



IntelliDecision.ai

User Manual

This document is prepared by Corestrat Pvt Ltd. Materials in this document shall only be for internal use and in no circumstances shall any extract of the document be replicated for external use.

Copyright @ 2025 Corestrat

Table of Contents

1.	Introduction	3
2.	Software Requirements Specifications	3
3.	Installation of ID.ai	4
4.	Getting Started	6
5.	Upload Data Section	8
5.1.1	Project Creation & File Input	8
5.1.2	Import from a Data Source	28
5.1.3	Existing Projects	32
5.1.4	Input Data Management	34
6.	Sampling and Target	59
6.1.1	Target Variable Selection	59
6.1.2	Stratified Sampling	61
6.1.3	Independent Variable Insights	62
7.	Train a Model Section	78
7.1.1	Model Settings and Root Node	79
7.1.2	Auto Grow	84
7.1.3	Run Model.ai	889
8.	Evaluate Your Model	111
9.	Build Your Decision	116
9.1.1	Reject Inferencing	122
9.1.2	Cut-off Decision – Overall	125
9.1.3	Cut-off Decisions – Segmented	126
10.	Auto Documentation	130
	Appendix A – Statistical Terms	132
	Frequently Asked Questions	134

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.

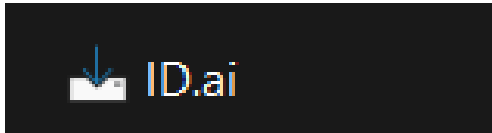
Operating System	Windows 10 & Higher
RAM	>=16 GB
Disk Space	>=200 GB
Software	MS Word 16 & Higher

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

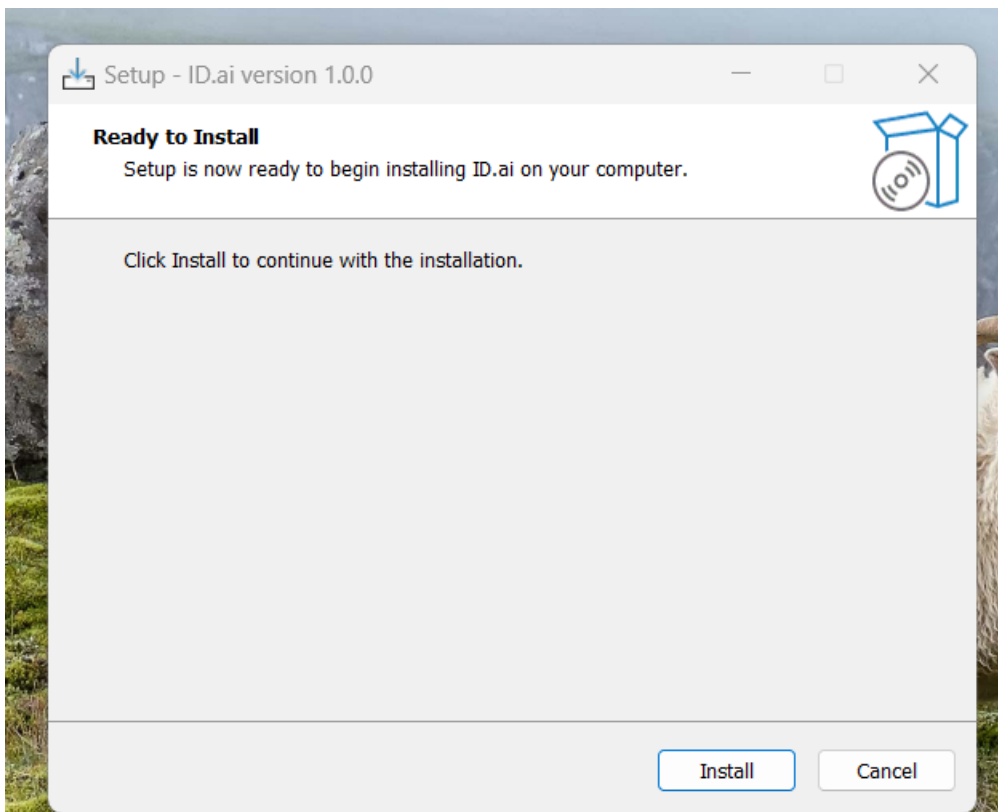
Installation of ID.ai

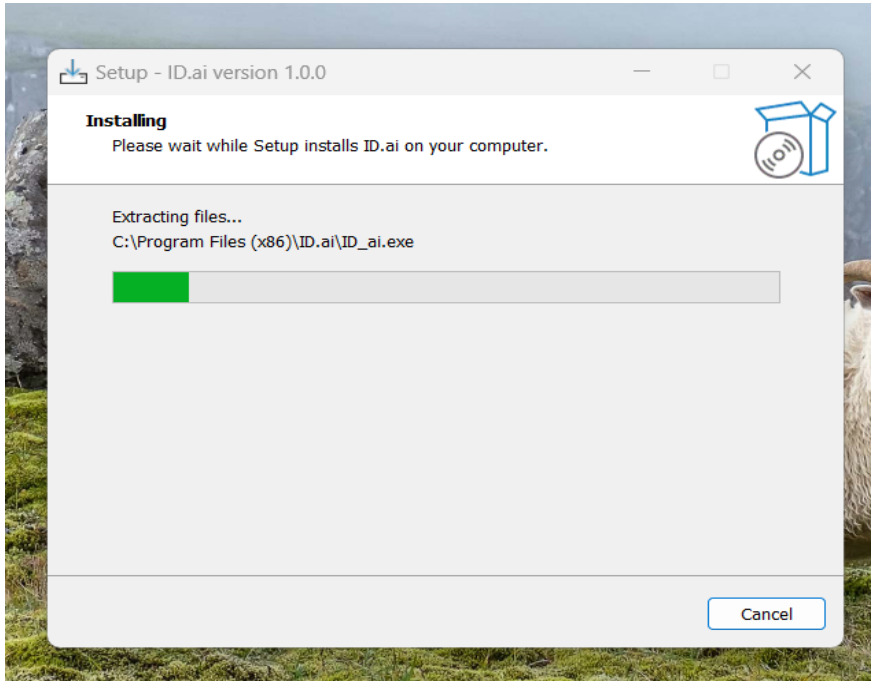
This section provides the steps to install the ID.ai application on the user's laptop.

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

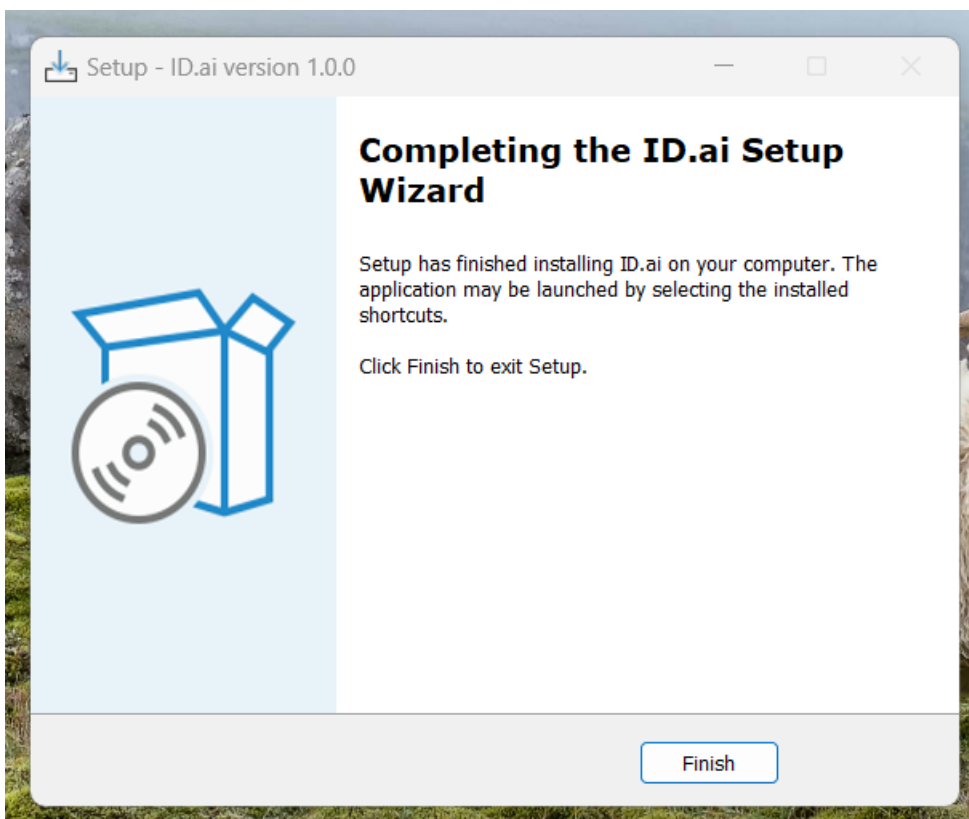


User should click on this exe file and will get the pop-up screen shown below requesting the user to proceed with the subsequent installation steps. Click on "Install". User should right click and "Run as Administrator" in case privileges are insufficient for direct installation.



