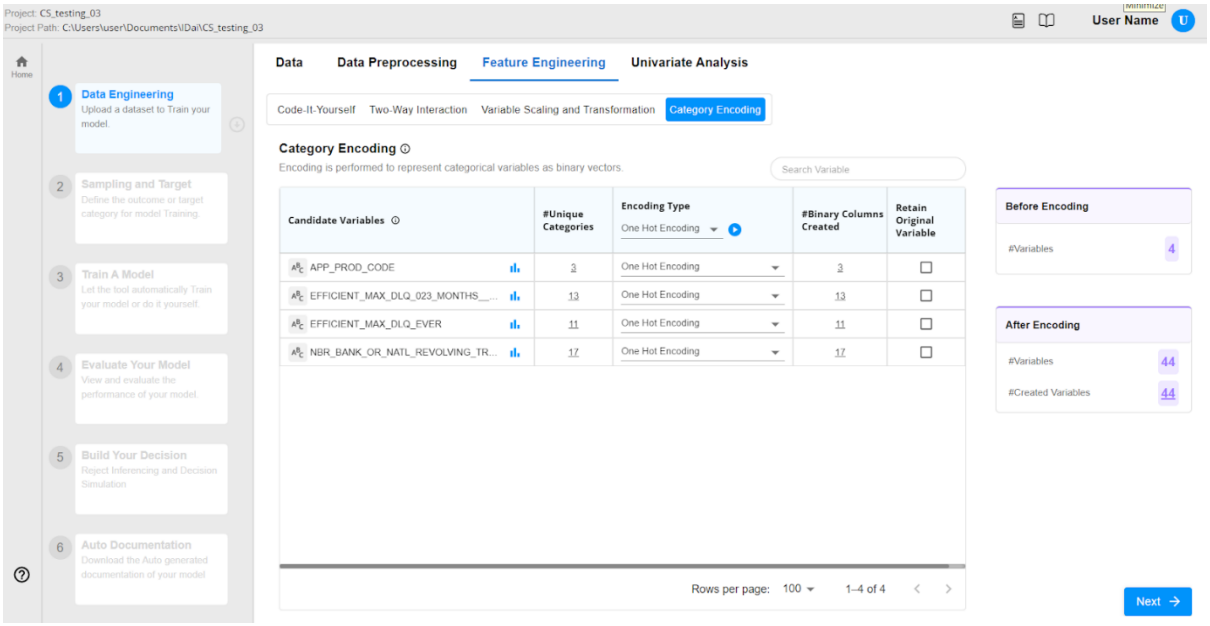


The user can perform **batch category encoding** by choosing an encoding type and clicking the blue '**Play**' button. This applies the selected transformation to all

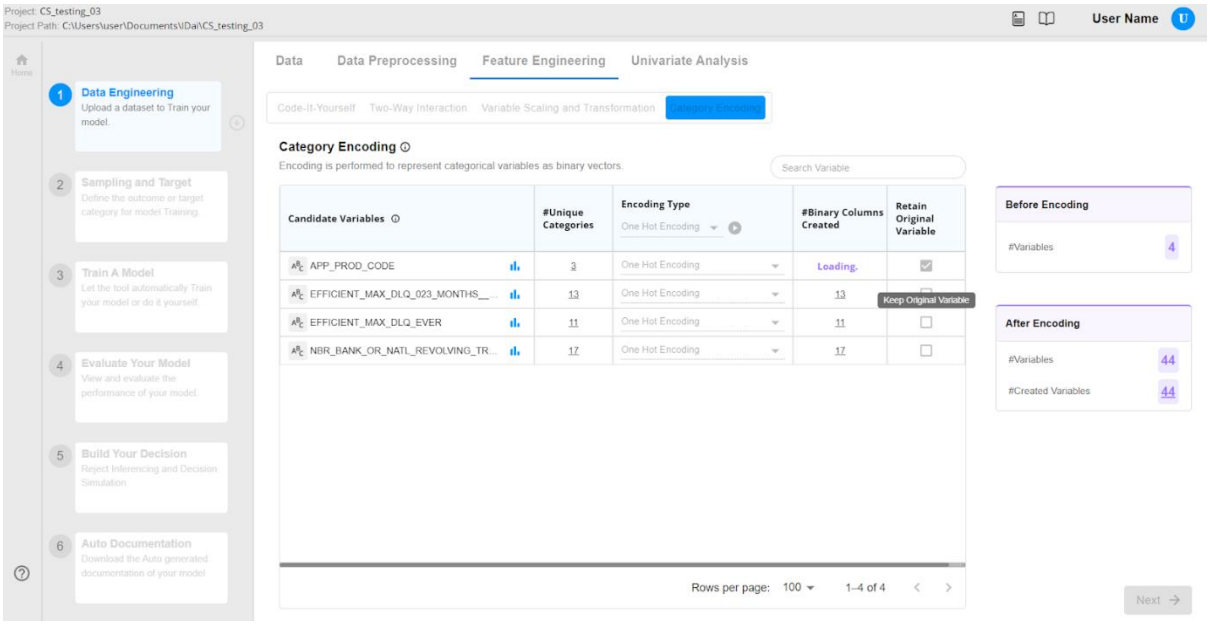
variables in the table.

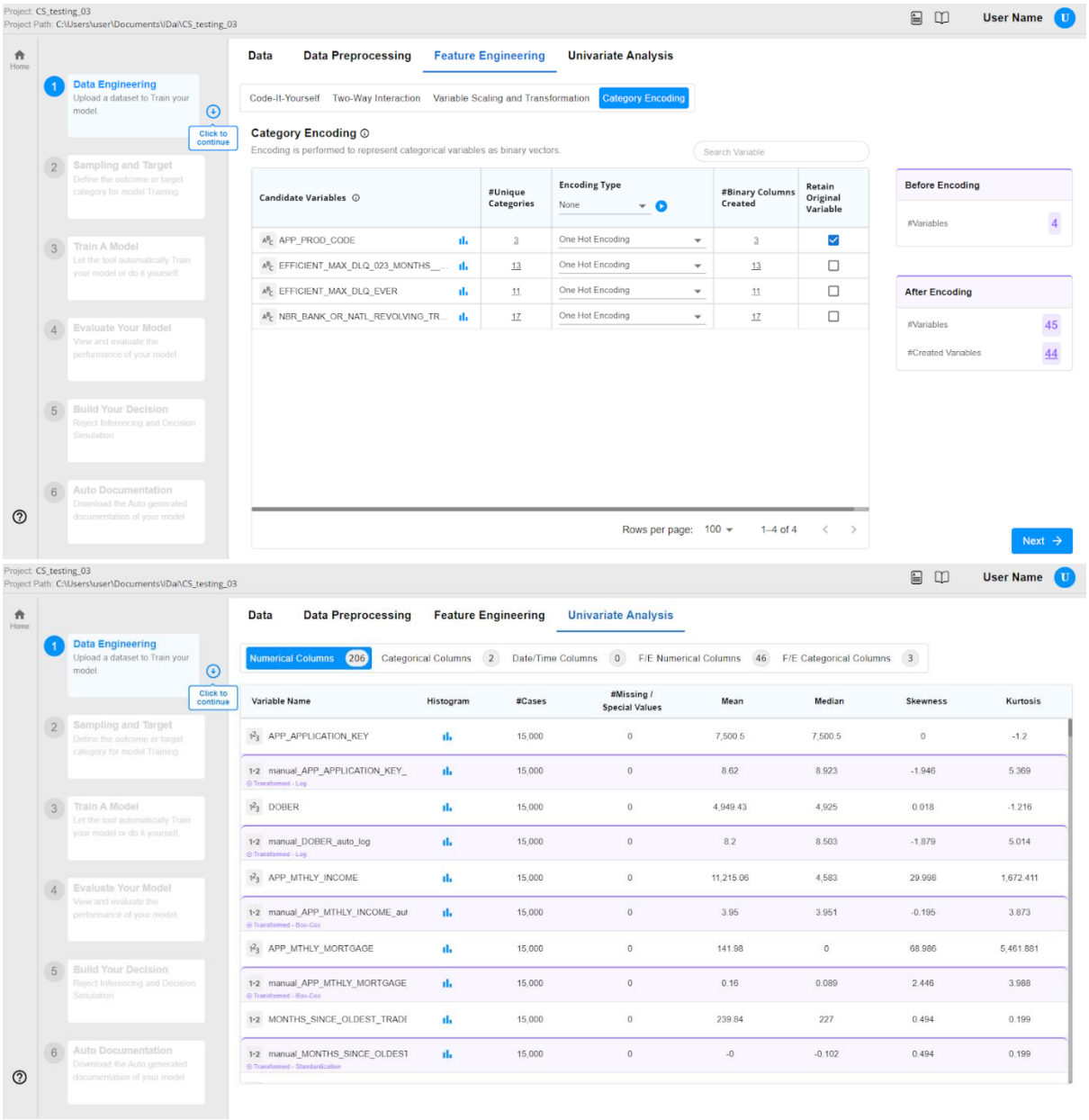




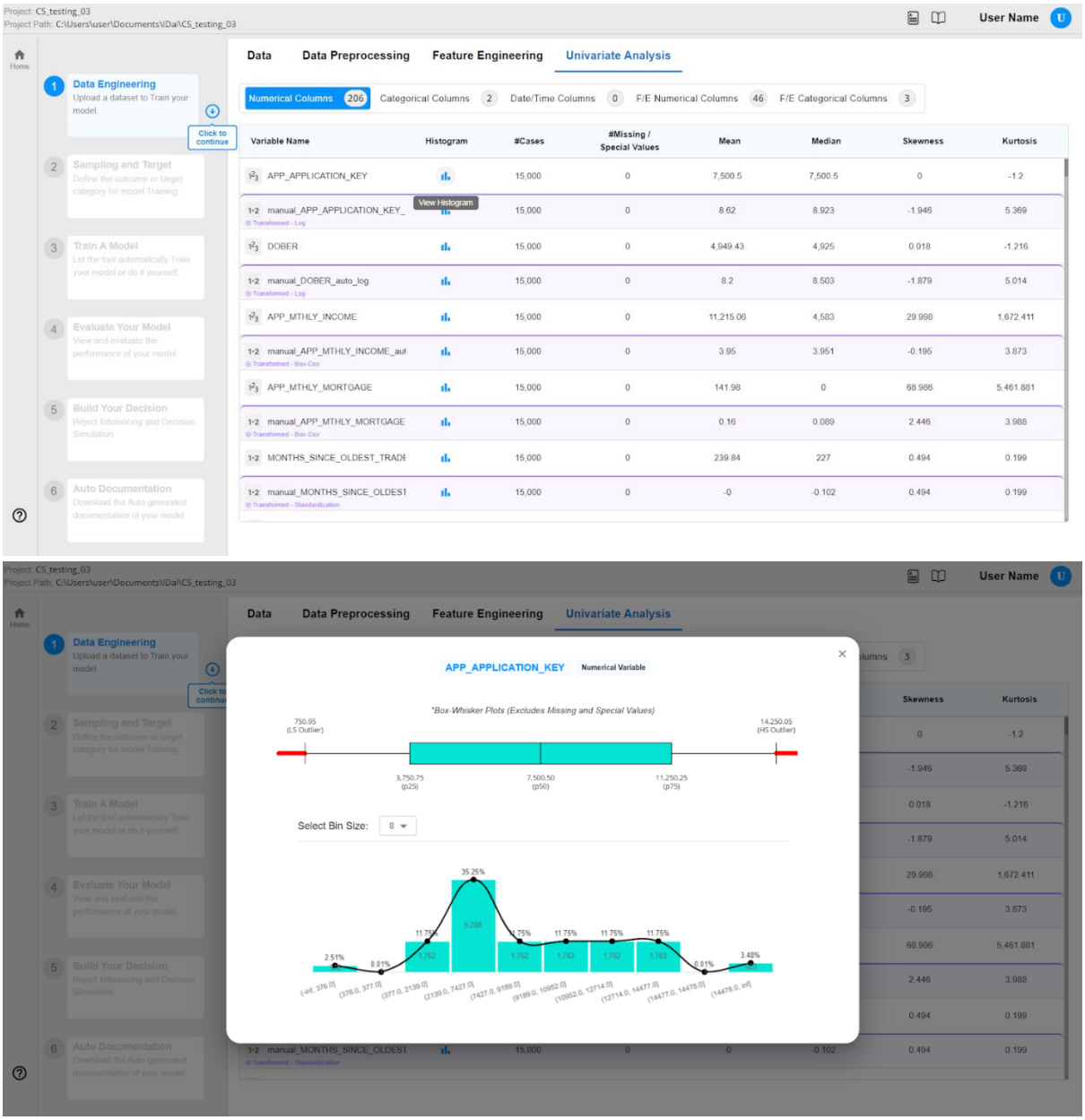


To retain both the original and transformed variables/columns, the user can check the corresponding boxes under **'Retain Original Variable'** for the variables they wish to preserve.



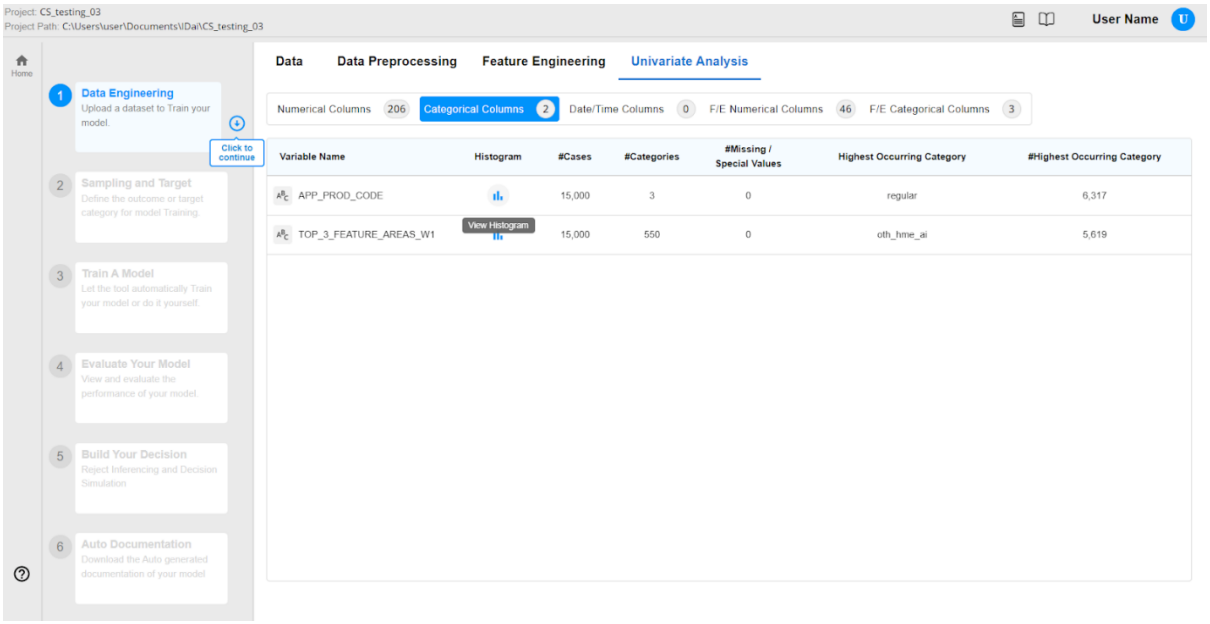


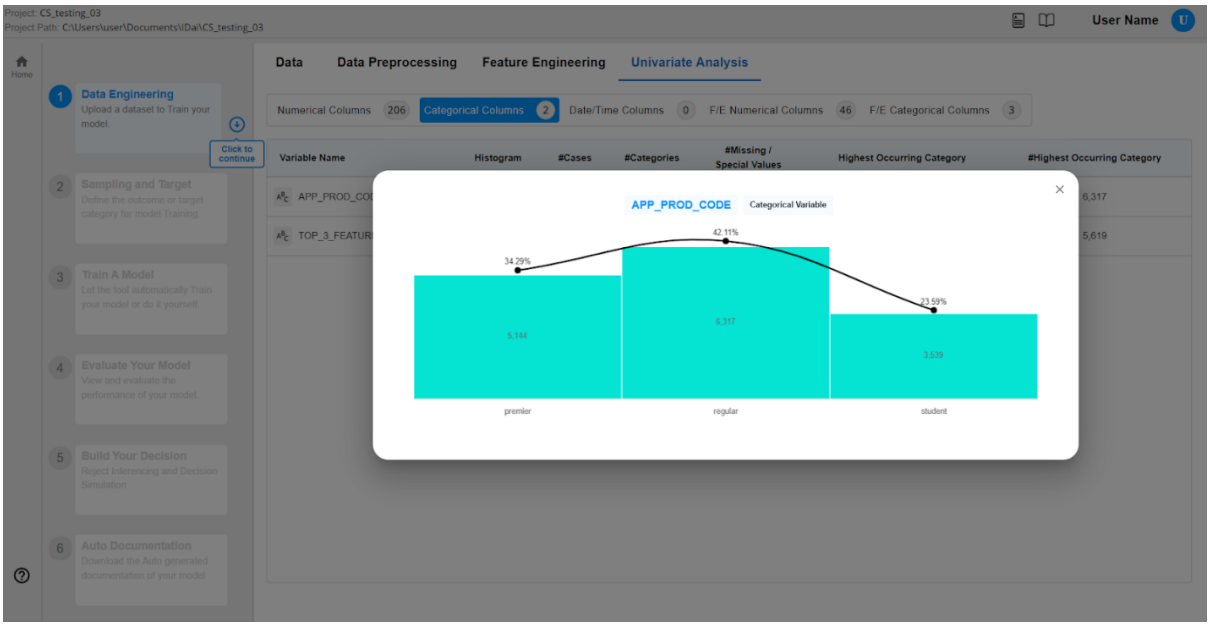
To get a visual representation of any variable's distribution, the user can click on the bar chart icon under the histogram column (blue oval in the screenshot above). The resultant chart provides the lower and higher bound values, the 25th /50th/75th percentile values as well as any outliers. The range is also split into 10 bins by default and the count and share in each bin is provided.

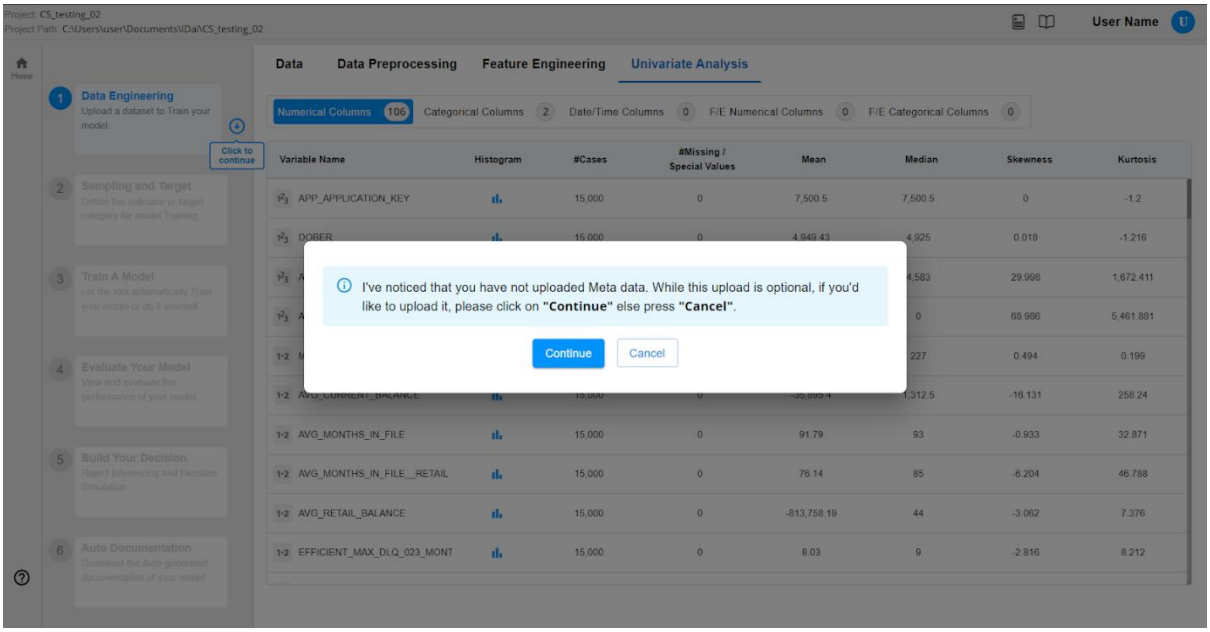


Users can adjust the bin size to increase the number of bars in the distribution chart, allowing for a more detailed view of how the data is spread across different value ranges — which can help in detecting subtle patterns or outliers

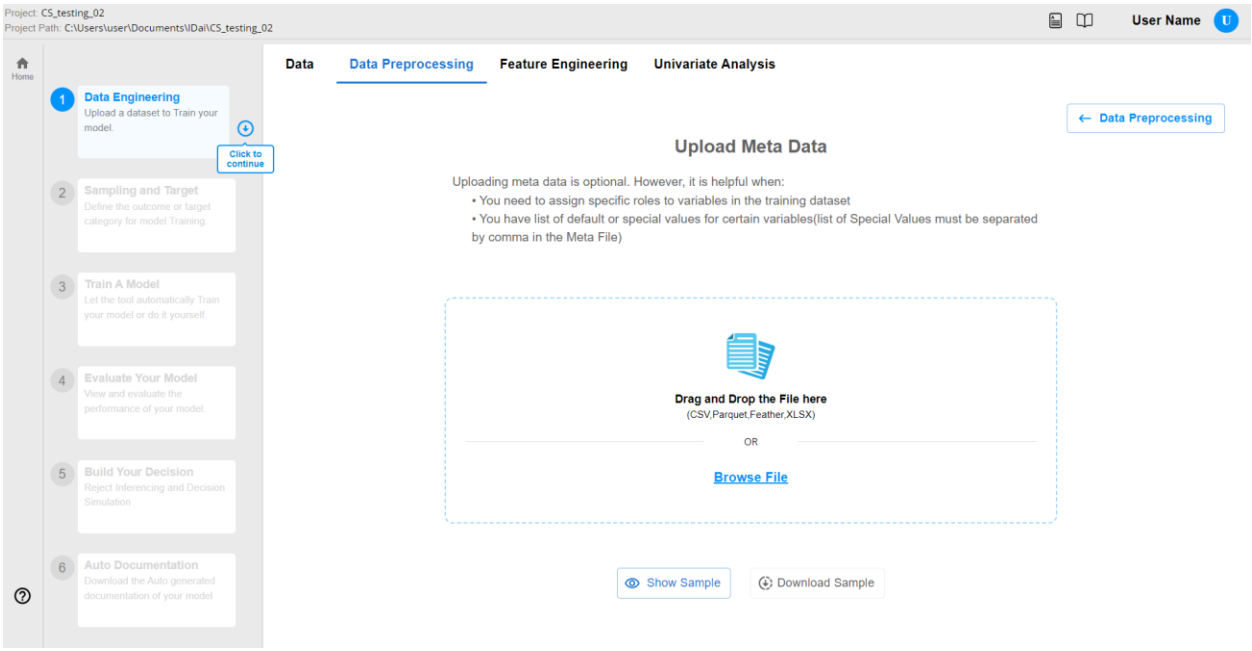
Similarly the user can view the histogram for categorical variables. This bar chart shows how a dataset is divided across different categories, helping users quickly compare their sizes or frequencies.


Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Home
1 Data Engineering
Upload a dataset to Train your model.
Click to continue
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Data
Data Preprocessing
Feature Engineering
Univariate Analysis
Numerical Columns 206
Categorical Columns 2
Date/Time Columns 0
F/E Numerical Columns 46
F/E Categorical Columns 3
Variable Name
Histogram
#Cases
#Categories
#Missing / Special Values
Highest Occurring Category
#Highest Occurring Category
APP_PROD_CODE
15,000
3
0
regular
6,317
TOP_3_FEATURE_AREAS_W1
View Histogram
15,000
550
0
oth_hme_ai
5,619

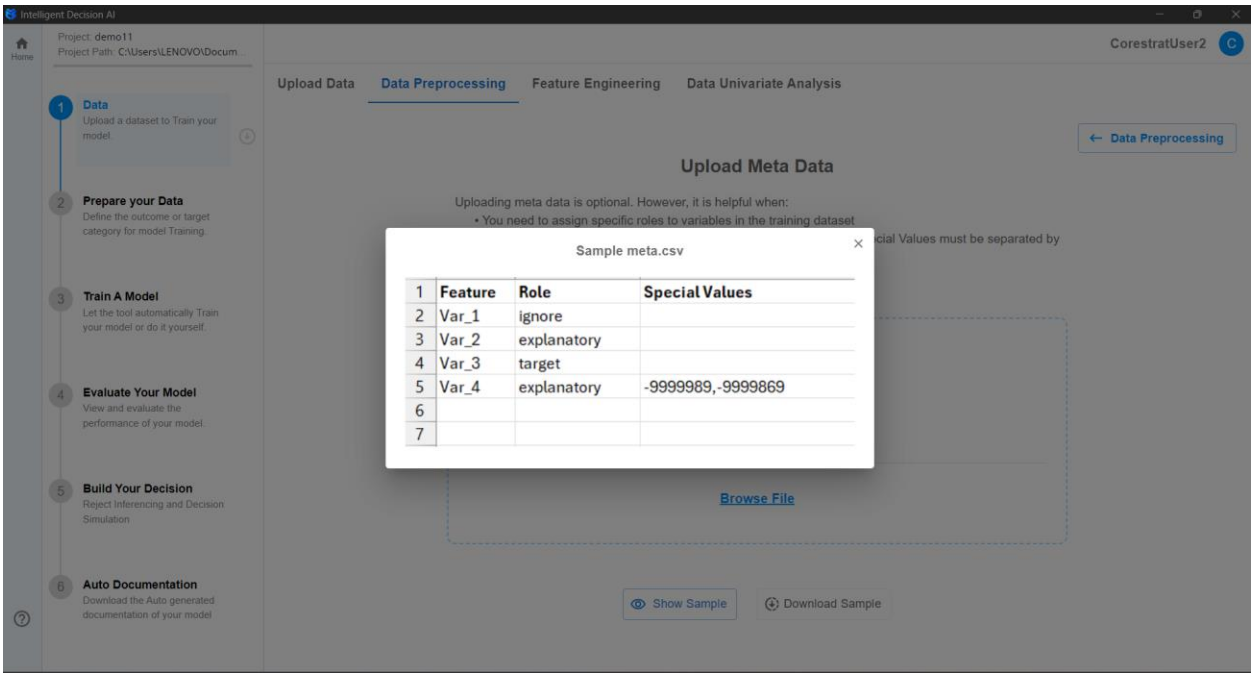

Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Data
Data Preprocessing
Feature Engineering
Univariate Analysis
APP_PROD_CODE
Categorical Variable
34.29%
42.11%
23.59%
5,144
6,317
3,539
premier
regular
student



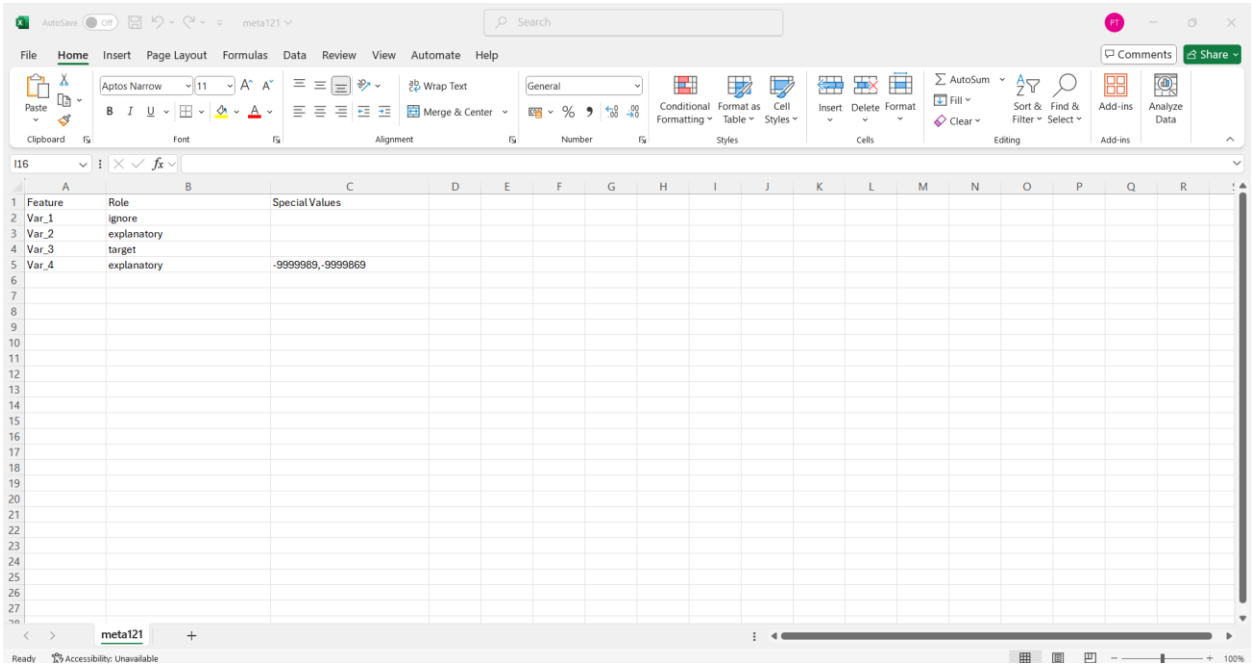
Upon clicking the downward arrow to the left of the screen, the user can add meta data by clicking '**Continue**'. This approach already has the ignore variables option already populated with likely exclusion candidates.



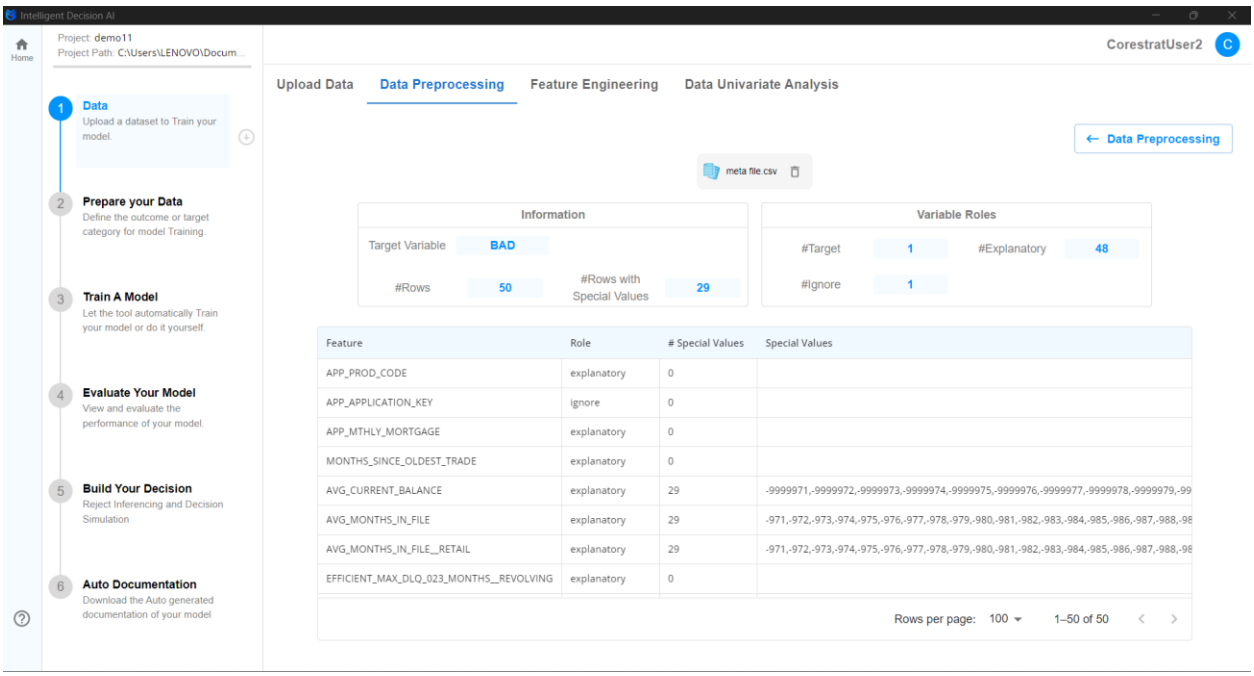
Once use clicks on "**Show sample**" a small pop-up displays the likely candidates for the role categories as shown below.



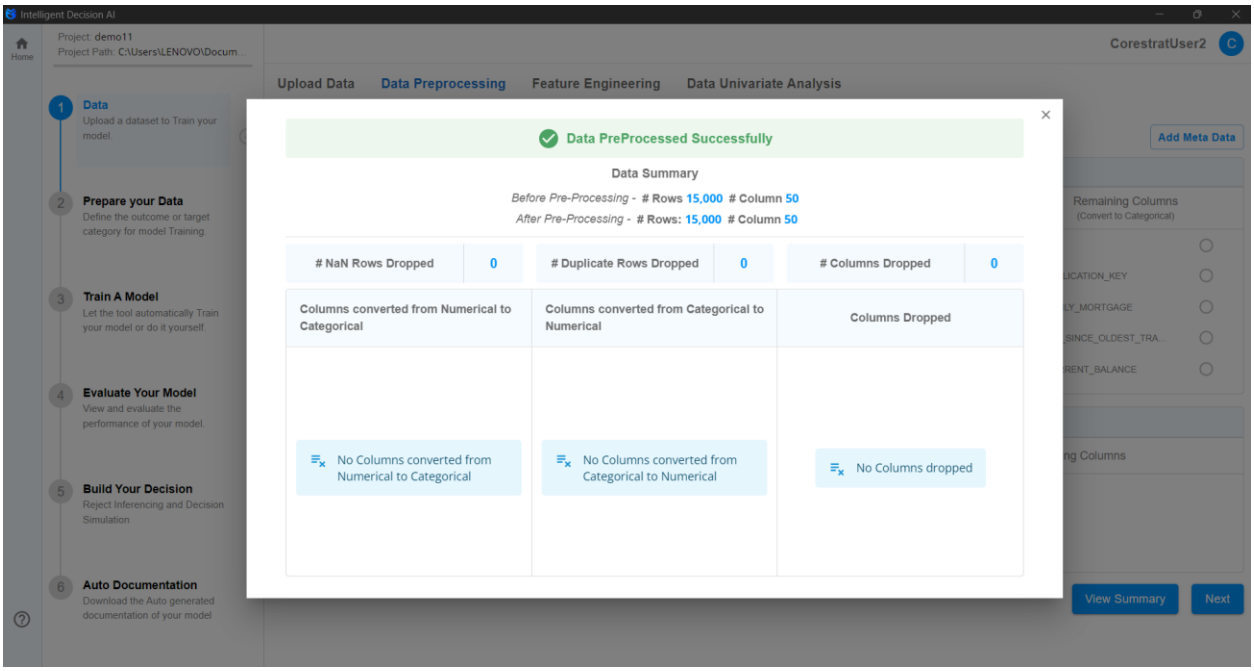
If user would like to view the above in Excel format, they need to click on "**Download Sample**" and the same will be displayed in Excel format as shown below.



Once the Meta file is uploaded, the key characteristics are displayed on screen as shown below.



Once the user applies the relevant changes by clicking on “**Apply Changes**”, the summary of these is displayed.



Sampling and Target

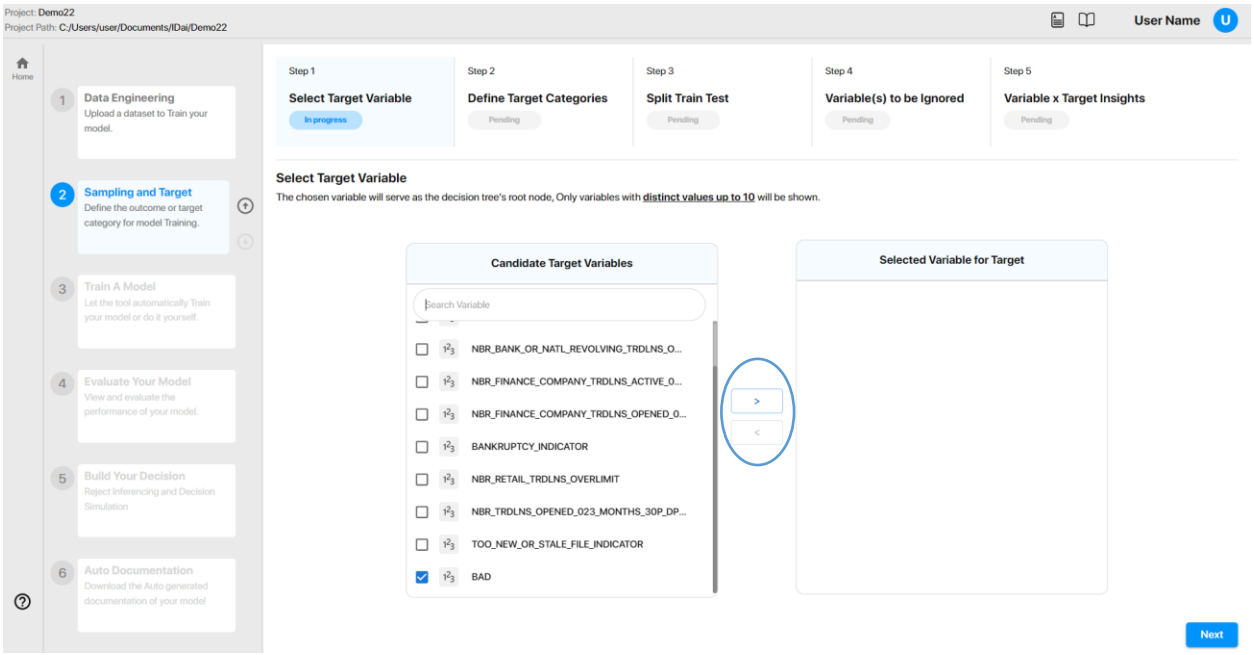
Target Variable Selection

The first step is to select the variable that represents the outcome being modelled. ID.ai provides the following:

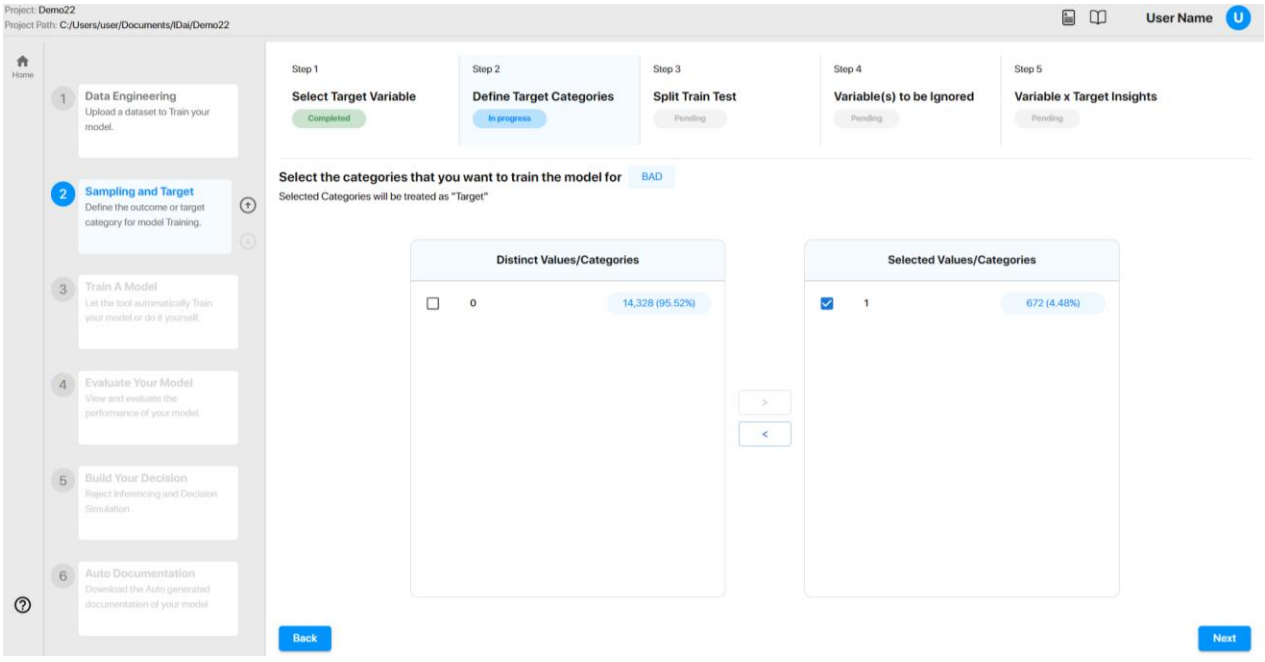
1. List of all variables in the dataset which have fewer than 10 distinct values called “candidate target variables”.