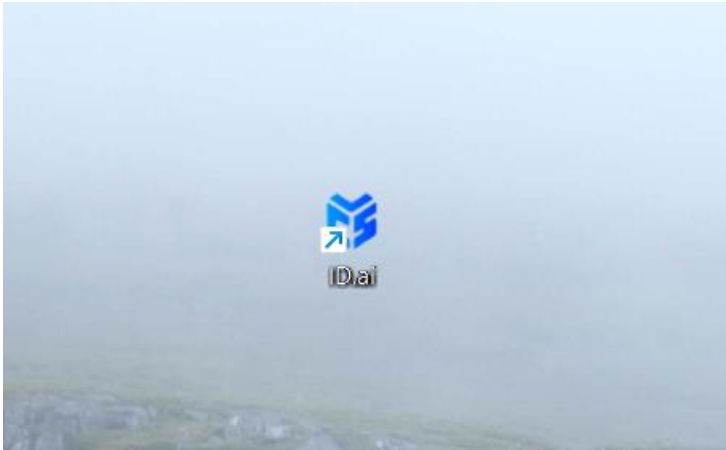


The user needs to click on “Finish” once installation is complete.



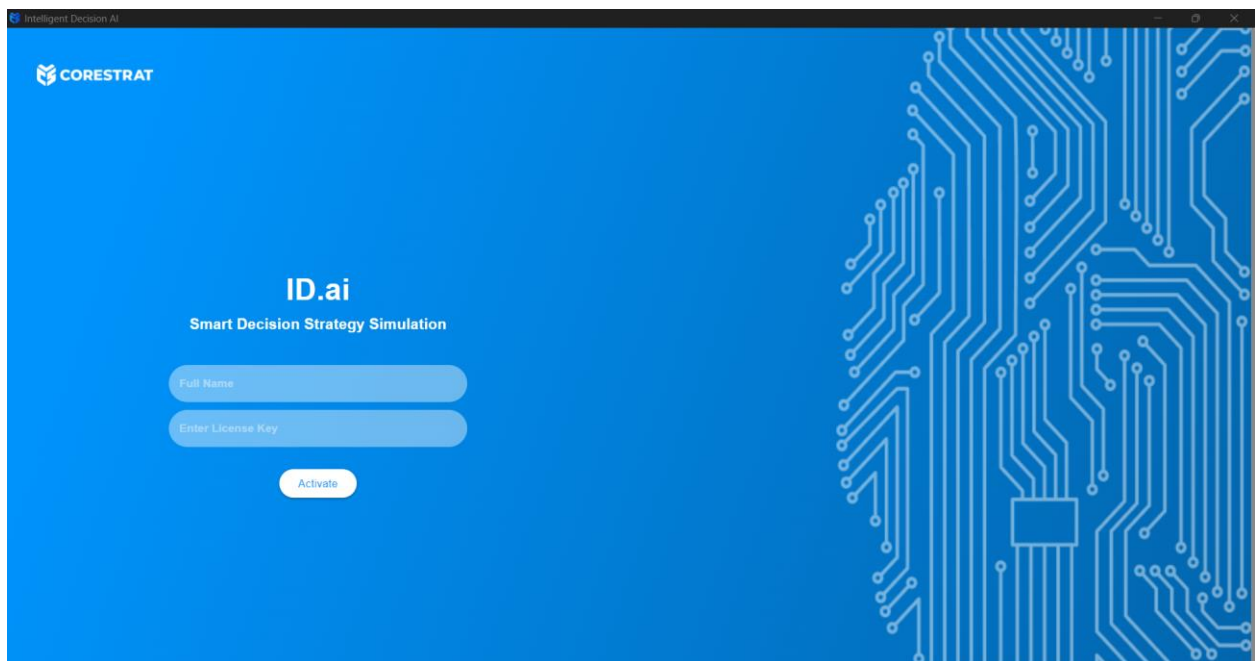
The user will get a shortcut to open the ID.ai application on the desktop.

Click on the ID.ai icon as shown below.



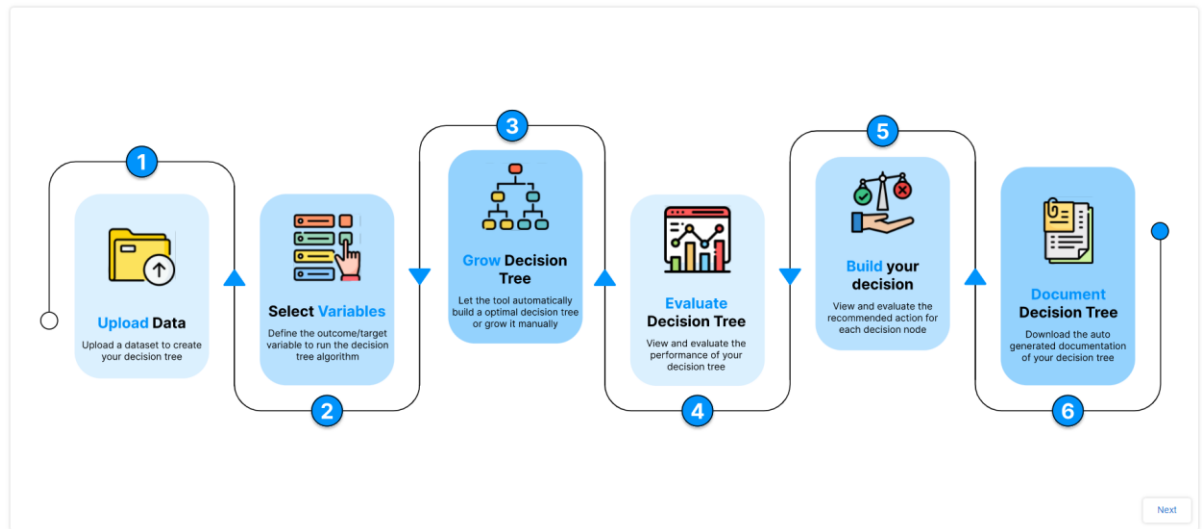
Getting Started

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.

How to build a decision tree using ID.ai

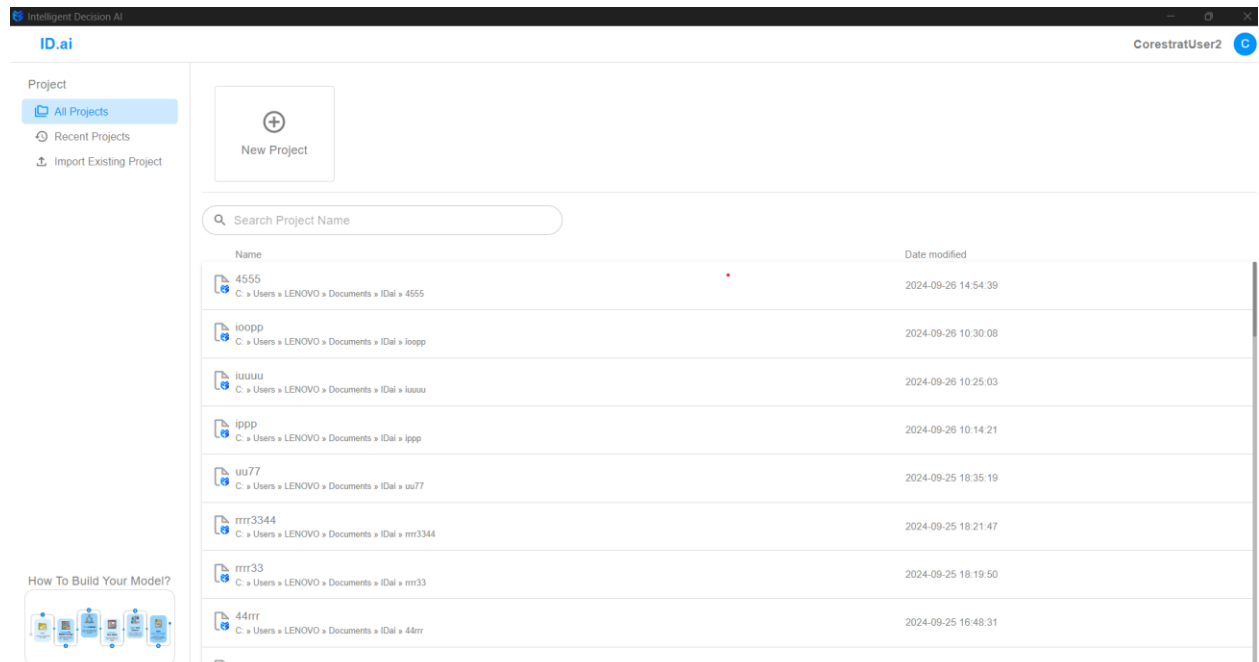


The steps within these six stages will be described in the subsequent sections.

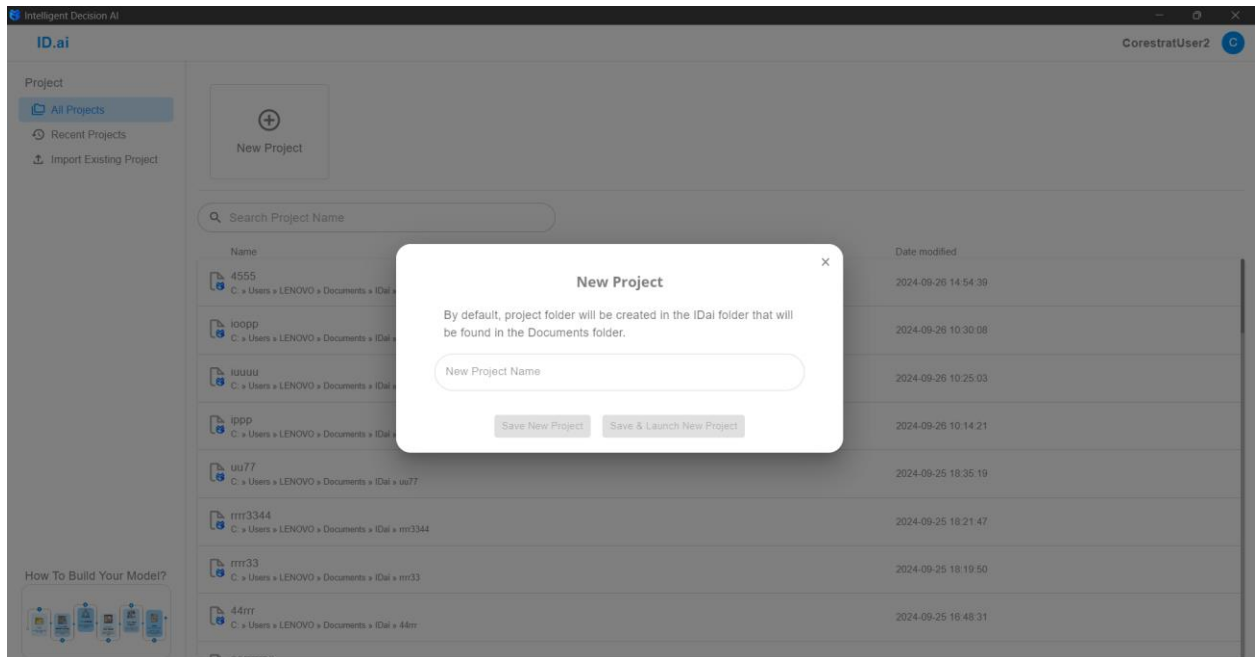
Upload Data Section

Project Creation & File Input

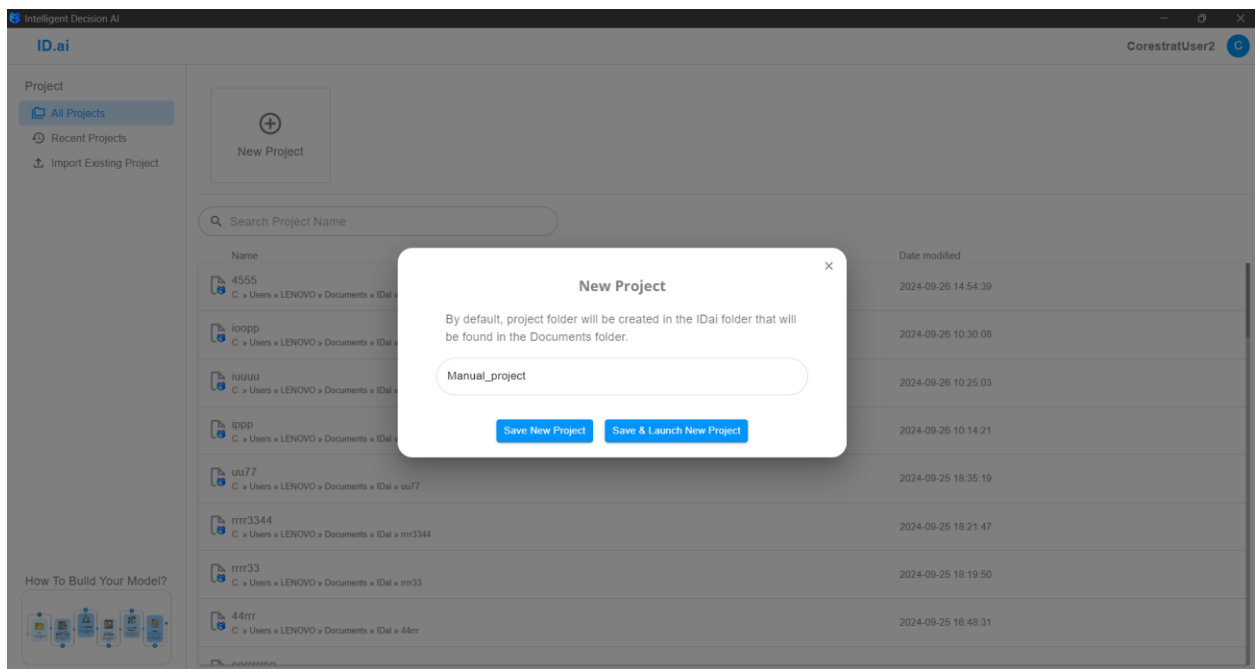
ID.ai can be used to build a new model from scratch or refine an existing model using new assumptions or additional data elements. Both options are illustrated in the screenshot below.