2. User can select one value from this list by moving it from the list on the left to the right using the “>” arrow.

3. To reverse the selection the user can move variables from the right to the left using the “<” arrow. (both this are indicated in the blue oval highlighted portion in the screenshot)

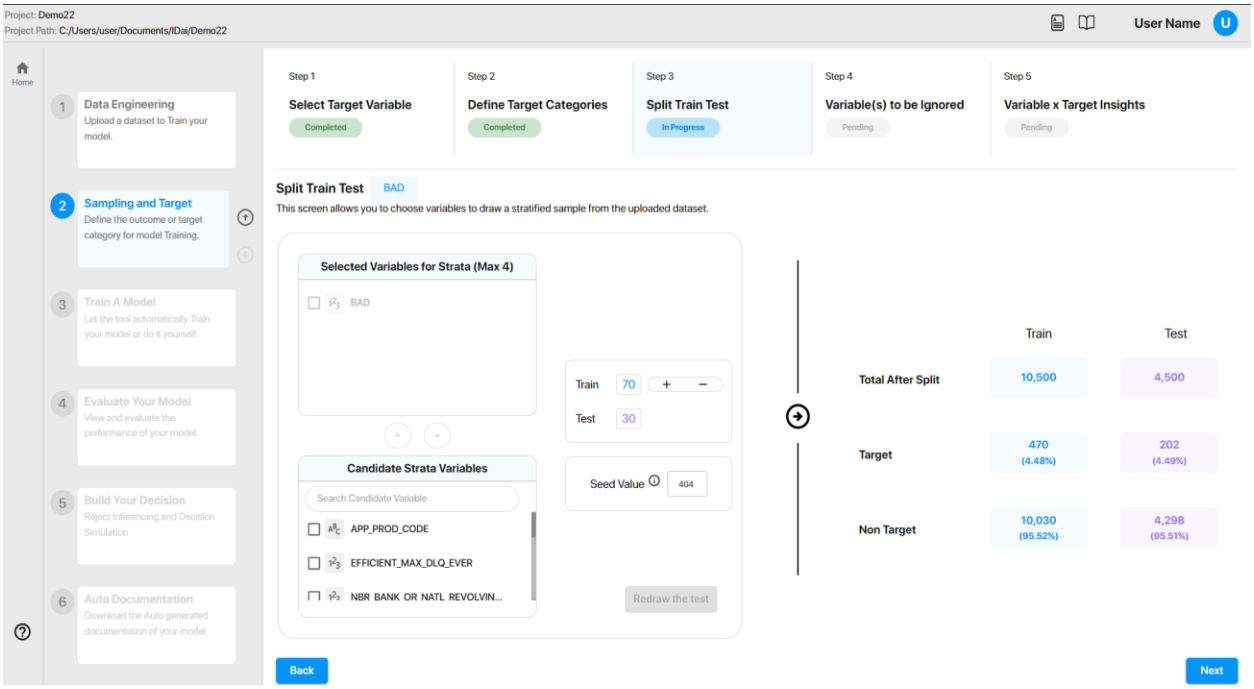


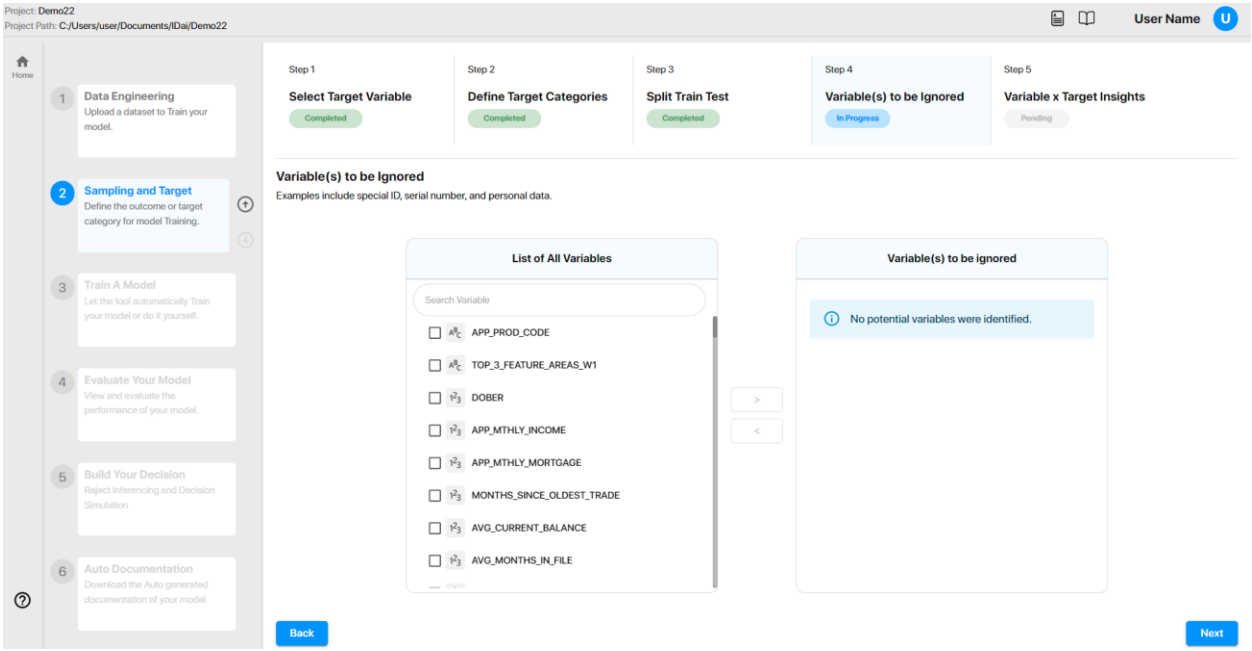
Once the target variable is selected, the next screen takes the user to the "**Define Target Categories**". This provides a list of all the distinct values within the selected target variable. The user needs to select which among these distinct values will be considered as desired outcomes and which do not.



Stratified Sampling

The next step is to split the input dataset into the “train” and “test” samples. The model will be built on the “train” sample and the results will be applied on the “test” sample to check the integrity of the model build. The default is set at 70/30 split between train and test respectively, although the user has the option to customize the split.





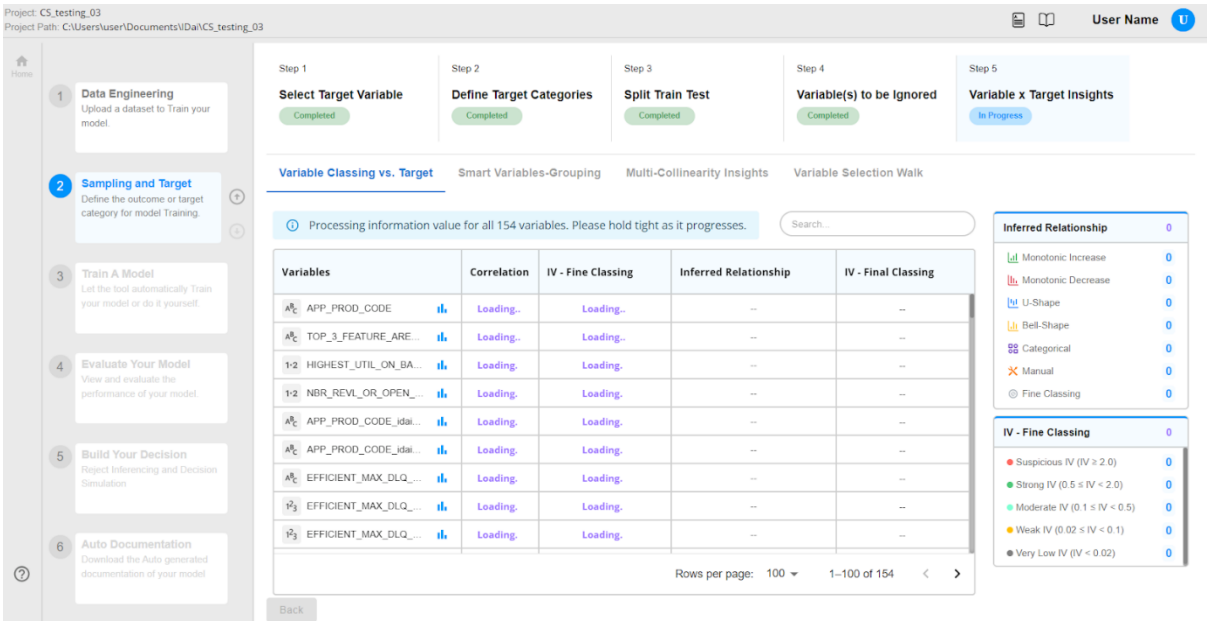
Independent Variable Insights

Users can combine multiple values of each categorical variable into custom value groups based on similar predictive power or business context. The next few steps indicate the procedures for the same.

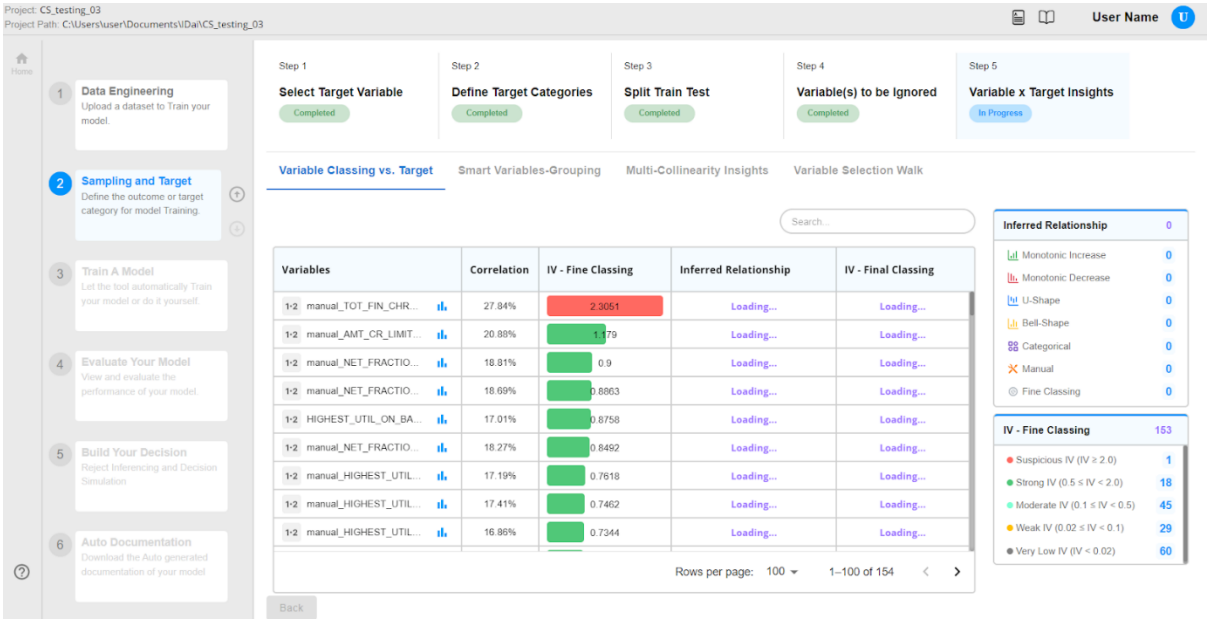
Information Value

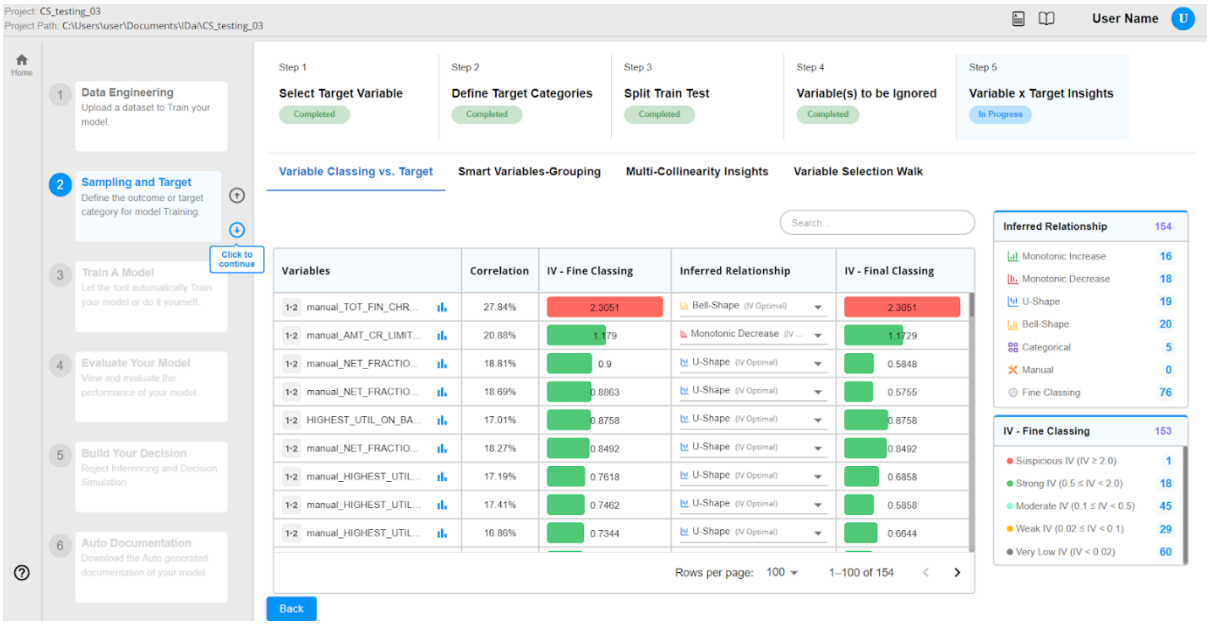
Information Value is a measure of the predictive power of the independent variable on the target variable. The IV value thresholds for suspicious, strong, moderate, weak are >2.0, 0.5 to 2.0, 0.1 to 0.5, 0.02 to 0.1 and < 0.02 respectively.

Upon reaching the **"Variable Classing vs Target"** screen and initiating IV calculation, the **Correlation** and **IV-Fine Classing** values are computed first.

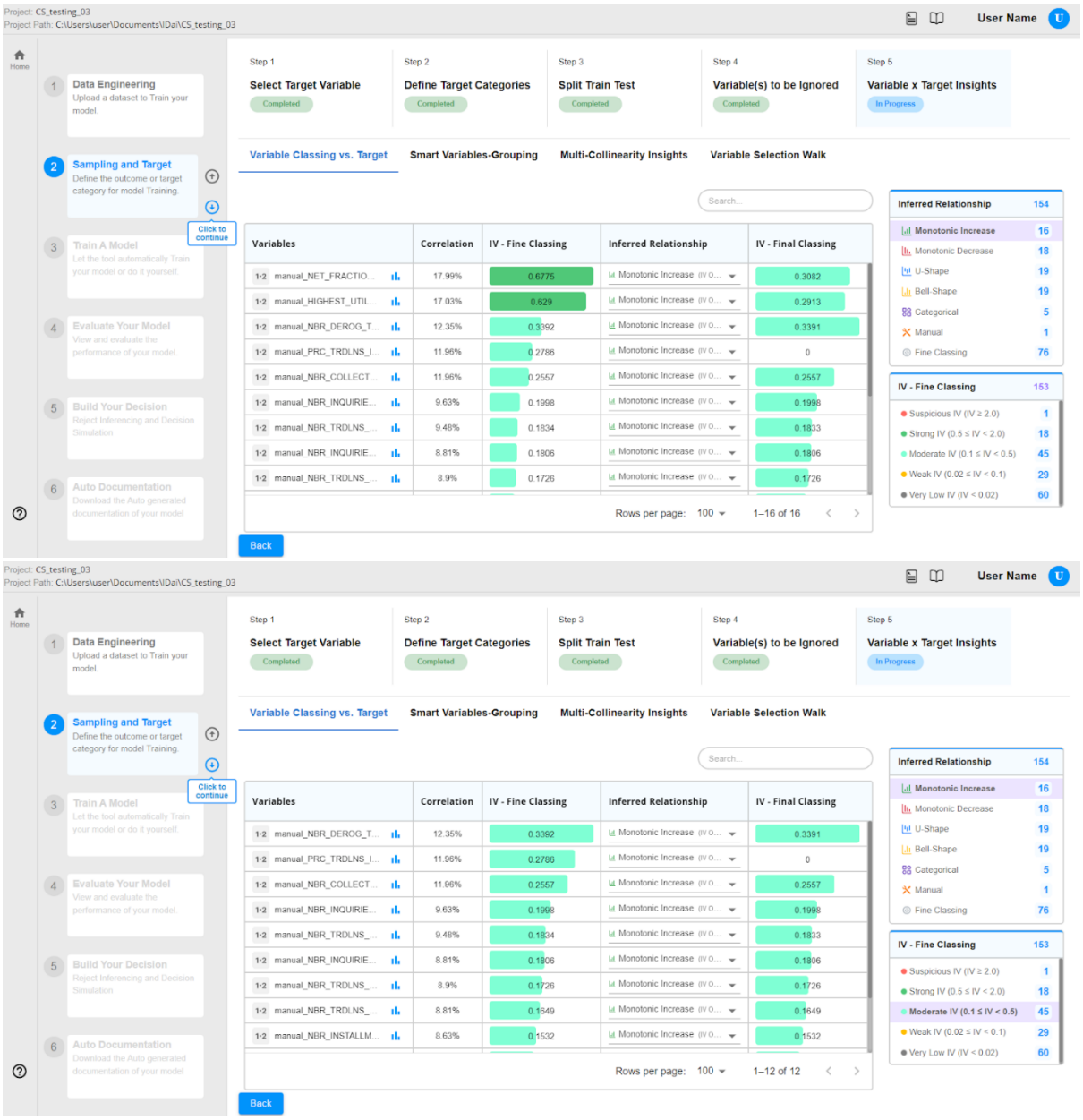


After calculating the Correlation and IV-Fine Classing values, the system then computes the **Inferred Relationship** and **IV-Final Classing** values.

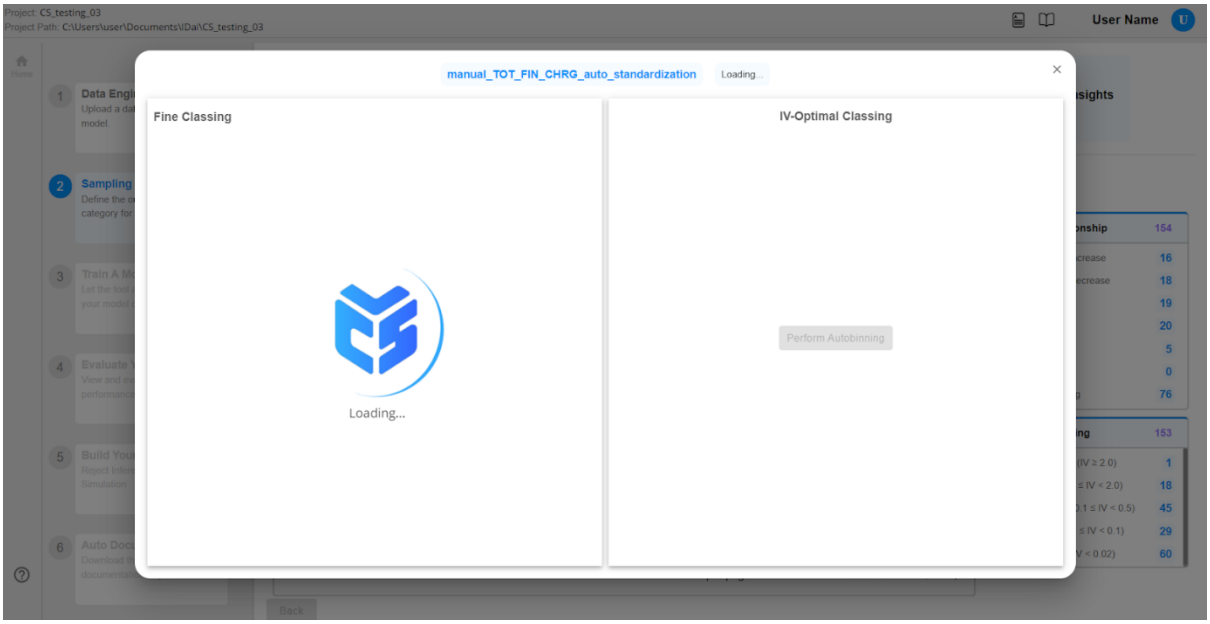




To view rows with a specific inferred relationship and/or IV-Fine Classing value range, the user can simply select the desired category, and only the matching rows will be displayed. (as depicted in the images below)



When the user clicks on the **Histogram** icon (next to the variable name), a detailed histogram screen appears.

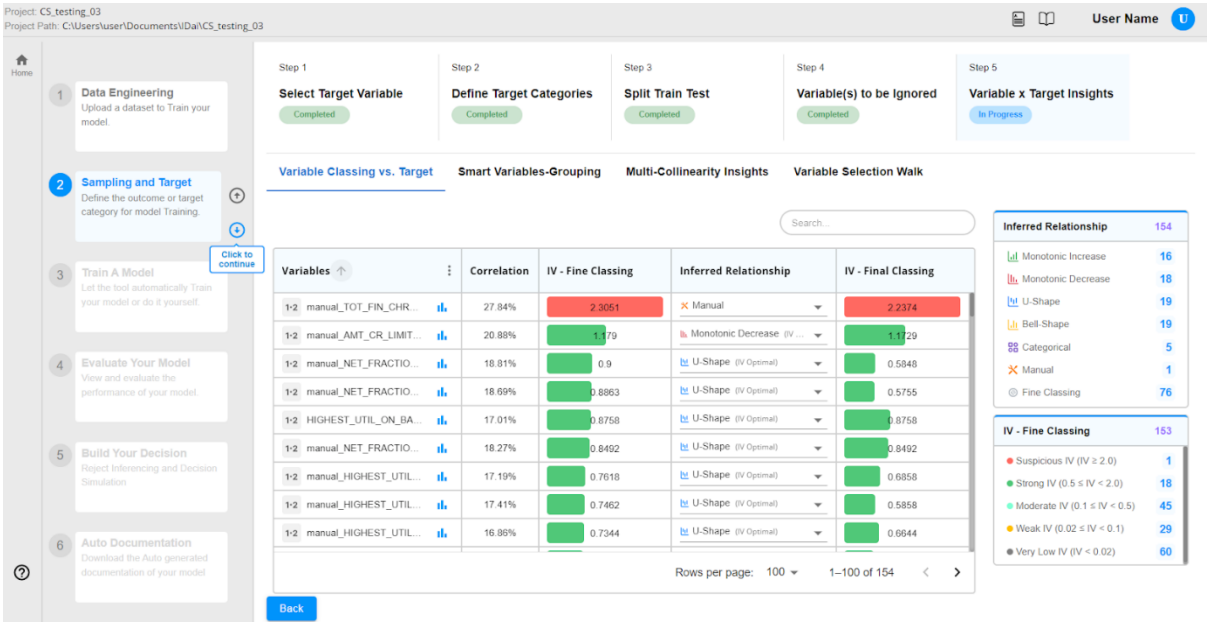


If user selects a numerical variable

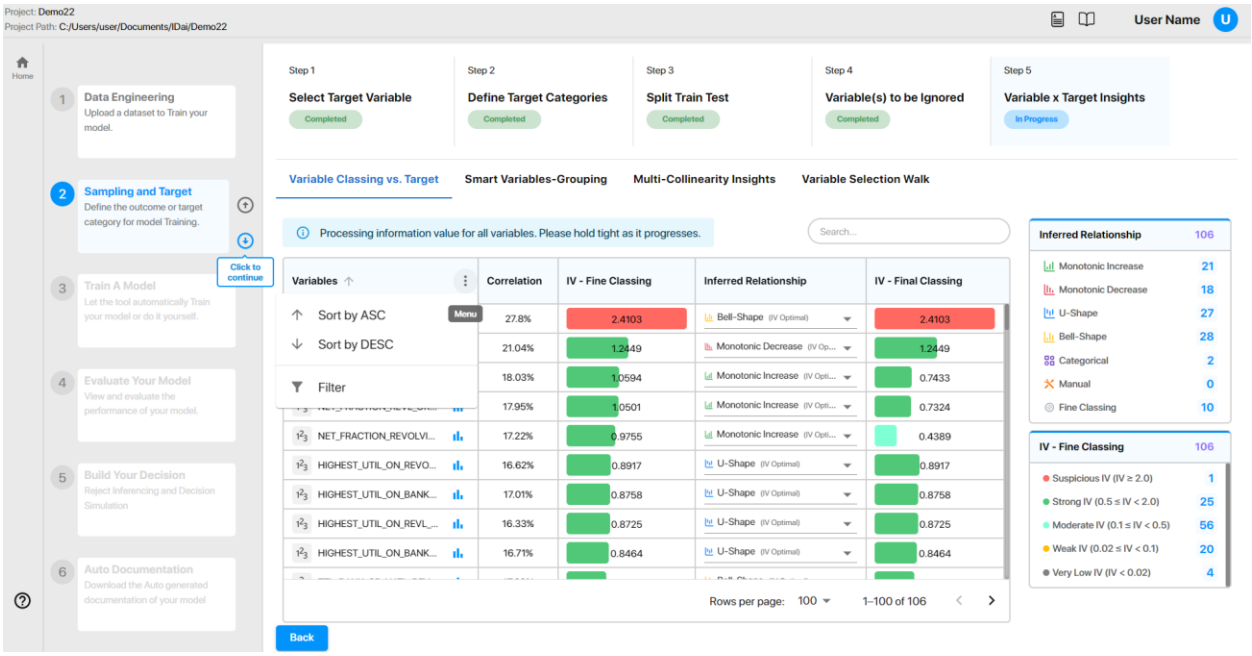
Users can define custom splits by entering values into the **'Key in Splits by Comma'** textbox. After clicking the **'Apply'** button, the customized graph will be shown in the component below.

s(he) can also choose between **'User Specified Classing'** and **'IV-Optimal Classing'** based on their preferred method.

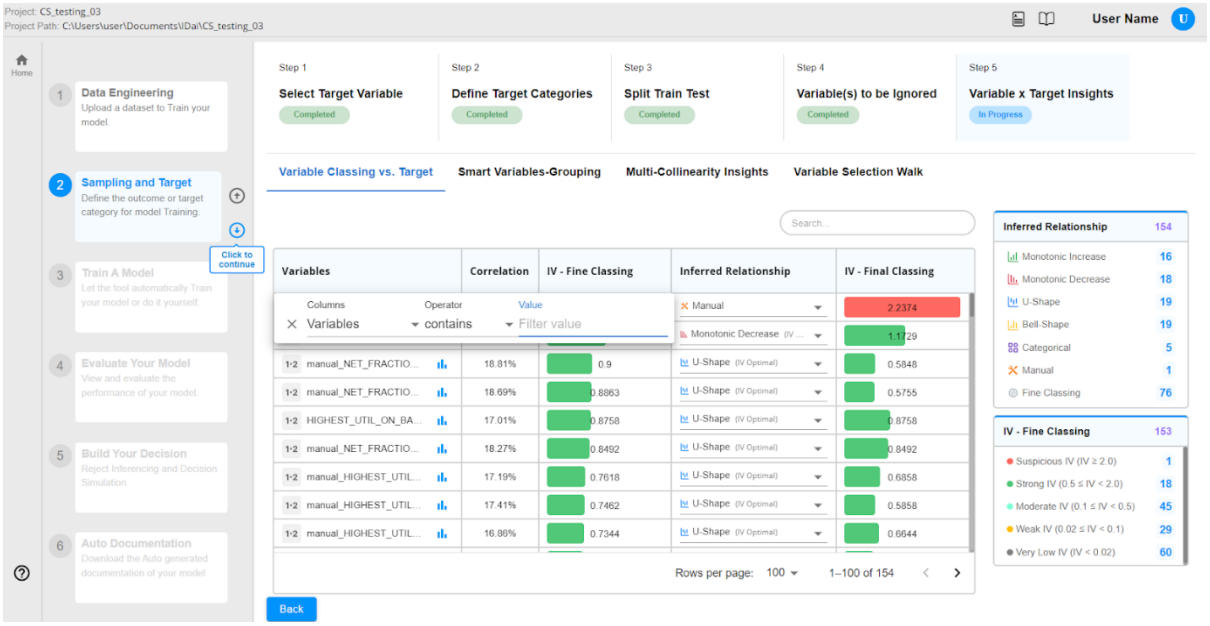
Once the user clicks **'Submit'**, the **Inferred Relationship** for that variable changes to **'Manual'**, replacing the previously inferred value from IV-Optimal Classing.



Users can control what data appears on the screen by clicking the **three dots** next to a column name. This opens options to sort in ascending/descending order and apply filters.



If the user selects **‘Filter’**, they can apply specific filters to refine the data displayed on the screen.



If user selects a categorical variable

To create custom bins, click on '**Perform Manual Binning**'. In the '**Key Your Splits**' column, group categories into the same bin by assigning them the same key value

Once you click '**Apply Manual Bins**', a new chart showing the custom binning will appear on the right.

After submitting, you can choose between the manually defined bins or the IV-Optimal classing.

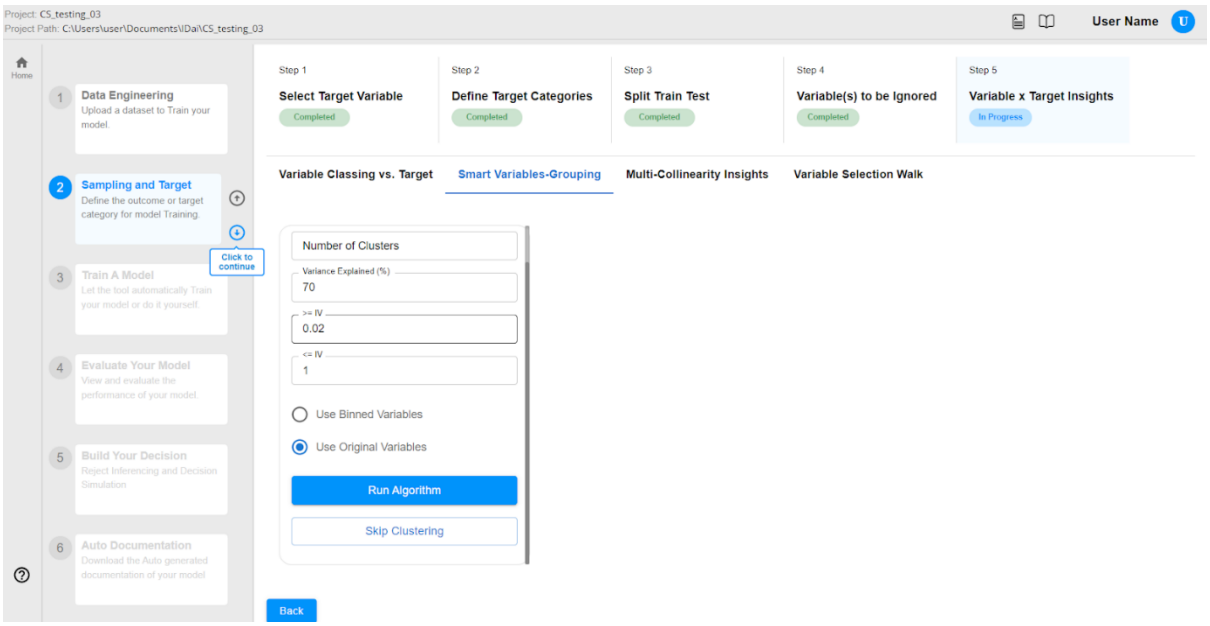

APP_PROD_CODE_idaix_EFFICIENT_MAX_DLQ_EVER
Fine Classing
Correlation with Target: 10.95
IV: 0.2549
IV-Optimal Classing
Categorical (IV Optimal)
IV: 0.2549
User Specified Classing
Manual
IV: 0.1846
Target Rate % (Train)
Target Rate % (Test)
% Count (Train)
% Count (Test)
Perform Manual Binning
Submit
User Name
Back

Clustering :

This screen allows users to configure the VarClus Clustering Algorithm:

1. Number of Clusters or Variance Retention:
• Specify the desired number of clusters or the proportion of variance to retain.
• The algorithm will stop splitting clusters once either condition is met.
2. IV Threshold:
• Define an Information Value (IV) threshold. Variables falling outside this threshold will be excluded from the clustering process.
3. Clustering Method:
• Choose between clustering on the Weight of Evidence (WoE) of binned variables or on the original values of the variables.

If user selects '**Use Original Variables**' :



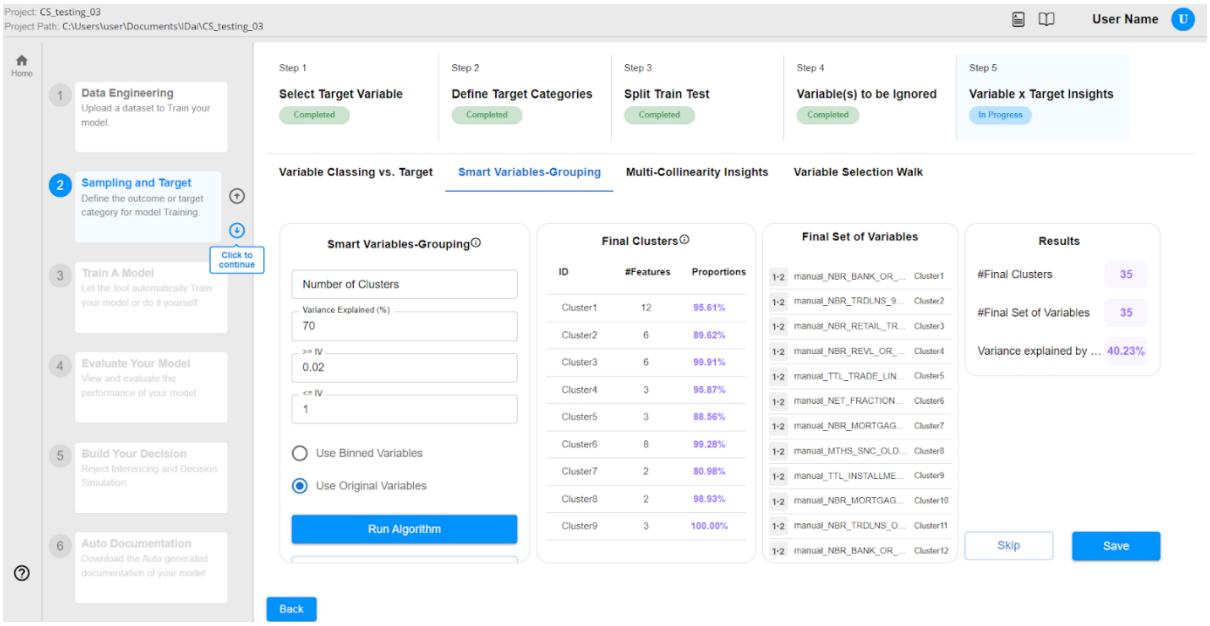
Once the algorithm is executed, this screen displays:

1. Cluster Summary Table (Middle):
• Displays the number of clusters formed and the proportion of variance explained by the first principal component (PC1) of each cluster.
• Users can click on any cluster to manually select or change the variable that represents it.

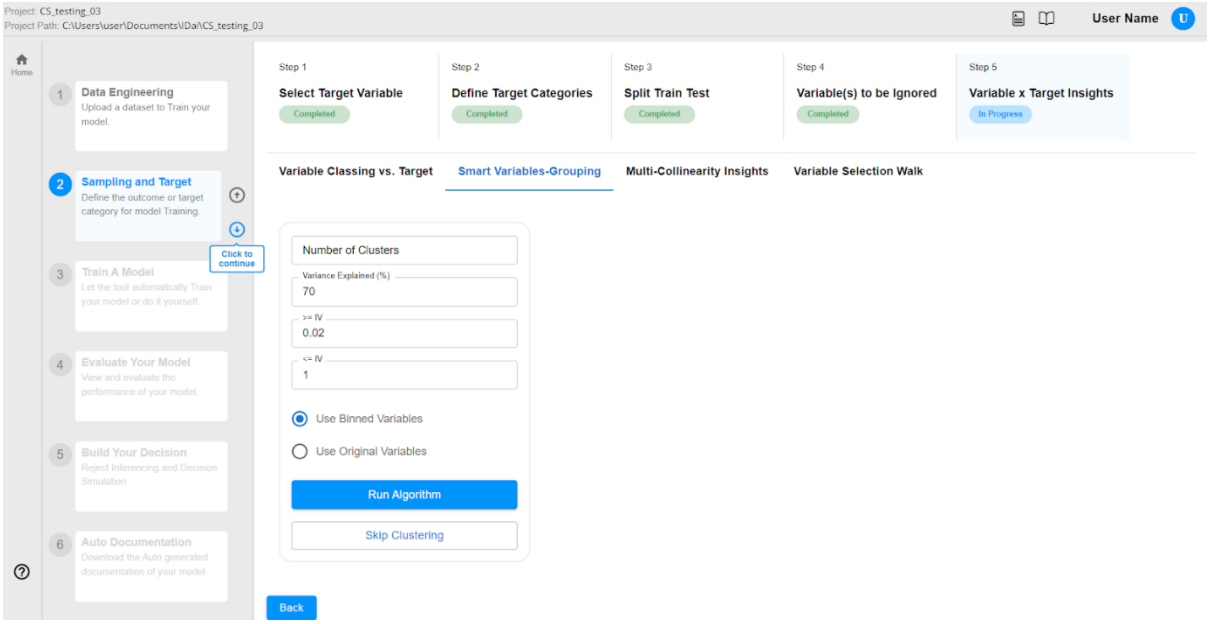
2. Final Variable Table (Table 2):
• Lists the final set of variables selected for model building. All other variables will be discarded if the user saves the clustering results.

3. Summary Text Box (Right):

• Shows the proportion of variance explained by the final set of variables across the overall dataset.



If user selects '**Use Binned Variables**':



Once the algorithm is executed, this screen displays:

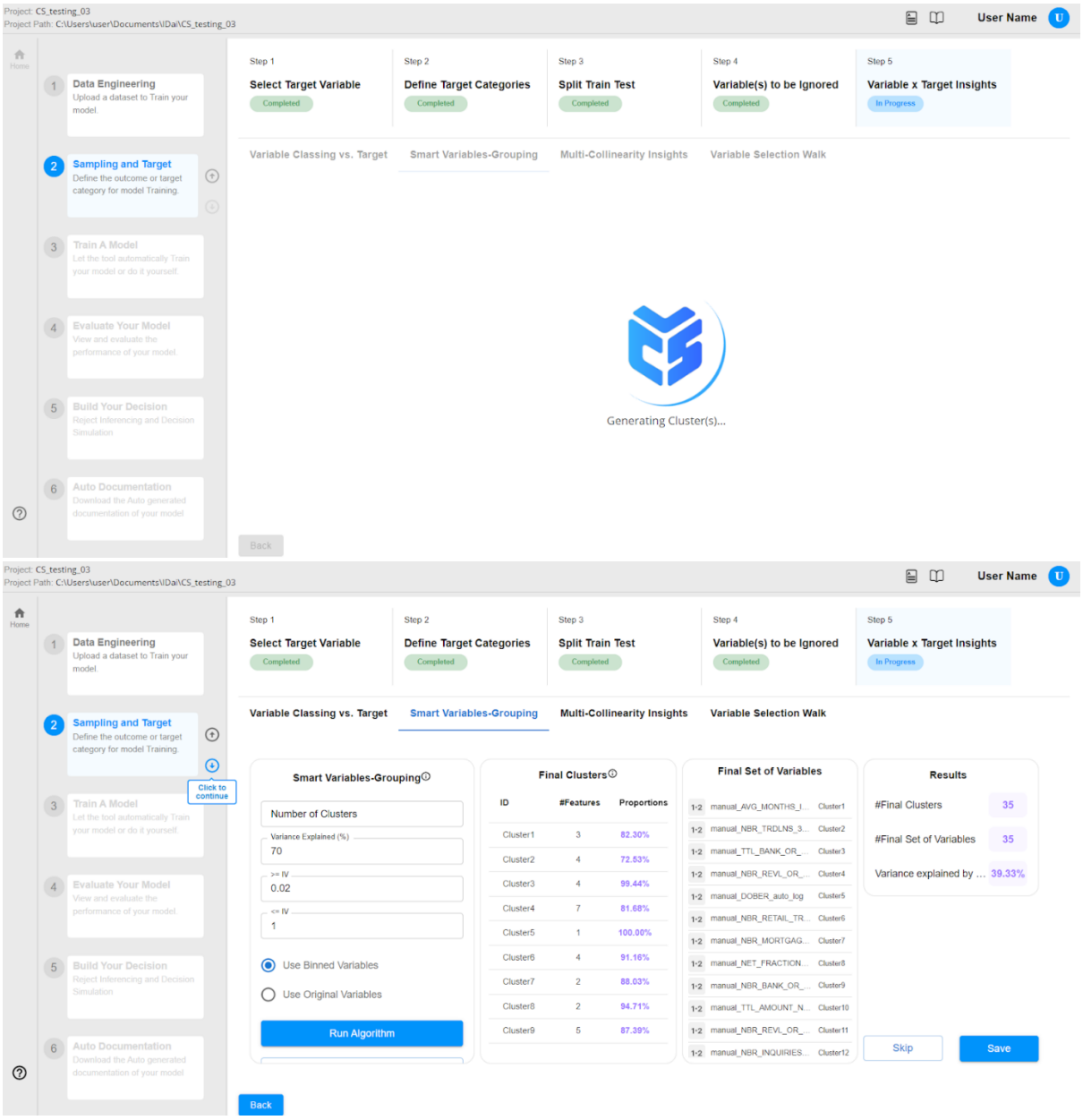
1. Cluster Summary Table (Middle):

• Displays the number of clusters formed and the proportion of variance explained by the first principal component (PC1) of each cluster.