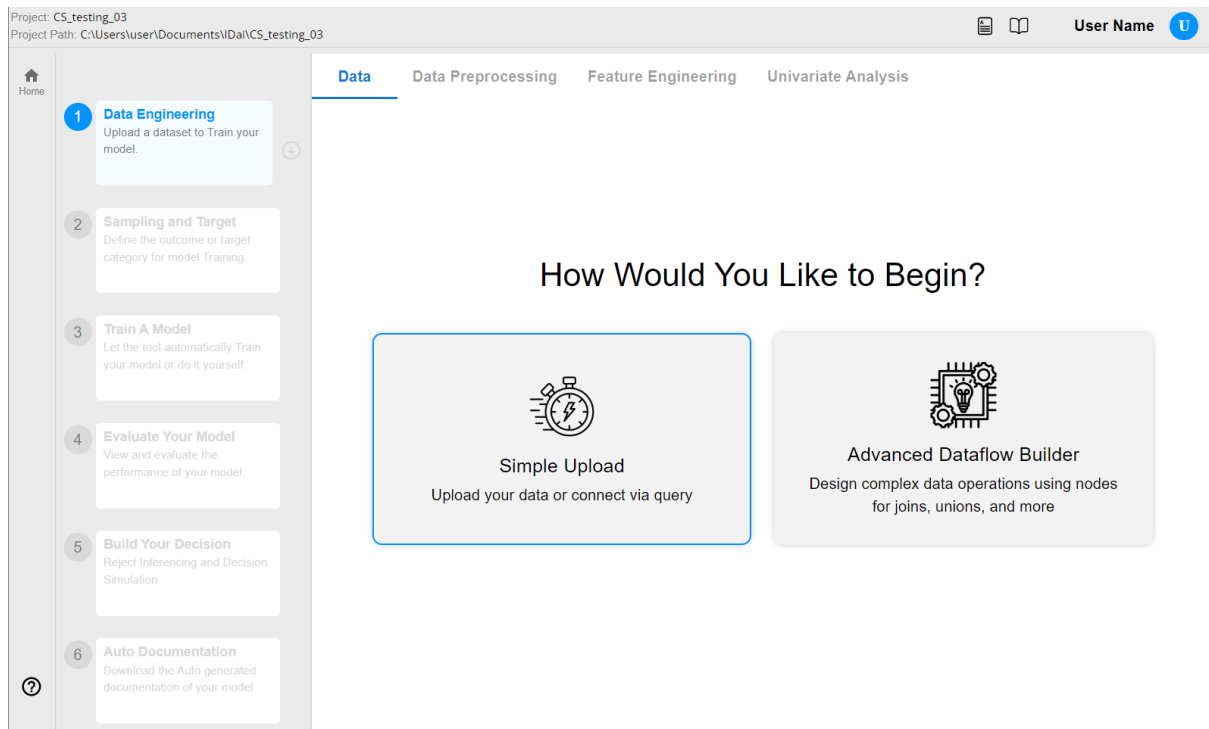
For building a new project, click on “**New Project**” and assign the project name in the pop-up screen as shown below.



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.



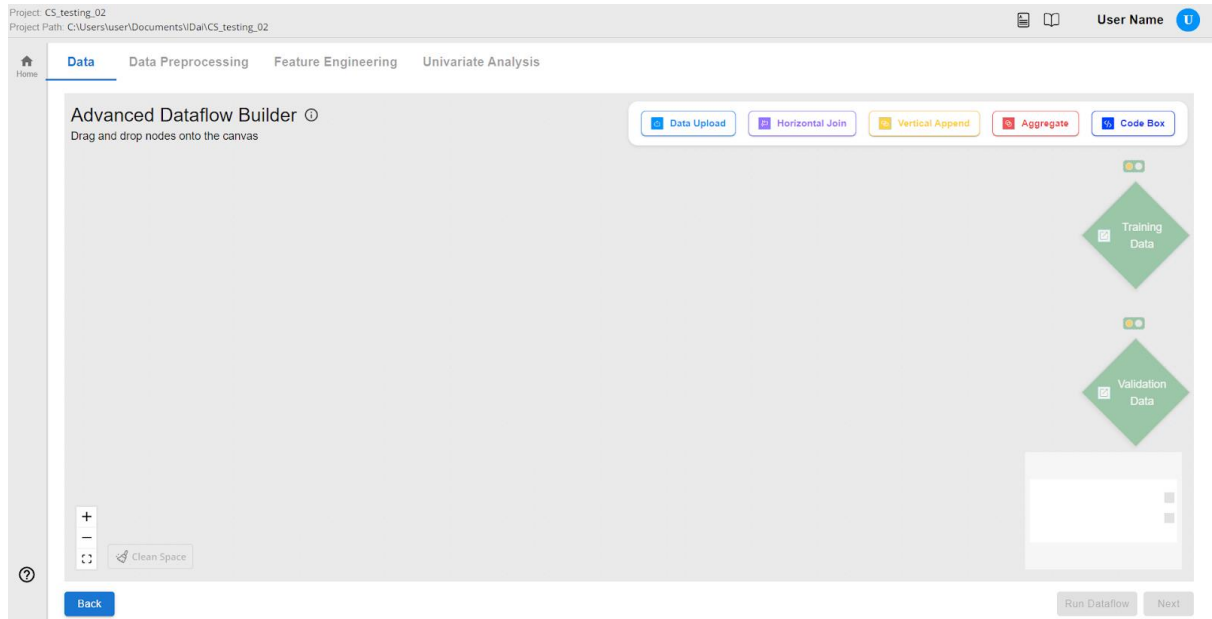
In the subsequent screen, the “project_name” and “project path” are displayed at top left side of the screen. For starting a new project, the user must point to the folder location and file name of the input data source.



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.

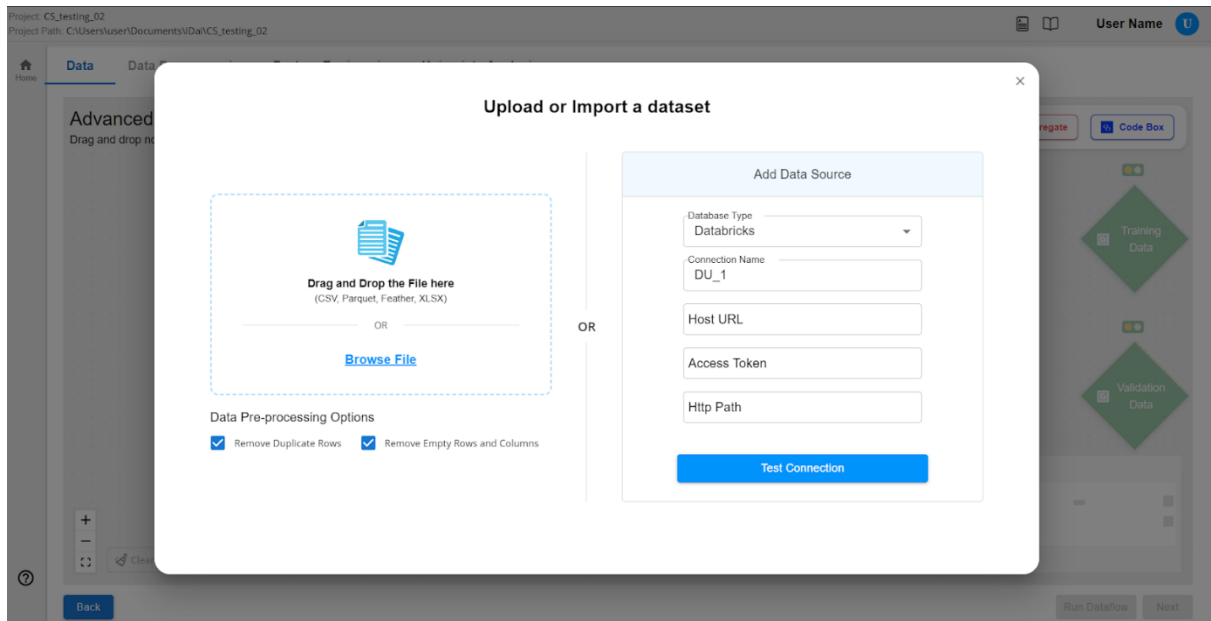
The **Advanced Dataflow Builder** enables users to create complex dataflows with ease by using drag-and-drop action nodes to perform powerful data transformations.

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



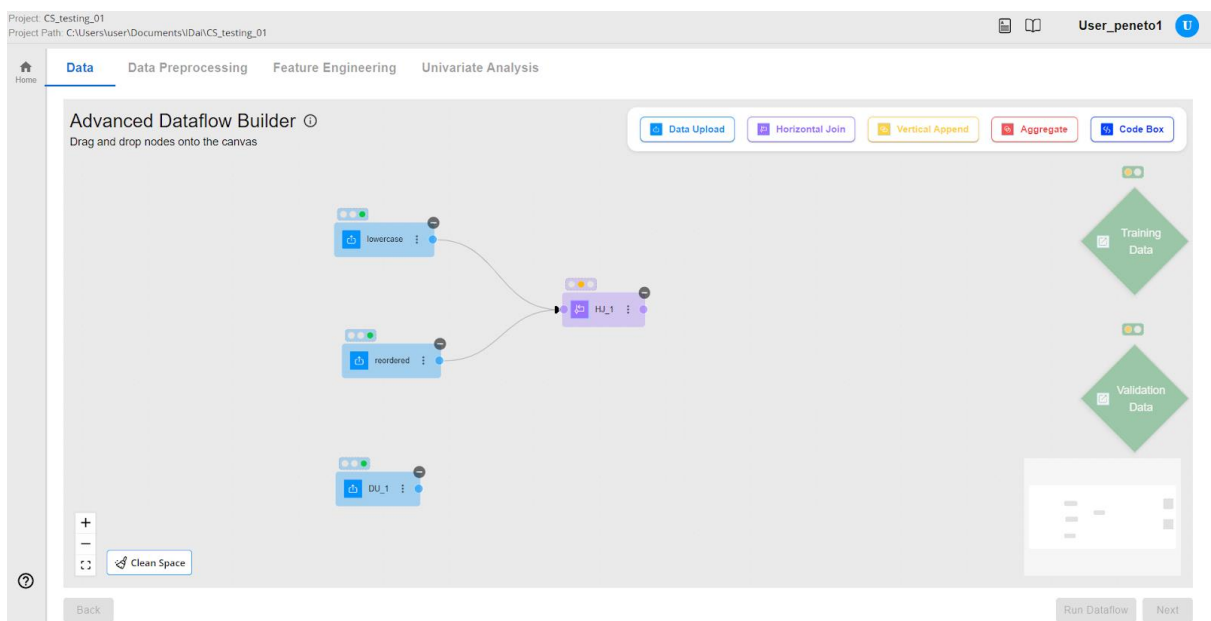
The first step is for the user to add **Data Upload** boxes to the whiteboard. Upon dragging the Data Upload component to the canvas, the screen shown below will appear.

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.

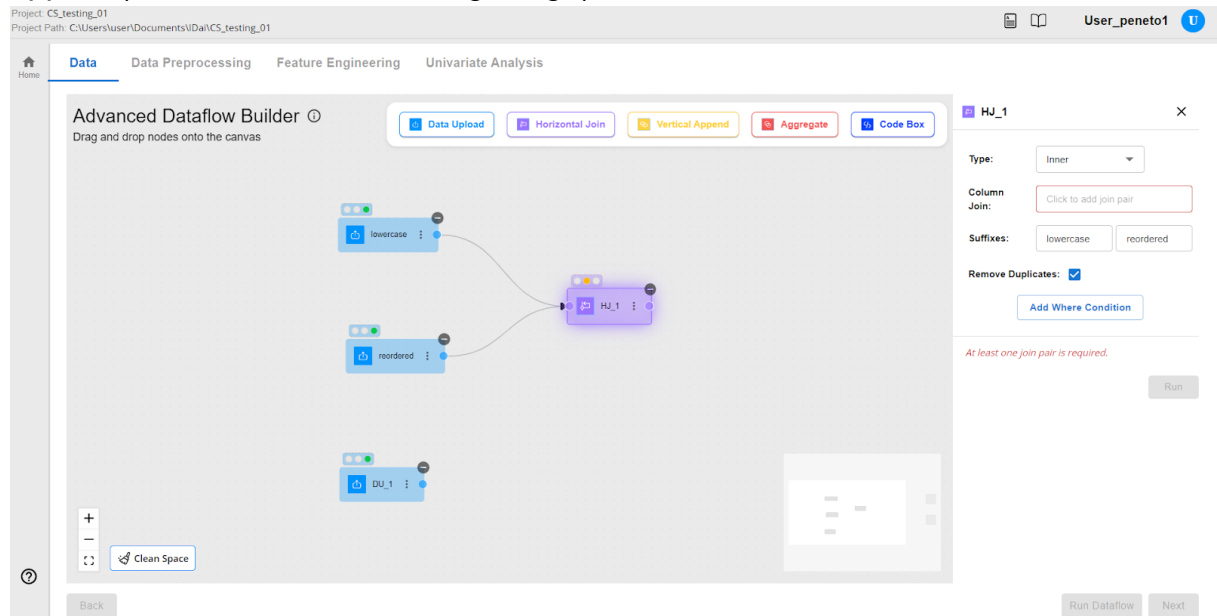


Once a Data Upload box has been successfully configured (indicated by the green light on top of each action box), the user can begin connecting it to action nodes.

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.

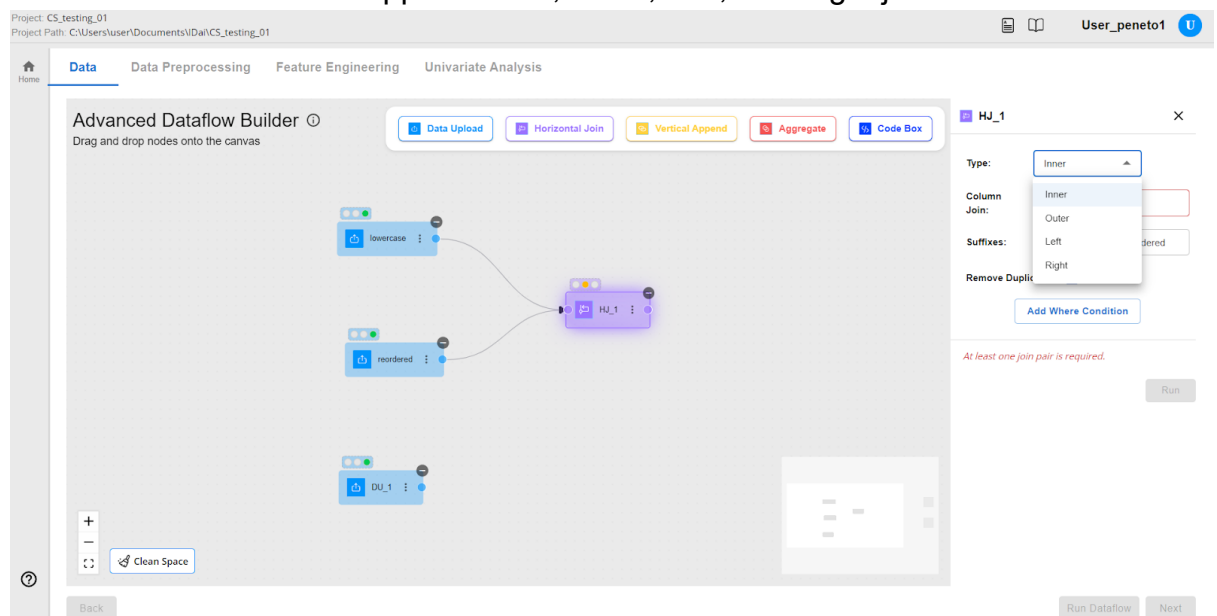


To use the **Horizontal Join** component, the user can drag and drop the box onto the canvas. Upon selecting the newly added box, a configuration sidebar will appear (as shown in the following image).