• Users can click on any cluster to manually select or change the variable that represents it.

2. Final Variable Table (Table 2):
• Lists the final set of variables selected for model building. All other variables will be discarded if the user saves the clustering results.

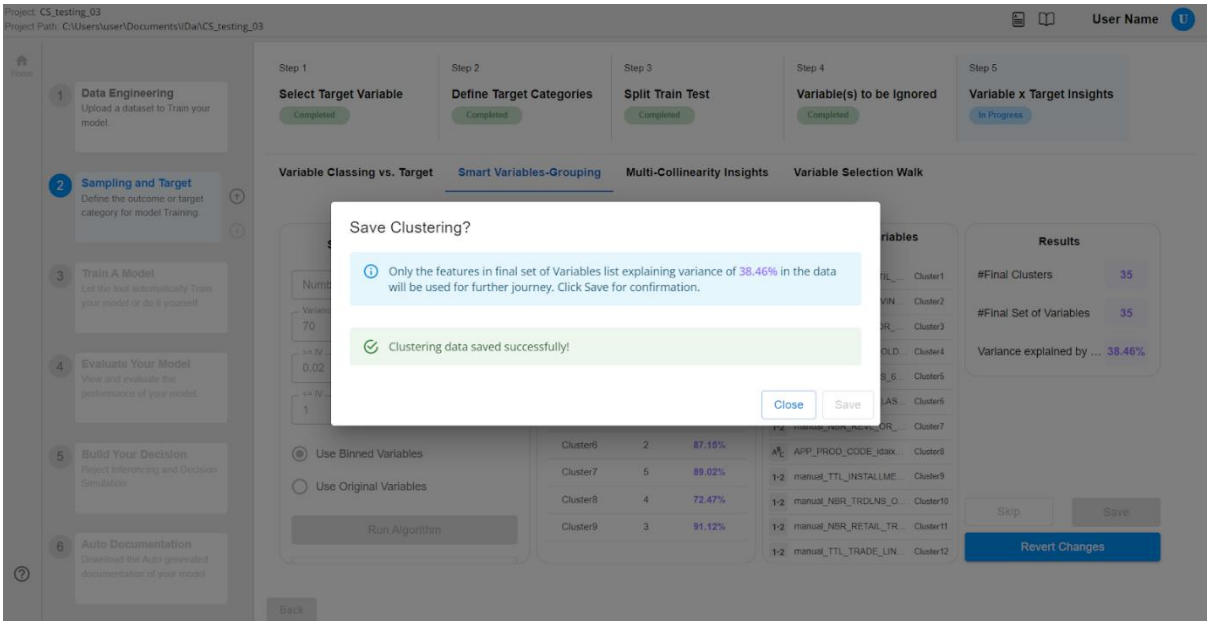
3. Summary Text Box (Right):
• Shows the proportion of variance explained by the final set of variables across the overall dataset.



Once you have chosen the required methodology with which to run the algorithm, you can:

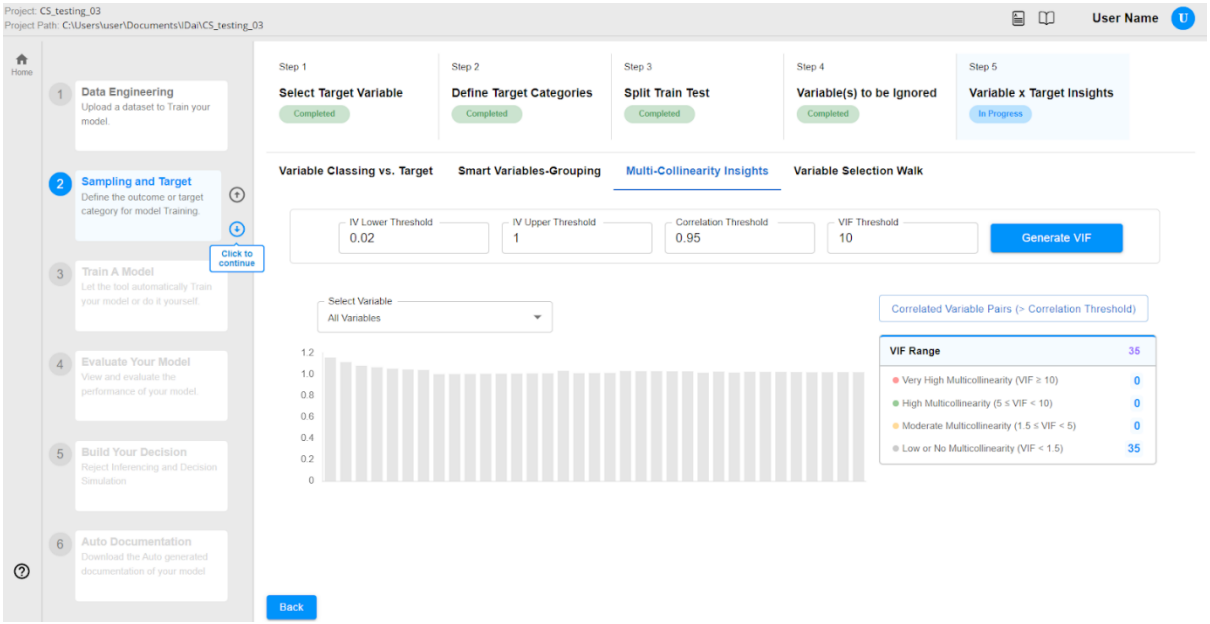
Click “**Save**” button, to save the clustering results and to proceed with the final variable set.

Click “**Skip**” button to discard the clustering results and to retain the original variable set.

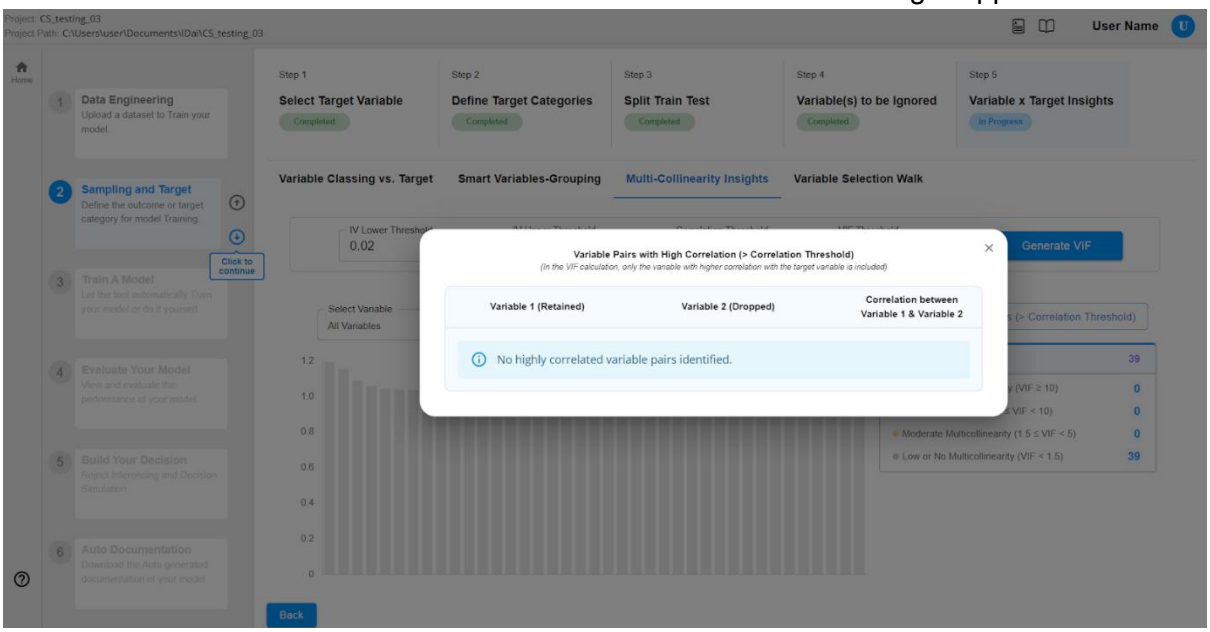


Correlation and multicollinearity

This tab provides the correlation values between the target variable and all the independent variables. Using the measure of Variance Inflation Factor (VIF) the variables that fall into very high (VIF > 10), high (VIF between 5 & 10), moderate (VIF between 1.5 & 5) and low (VIF < 1.5) are listed.

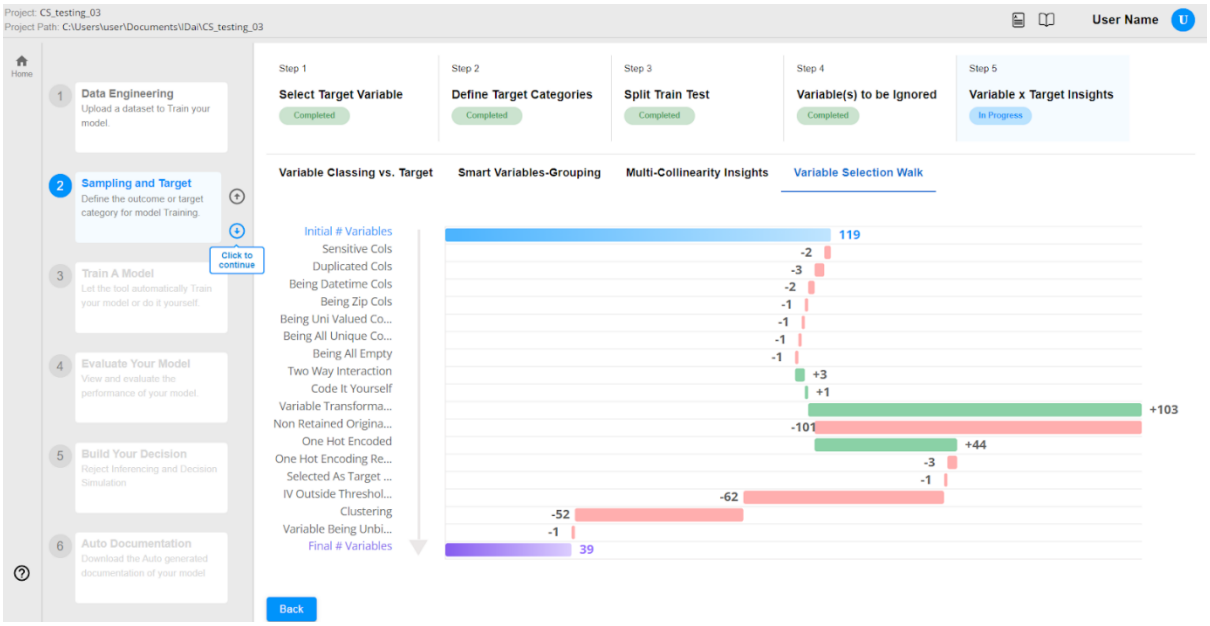


The button “**Correlated Variable Pairs (> Correlation Threshold)**” provide the correlation between the two variables with correlation above the threshold amount, and will also show the variable that is retained and the variable that is being dropped



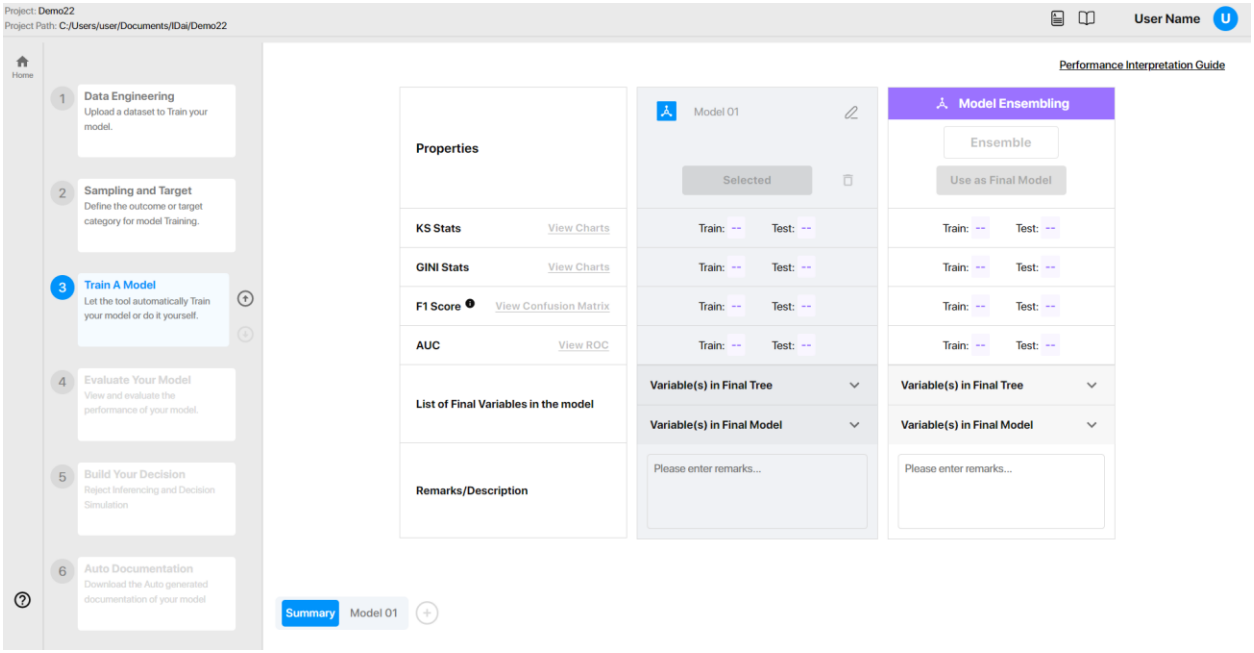
The page below provides a step-by-step overview of how the initial dataset was processed and transformed through various stages of variable selection. It visually tracks:

- Initial number of variables at the start of the pipeline.
- Removal steps, such as dropping sensitive, duplicate, empty, or irrelevant variables
- Addition steps, like creating new variables through interactions, transformations, and one-hot encoding
- Final number of variables available for modeling after all selections and transformations. Each step shows how many variables were added or removed, allowing you to understand the evolution of the data

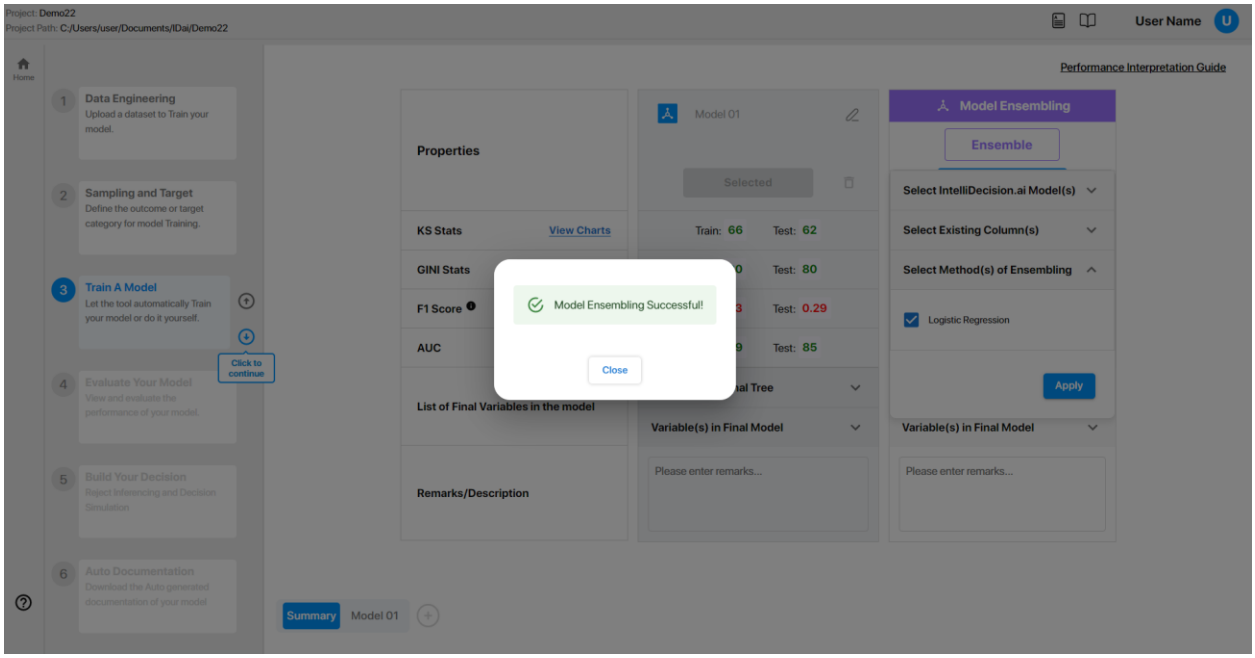


Model Comparison: Summary

- User can create upto 3 model max and compare them



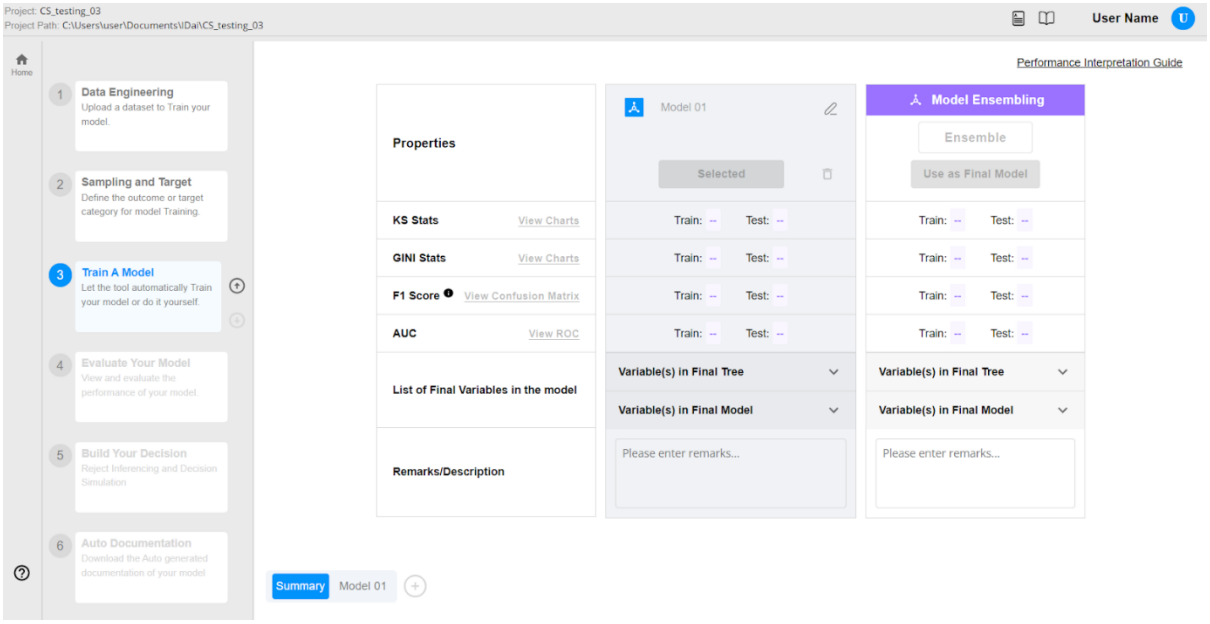
- Perform Ensembling, Compare Models with KS & Gini and Select Final Model to be used.



Train a Model Section

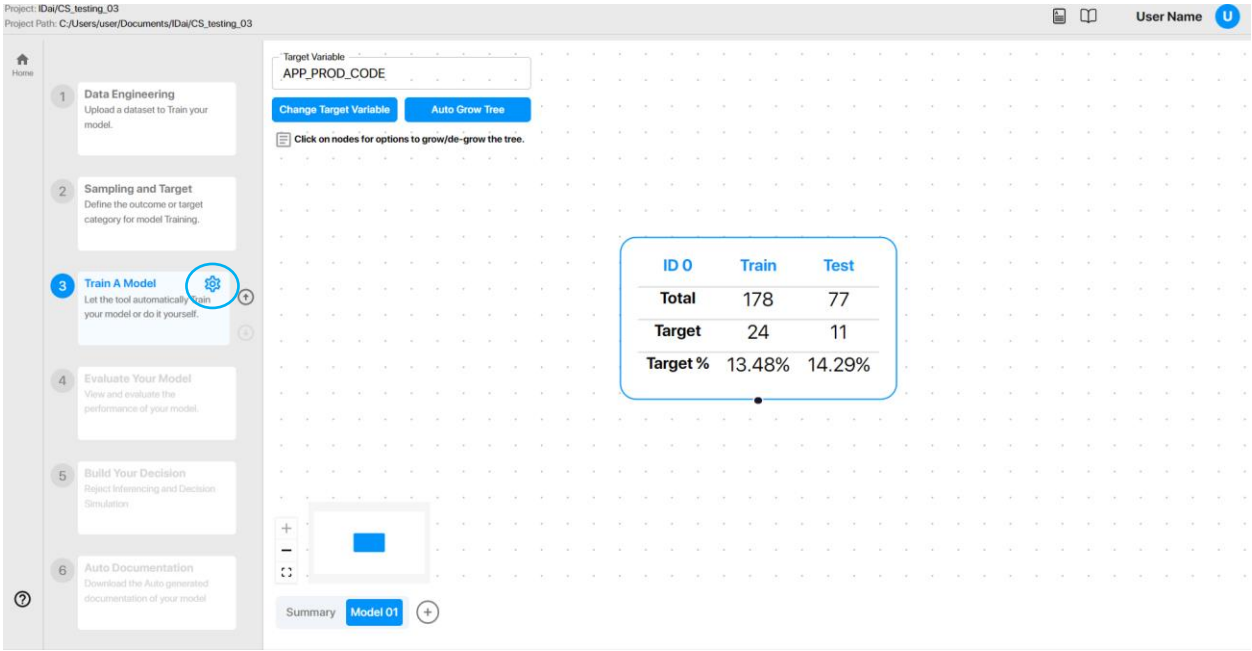
Model Settings and Root Node

Once all the input data has been finalized, the next stage is to build the model. The first screen (Summary of all models) is shown below. (This is what it looks like initial, before any models have been developed - by default model 01 is selected as the Final Model')



This guide helps users understand model performance at a glance by assigning color-coded scores to four key evaluation metrics: KS Statistic, Gini Coefficient, F1 Score, AUC (Area Under the Curve)

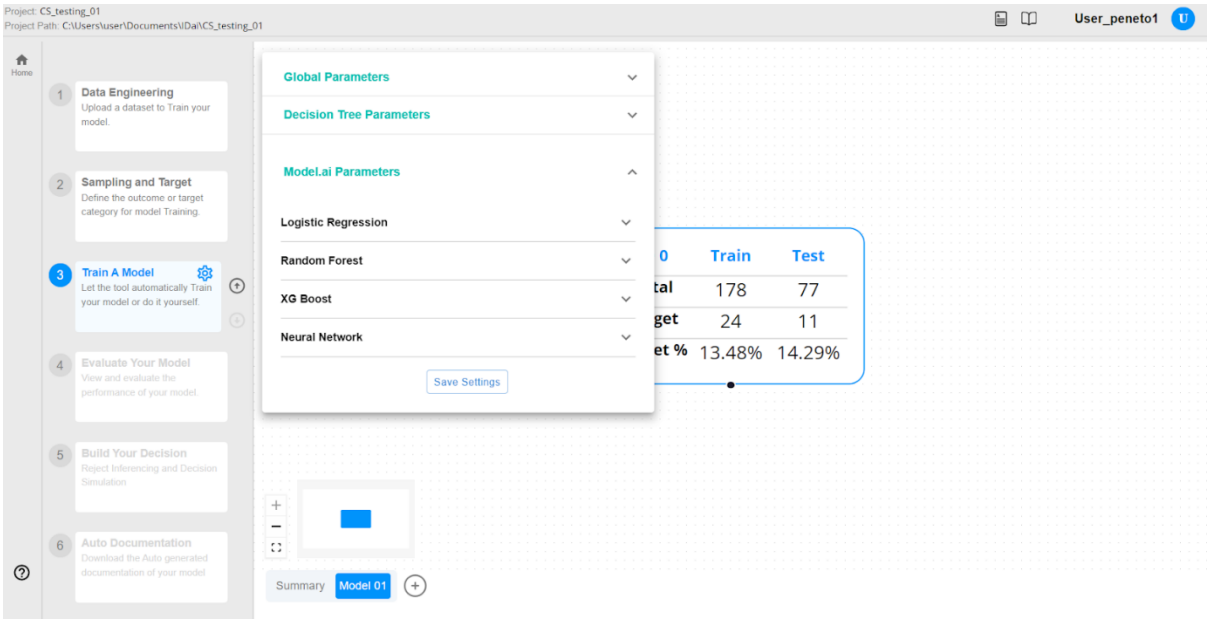
User can begin model creation by selecting a '**Model 01**' next to the '**Summary**' button

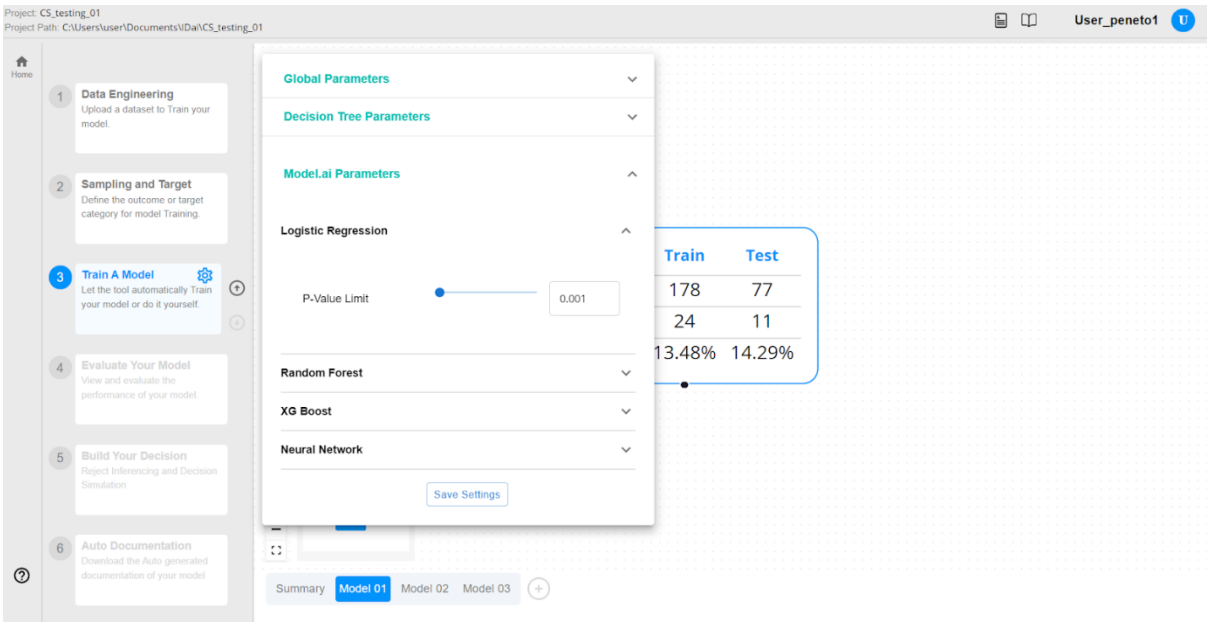


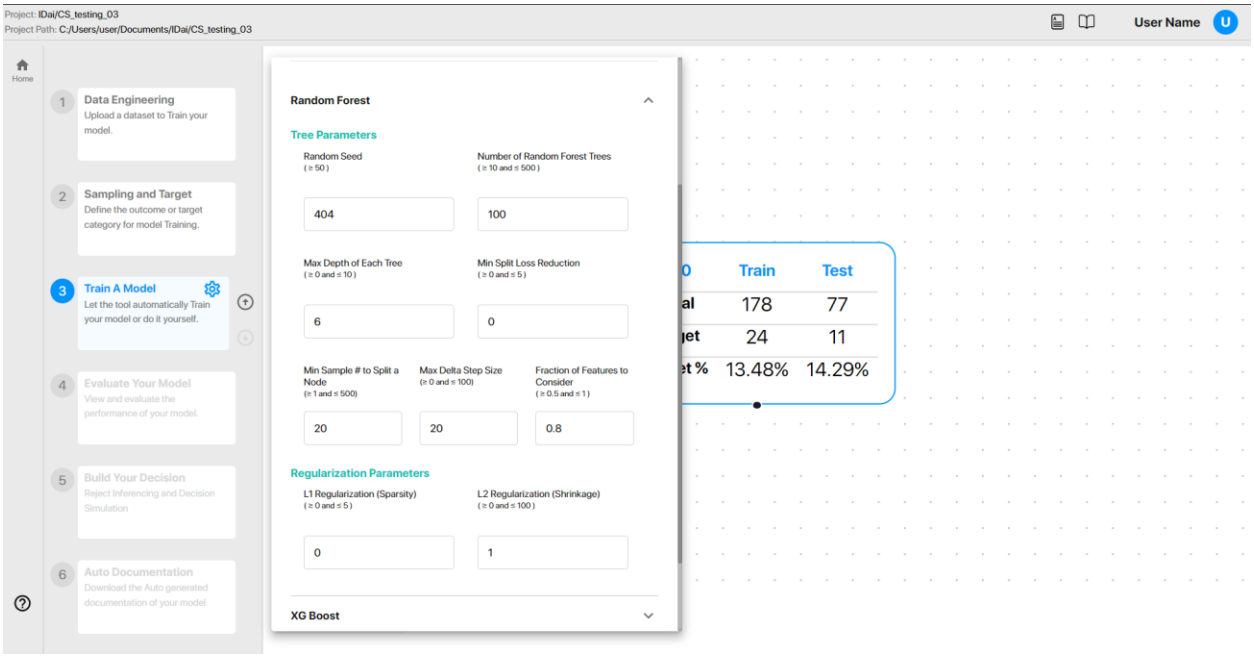
Before starting a model build, the user has the option to customize the following parameters by clicking on the settings symbol on the top right corner of the '**Train A Model**' tab (highlighted using a small blue circle above):

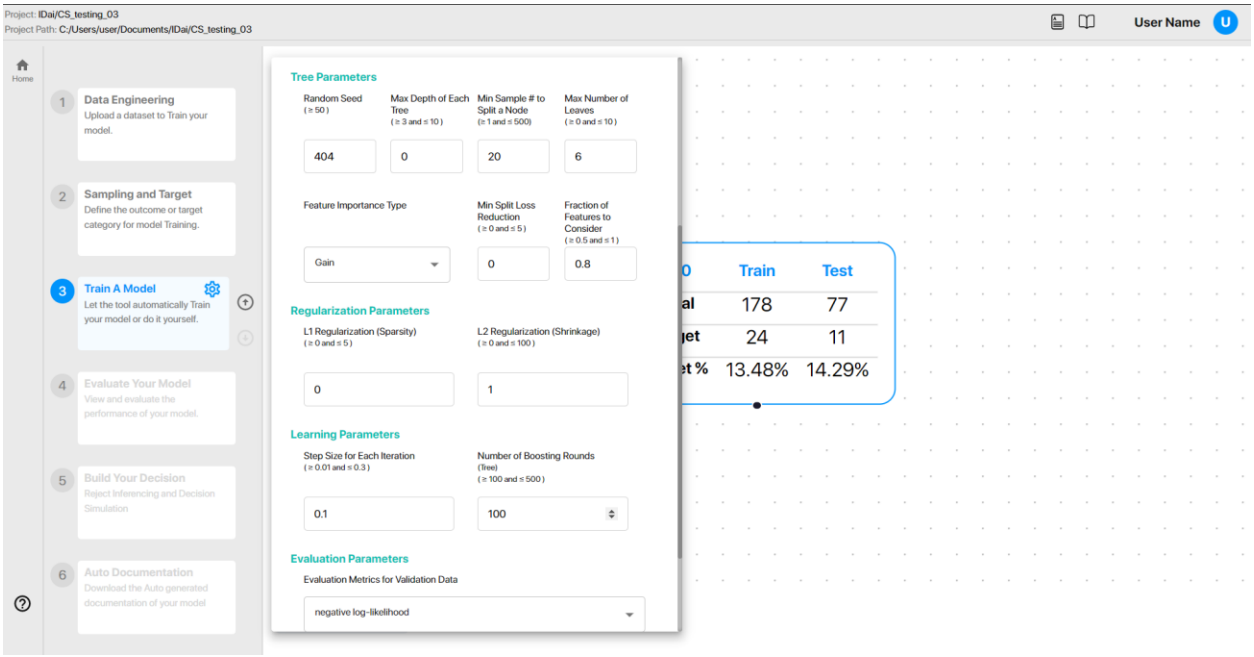
1. Global parameters: node size, maximum split levels, pairwise correlation limits, VIF and IV
2. Score scaling: Base Score, Base Odds and PDO (Probability of double odds)
3. Decision tree parameters: Minimum cases (#) or targets (# or %) in a node
4. Logistic model parameters: p-value limit
5. Random Forest and XGBoost parameters

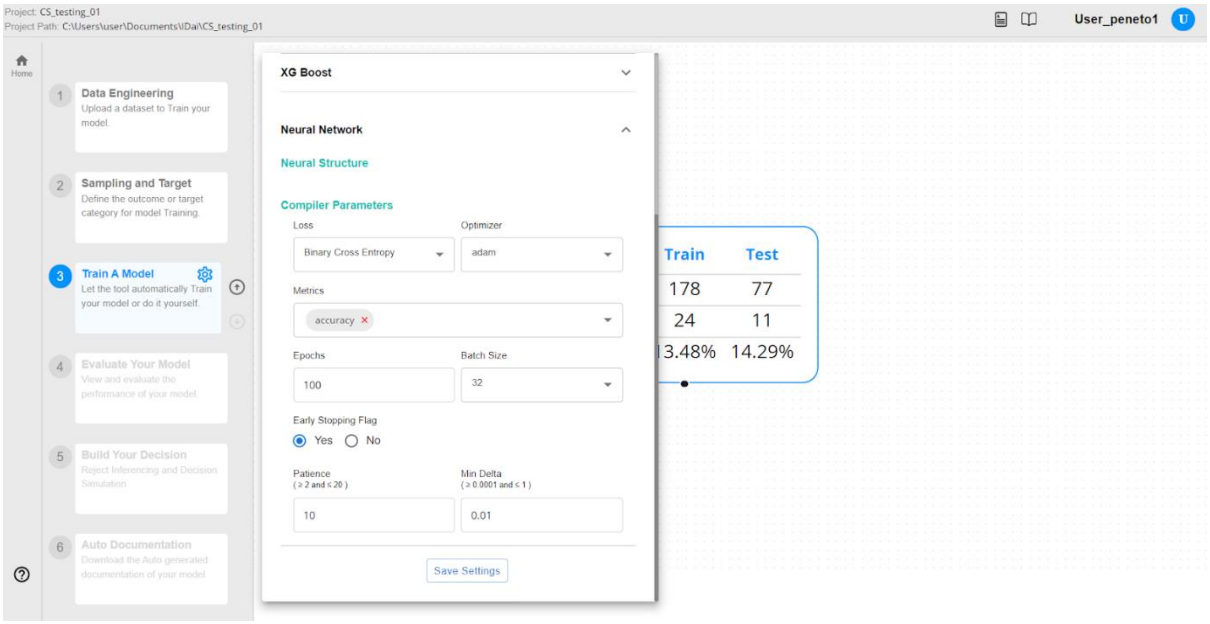
Once the user has input the desired settings, these are saved by clicking on the "**Update**" and "**Save Settings**" button on the respective pop-up boxes.


Project: CS_testing_01
Project Path: C:\Users\user\Documents\IDai\CS_testing_01
User_peneto1
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Global Parameters
Decision Tree Parameters
Model.ai Parameters
Logistic Regression
Random Forest
XG Boost
Neural Network
Save Settings
0
Train
Test
tal
178
77
get
24
11
et %
13.48%
14.29%
Summary
Model 01

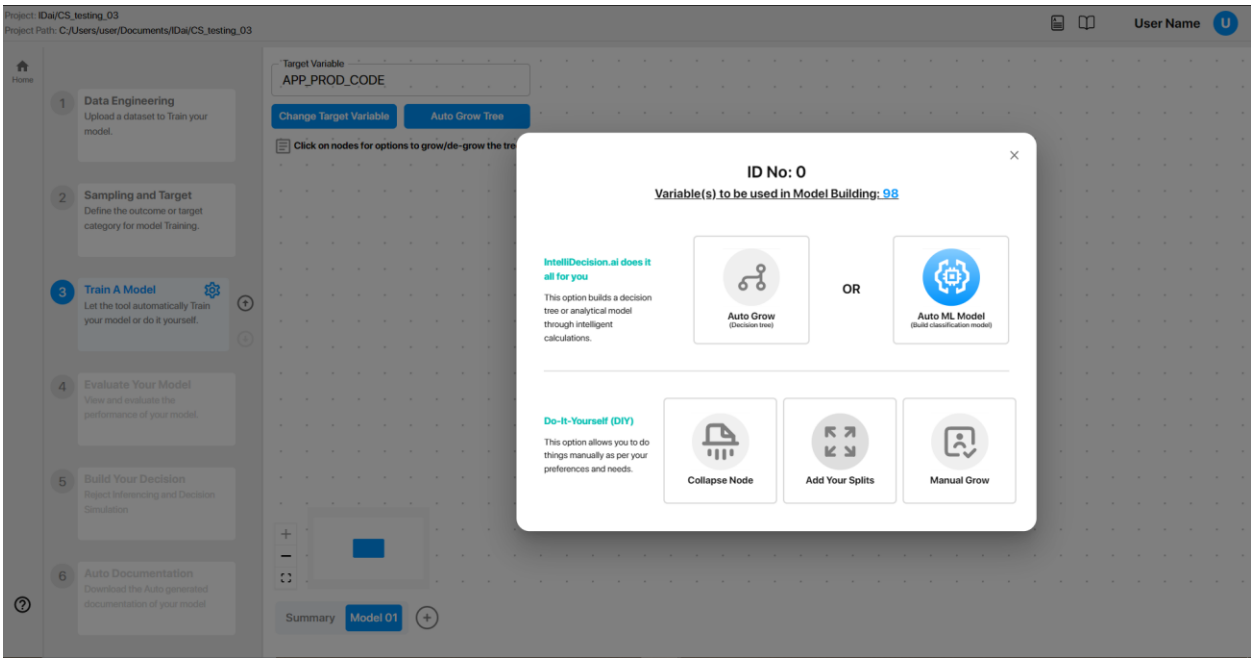

Project: CS_testing_01
Project Path: C:\Users\user\Documents\IDai\CS_testing_01
User_peneto1
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Global Parameters
Decision Tree Parameters
Model.ai Parameters
Logistic Regression
P-Value Limit
0.001
Random Forest
XG Boost
Neural Network
Save Settings
Train
Test
178
77
24
11
13.48%
14.29%
Summary
Model 01
Model 02
Model 03


Project: IDai/CS_testing_03
Project Path: C:/Users/user/Documents/IDai/CS_testing_03
User Name
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Random Forest
Tree Parameters
Random Seed (≥ 50)
404
Number of Random Forest Trees (≥ 10 and ≤ 500)
100
Max Depth of Each Tree (≥ 0 and ≤ 10)
6
Min Split Loss Reduction (≥ 0 and ≤ 5)
0
Min Sample # to Split a Node (≥ 1 and ≤ 500)
20
Max Delta Step Size (≥ 0 and ≤ 100)
20
Fraction of Features to Consider (≥ 0.5 and ≤ 1)
0.8
Regularization Parameters
L1 Regularization (Sparsity) (≥ 0 and ≤ 5)
0
L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)
1
XG Boost
0
Train
Test
al
178
77
get
24
11
et %
13.48%
14.29%


Project: IDai/CS_testing_03
Project Path: C:/Users/user/Documents/IDai/CS_testing_03
User Name
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Tree Parameters
Random Seed (≥ 50)
404
Max Depth of Each Tree (≥ 3 and ≤ 10)
0
Min Sample # to Split a Node (≥ 1 and ≤ 500)
20
Max Number of Leaves (≥ 0 and ≤ 10)
6
Feature Importance Type
Gain
Min Split Loss Reduction (≥ 0 and ≤ 5)
0
Fraction of Features to Consider (≥ 0.5 and ≤ 1)
0.8
Regularization Parameters
L1 Regularization (Sparsity) (≥ 0 and ≤ 5)
0
L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)
1
Learning Parameters
Step Size for Each Iteration (≥ 0.01 and ≤ 0.3)
0.1
Number of Boosting Rounds (Tree) (≥ 100 and ≤ 500)
100
Evaluation Parameters
Evaluation Metrics for Validation Data
negative log-likelihood
0
Train
Test
al
178
77
get
24
11
et %
13.48%
14.29%



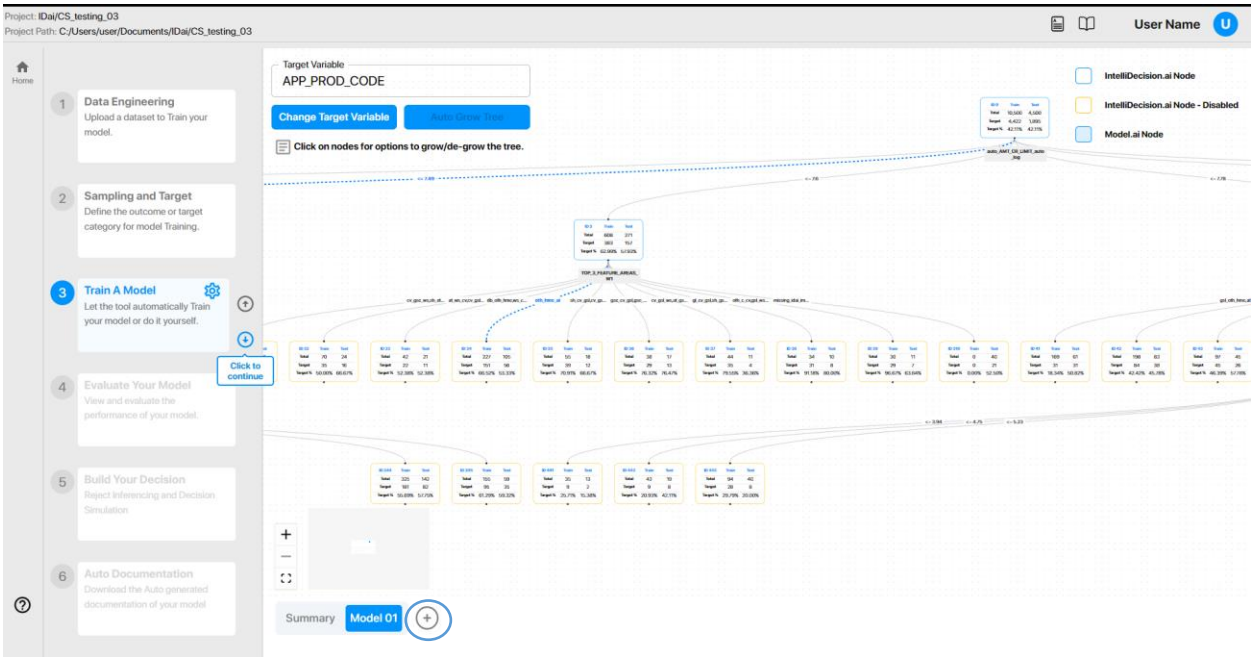
Once the settings are finalized, the user can click on the root node to get multiple options to either grow the decision tree ("**Auto Grow**") or build an AI based model ("**Run logistic** or **Random Forest** or **XGBoost**"). In both these cases, the user does not have any control over the resulting tree and segmentation.



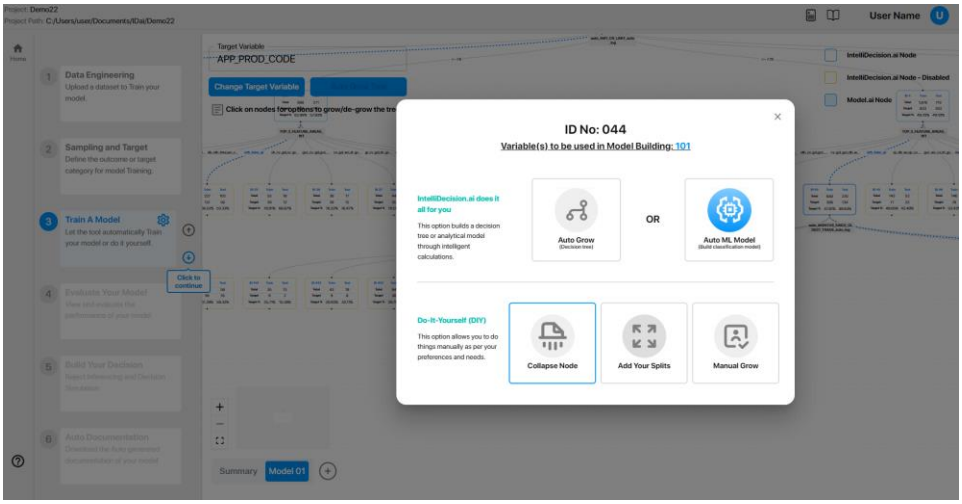
Auto Grow

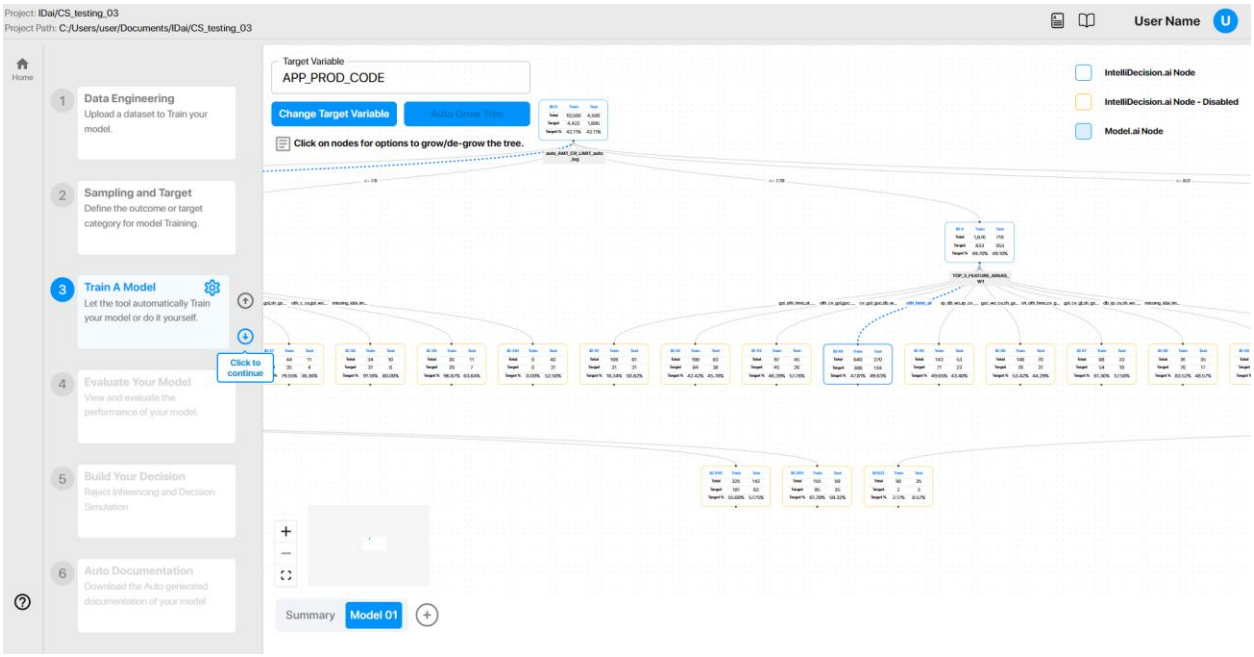
When the "**Auto Grow**" option is selected, a tree is developed starting from the root node and progressing to subsequent levels based on robust separation within each level and across different levels.

By Using '**+**' button we will get new screen of root node of tree, here we can run different operations based on our requirement.

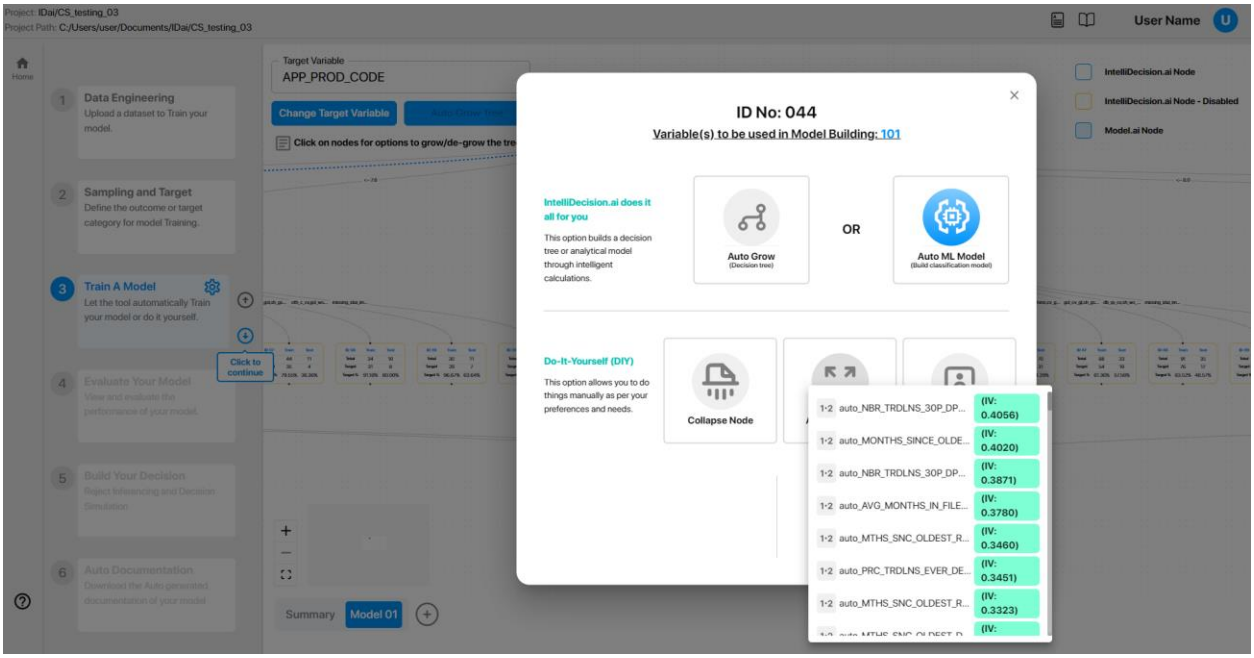


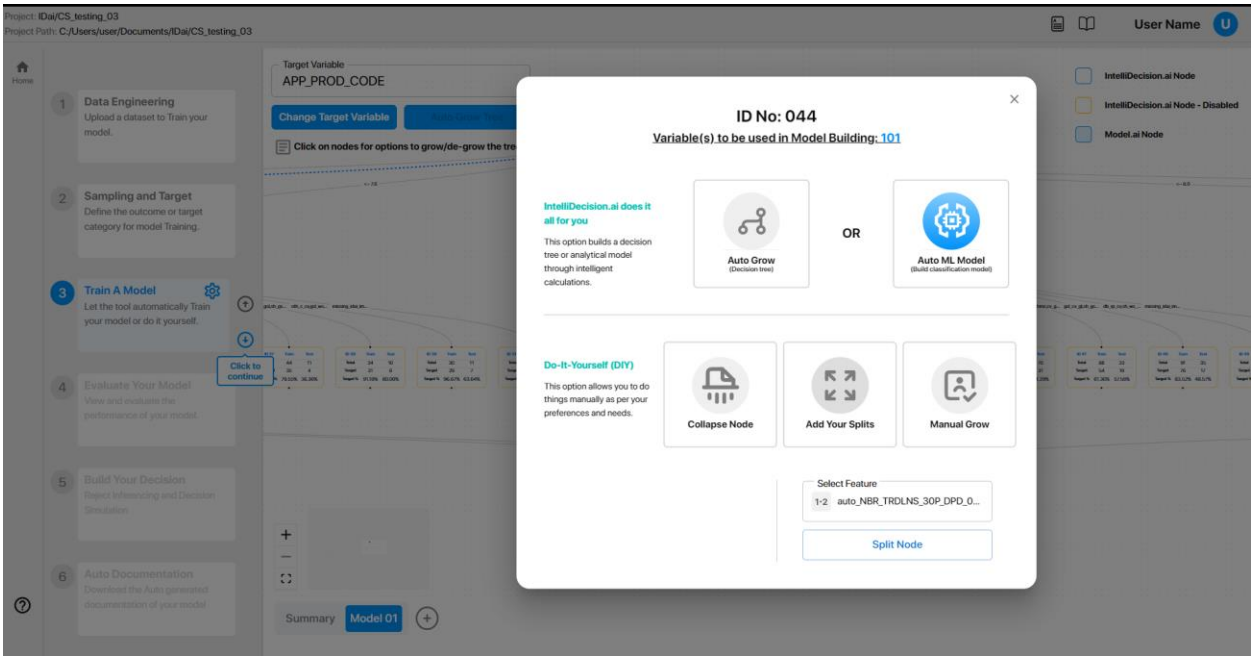
If user does not wish to split a particular node into sub-classes, they can go back to that node and click on "**Collapse Node**" option.

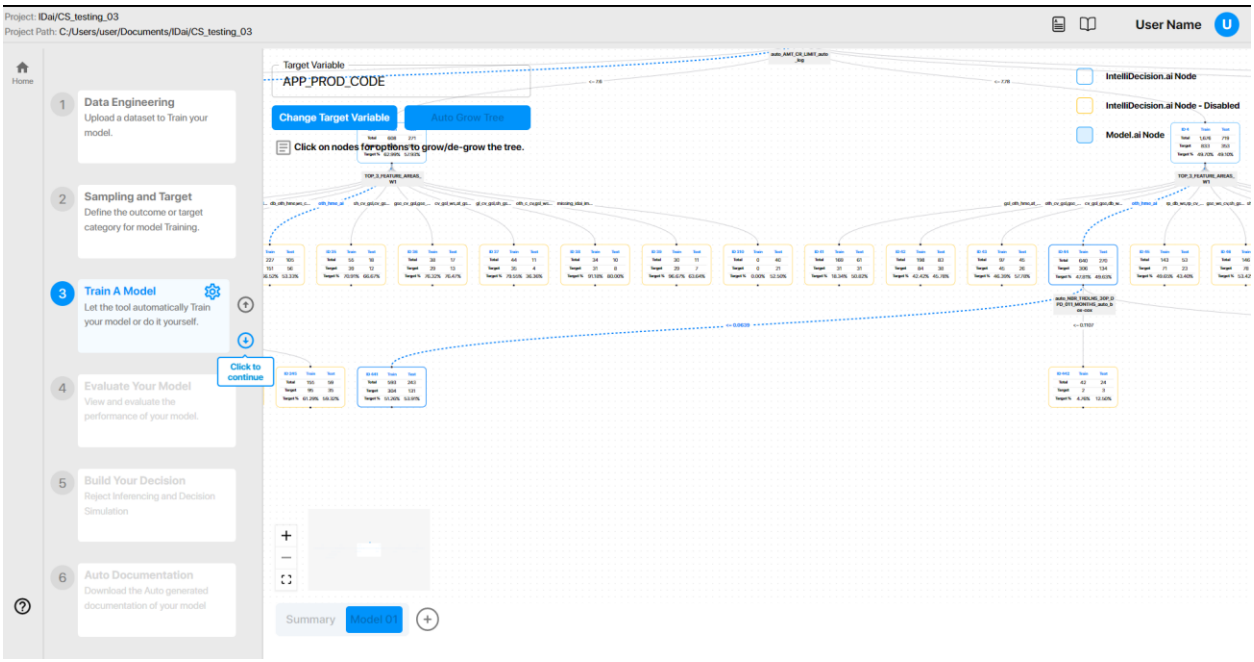




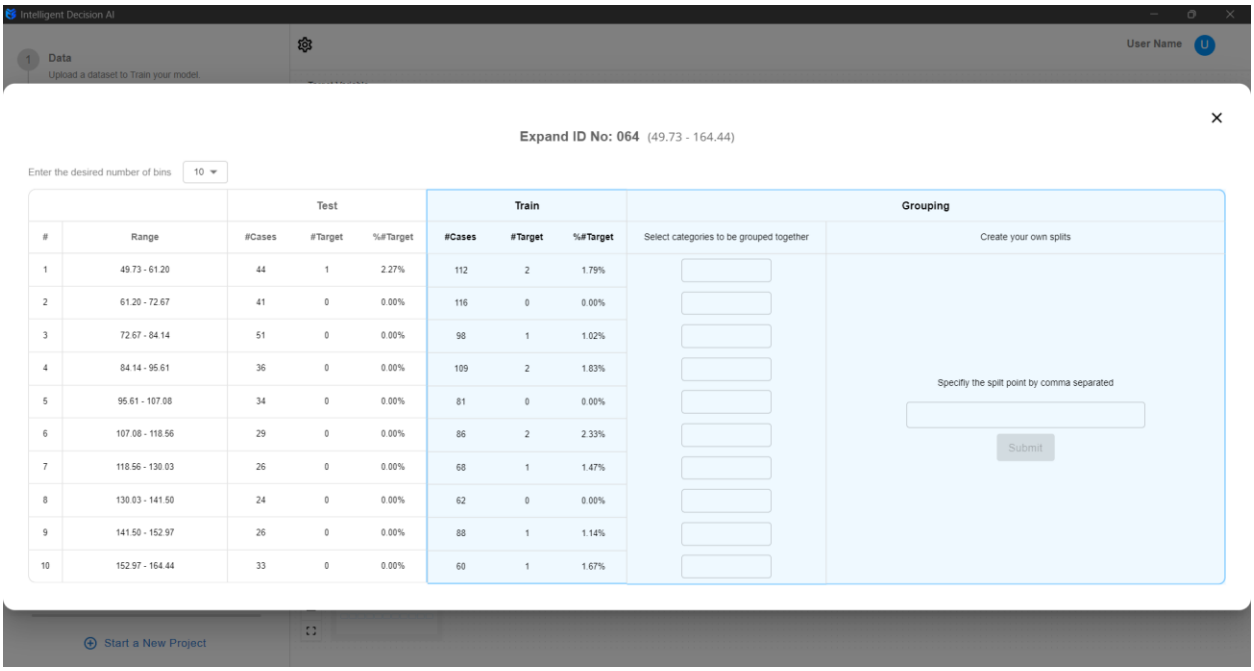
To insert an additional split within a node, user can click on "**Manual Grow**" within that node. User gets the options to choose any variable to be split then click on "**Split Node**"; the IV values are provided to make an informed decision. Do note that the variables with the highest IVs are placed at the top.

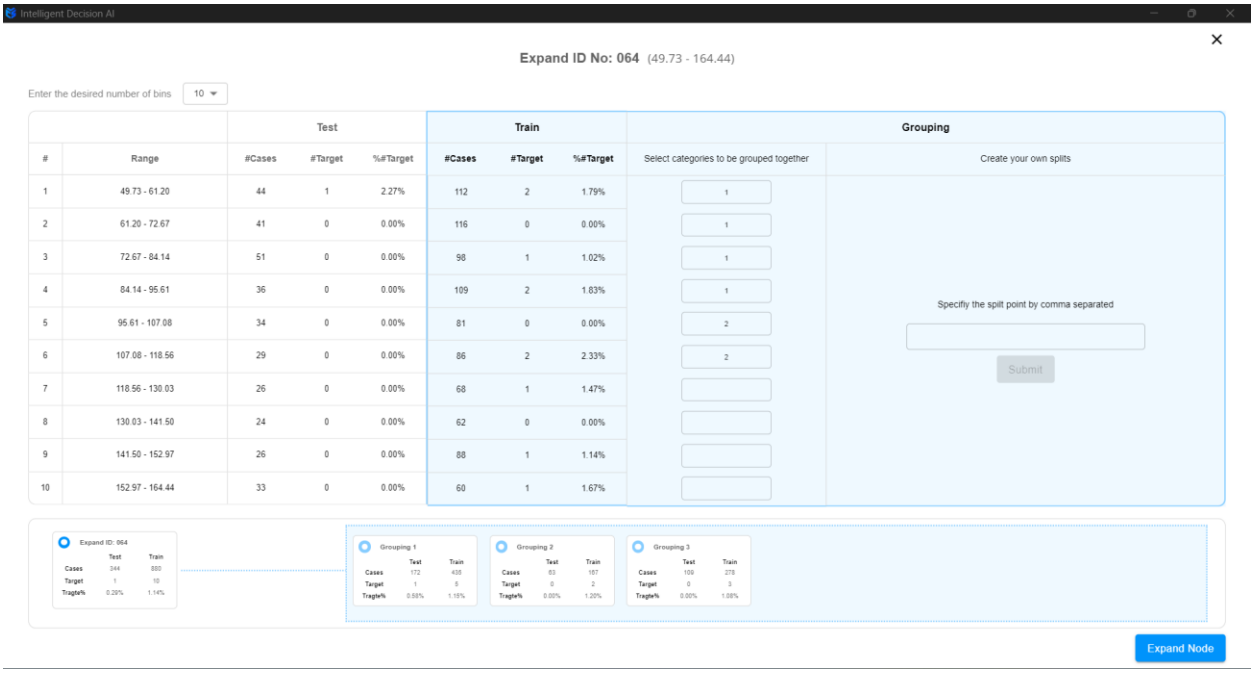




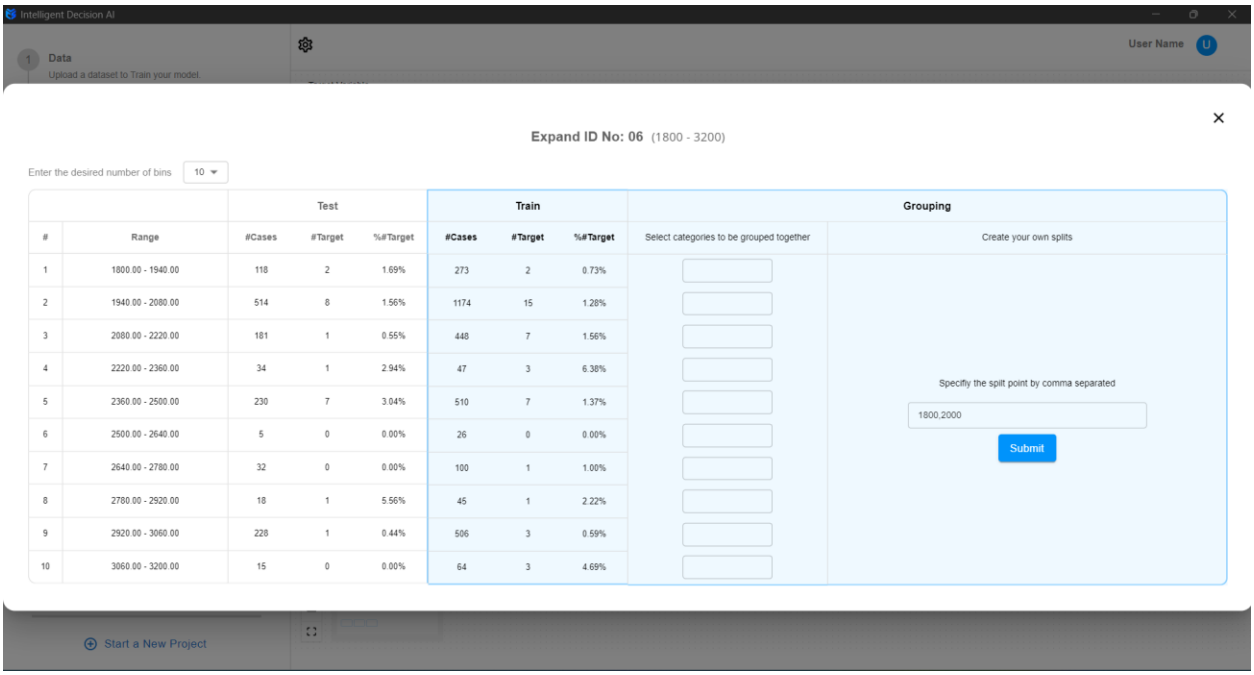


Once the automatically generated splits are available, the user can generate custom splits within this variable by clicking on "**Add your splits**". This leads to a dialog box where user can assign the variable ranges in integer-based groups.



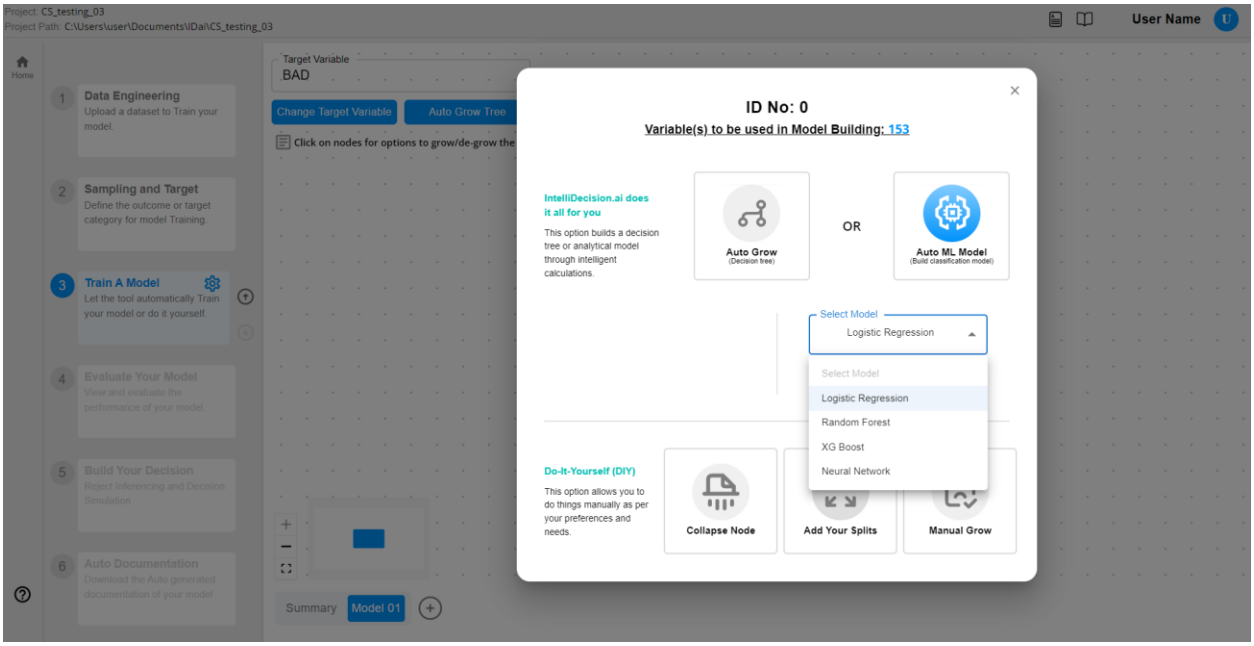


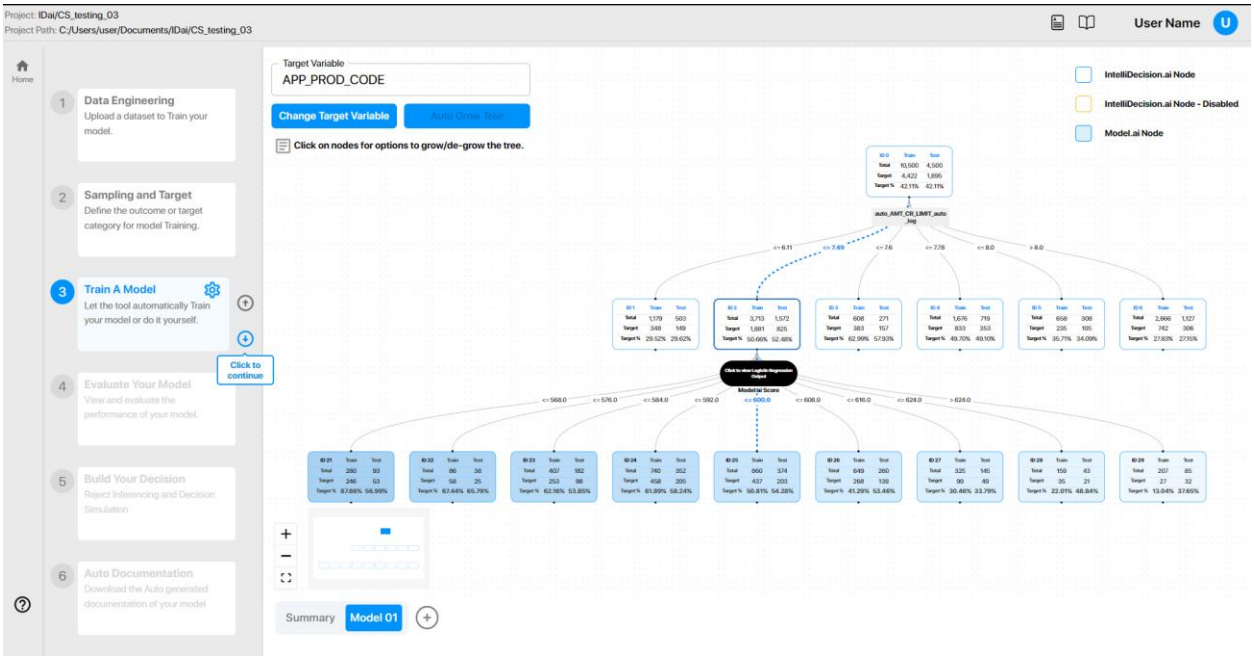
Another way to split is to “Specify the split point by comma-separated” values in the right of the screen. Press “**Submit**” after that. Click on **Expand Node**.


Intelligent Decision AI
User Name
Data
Upload a dataset to Train your model.
Expand ID No: 06 (1800 - 3200)
Enter the desired number of bins 10
Test
Train
Grouping
Range #Cases #Target %#Target #Cases #Target %#Target Select categories to be grouped together Create your own splits
1 1800.00 - 1940.00 118 2 1.69% 273 2 0.73%
2 1940.00 - 2080.00 514 8 1.56% 1174 15 1.28%
3 2080.00 - 2220.00 181 1 0.55% 448 7 1.56%
4 2220.00 - 2360.00 34 1 2.94% 47 3 6.38%
5 2360.00 - 2500.00 230 7 3.04% 510 7 1.37%
6 2500.00 - 2640.00 5 0 0.00% 26 0 0.00%
7 2640.00 - 2780.00 32 0 0.00% 100 1 1.00%
8 2780.00 - 2920.00 18 1 5.56% 45 1 2.22%
9 2920.00 - 3060.00 228 1 0.44% 506 3 0.59%
10 3060.00 - 3200.00 15 0 0.00% 64 3 4.69%
Specify the split point by comma separated
1800,2000
Submit
Start a New Project

Run Model.ai

In this option, the user just clicks on "**Auto ML Model**" button, select the model type and the results are automatically generated. User does not have the option to split or collapse these nodes.

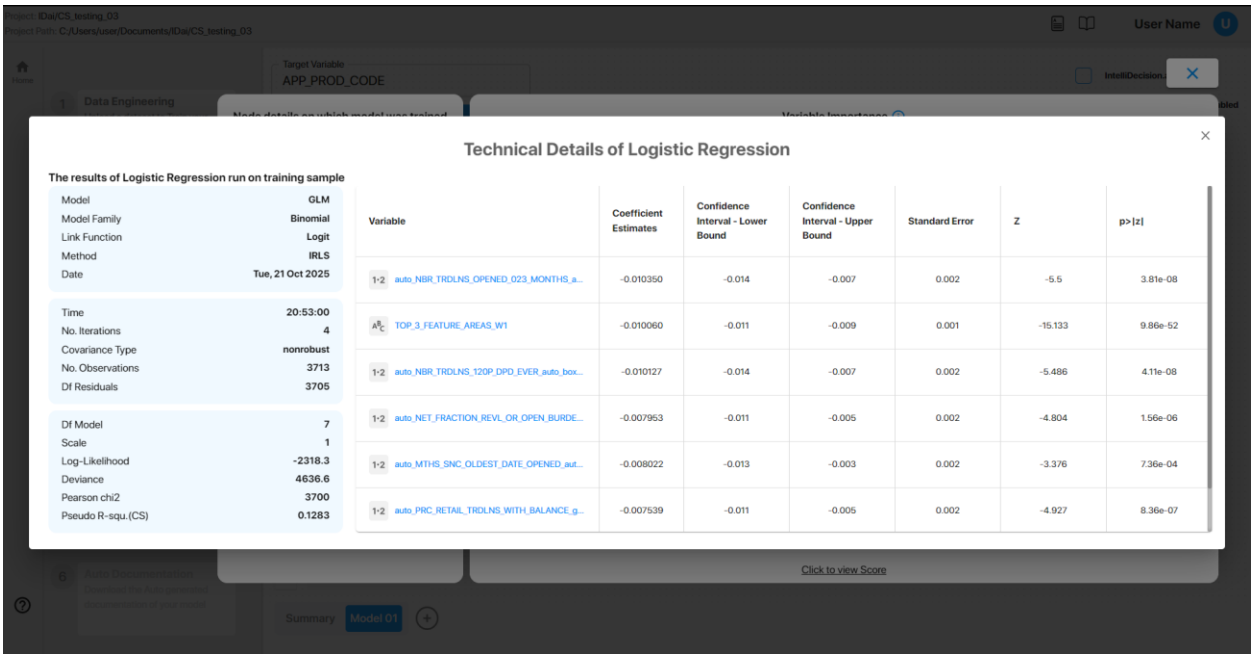




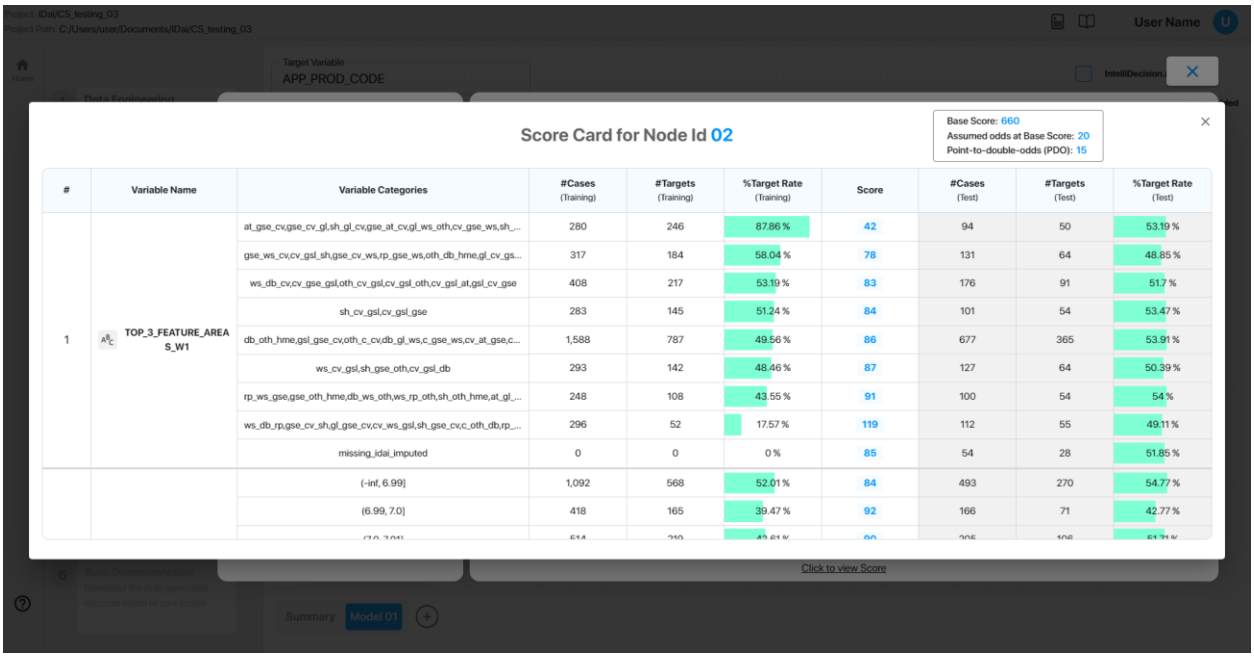
User can click on "**Click to view Logistic Regression Output**" to see the following results for the node on which model.ai was run:

1. Node details, variable importance, model performance metrics & target rate by score.

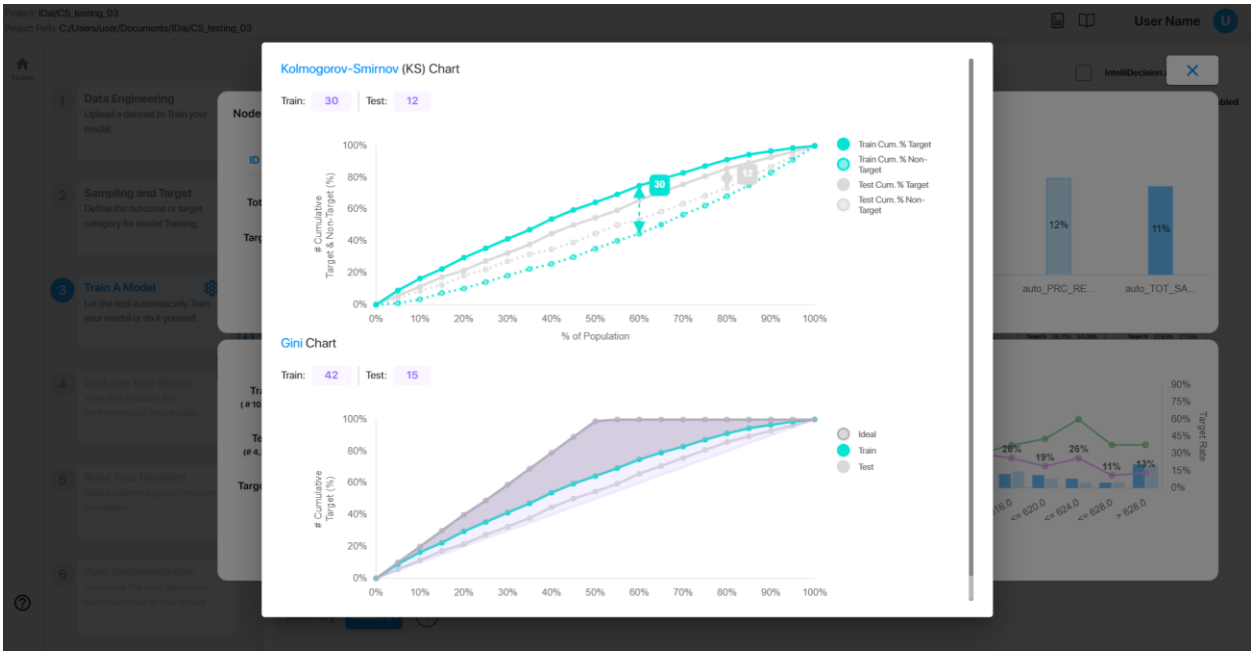
2. Logistic regression technical details



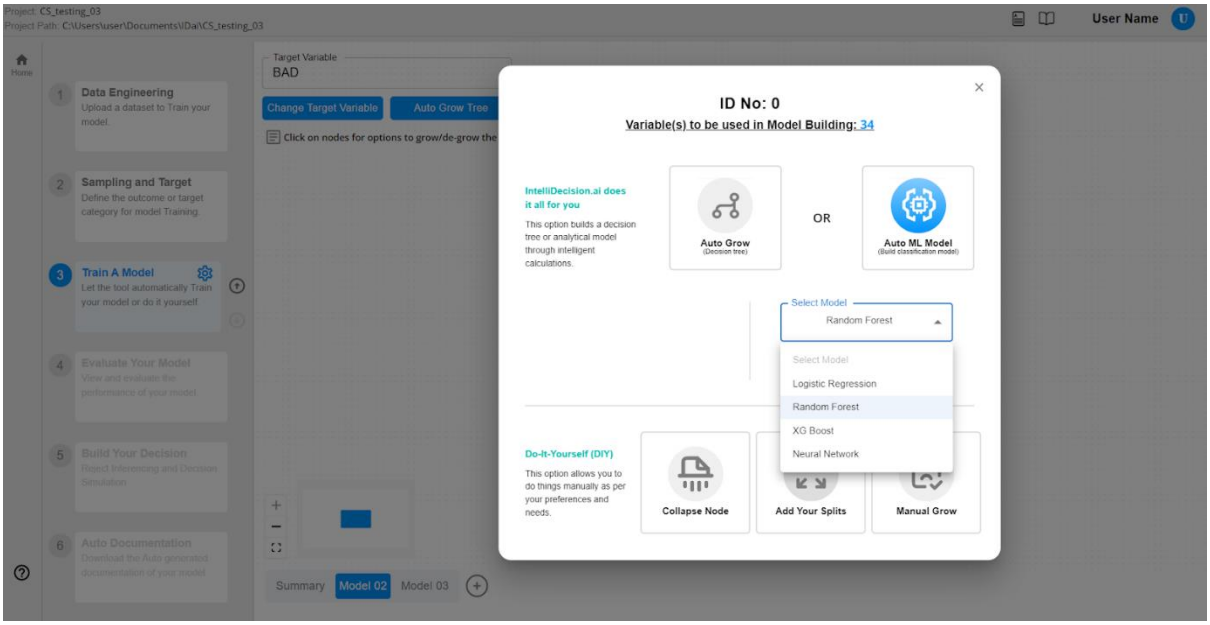
3. Scorecard details for that node



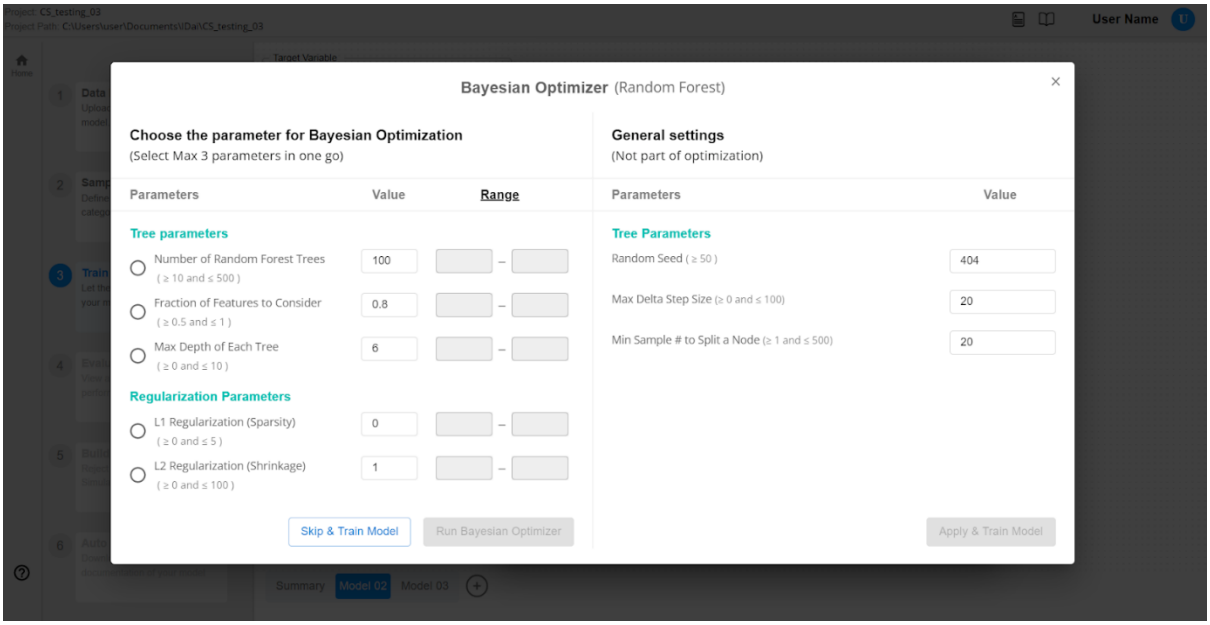
4. User can click on "**Click to see the graph**" in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.