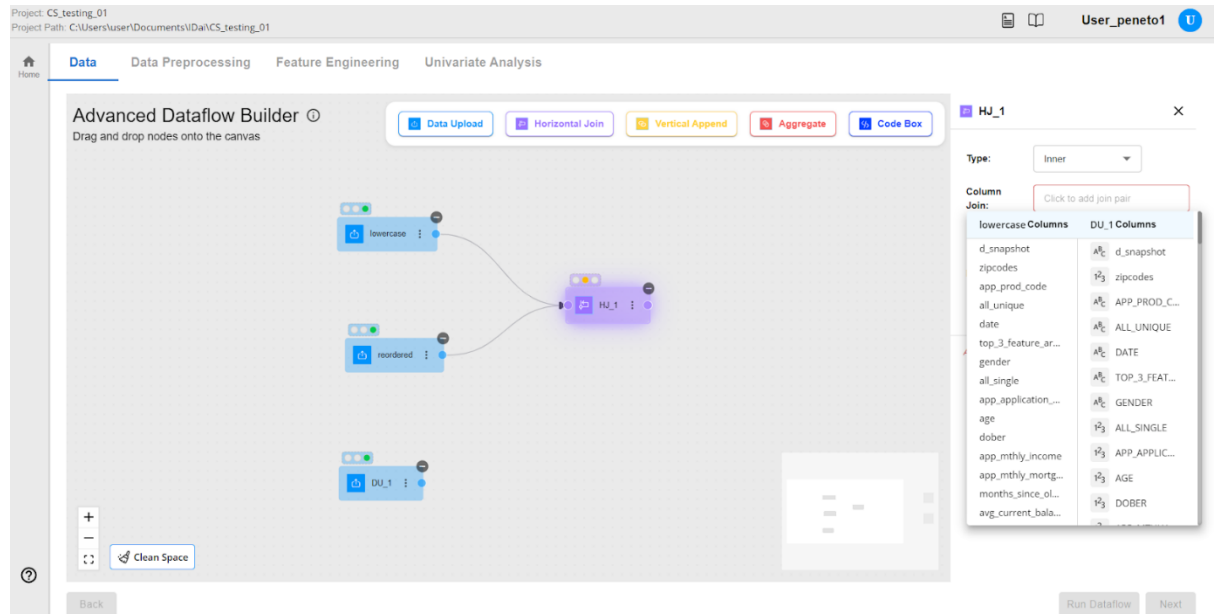
From this sidebar, the user can choose the type of join, select the columns for the join, customize table suffixes, and optionally add a custom WHERE condition to further filter the data.

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



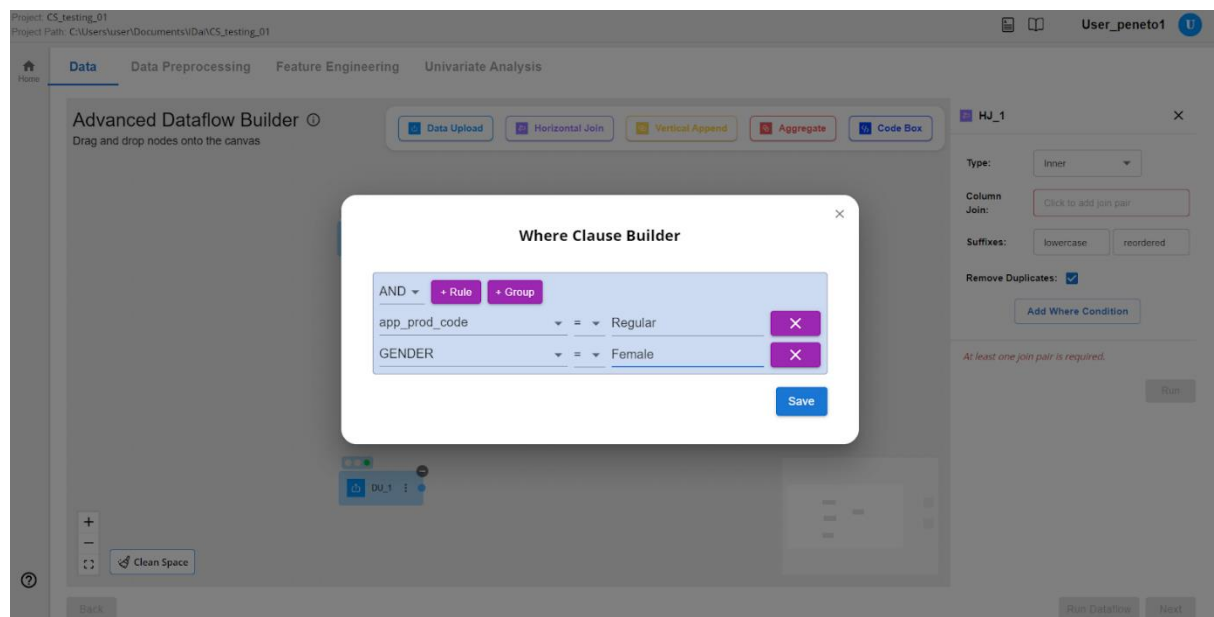
The column join section displays all columns from the datasets connected to the

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.

Project: CS_testing_01
Project Path: C:\Users\User\Documents\IDa\CS_testing_01

Home | Data | Data Preprocessing | Feature Engineering | Univariate Analysis

Advanced Dataflow Builder Ⓞ
Drag and drop nodes onto the canvas

Buttons: Data Upload, Horizontal Join, Vertical Append, Aggregate, Code Box

Canvas nodes: lowercase, reordered, DU_1, HJ_1

HJ_1 Configuration Panel:

- Type: Inner
- Column Join: Click to add join pair
- Selected Pairs: app_prod_code → APP_PROD_C...
- Suffixes: lowercase, reordered
- Remove Duplicates: ☒
- Where condition: app_prod_code = 'regular' and GENDER = 'female'
- Buttons: Modify Where Condition, Run

Bottom buttons: Back, Run Dataflow, Next

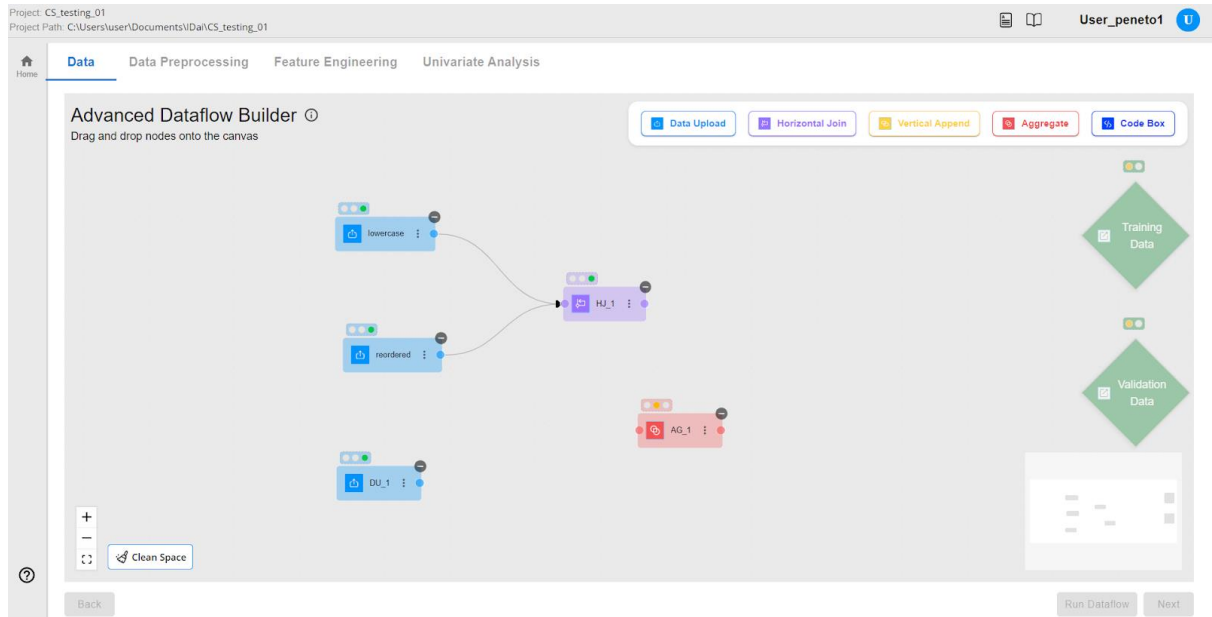
After configuring the necessary settings, the user can click **Run** to execute the join node, after which the output will be displayed as shown.