1. Random Forest Model Tree

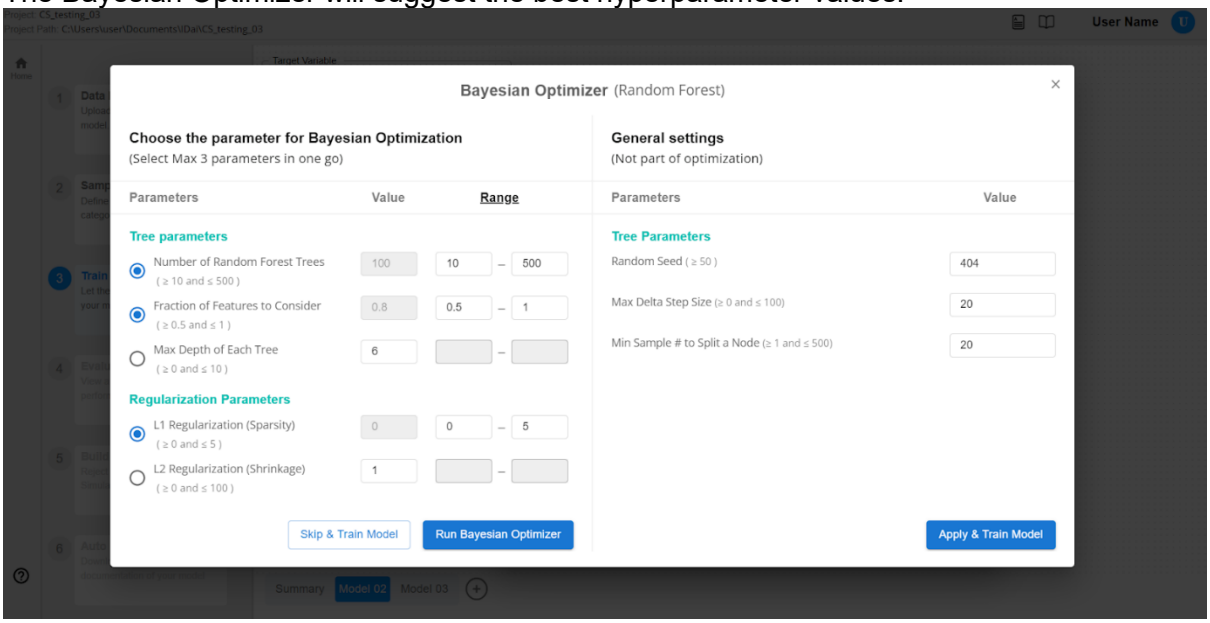


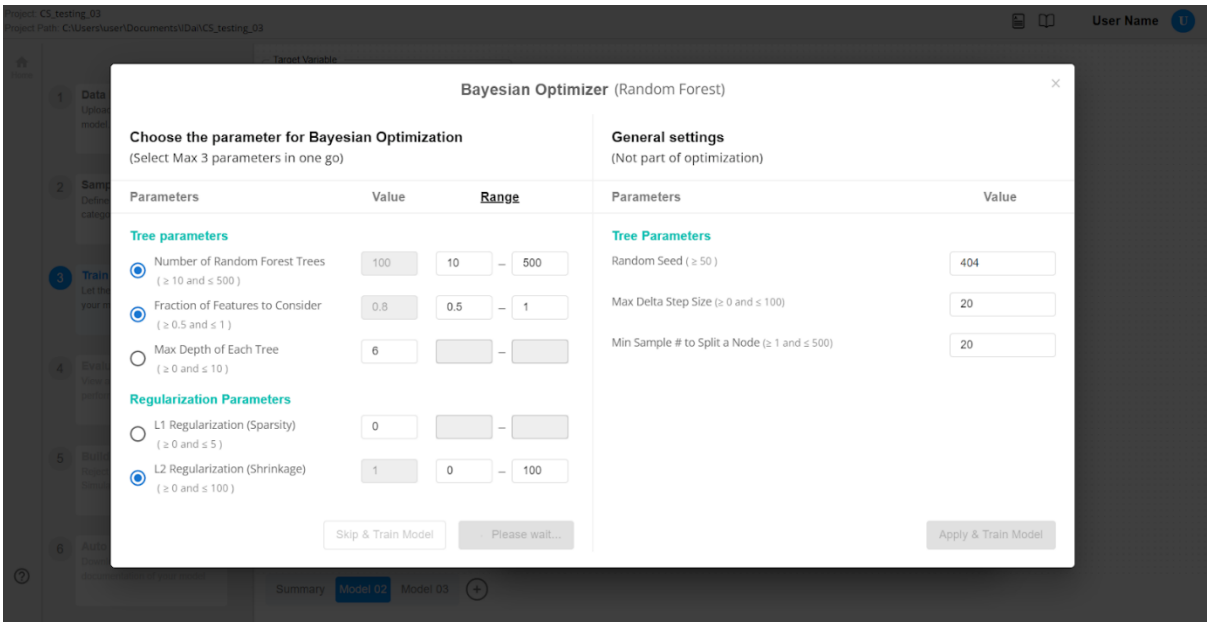
2. Run Bayesian Optimization

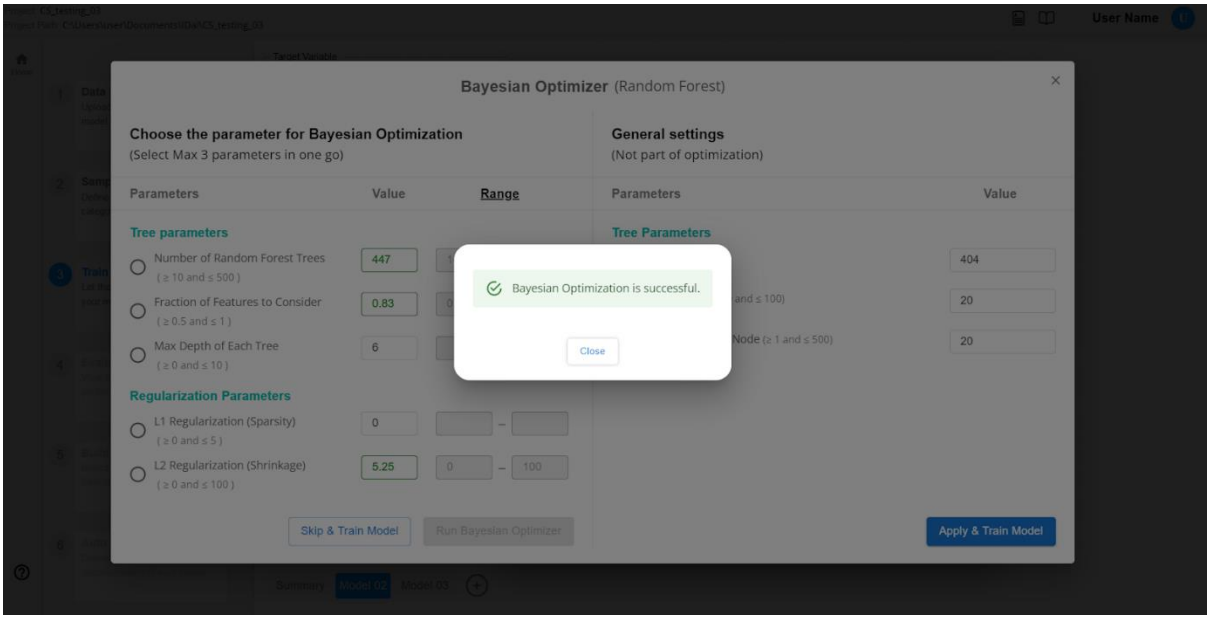


The user can either select "**Skip & Train Model**" or configure options under "**Tree Parameters**" and/or "**Regularization Parameters**", then click "**Run Bayesian Optimizer**".

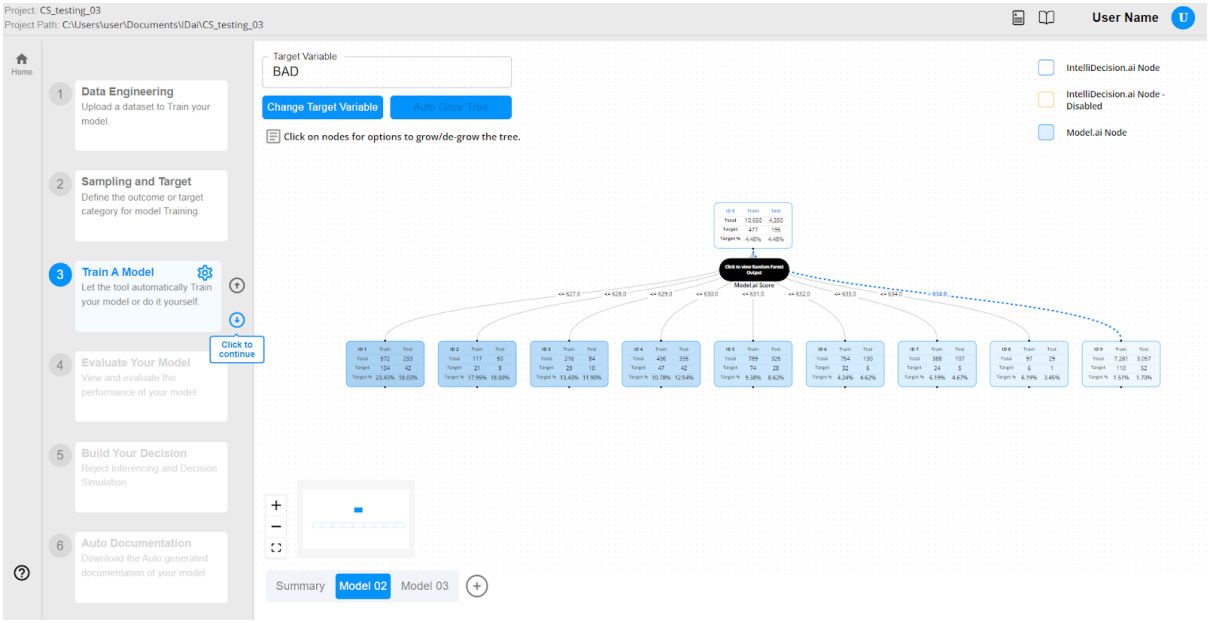
The Bayesian Optimizer will suggest the best hyperparameter values.







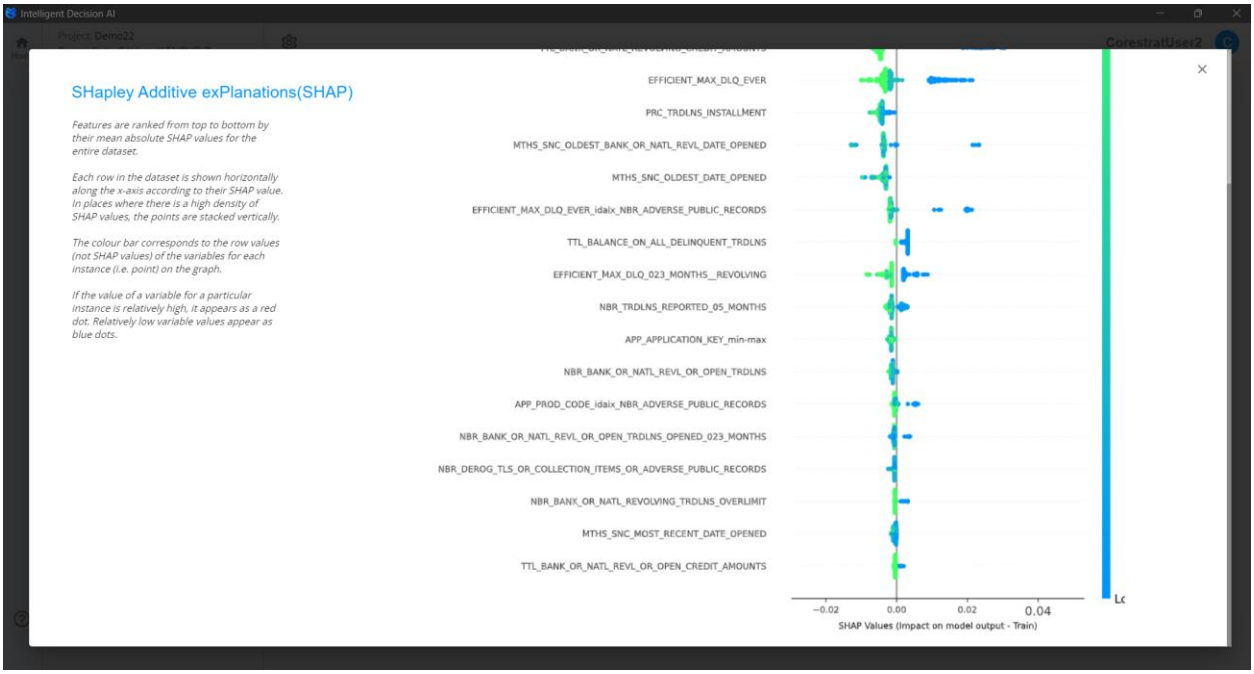
After this, the user can click "**Apply & Train Model**".



3. Random Forest provides a bar graph of variable importance and a SHAP chart along with KS and GINI charts. which can be viewed by clicking on **'Click to view Random Forest Output'**.

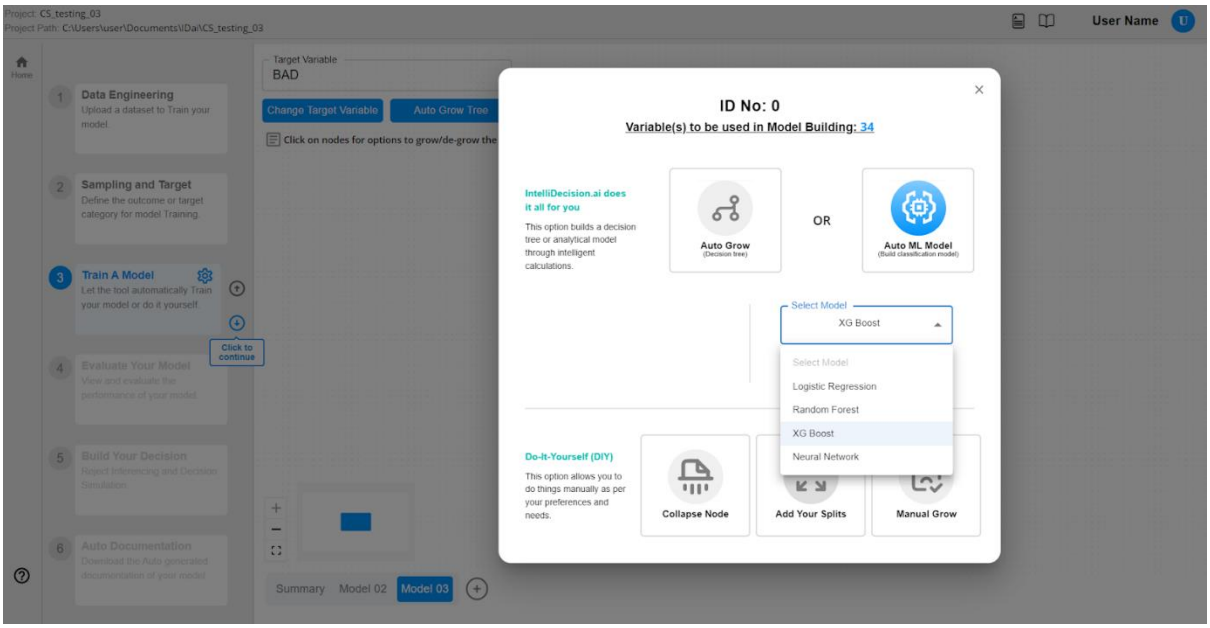

Intelligent Decision AI
Project: Demo22
Project Path: C:\Users\LENOVO\Docum...
CorestratUser2
Target Variable
BAD
ID.ai Node
1 Data
Upload a dataset to Train your model
2 Prepare your Data
Define the outcome or target category for model Training
3 Train A Model
Let the tool automatically Train your model or do it yourself
4 Evaluate Your Model
View and evaluate the performance of your model
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Node details on which model was trained using Model.ai
ID 4 Train Test
Total 664 292
Target 35 18
Variable Importance
14% 14% 11% 9% 8% 7% 6% 5% 5% 4% 3% 2% 2% 2% 1% 1% 1% 1% 1%
HIGHEST_UTI...
TTL_TRADE_L...
NBR_BANK_OR...
MTHS_SNC_OL...
MTHS_SNC_OL...
PRC_TRDLNS_...
TTL_BANK_OR...
EFFICIENT_M...
APP_APPLICA...
EFFICIENT_M...
EFFICIENT_M...
NBR_TRDLNS_...
NBR_BANK_OR...
NBR_BANK_OR...
NBR_DEROG_T...
TTL_BANK_OR...
MTHS_SNC_MO...
APP_PROD_CO...
NBR_TRDLNS_...
Click to view SHAP Chart
Model Performance Metrics
Train (# 10,500) 76 KS 90 GINI
Test (# 4,500) 13 KS -10 GINI
Target % 5.3% 6.2%
Click to see the graph
Target Rate by Model Score
Population %
Target Rate
100% 77% 28% 21% 15% 5% 5% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0%
<= 617.0 <= 623.0 <= 630.0 <= 636.0 <= 643.0 <= 649.0 <= 655.0 <= 662.0 <= 668.0 <= 675.0 <= 681.0 <= 688.0 <= 694.0 <= 700.0 <= 707.0 <= 713.0 <= 719.0 <= 725.0 > 725.0
% Target Train
% Target Test
% Cases Train
% Cases Test
Summary
Model 01
Model 02
Model 03


SHapley Additive exPlanations(SHAP)
Features are ranked from top to bottom by their mean absolute SHAP values for the entire dataset.
Each row in the dataset is shown horizontally along the x-axis according to their SHAP value. In places where there is a high density of SHAP values, the points are stacked vertically.
The colour bar corresponds to the row values (not SHAP values) of the variables for each instance (i.e. point) on the graph.
If the value of a variable for a particular instance is relatively high, it appears as a red dot. Relatively low variable values appear as blue dots.
HIGHEST_UTIL_ON_BANK_OR_NATL_REVL_TRDLNS
TTL_TRADE_LINE_BALANCES
NBR_BANK_OR_NATL_REVOLVING_TRDLNS_WITH_BALANCE_gt2000
TTL_BANK_OR_NATL_REVOLVING_CREDIT_AMOUNTS
EFFICIENT_MAX_DLQ_EVER
PRC_TRDLNS_INSTALLMENT
MTHS_SNC_OLDEST_BANK_OR_NATL_REVL_DATE_OPENED
MTHS_SNC_OLDEST_DATE_OPENED
EFFICIENT_MAX_DLQ_EVER_idaix_NBR_ADVERSE_PUBLIC_RECORDS
TTL_BALANCE_ON_ALL_DELINQUENT_TRDLNS
EFFICIENT_MAX_DLQ_023_MONTHS__REVOLVING
NBR_TRDLNS_REPORTED_05_MONTHS
APP_APPLICATION_KEY_min-max
NBR_BANK_OR_NATL_REVL_OR_OPEN_TRDLNS
APP_PROD_CODE_idaix_NBR_ADVERSE_PUBLIC_RECORDS
NBR_BANK_OR_NATL_REVL_OR_OPEN_TRDLNS_OPENED_023_MONTHS
NBR_DEROG_TLS_OR_COLLECTION_ITEMS_OR_ADVERSE_PUBLIC_RECORDS

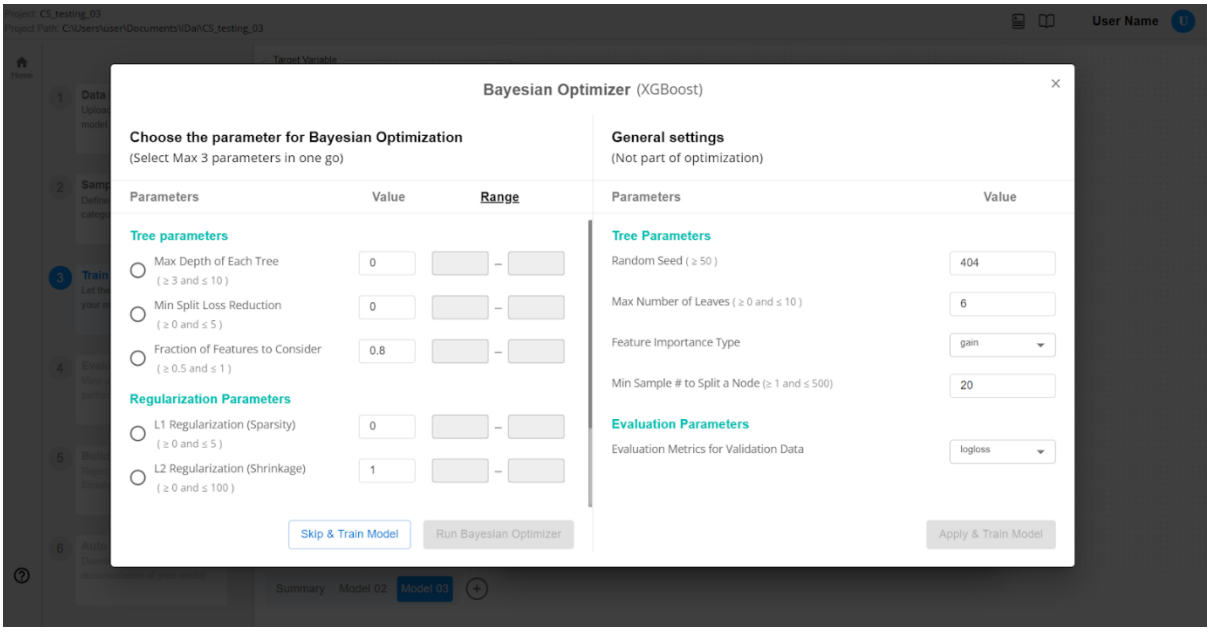


4. User can click on “**Click to view SHAP chart**” to see the SHAP chart, s(he) can also select “**Click to see the graph**” in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.

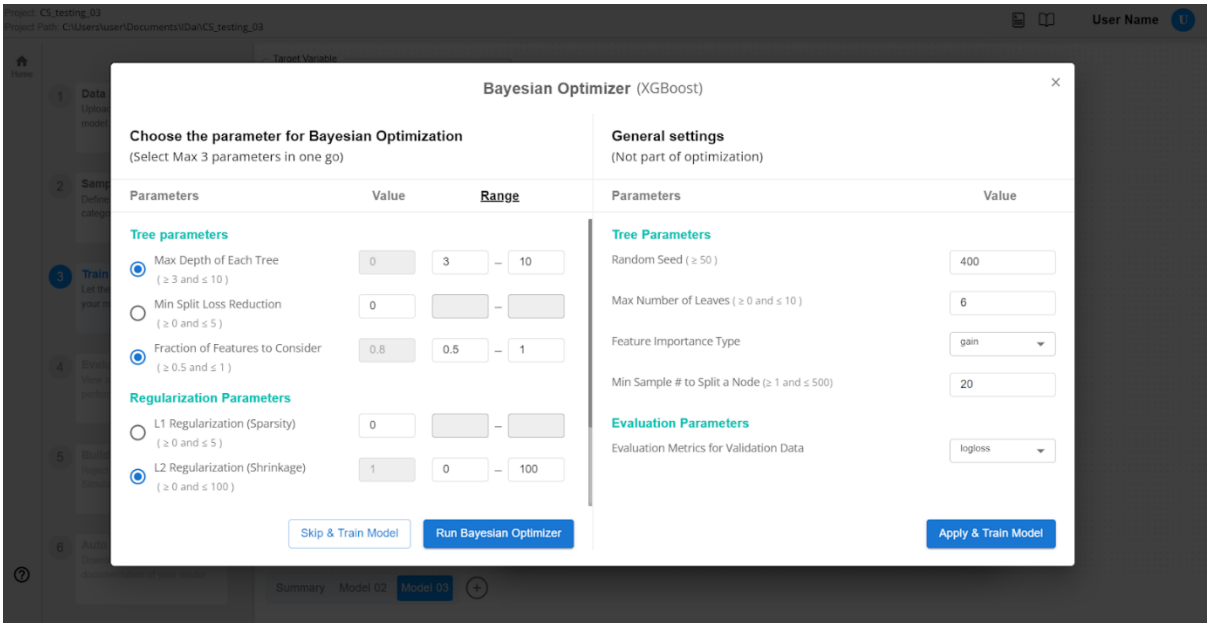
1.**XGBoost model Tree**

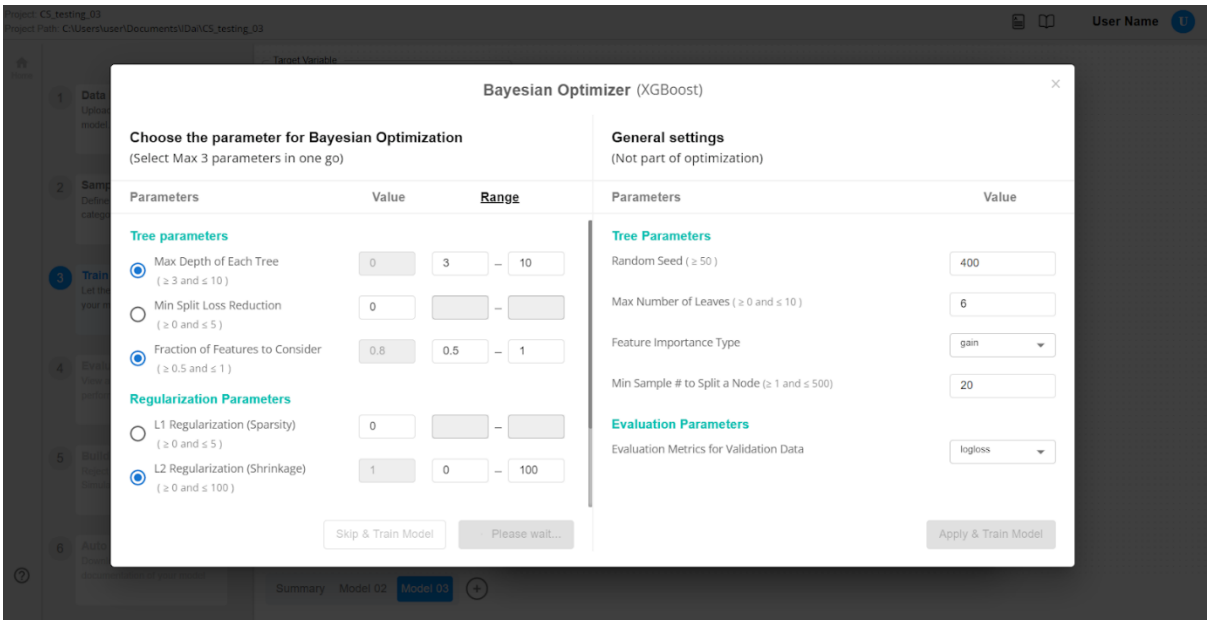


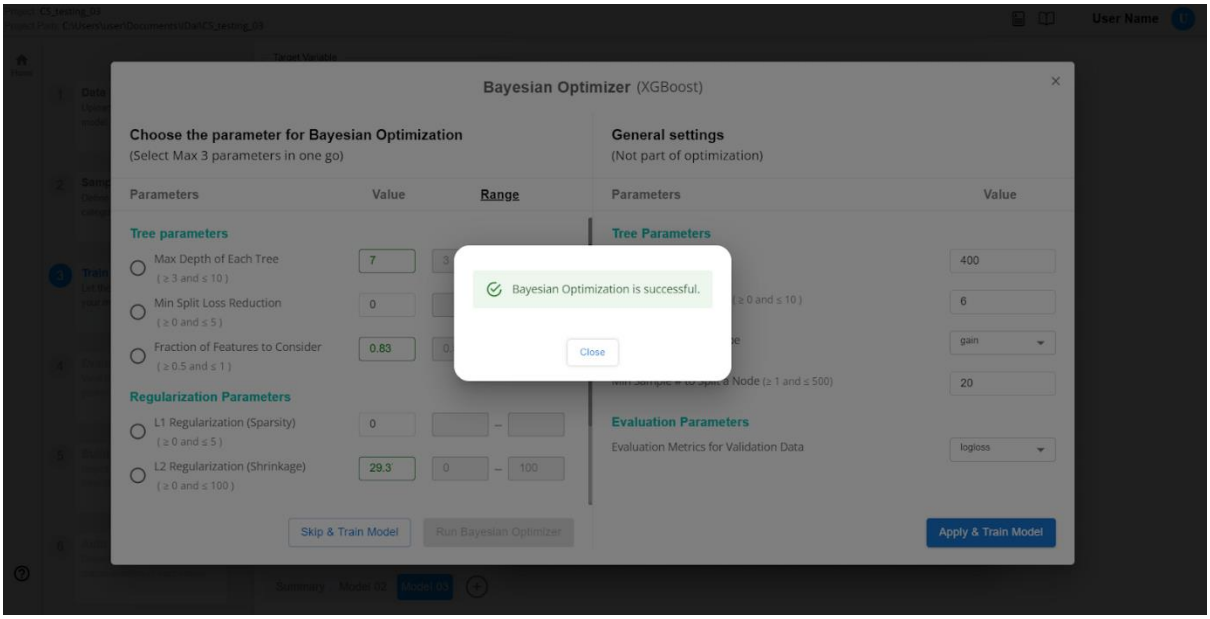
2.**Run Bayesian Optimization**



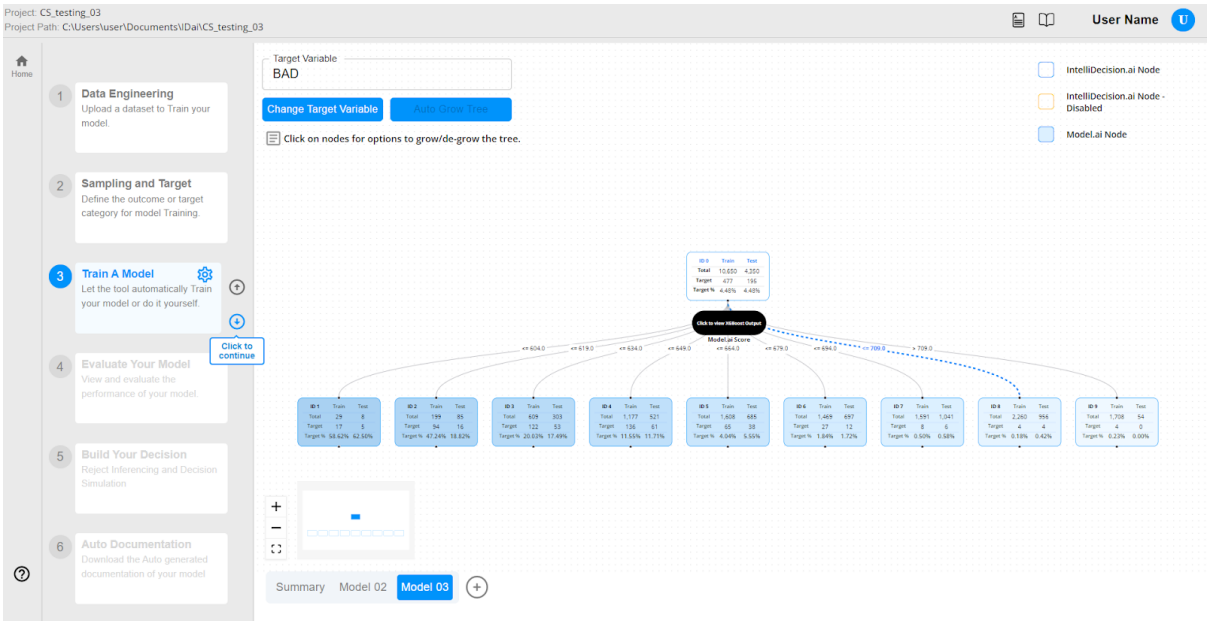
The user can either choose "**Skip & Train Model**" or select the hyperparameters to optimize and run the Bayesian Optimizer.


Bayesian Optimizer (XGBoost)
Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)
General settings
(Not part of optimization)
Parameters
Value
Range
Tree parameters
Max Depth of Each Tree
Min Split Loss Reduction
Fraction of Features to Consider
Regularization Parameters
L1 Regularization (Sparsity)
L2 Regularization (Shrinkage)
Tree Parameters
Random Seed
Max Number of Leaves
Feature Importance Type
Min Sample # to Split a Node
Evaluation Parameters
Evaluation Metrics for Validation Data
Skip & Train Model
Run Bayesian Optimizer
Apply & Train Model


Bayesian Optimizer (XGBoost)
Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)
General settings
(Not part of optimization)
Skip & Train Model
Please wait...
Apply & Train Model

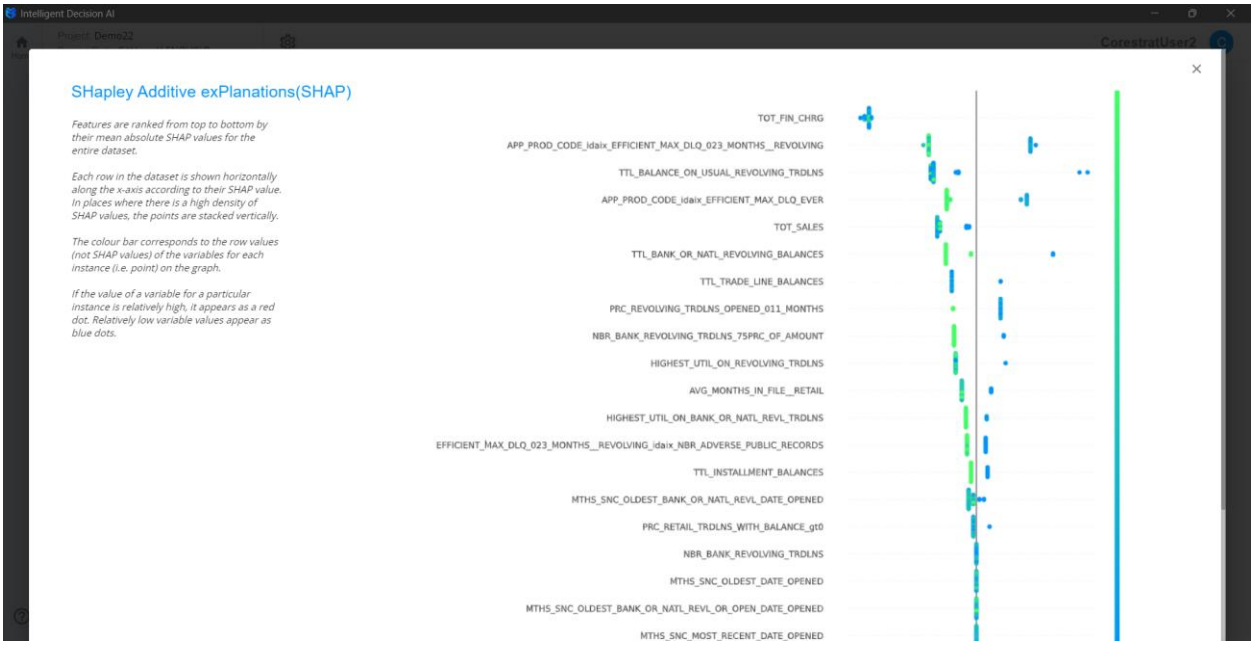


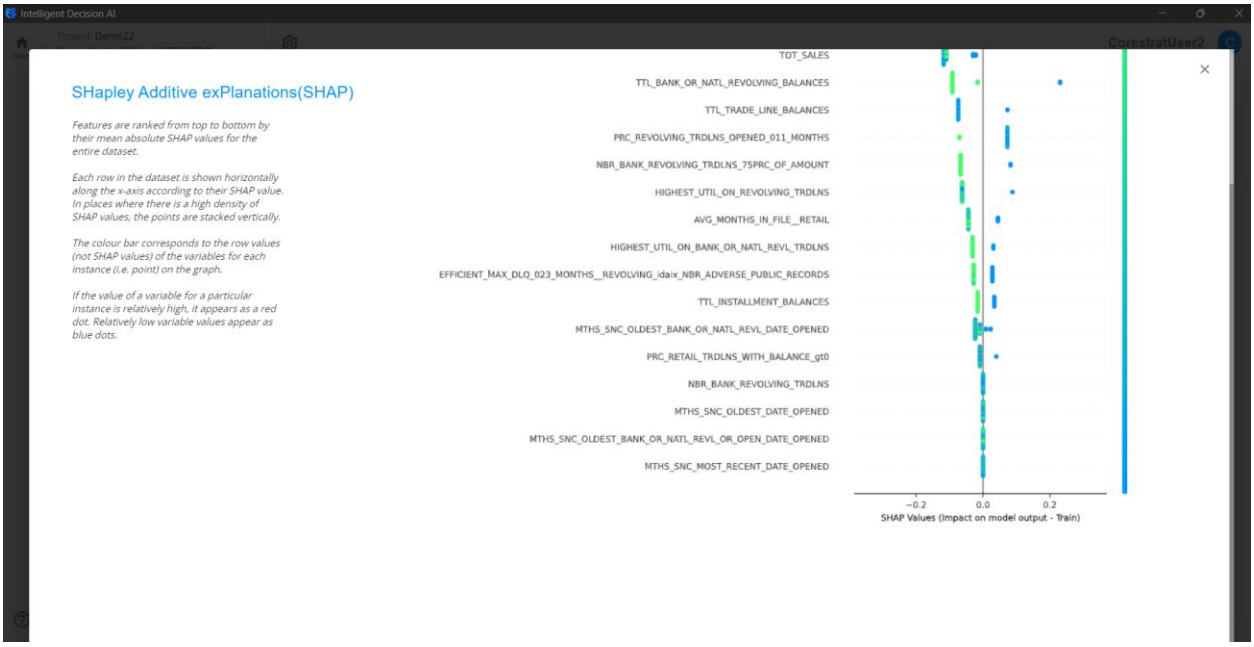
Once hyperparameters are optimized using Bayesian Optimization, the user can click "**Apply & Train Model**" to train an optimized model.



3. XGBoost provides a bar graph of variable importance and a SHAP chart along with KS and GINI charts. which can be viewed by clicking on **'Click to view XGBoost Output'**.

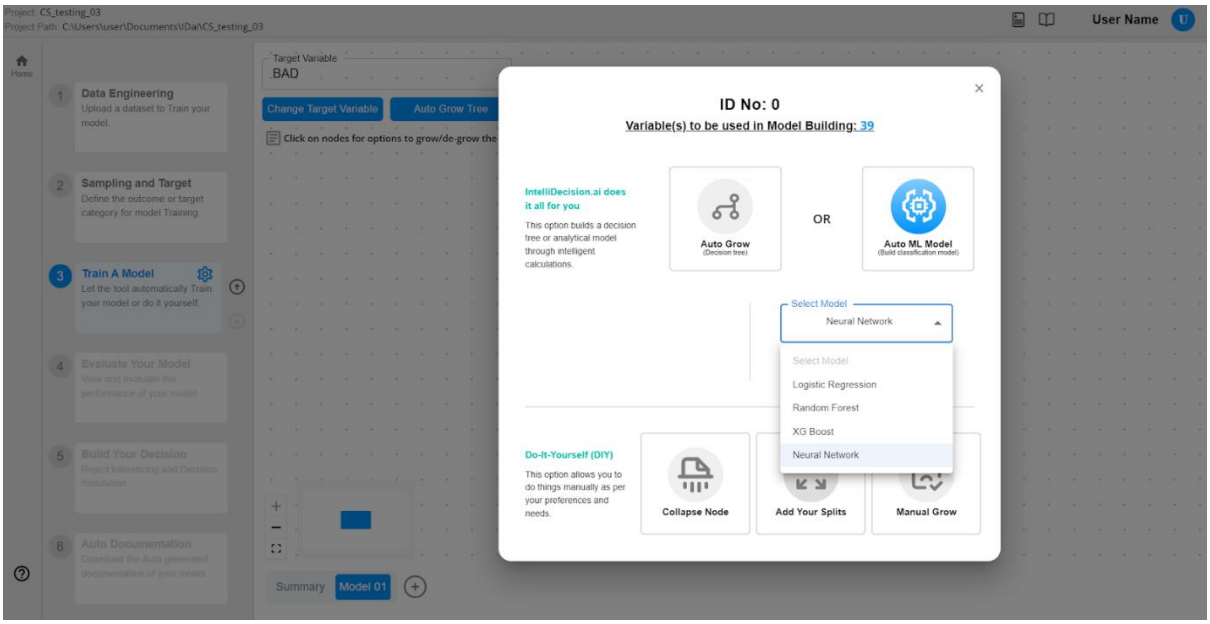

Intelligent Decision AI
Project: Demo22
Project Path: C:\Users\LENOVO\Docum...
CorestratUser2
Target Variable
BAD
ID.ai Node
1 Data
Upload a dataset to Train your model.
2 Prepare your Data
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Node details on which model was trained using Model.ai
ID 4 Train Test
Total 664 292
Target 35 18
Variable Importance
10% 10% 9% 8% 8% 7% 7% 7% 7% 7% 6% 5% 4% 2% 2% 1%
APP_PROD_CO... HIGHEST_UTI... NBR_BANK_RE... APP_PROD_CO... TOT_SALES TTL_TRADE_L... AVG_MONTHS_... TTL_BALANCE... PRC_REVOLVI... TOT_FIN_CHR... TTL_BANK_OR... HIGHEST_UTI... EFFICIENT_M... TTL_INSTALL... PRC_RETAIL_... MTHS_SNC_OL...
Click to view SHAP Chart
Model Performance Metrics
Train (# 10,500) 74 KS 87 GINI
Test (# 4,500) 39 KS 32 GINI
Target % 5.3% 6.2%
Click to see the graph
Target Rate by Model Score
Population %
Target Rate
% Target Train
% Target Test
% Cases Train
% Cases Test
Summary
Model 01
Model 02
Model 03

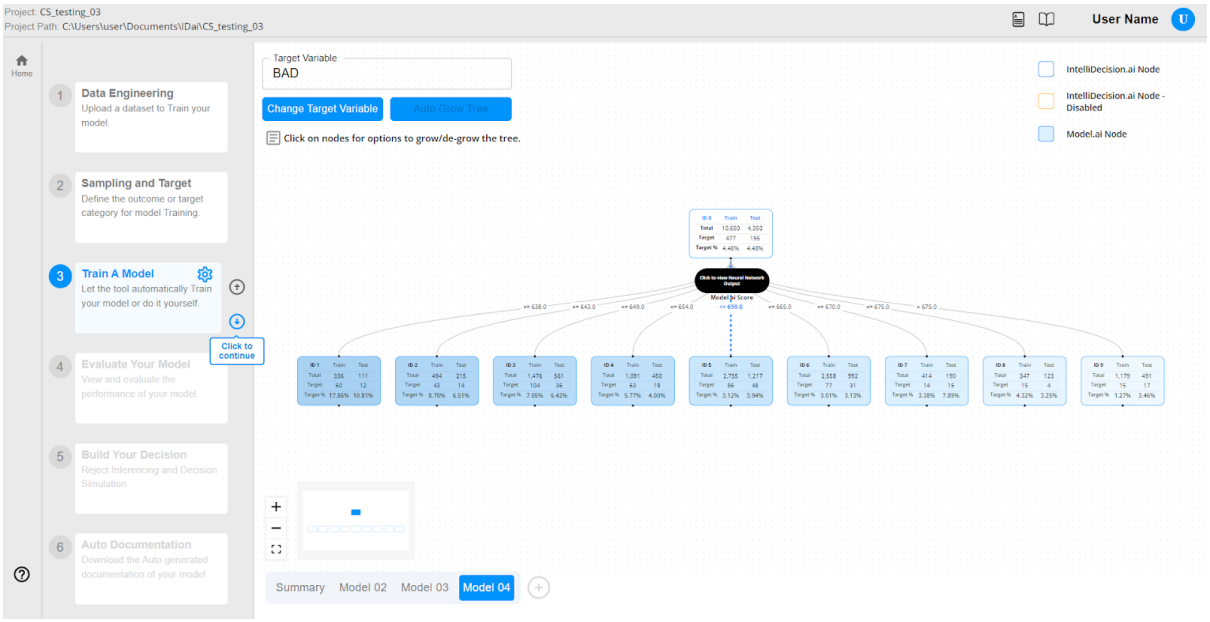

SHapley Additive exPlanations(SHAP)
Features are ranked from top to bottom by their mean absolute SHAP values for the entire dataset.
Each row in the dataset is shown horizontally along the x-axis according to their SHAP value. In places where there is a high density of SHAP values, the points are stacked vertically.
The colour bar corresponds to the row values (not SHAP values) of the variables for each instance (i.e. point) on the graph.
If the value of a variable for a particular instance is relatively high, it appears as a red dot. Relatively low variable values appear as blue dots.
TOT_FIN_CHRG
APP_PROD_CODE_idaix_EFFICIENT_MAX_DLQ_023_MONTHS__REVOLVING
TTL_BALANCE_ON_USUAL_REVOLVING_TRDLNS
APP_PROD_CODE_idaix_EFFICIENT_MAX_DLQ_EVER
TOT_SALES
TTL_BANK_OR_NATL_REVOLVING_BALANCES
TTL_TRADE_LINE_BALANCES
PRC_REVOLVING_TRDLNS_OPENED_011_MONTHS
NBR_BANK_REVOLVING_TRDLNS_75PRC_OF_AMOUNT
HIGHEST_UTIL_ON_REVOLVING_TRDLNS
AVG_MONTHS_IN_FILE__RETAIL
HIGHEST_UTIL_ON_BANK_OR_NATL_REVL_TRDLNS
EFFICIENT_MAX_DLQ_023_MONTHS__REVOLVING_idaix_NBR_ADVERSE_PUBLIC_RECORDS
TTL_INSTALLMENT_BALANCES
MTHS_SNC_OLDEST_BANK_OR_NATL_REVL_DATE_OPENED
PRC_RETAIL_TRDLNS_WITH_BALANCE_gt0
NBR_BANK_REVOLVING_TRDLNS
MTHS_SNC_OLDEST_DATE_OPENED
MTHS_SNC_OLDEST_BANK_OR_NATL_REVL_OR_OPEN_DATE_OPENED
MTHS_SNC_MOST_RECENT_DATE_OPENED



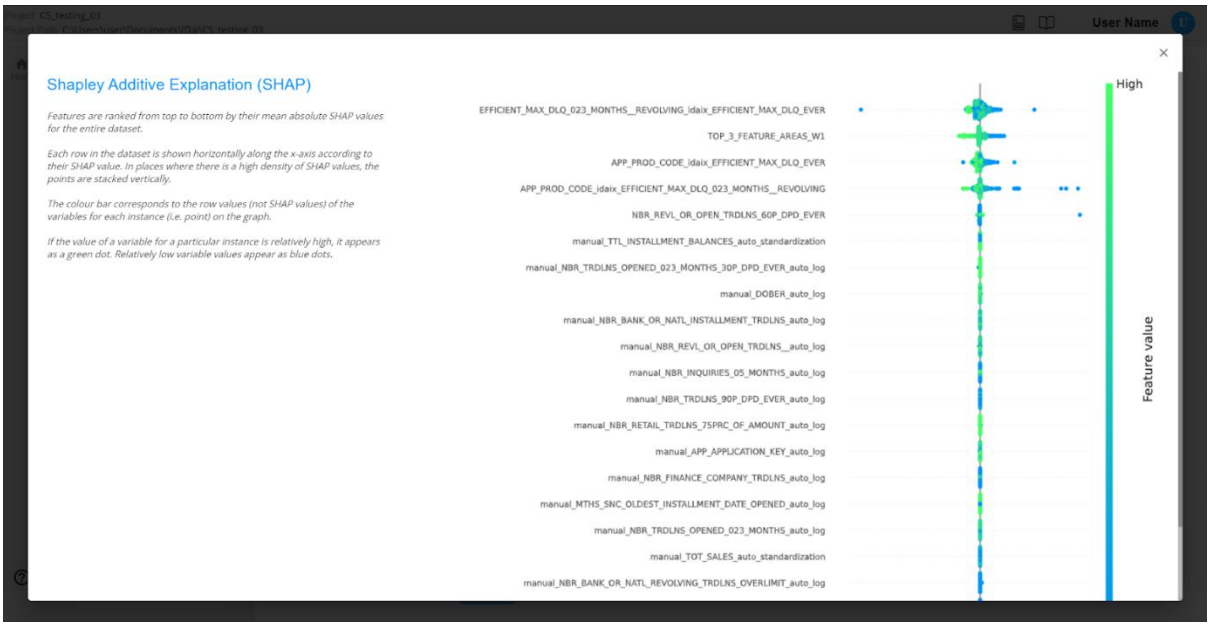
4.User can click on “**Click to view SHAP chart**” to see the SHAP chart, s(he) can also select “**Click to see the graph**” in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.

1. **Neural Networks Model**



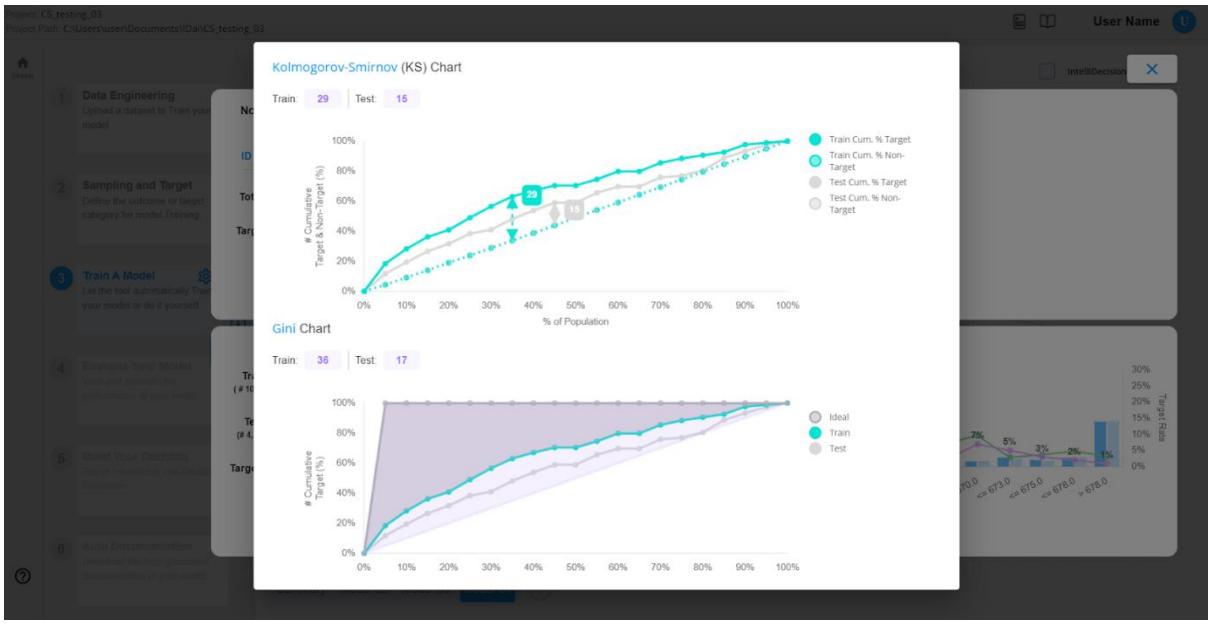


2.Neural Network provides a bar graph of variable importance and a SHAP chart along with KS and GINI charts. which can be viewed by clicking on **'Click to view Neural Network Output'**.

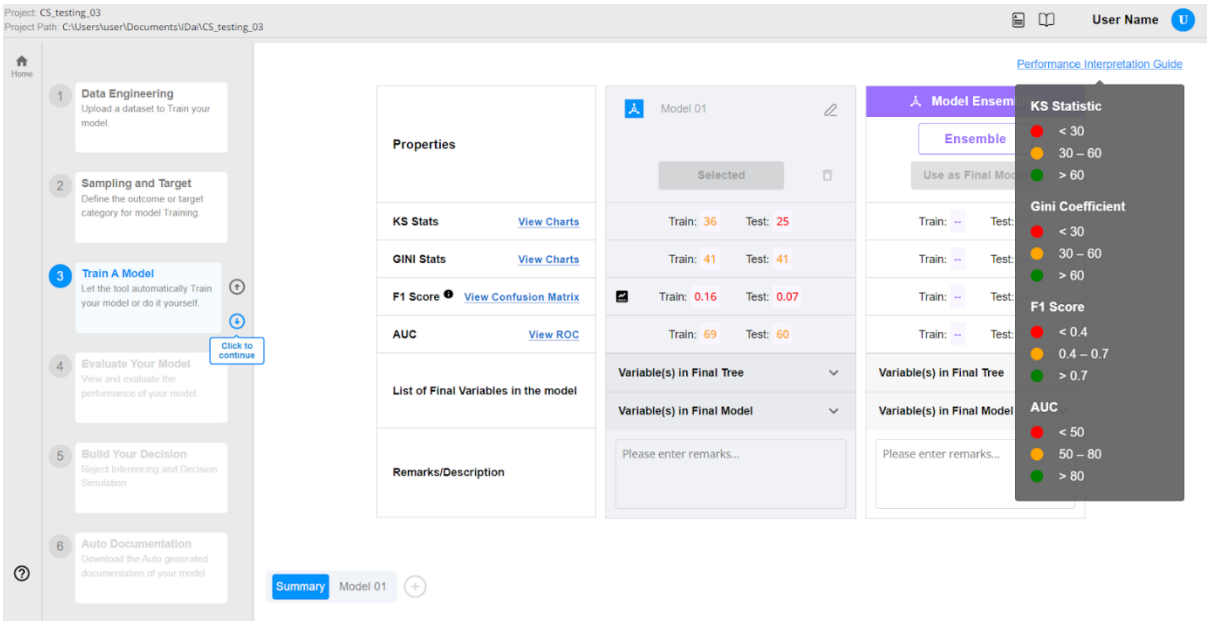


3. User can click on “**Click to view SHAP chart**” to see the SHAP chart, s(he) can also select “**Click to see the graph**” in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test

samples.



On the '**Summary**' screen, there is a '**Performance Interpretation Guide**' provided to enable the users to make sense of the color depictions used on this screen



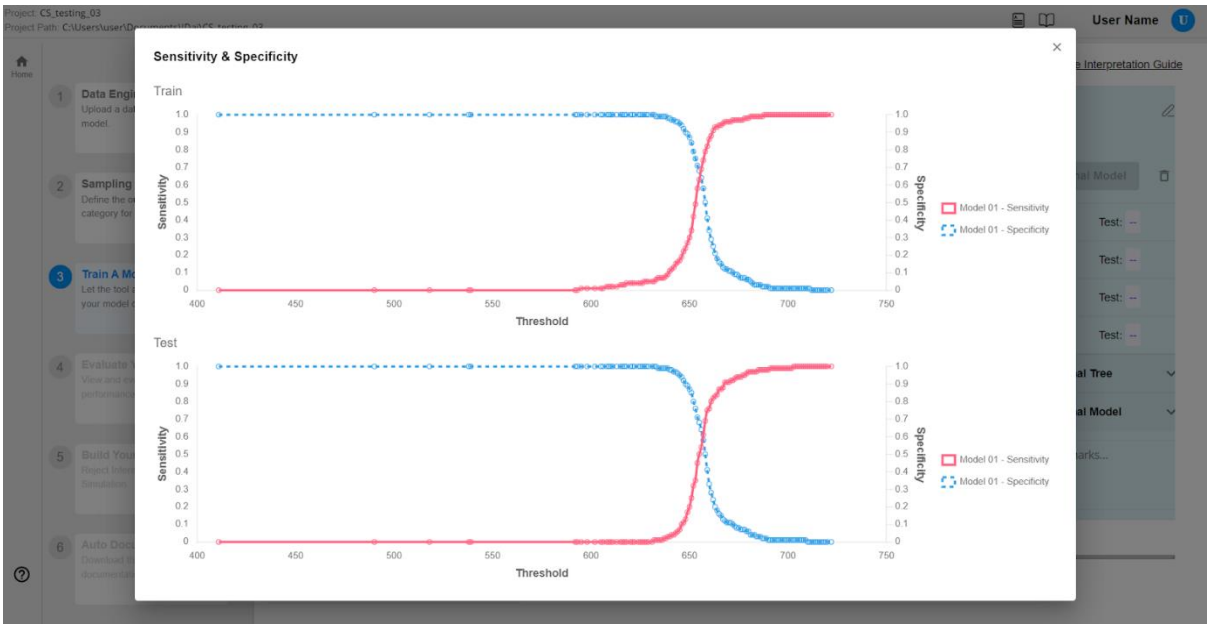
Each metric is rated using a **traffic light system**:

- **Red** indicates poor performance

- **Orange** indicates moderate performance

- **Green** indicates good performance

The thresholds for each color are displayed next to the metric, allowing users to quickly interpret and compare model quality based on standard benchmark ranges.



This page helps users visualize how the model's **sensitivity** (true positive rate) and **specificity** (true negative rate) change with different threshold values.

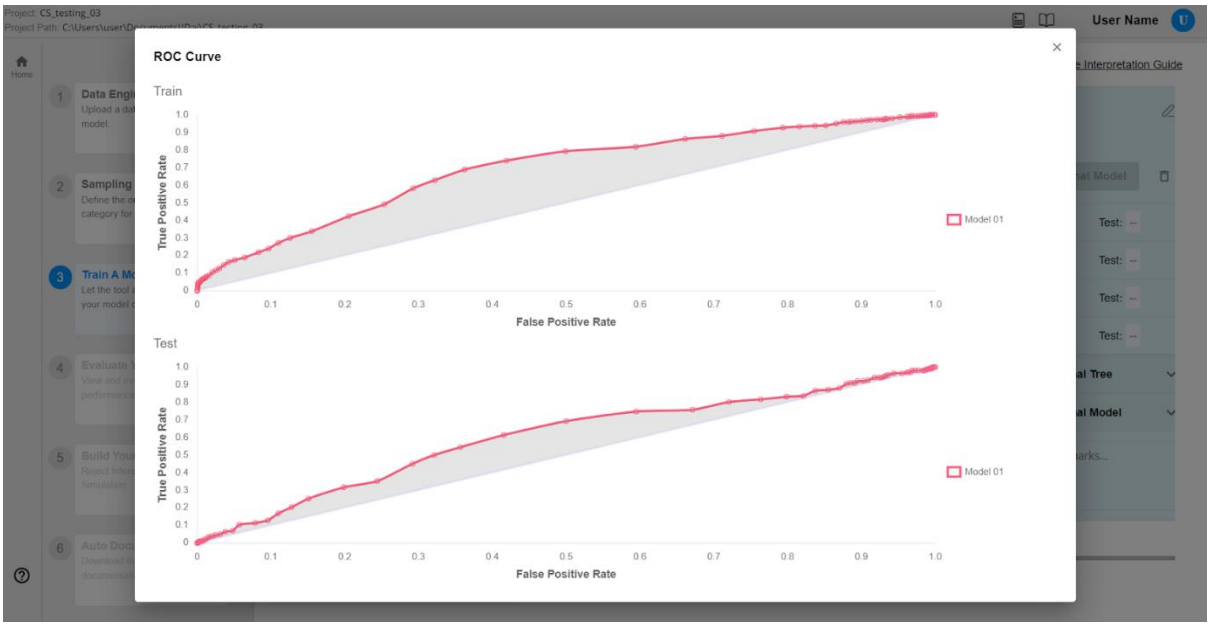
- The **top graph** shows results from the **training data**.

- The **bottom graph** shows results from the **test data**.

- The **pink line** shows how well the model is catching positive cases (called **sensitivity**).

- The **blue dotted line** shows how well it's avoiding false positives (called **specificity**).

By looking at where the two lines are balanced, you can choose a threshold that works best for your needs



This screen shows how well your model is able to distinguish between positive and negative outcomes.

- The **top graph** shows the performance on the **training data**.

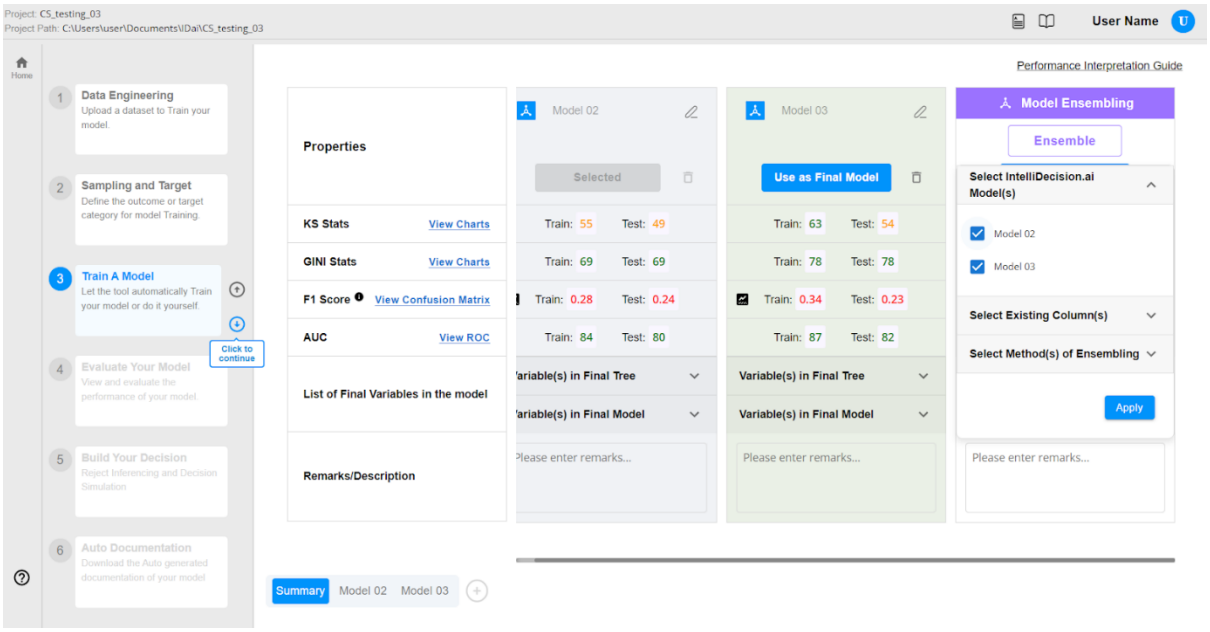
- The **bottom graph** shows the performance on the **test data**.

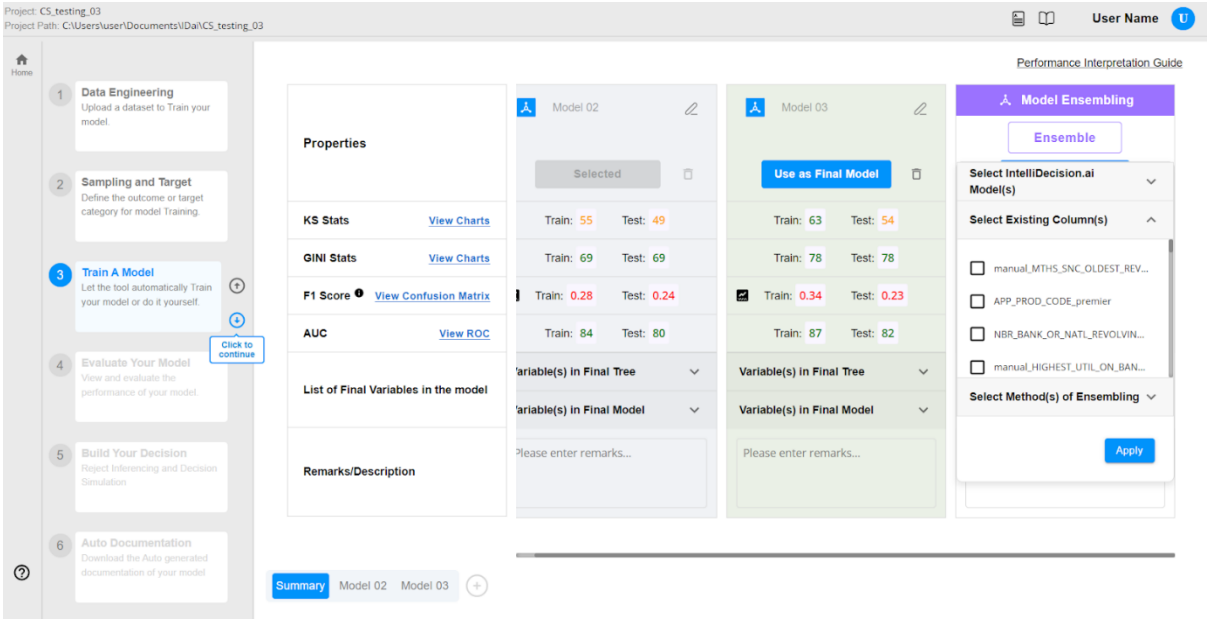
- The **ROC (Receiver Operating Characteristic) curve** plots:

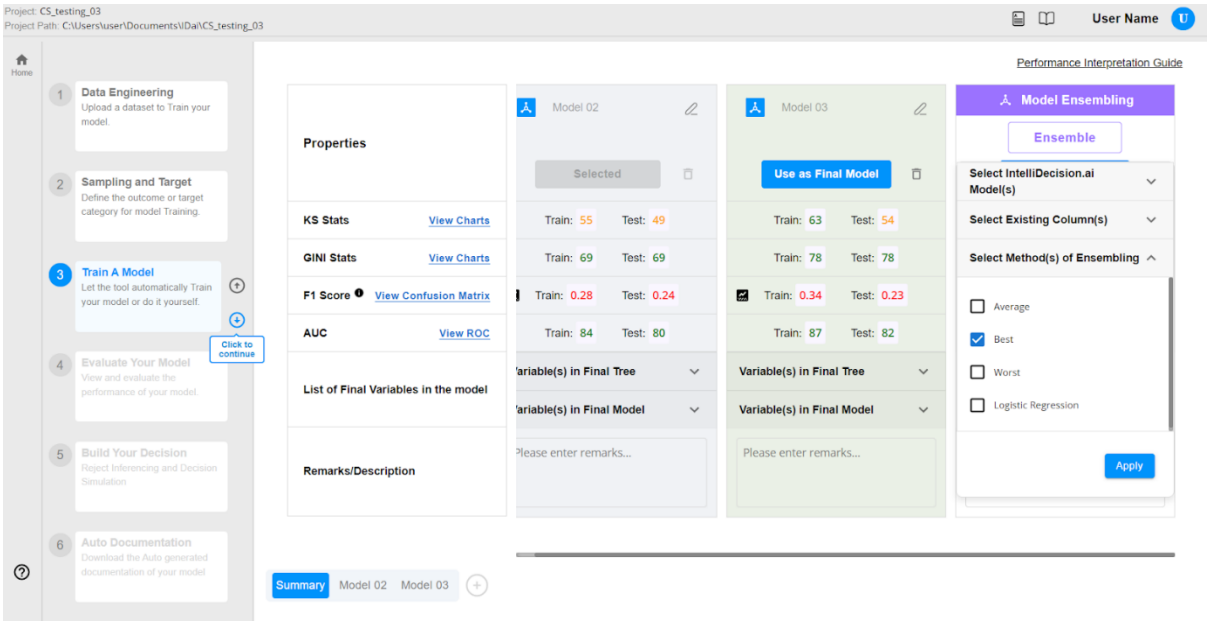
    - **True Positive Rate** (how many actual positives your model correctly identifies) on the Y-axis

    - **False Positive Rate** (how many actual negatives are wrongly flagged as positives) on the X-axis

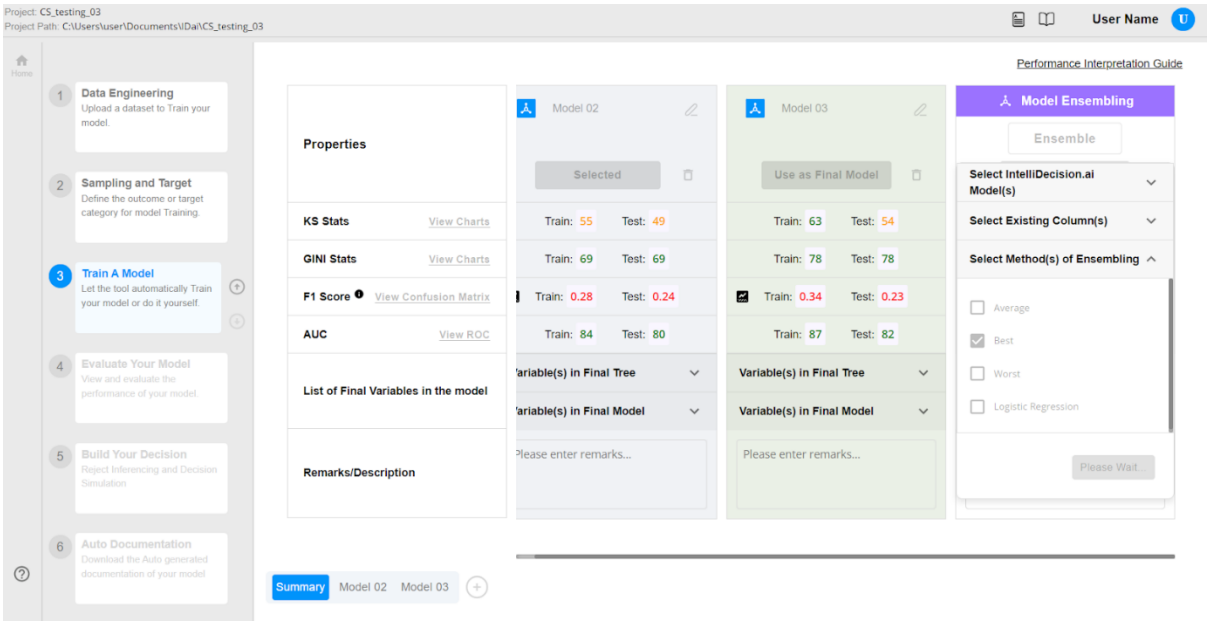
When performing Model Ensemble, first select the models to include, optionally choose the columns to be used, and specify the ensembling method.

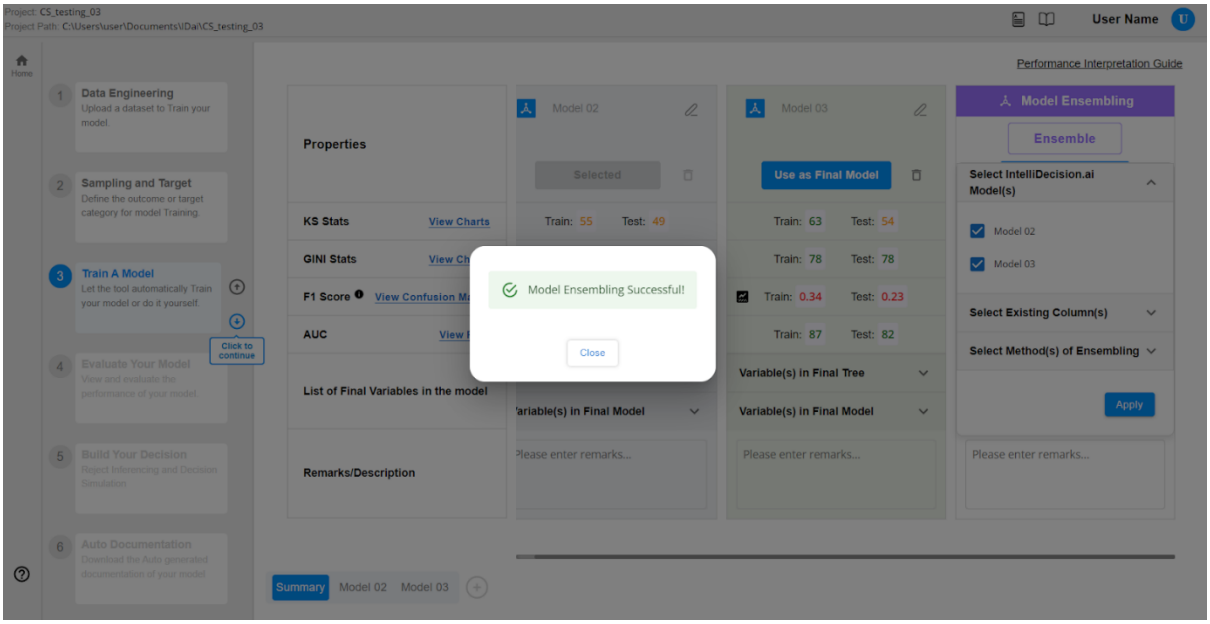

Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Performance Interpretation Guide
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
Click to continue
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Properties
KS Stats View Charts
GINI Stats View Charts
F1 Score View Confusion Matrix
AUC View ROC
List of Final Variables in the model
Remarks/Description
Model 02
Selected
Train: 55 Test: 49
Train: 69 Test: 69
Train: 0.28 Test: 0.24
Train: 84 Test: 80
Variable(s) in Final Tree
Variable(s) in Final Model
Please enter remarks...
Model 03
Use as Final Model
Train: 63 Test: 54
Train: 78 Test: 78
Train: 0.34 Test: 0.23
Train: 87 Test: 82
Variable(s) in Final Tree
Variable(s) in Final Model
Please enter remarks...
Model Ensembling
Ensemble
Select IntelliDecision.ai Model(s)
Model 02
Model 03
Select Existing Column(s)
Select Method(s) of Ensembling
Apply
Please enter remarks...
Summary
Model 02
Model 03


Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDai\CS_testing_03
User Name
Performance Interpretation Guide
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
Click to continue
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Properties
KS Stats View Charts
GINI Stats View Charts
F1 Score View Confusion Matrix
AUC View ROC
List of Final Variables in the model
Remarks/Description
Model 02
Selected
Train: 55 Test: 49
Train: 69 Test: 69
Train: 0.28 Test: 0.24
Train: 84 Test: 80
Variable(s) in Final Tree
Variable(s) in Final Model
Please enter remarks...
Model 03
Use as Final Model
Train: 63 Test: 54
Train: 78 Test: 78
Train: 0.34 Test: 0.23
Train: 87 Test: 82
Variable(s) in Final Tree
Variable(s) in Final Model
Please enter remarks...
Model Ensembling
Ensemble
Select IntelliDecision.ai Model(s)
Select Existing Column(s)
manual_MTHS_SNC_OLDEST_REV...
APP_PROD_CODE_premier
NBR_BANK_OR_NATL_REVOLVIN...
manual_HIGHEST_UTIL_ON_BAN...
Select Method(s) of Ensembling
Apply
Summary
Model 02
Model 03



Clicking ‘Apply’ will initiate the ensembling process.





Variables in Final Tree lists all variables used in the decision tree (if decision trees were part of the ensemble).

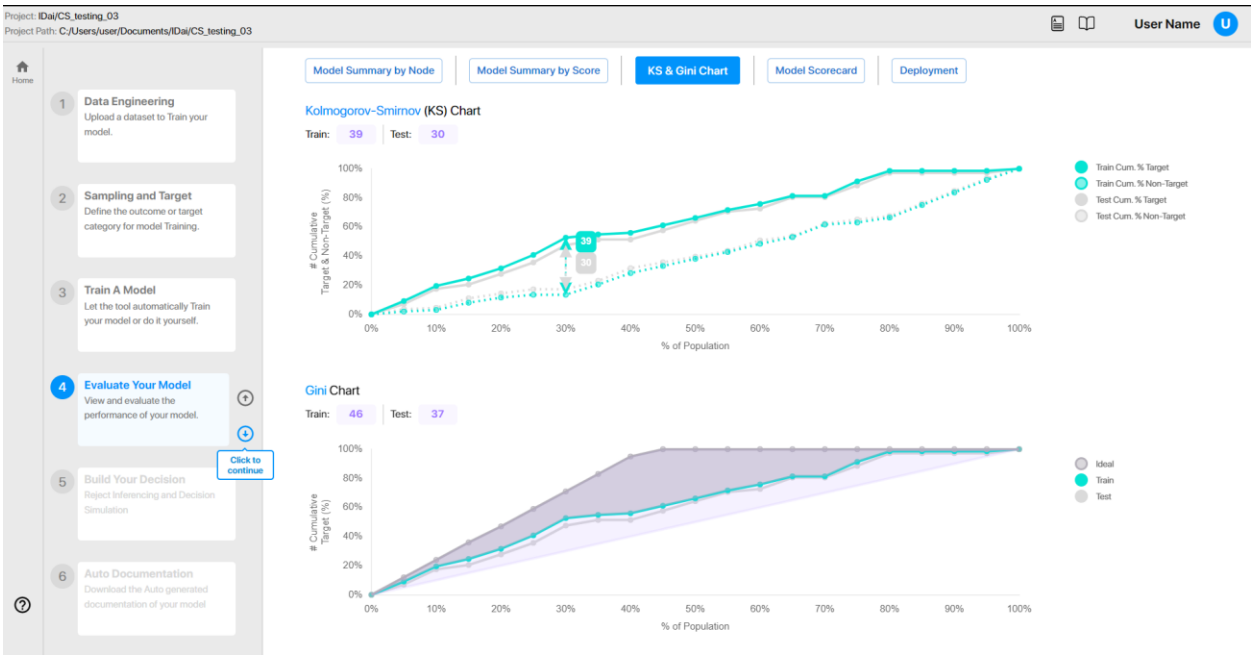
Variables in Final Model lists all variables used to build the models (if AutoModels were included in the ensemble).