Preview: HJ_1

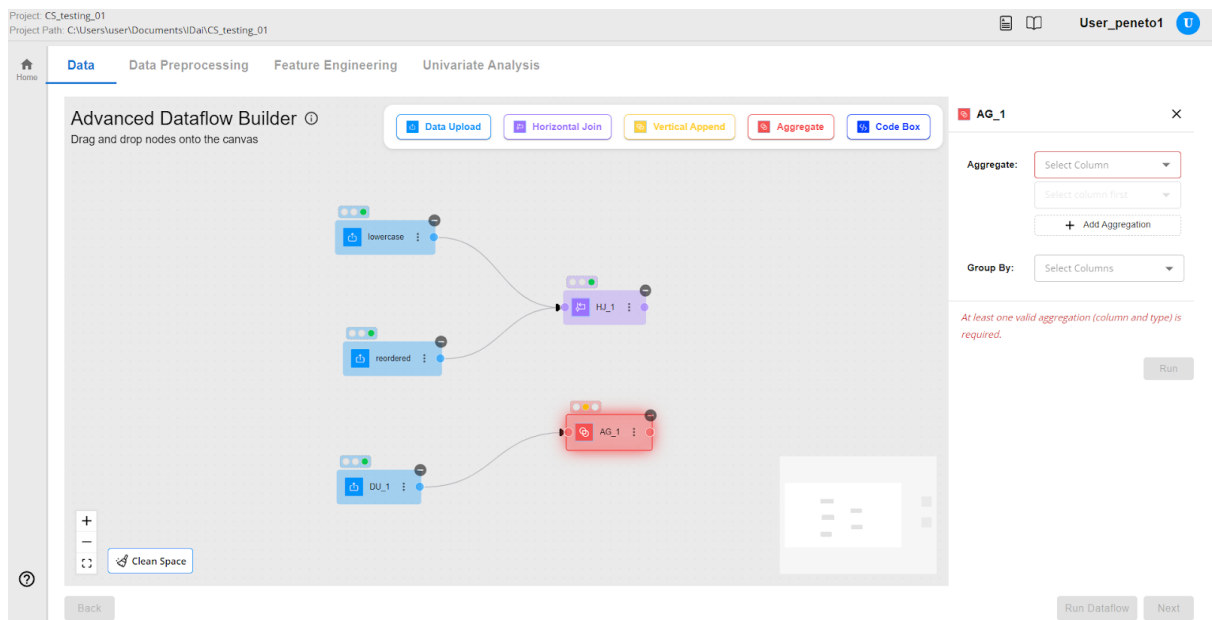
C:\Users\User\Documents\IDa\CS_testing_01\others\data\HJ_1

Rows				Columns			
Total	15,000	#Empty	0	Total	236	#Empty	0
#Distinct	15,000	#Duplicate	0	#Numerical	224	#Categorical	12

app_prod_code	all_single	all_unique	date	top_3_feature_areas_w1	app_a
REGULAR	1	026347df8155d73c15c328c9f...	28/5/2020	OTH_HME_AI	1
REGULAR	1	8a642af14ac909a353973e4b...	24/3/2023	OTH_HME_AI	2
REGULAR	1	940dfbcea633bcf78b120bf93...	25/2/2020	GSL_CV_GL	3
REGULAR	1	96a4d69280f3a255628964f2b...	31/8/2023	OTH_HME_AI	4
REGULAR	1	9de2c83009ca37a21684dd14...	13/12/2023	CV_GSL_DB	5
REGULAR	1	a3d114cdd72c7b7f3655b43d...	6/4/2023	CV_GSL_GSE	6
REGULAR	1	a44ecb7c6176e1b9bb89715f...	2/1/2019	SH_OTH_HME	7
REGULAR	1	a9cf8c9354ccd11b10de19623...	3/3/2022	OTH_HME_AI	8
PREMIER	1	b2cf8ee5786f68592ad3b096e...	22/10/2021	GSE_CV_GSL	9
STUDFNT	1	c7ba343dae3656e952ff7bde...	13/9/2022	OTH_HMF_AI	10



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.

Project: CS_testing_01
Project Path: C:\Users\user\Documents\Dan\CS_testing_01

User_peneto1

Home Data Data Preprocessing Feature Engineering Univariate Analysis

Advanced Dataflow Builder Drag and drop nodes onto the canvas

Buttons: Data Upload Horizontal Join Vertical Append Aggregate Code Box

Canvas nodes: lowercase, reordered, DU_1, HJ_1, AG_1

AG_1 configuration:

- Aggregate: APP_MTHLY_INC...
 - sum mean median min
- Group By:
 - sum mean median min max std var count unique first last prod

Buttons: Back Clean Space Run Dataflow Next

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

Project: CS_testing_01
Project Path: C:\Users\user\Documents\Dan\CS_testing_01

User_peneto1

Home Data Data Preprocessing Feature Engineering Univariate Analysis

Advanced Dataflow Builder Drag and drop nodes onto the canvas

Buttons: Data Upload Horizontal Join Vertical Append Aggregate Code Box

Canvas nodes: lowercase, reordered, DU_1, HJ_1, AG_1

AG_1 configuration:

- Aggregate: APP_MTHLY_INC...
 - sum mean median min
- Group By: APP_PRO...
 - count unique first
 - count unique first last min max

Buttons: Back Clean Space Run Dataflow Next

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

Project: CS_testing_01
Project Path: C:\Users\User\Documents\IDa\CS_testing_01

Home | Data | Data Preprocessing | Feature Engineering | Univariate Analysis

Advanced Dataflow Builder
Drag and drop nodes onto the canvas

Nodes: lowercase, reordered, DU_1, AG_1

AG_1 Settings:

- Aggregate: APP_MTHLY_INC...
- sum, mean, median, min
- APP_PROD...
- count, nunique, first
- + Add Aggregation
- Group By: Select Columns
- Run

Buttons: Back, Run Dataflow, Next

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

Preview: AG_1

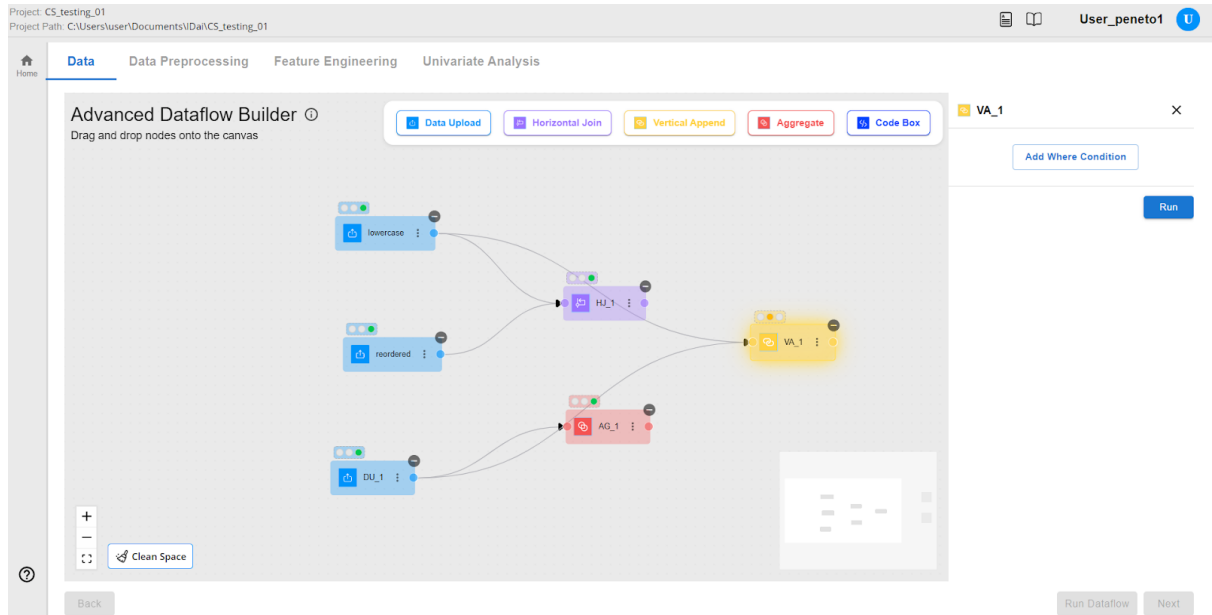
C:\Users\User\Documents\IDa\CS_testing_01\others\data\AG_1