Evaluate Your Model

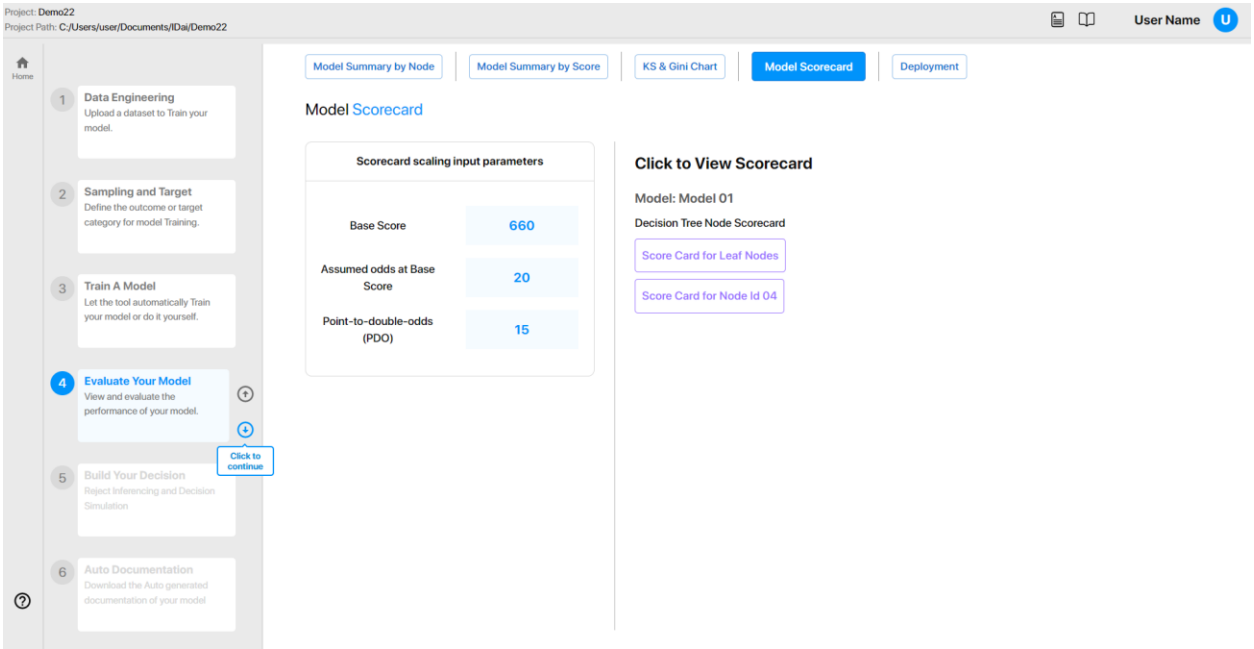
Once the model has been built, the next stage involves analysing the performance of the model. This section provides a description of the various performance metrics available in ID.ai to evaluate the model. The headings in the bullets refer to the buttons within the tool.

- Model summary by Node – Information on all nodes used in the model, the counts and bad rates in both Train and Test samples.

- Model summary by score – counts and bad rates for each score bin and raw score for both Train and Test samples.
- KS & Gini chart – visual representation of KS and Gini values for both Train and Test samples



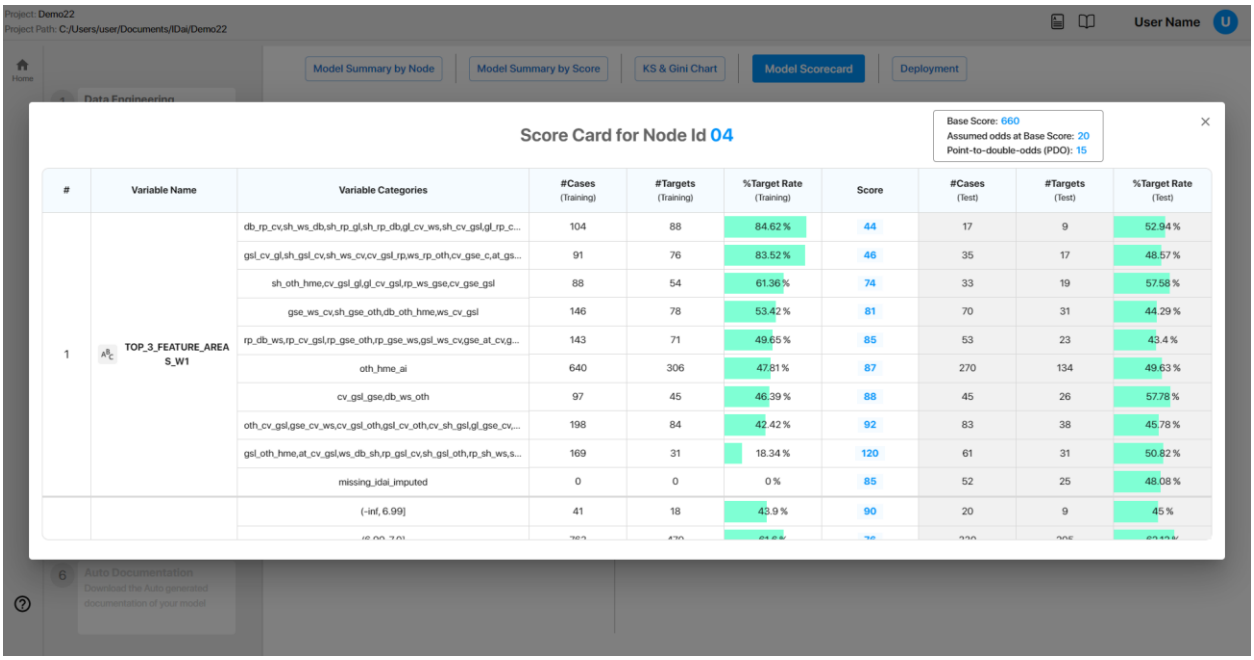
- Model scorecard – this is used to scale the scores for the various nodes based on predefined base score values, odds and base score and probability of double odds.



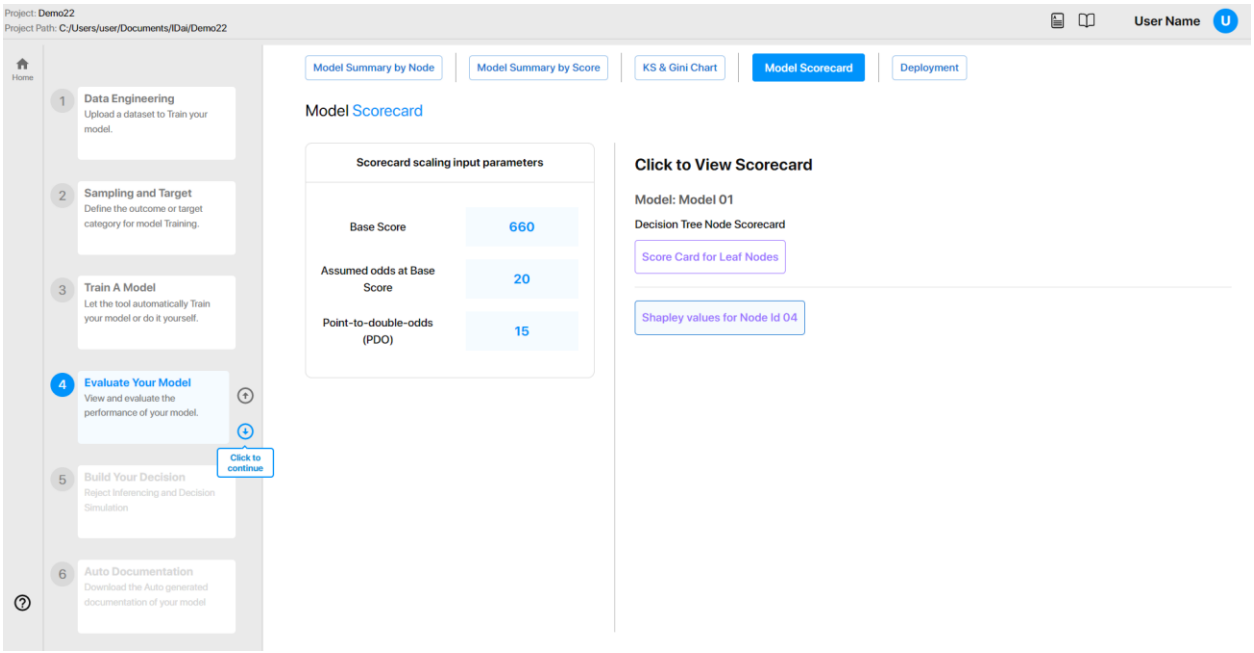
- User can either see the scores generated for the leaf nodes by clicking on the “**Score card for Leaf Nodes**” button

OR

- Scores for the AI Logistic Regression Model node by clicking on the “**Score card for Node ID 04**” button

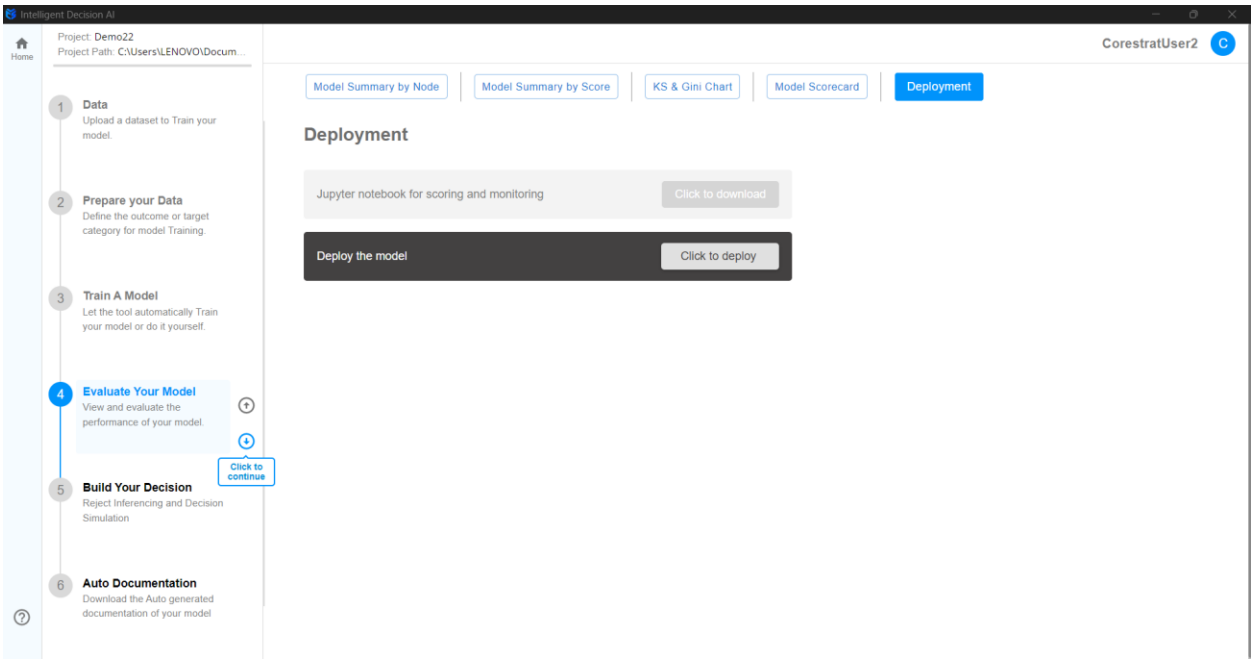


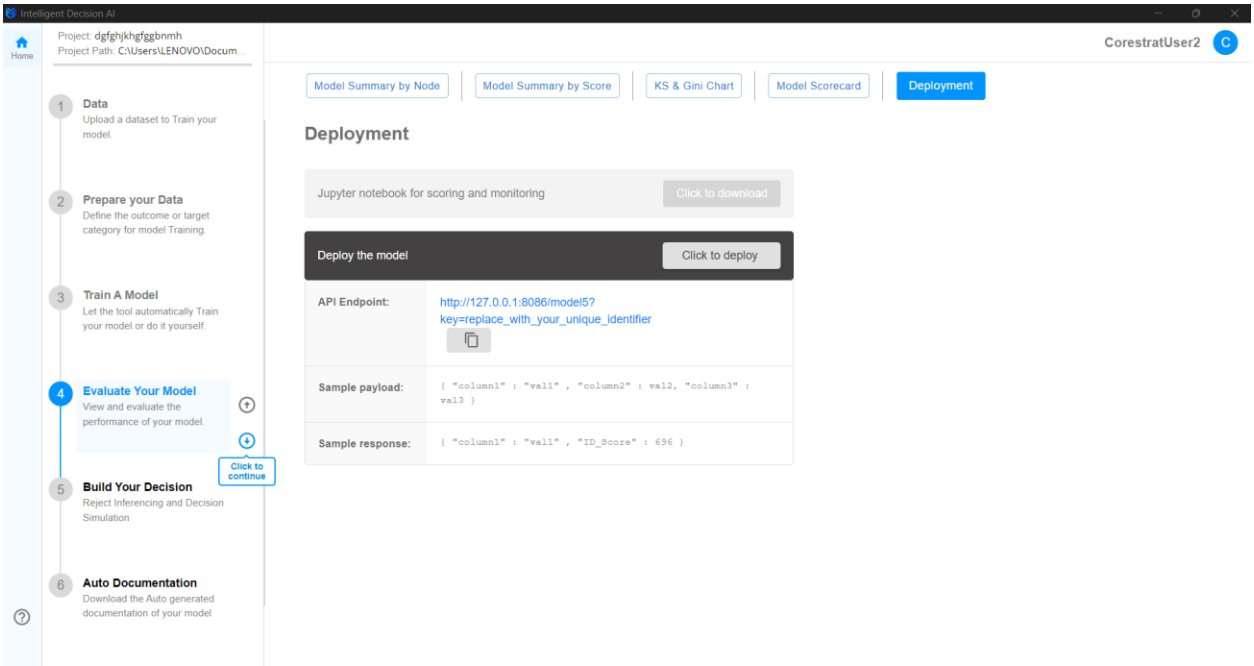
- To view shapely values for the AI Random Forest or XGBoost model, click on the “**Shapely values for Node ID 04**” button


Project: Demo22
Project Path: C:/Users/user/Documents/IDai/Demo22
User Name
Home
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
Click to continue
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
Model Summary by Node
Model Summary by Score
KS & Gini Chart
Model Scorecard
Deployment
Model Scorecard
Scorecard scaling input parameters
Base Score 660
Assumed odds at Base Score 20
Point-to-double-odds (PDO) 15
Click to View Scorecard
Model: Model 01
Decision Tree Node Scorecard
Score Card for Leaf Nodes
Shapley values for Node Id 04


Intelligent Decision AI
Project: Demo22
Project Path: C:\Users\LENOVO\Docum...
CorestratUser2
Model Summary by Node
Model Summary by Score
KS & Gini Chart
Model Scorecard
Deployment
Variable importance by shapely values
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
HIGHEST_UTI...
TTL_TRADE_L...
NBR_BANK_OR...
TTL_BANK_OR...
EFFICIENT_M...
PRC_TRDLNS_...
MTHS_SNC_OL...
MTHS_SNC_OL...
EFFICIENT_M...
TTL_BALANCE...
EFFICIENT_M...
NBR_TRDLNS_...
APP_APPLICA...
NBR_BANK_OR...
APP_PROD_CO...
NBR_BANK_OR...
NBR_DEROG_T...
NBR_BANK_OR...
MTHS_SNC_MO...

- **Deployment:** When user clicks on “Click to deploy”, they will get “API Endpoint” along with sample payload and sample response. They can run “API Endpoint” to get Json response.

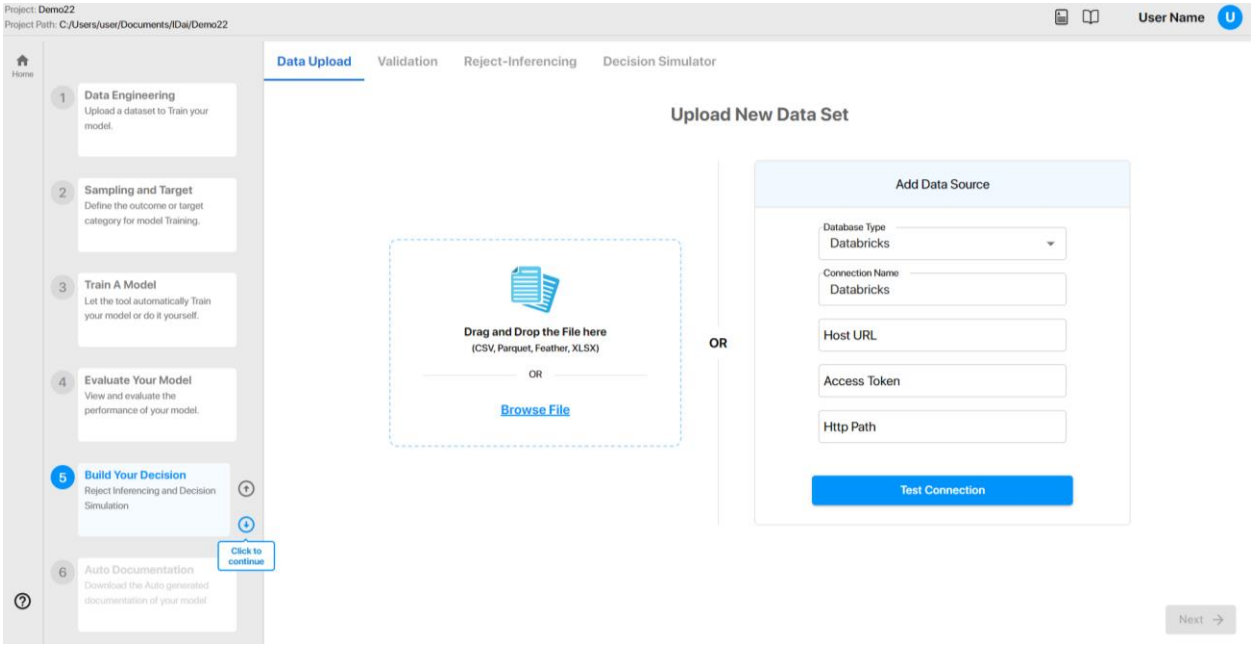




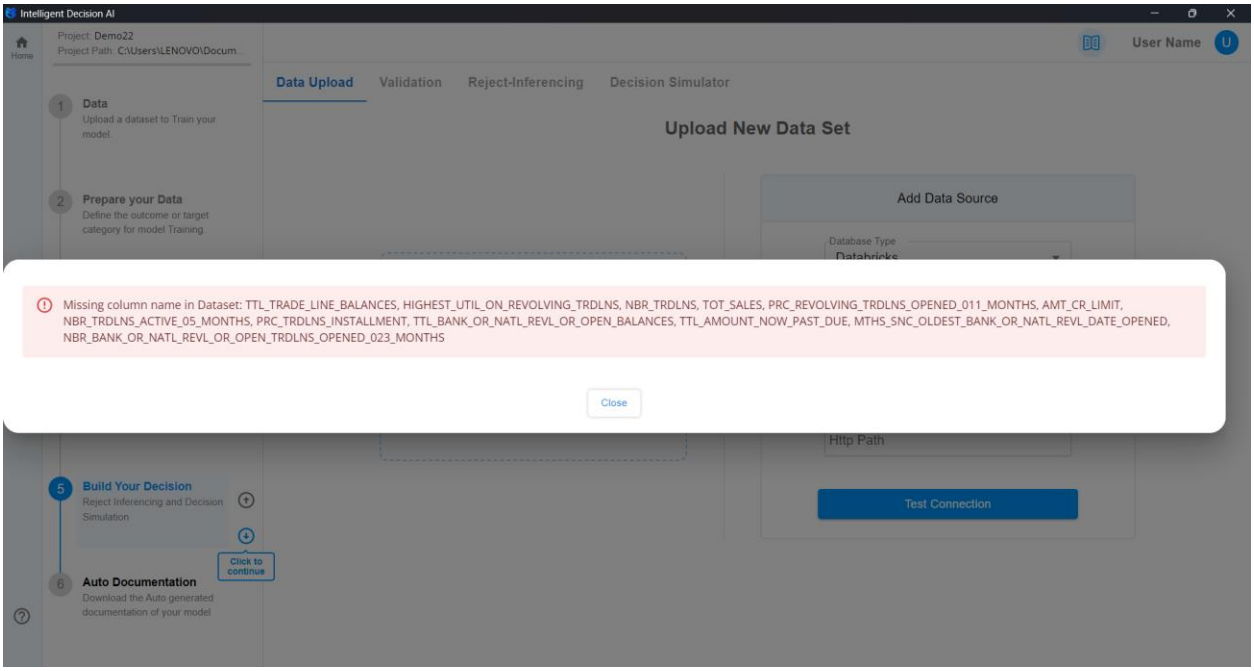
Build Your Decision

Once the scorecard has been developed, the user can upload an OOT (out-of-time) dataset to simulate the decision based on the previously developed model. Simulation(s) can be done on the overall score or a particular sub-segment.

The first step is to upload a fresh OOT/unseen dataset which should contain ALL the variables in the model build stage. User needs to upload this file from the location to the tool as shown in the screenshot below. (The file specifications are the same as described earlier in Section 4.1.1)

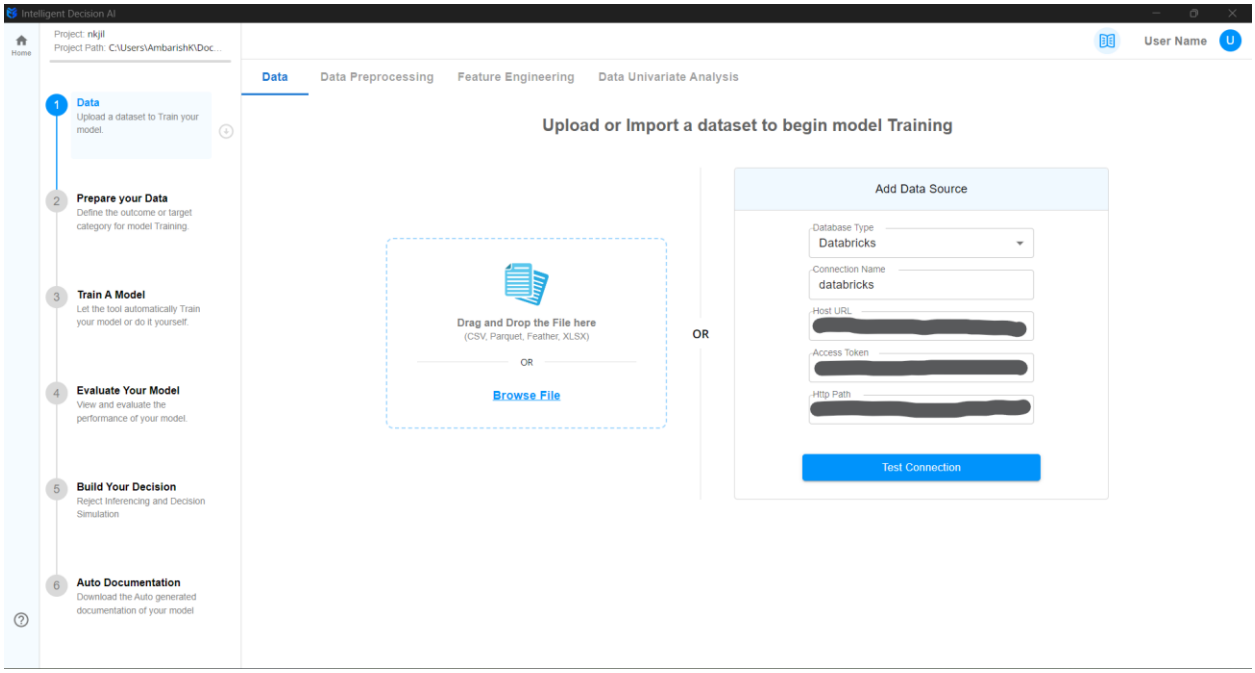


If the OOT/unseen dataset uploaded does not contain ALL the columns from the model that was built and selected, it would display an error message as shown below. Kindly upload an appropriate dataset on which to build a decision using the model built in Step 3.

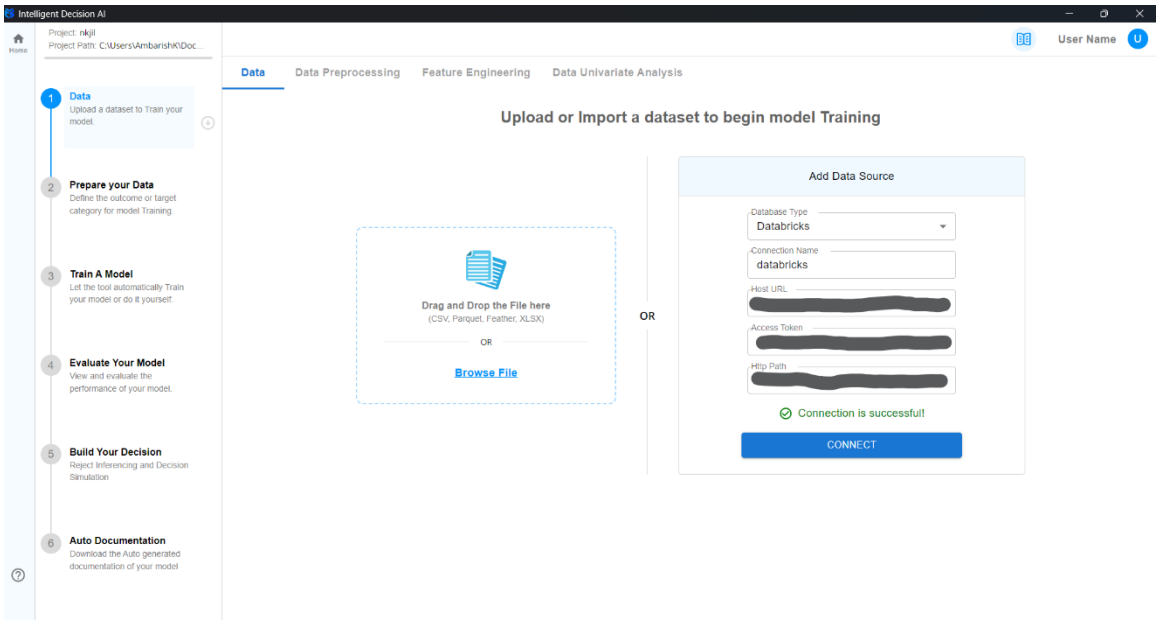


Data can also be added using Databricks. User needs to follow the below steps:

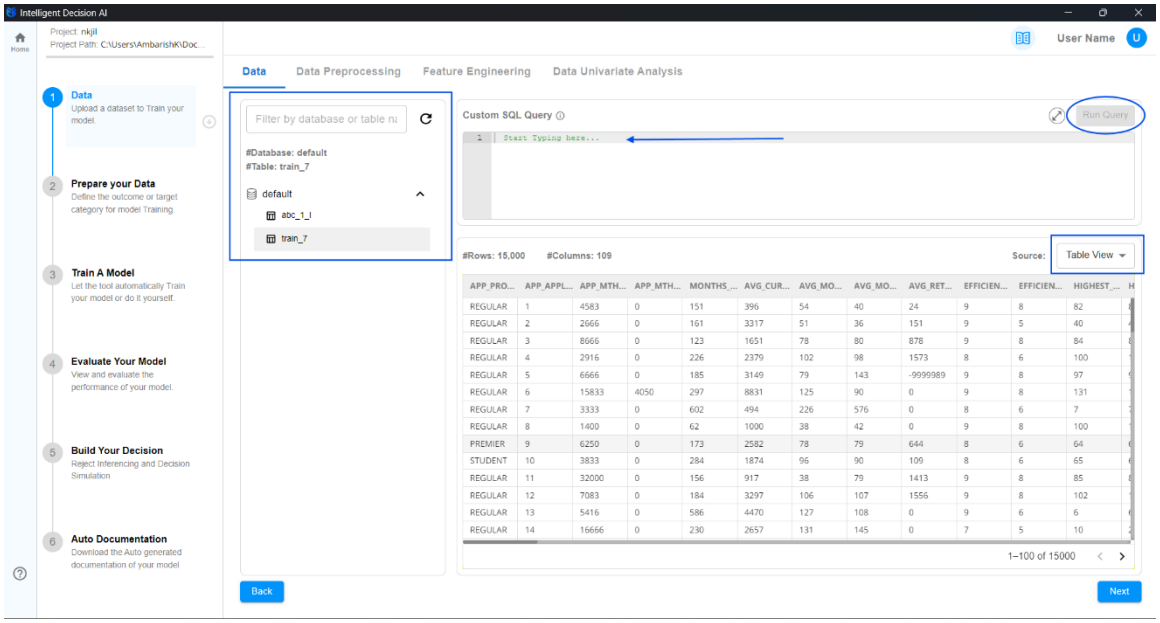
1. Enter the Host URL, Access Token and the Http Path in the allocated fields.
2. Click on Test Connection



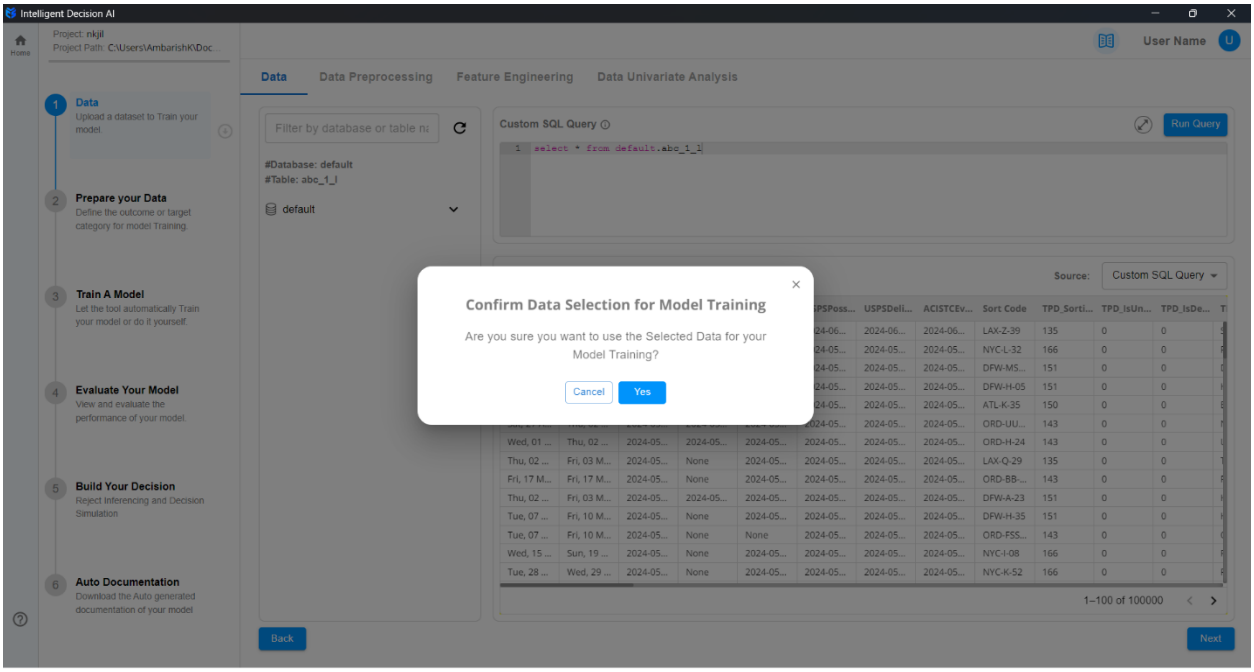
3. Once it shows “**Connection successful**”, click on “**Connect**”.



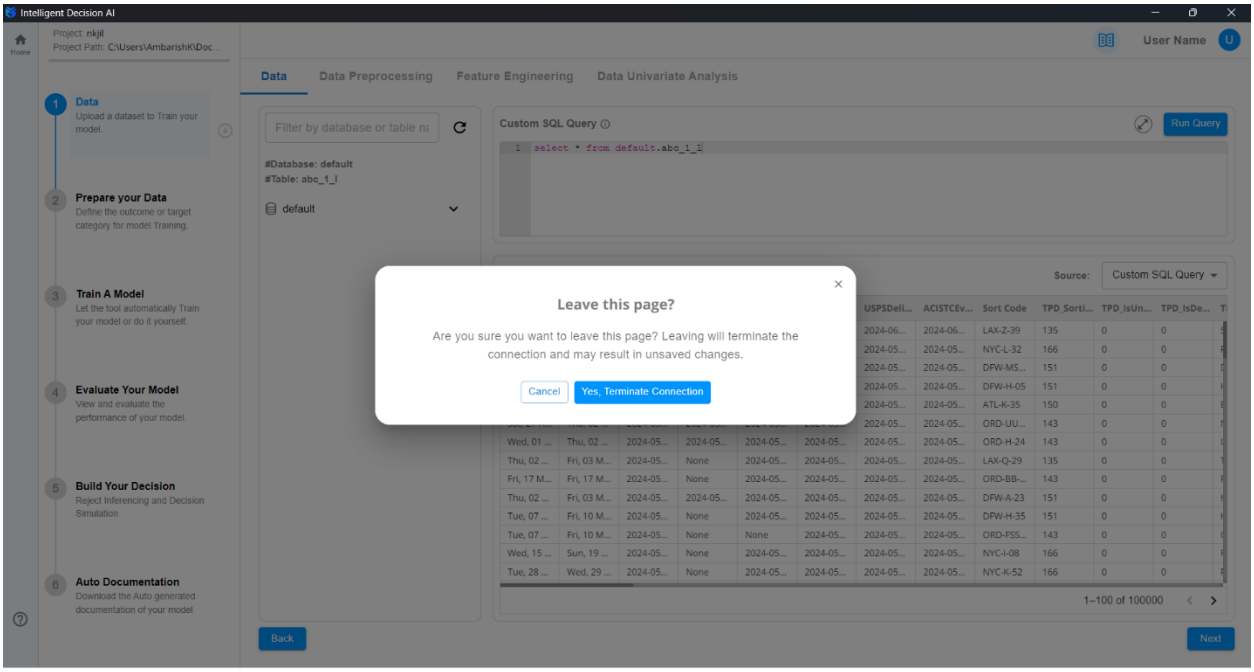
4. User can choose the table from the left side from the database of need (refer image).
5. Write a custom SQL Query in the editor provided to get data from the table based on user criteria. Once done, click on “**Run Query**” (refer image). This is an optional step.
6. The source of the data shown can be changed through the source dropdown.
7. Once satisfied with the data extracted, click on “**Next**” button.



8. The below shown dialog box shall appear. If want to proceed further with this selection, click “**Yes**”, else click on “**Cancel**”.



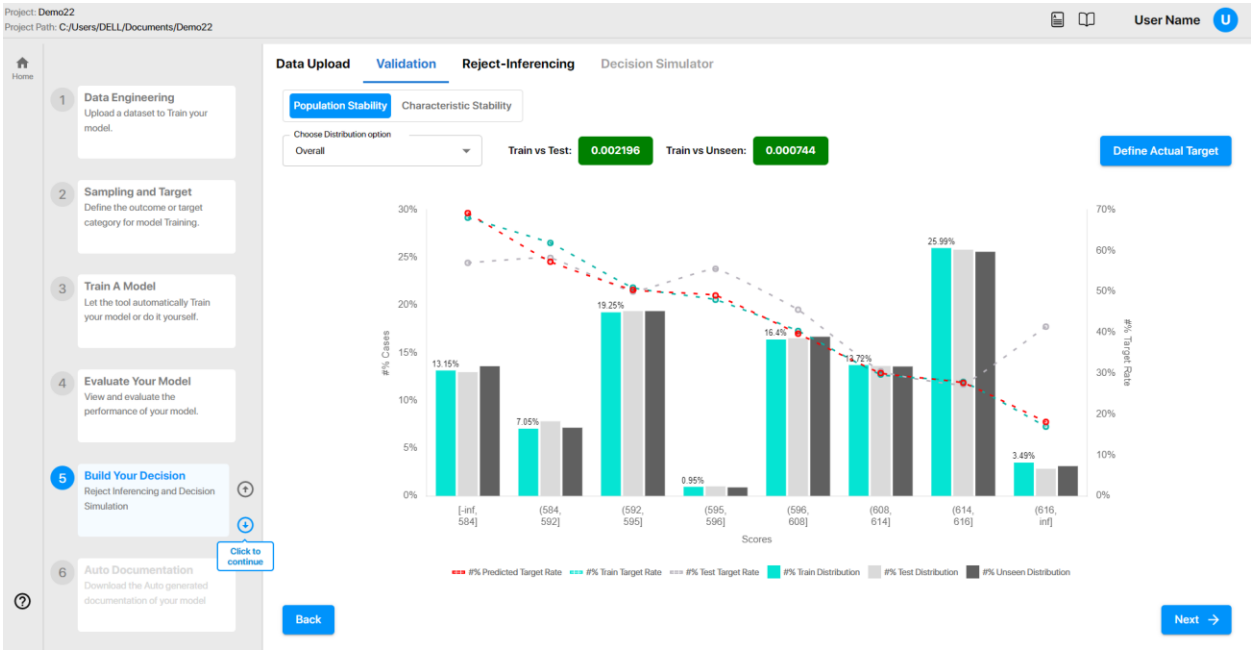
If you want to terminate the connection, click on the "**Back**" button as shown in the image and click on "**Yes Terminate Connection**" when prompted.



The next screen provides the population stability and characteristic stability.

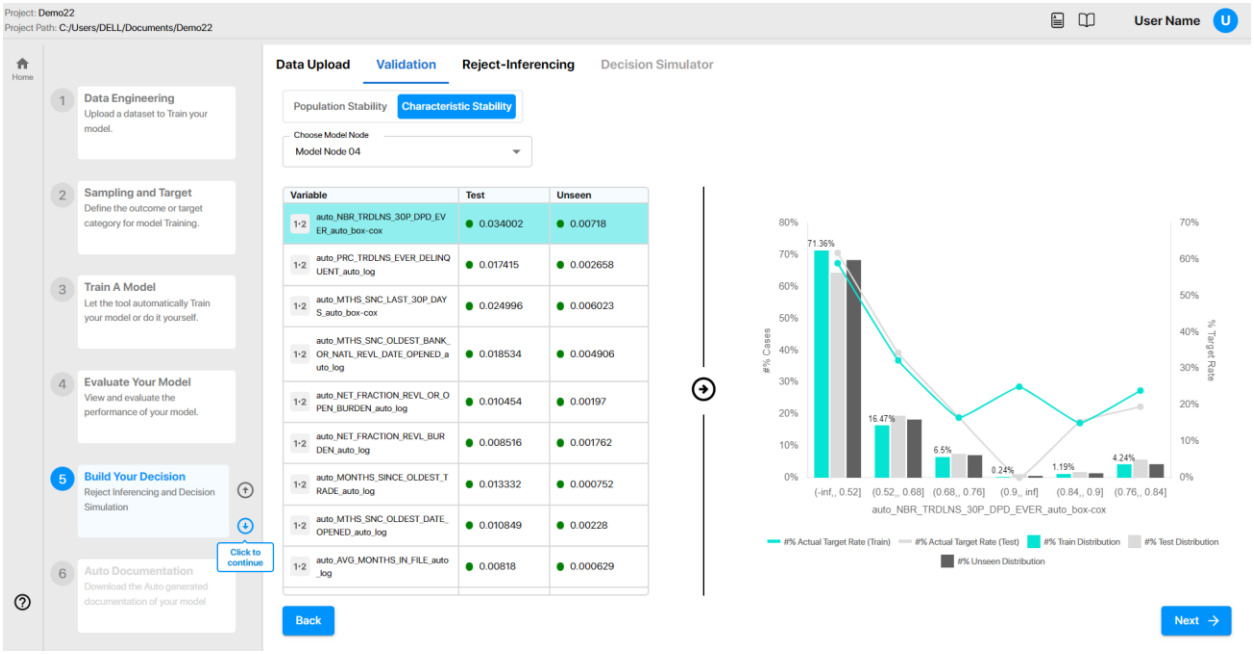
The Population Stability Index (PSI) measures how much the data used to train the model differs from new, unseen data. It helps check if the model's predictions are still reliable over time. A high PSI means the data has changed a lot, which could affect the model's performance.

User has the option to choose a distribution option between Overall, Decision Tree Leaf Nodes and the Specific Node using the drop-down menu in "**Choose Distribution Option**".



The Characteristic Stability Index (CSI) measures how much a variable's behavior/distribution has changed between the data used to train the model and new, unseen data. It helps check if the patterns the model learned are still consistent or if they've shifted over time. A high CSI means there's a big difference, which could affect the model's performance.

The equivalent drop-down menu option for characteristic stability is available only when Model.ai is run.

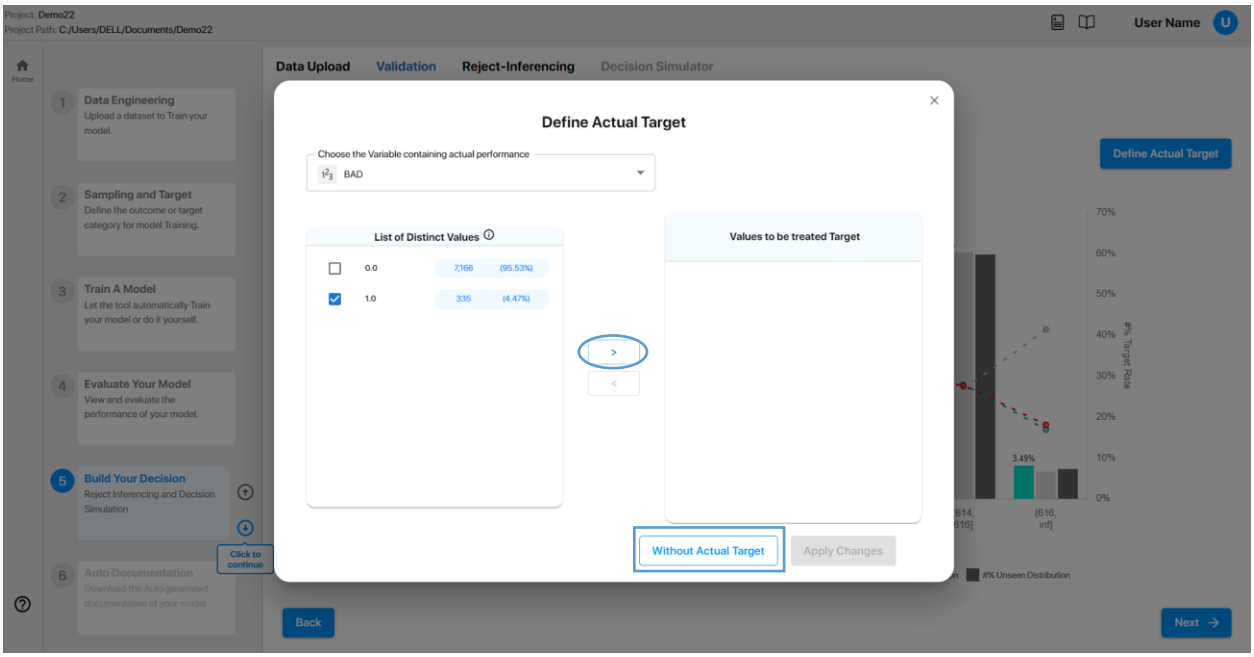


<u>"**Define Actual Target**" button in Population stability section</u>

This approach enables the original model build dataset's target variable to be used as the target variable in the newly uploaded dataset also.

1. Click on "**Define Actual Target**" button located at top right of the screen
2. Click on "**Choose the variable containing actual performance**" to select the target variable **OR Goto Step 5**
3. Then select the checkbox in "**List of Distinct Values**" to specify which values within the target variables should be considered as the target value and shift to "**Values to be treated Target**" section using "**>**" button
4. Then click on "**Apply Changes**" for validation
5. Click on "**Without Actual Target**" (refer image, blue rectangle)

   In this approach, the OOT sample dataset does not have the target variable. The target is predicted by applying the model on the OOT dataset and predicting the target variable for each record based on the values of the independent variables.



Reject Inferencing

The next screen provides Reject Inferencing. It is a method for improving the quality of a model by incorporating data from previously rejected/unavailable records. Bias can result if a credit scorecard model is built only on accepts and does not account for applications rejected that were rejected and hence have an unknown target status.

- If you want to proceed to next screen, click "**Without Reject Inferencing**" button (refer Image, blue rectangle).

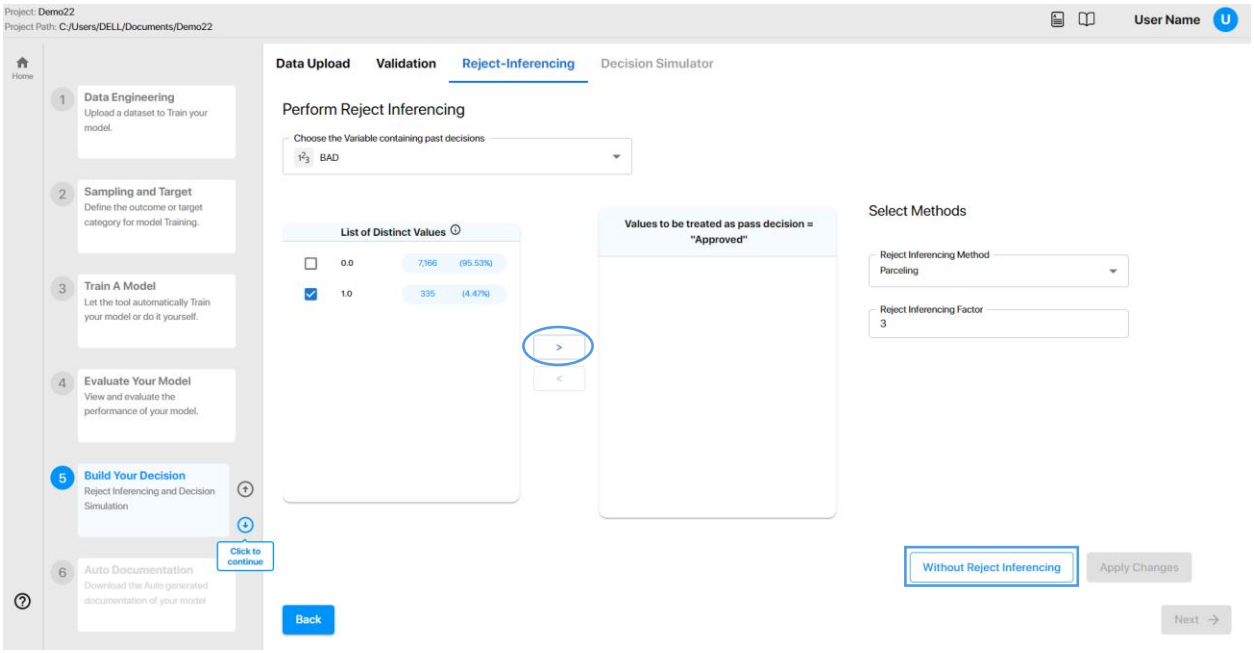
  OR

- Steps to perform:
  1. Click on "**Choose the variable containing past decisions**" and select the appropriate variable

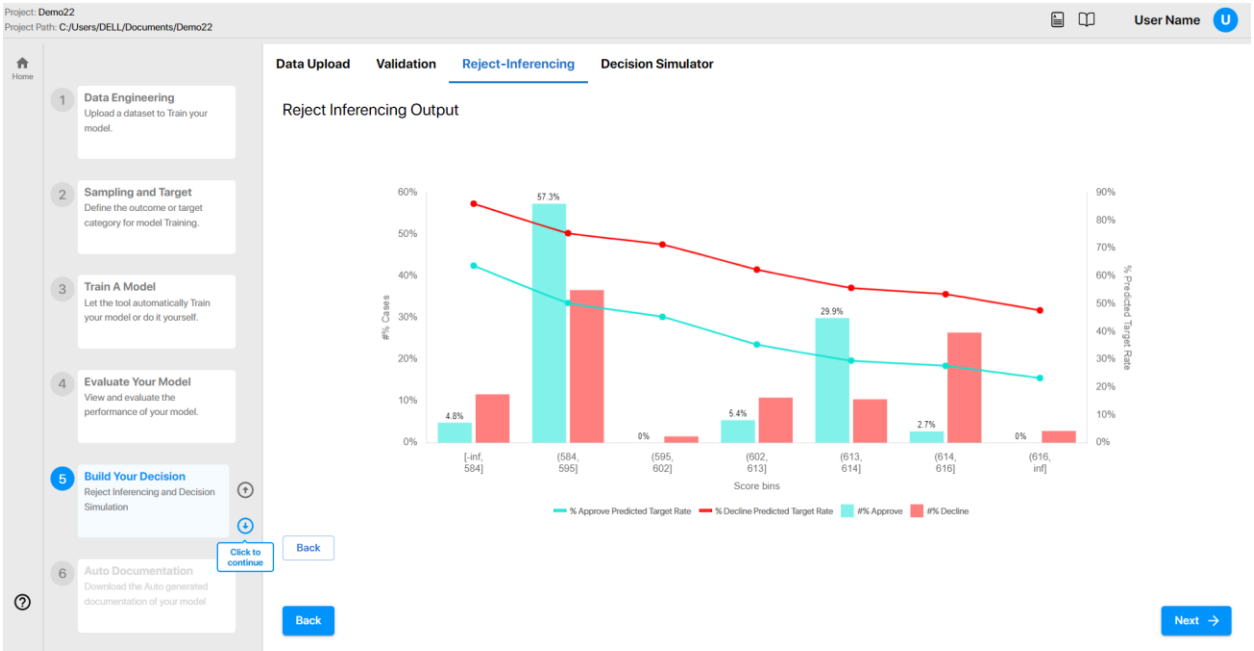
  2. After that, select the value from "**List of Distinct Values**" which should be considered as approved and move to the right side using the '>' button to 'Values to be treated as pass decision = "Approved"' section (refer image, blue oval)

3. You can modify the value of '**Reject Inferencing Factor**' to an integer between 1 and 10 based on your specific needs. Default is kept at 3.

4. Click on "**Apply Changes**" button to perform Reject Inferencing.



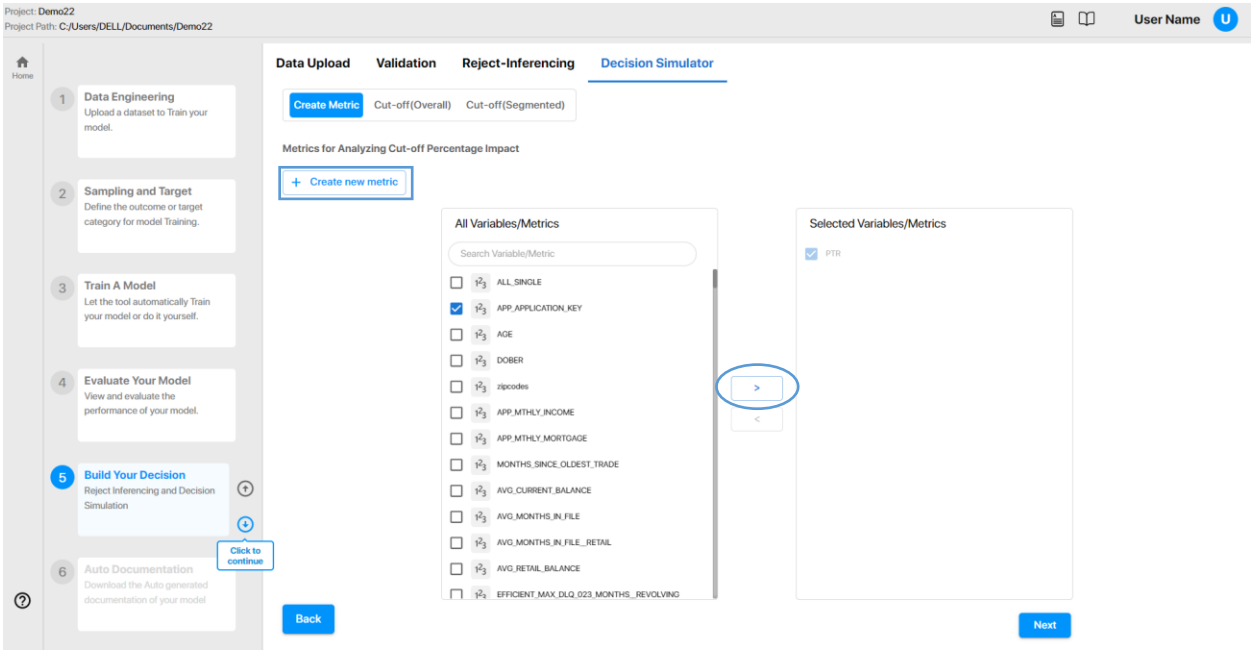
It would display the output as represented in below image. Click on the "**Back**" button on bottom left corner to return to "**Reject Inferencing**" screen.



Decision Simulator

Create Metric

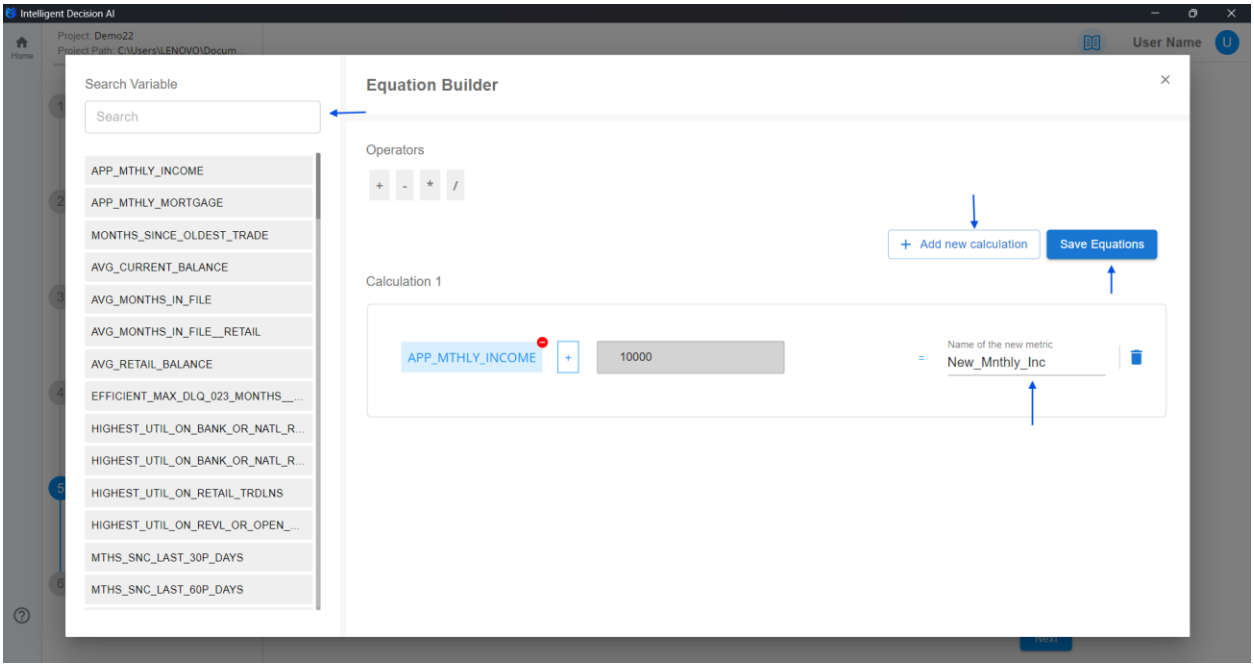
- User can select the metrics from the list of "**All Variables/Metrics**" from their uploaded dataset (refer image, blue oval)
  AND
- User can click on **"+ Create new metric**" to create new metrics of their own (refer image, blue rectangle)
  OR
- Click on "**Next**" button to skip this and proceed further.



"+ Create new metric"

- Drag the operator into the canvas that is to be performed
- Can search for the name of the feature that is wanted for operating on from the "**Search Variable**" search box (top-left)
  Note* : Only numerical features from the dataset will be visible here
- Select the variable from the list on the left of the screen and drag onto the appropriate side of the operator. Numerical values can be used too along with the features available (as shown in the image)
- Give a distinct name to the variable which doesn't conflict with existing variables from the dataset or the newly created metric.
- Click on "**Add new calculation**" to create another metric.

- The equation can be discarded by clicking on the small trash icon located beside the new metric name.
- Click on "**Save Equations**" to save the newly created metrics for further use.



Cut-off Decision – Overall

This tab provides a summary of the scorecard performance when different cut-off thresholds are selected. This overall approach does not allow segmented cut-offs, which will be described in the next section.

This screen allows the user to view the score cut-offs and their impact on two pre-defined outcomes when compared to existing scenario:

1. Maintaining same target rate
2. Maintaining same approval rate

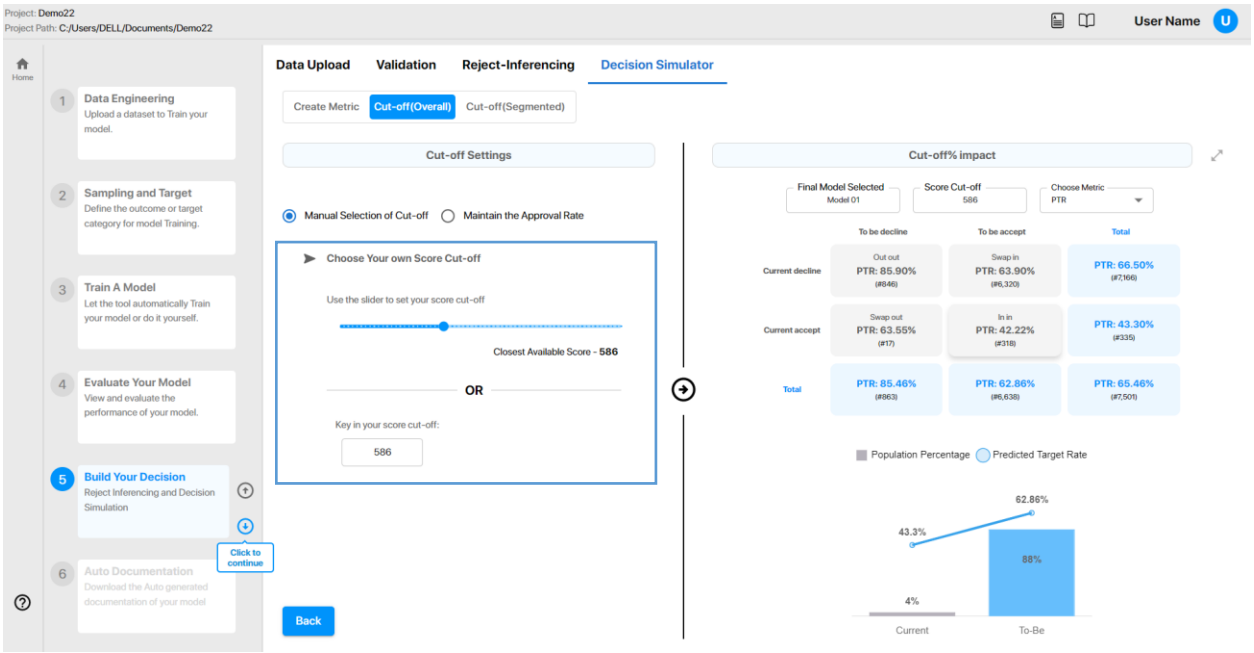
Users can also choose actual score point values to analyse the impact of this on target and approval rates.

The impact of the above selections can be seen in terms of:

1. Accept/Decline share before and after cut-off selection
2. Bad rates of Accept/Decline before and after cut-off selection

Impact can be seen graphically from the bar chart to view the population percentage and the metric

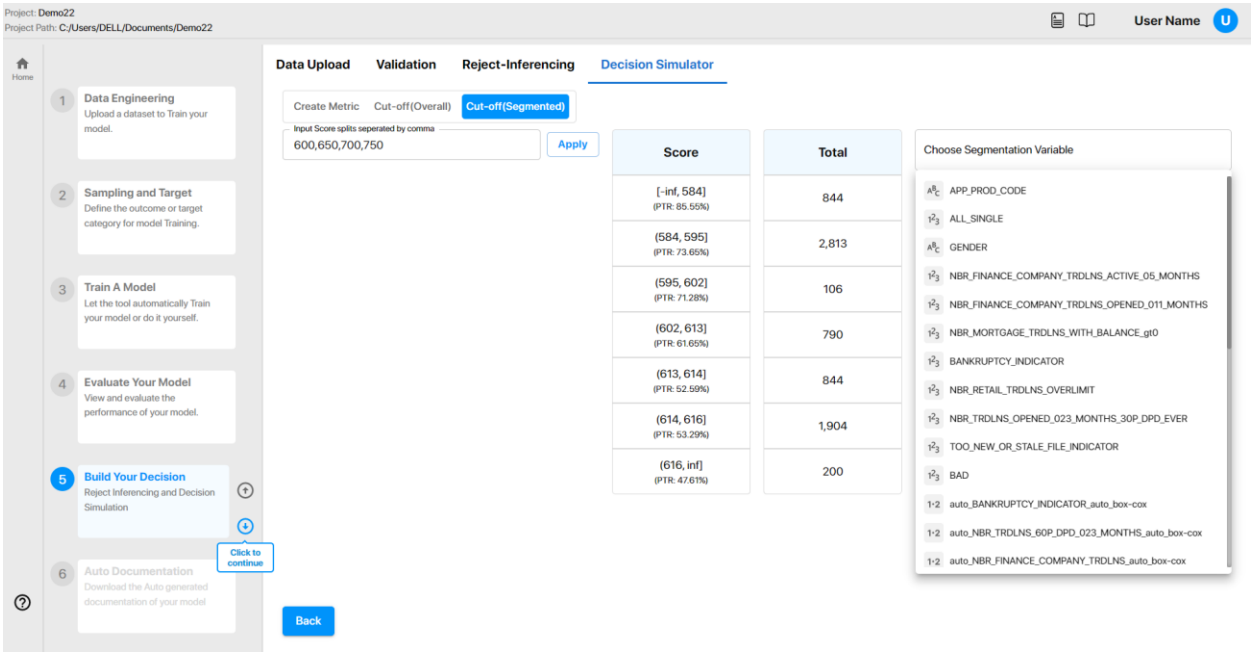
New Metrics created or added can be selected from “Choose Metric” dropdown to view their impact on the cut-off.



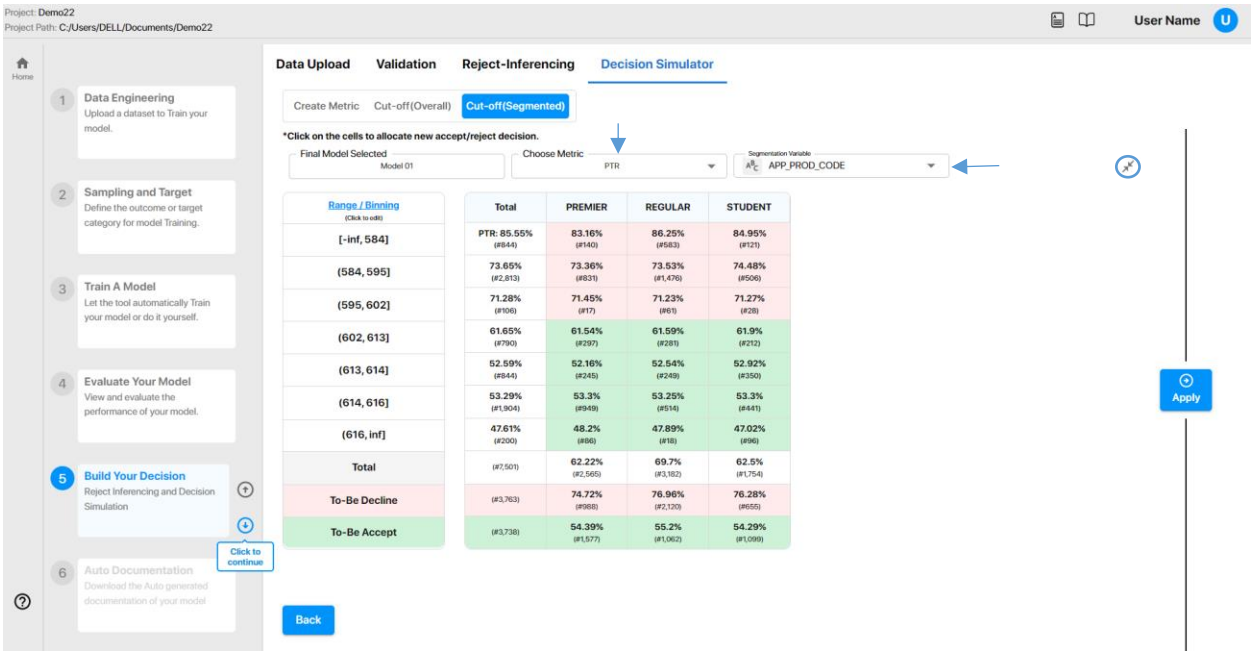
Cut-off Decisions – Segmented

This screen displays the total population or target that belongs to individual score bins. This option enables the user to set tailored cut-off decisions for specific segments by selecting a segmentation variable for individual score bins. This allows for decision making based on distinct groups within the chosen segmentation variable. To customise the score bins, user should input custom score values that they want to create bins for and click “Apply” (refer below image). This step can however be skipped and can proceed further with only the selection of segmentation variable.

The following screenshot shows the distribution of all the cases by score bin.



Let us assume that the user would like to have a segmented cut-off decision using a variable named “APP_PROD_CODE”. In this scenario, the user selects this variable in the Drop-down menu named “**Choose segmented variable**” available on the right side of the table (refer above image).



The count of records within each score bin split into the different categories of the APP_PROD_CODE is displayed here.

Option 1:

- ID.ai provides with pre-analysed recommendation based on the model built on this unseen dataset for the “To-Be Accept” and “To-Be Decline” in shade of green and red respectively. By default, the Predicted Target Rate (PTR) is taken to be 4.30%.

- User can modify this selection by clicking to deselect and select again based on their purpose.

- Click on “**Apply**” button to view the “Cut-off Impact” of the new segmentation.

- The results of this decision will be displayed including the population decline/accept counts and the bad rates.

- User can do multiple iterations of the accept/decline combinations and every time the impacts will be available.

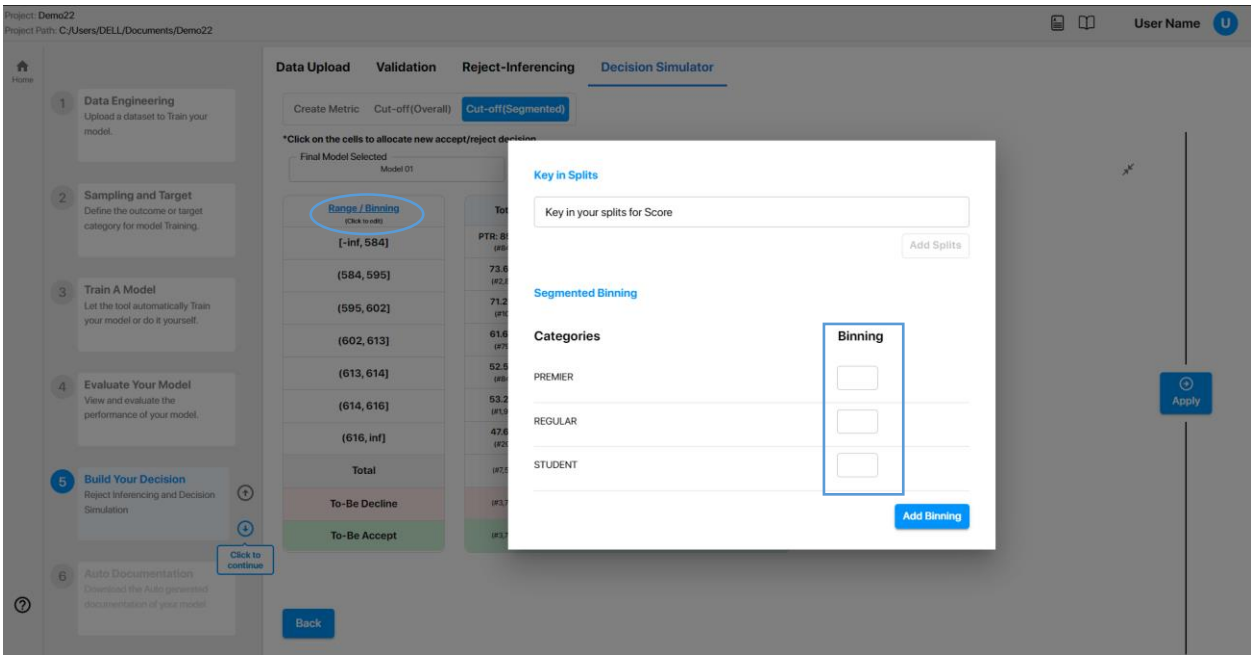
- Once satisfied with the segmented cut-off decision, user can click on “**Save Decision**” to store this decision.

The  option can be used to toggle between expanded and compact view.

Option 2:

This option is to combine multiple categories of the segmented variable.

1. Click on “**Click to edit**” under Range/Binning.
2. In the ensuing pop-up screen assign the same integer values to those categories of the segmented variable that need to be combined.
3. Click on “**Add Binning**”



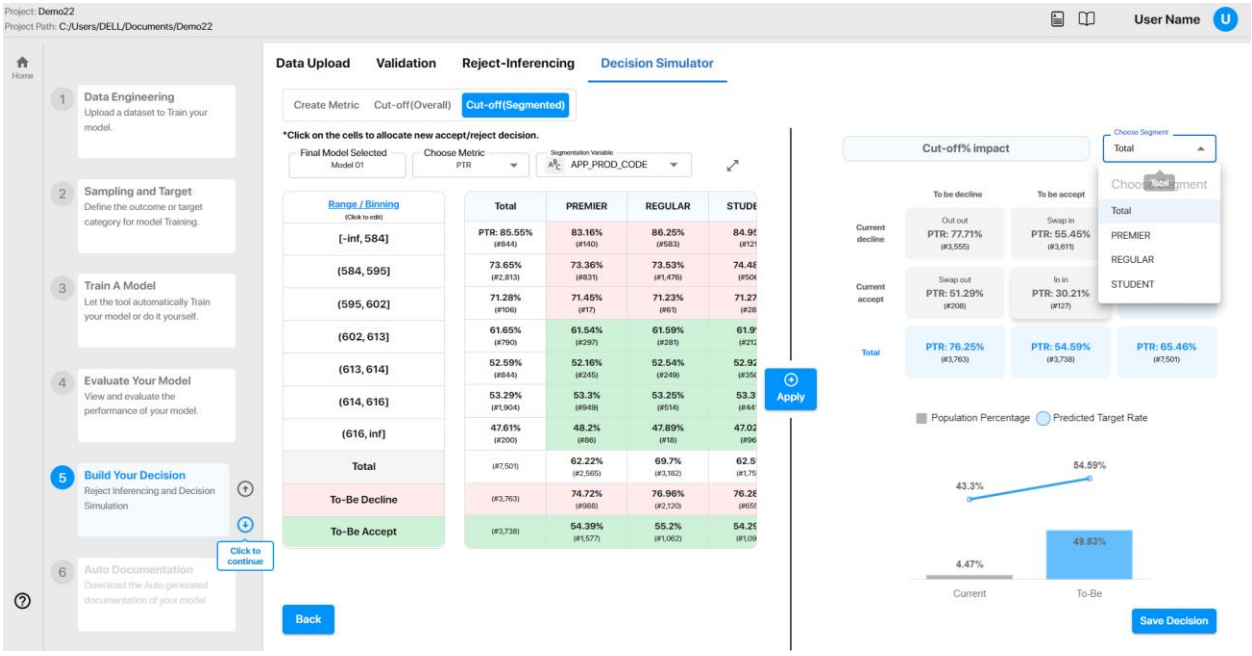
1. In the subsequent screen, click only on the specific boxes which will be accepted.

2. Once user clicks on "**Apply**" after doing the above, the results of this decision will be displayed including the population decline/accept counts and the bad rates.

3. User can do multiple iterations of the accept/decline combinations and every time the impacts will be available.

4. The impact can be viewed for different binning/segmentation created by changing "**Choose Segment**".

5. Once satisfied with the segmented cut-off decision, user can click on "**Save Decision**" to store this decision.



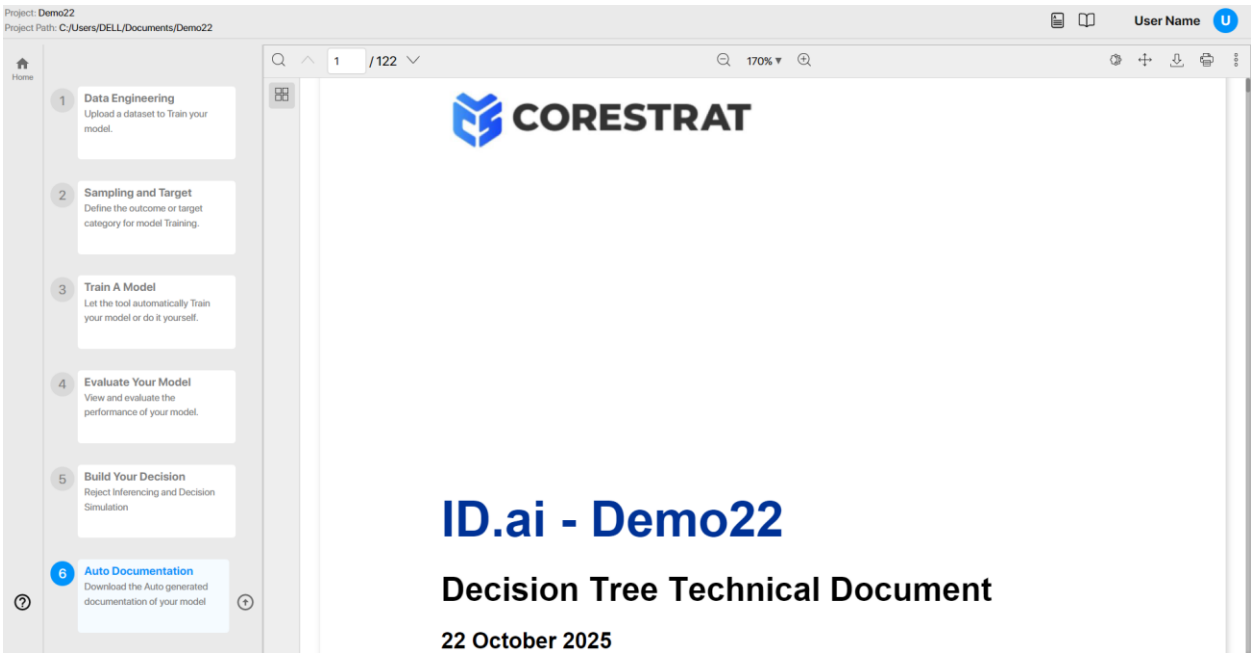
Auto Documentation

Once the model-build and cut-off decisions have been completed, the final stage is the generation of the “Decision Tree Technical Document” which provides a comprehensive document that will be invaluable for audit trail purposes. The user just needs to click on “Generate document” and an PDF version of the document will be available in 2-3 minutes.

While almost the entire document is pre-filled with the relevant information, the following three sections must be filled by the user before dissemination.

1. **Executive Summary**
2. **Data Sources and Sampling**
3. **Decision Tree Fact Sheet**

Since these sections involve the business context and user’s knowledge of the data, these ideally should be filled by the user.


Project: Demo22
Project Path: C:/Users/DELL/Documents/Demo22
User Name
Home
1 / 122
170%
1 Data Engineering
Upload a dataset to Train your model.
2 Sampling and Target
Define the outcome or target category for model Training.
3 Train A Model
Let the tool automatically Train your model or do it yourself.
4 Evaluate Your Model
View and evaluate the performance of your model.
5 Build Your Decision
Reject Inferencing and Decision Simulation
6 Auto Documentation
Download the Auto generated documentation of your model
CORESTRAT
ID.ai - Demo22
Decision Tree Technical Document
22 October 2025

Appendix A – Statistical Terms

1. **p-value:** Measure that quantifies the evidence against a null hypothesis. A low p-value suggests data is inconsistent with the null, potentially favouring an alternative hypothesis. Common significance thresholds are 0.05 or 0.01.

2. **IV (Information Value)**: A numerical value that quantifies the predictive power of an independent continuous variable x in capturing the binary dependent variable y. IV is helpful for reducing the number of variables as an initial step in preparing for Logistic Regression, especially when there are a large number of potential variables. IV is based on an analysis of each individual independent variable in turn without considering other predictor variables.

3. **WOE (Weight of evidence):** Closely related to the IV value, WOE measures the strength of each grouped attribute in predicting the desired value of the Dependent Variable.

4. **VIF (Variance Inflation Factor):** A measure of multicollinearity among the independent variables in a multiple regression model.

5. **OOT (Out of Time) Sample:** Used to indicate a dataset from a period outside the original model build window; used to validate the accuracy of the model in other time periods.

6. **Gini coefficient:** Gini coefficient, commonly known as Gini, is a metric widely used to evaluate classification models. It ranges from 0 to 1, with zero representing perfect equality (no discrimination) and one representing perfect inequality (perfect discrimination). In the context of credit risk modelling, a higher Gini coefficient indicates better model performance in terms of its ability to accurately rank borrowers based on their creditworthiness.

7. **K-S ("Kolmogorov-Smirnov") Value:** The KS value provides a measure of the discriminatory power of a model. It looks at the maximum difference between the distribution of cumulative events and cumulative non-events and is a way of comparing the cumulative sum of the positive and negative classes. It measures the maximum difference between the two over the range of predicted probabilities. A high KS score indicates that the model has a better separation between the positive (goods) and negative classes (bads).

8. **Skewness**: Measure the degree of asymmetry of a distribution.

9. **Kurtosis**: Measure of the peak height of a distribution.

10. **Base score:** The actual score point in the scaled scorecard which gives the base odds for the target variable to go into the desired state.

11. **Base odds:** The odds of the target variable to go into the desired state at the base score.

12. **PDO (Probability of Double Odds):** The actual score points difference needed to increase the odds of the target variable's target rate twice.

Frequently Asked Questions

Q. What are the minimum specifications for a machine to install and run ID.ai?

A. The minimum configurations a system hosting ID.ai is >=16 GB RAM minimum, 200 GB Free disk space with MS Word Installed

Q. Can I save the project in a custom folder other than the default folder path provided in ID.ai?

A. No, currently this facility is not available; it will be enabled in a future version of ID.ai.

Q. Does ID.ai run on MacBook also?

A. The current version of ID.ai runs only in the Windows environment. Future versions will be Apple OS compatible also.

Q. What statistical technique is used for building the model?

A. Logistic regression

Q. What should user do if activation is unsuccessful?

A. Please reach out to your company's system administrator who purchased the license keys from Corestrat or drop a mail to solutions@corestrat.ai with your license key

Q. Where can I find the current project saved?

A. The default location is saved in the default local system path usually having the following location "C:\users\<machinename>\documents\Idai\<projectname>"

Q Where can I find current autodocument to be saved.

A. The default location is saved in the default local system path usually having the following location
"C:\users\<machinename>\documents\Idai\<projectname>\documents

Q. How can I start a new project while on one project?

A. In the home screen: Click on Home > All Projects > New Project

Q. How can I delete the project or dataset I uploaded?

A. There is an option to "delete" (trash can icon) under each project in the path above

Q. How is the performance of the model evaluated?

A. Model Evaluation can be done using the KS and Gini Statistics provided by the app.

Q. Can I make predictions with new data after training the model?

A. Yes you can make predictions with the “Build Your Decision” screen by uploading an unseen dataset and using the trained model on it.

Q. Can I use the model to score new customers or cases?

A. Yes, that can be done using the trained model on the unseen dataset in “Build Your Decision” screen.

Q. Can I save and export my trained model?

A. Yes. The trained model is saved in the default folder path provided in ID.ai as a file in ‘.pkl’ format.

Q. Can I generate scorecards from the trained model?

A. Yes. The scorecard generated from the trained model is available in the 'Model Scorecard' tab of the 'Evaluate Your Model' section.

Q. Does ID.ai automatically detect the target variable?

A. ID.ai suggests a list of potential target variables; user has the option to accept from those or use another target variable. Target Variable Selection

Q. Can the user select the features to be included in the model?

A. Yes, it can be done through various steps of preparing data (variables to be ignored, clustering and keeping/discarding variables) and while building the model (Manual Grow in “Train a Model”)

Q. Where can I get the template for metadata?

A. Under “Data Preprocessing” tab, on the top right use the button named “Add Meta Data”