Rows				Columns			
Total	3	#Empty	0	Total	9	#Empty	0
#Distinct	3	#Duplicate	0	#Numerical	7	#Categorical	2

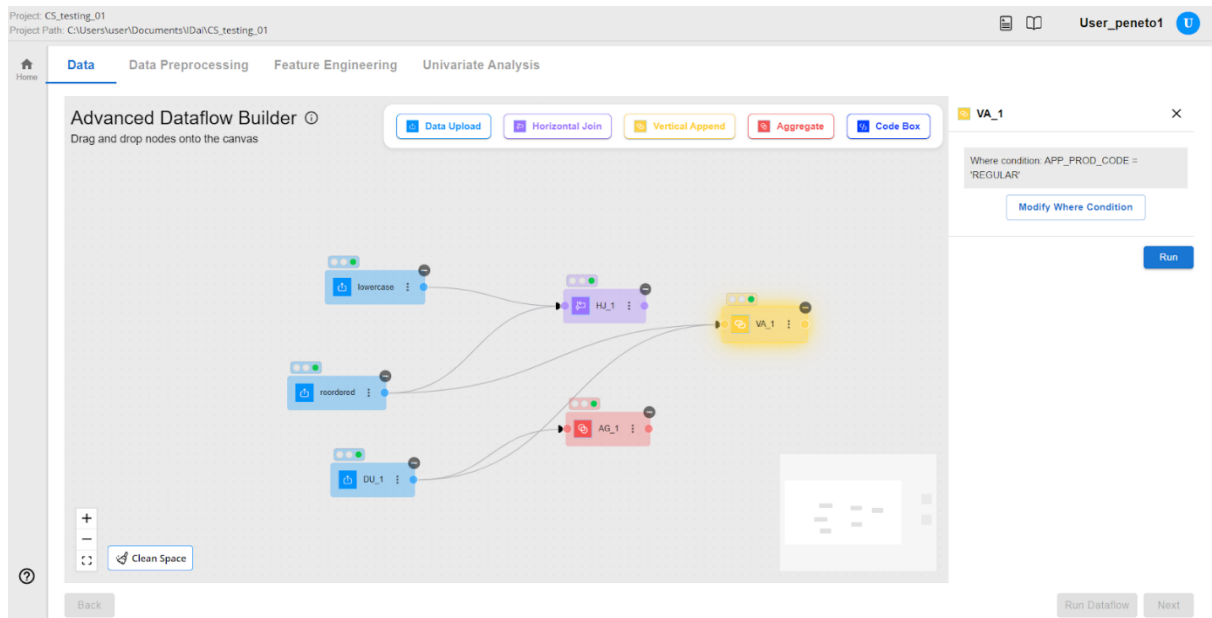
APP_PROD_CODE_first	APP_PROD_CODE_last	APP_PROD_CODE_nunique	APP_PROD_CODE_count	APP_MTHLY_INCOME_sum	APP_MTHLY_INCOME_mean
PREMIER	PREMIER	1	5,144	66,237,712	12,876.694
REGULAR	REGULAR	1	6,317	68,031,686	10,769.619
STUDENT	STUDENT	1	3,539	33,956,462	9,594.931

Buttons: Back, Run Dataflow, Next

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.



Project: CS_testing_01

Project Path: C:\Users\User\Documents\Da\CS_testing_01

Home

Data

Data

Advanced

Drag and drop

Preview: VA_1

C:\Users\User\Documents\Da\CS_testing_01\other\data\VA_1

Data Cleaned: Rows Filtered Out 17366

Rows

Total12,634

#Empty0

#Distinct6,317

#Duplicate6,317

Columns

Total118

#Empty0

#Numerical112

#Categorical6

1-2MTHS_SNC_OLDEST_BANI1-2AVG_CURRENT_BALANCE1-2NBR_DEROG_TLS_OR_CO1-2NBR_INQUIRIES_011_MON1-2MTHS_SNC_OLDEST_INST1-2NBR

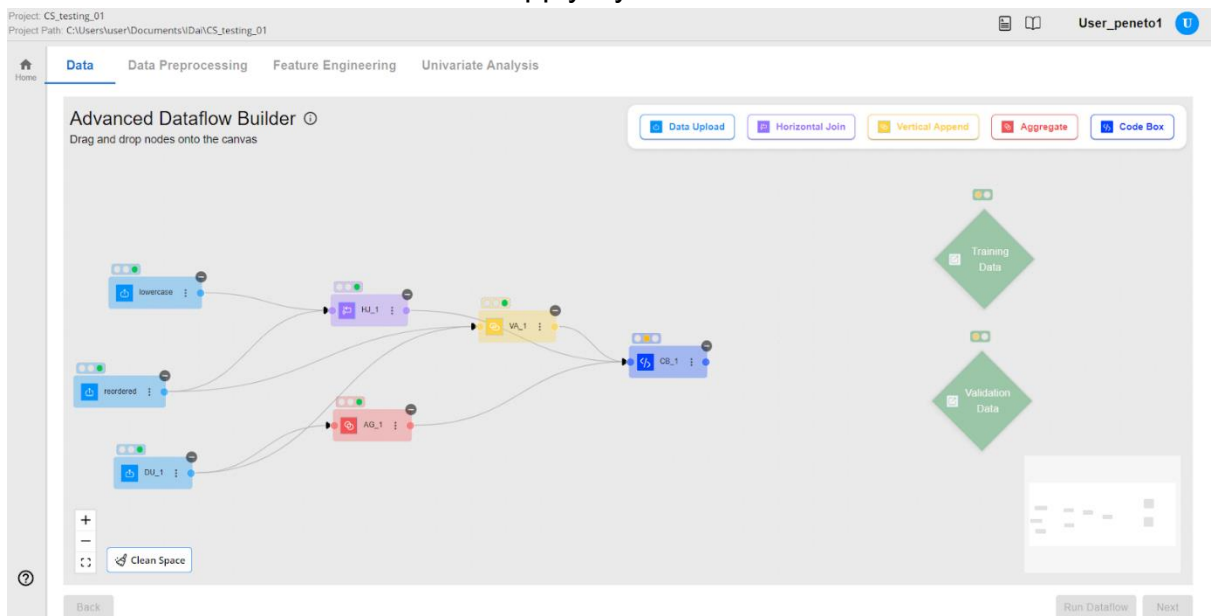
98	396	1	2	151	1
30	3,317	1	0	161	0
109	1,651	0	0	123	0
220	2,379	0	1	209	0
185	3,149	1	0	69	0
297	8,831	0	3	277	0
602	494	0	2	143	0
62	1,000	0	1	-984	0
16	917	0	2	-984	0
181	3,297	0	0	156	0

Back

Run Dataflow

Next

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.

The screenshot displays the CORESTRAT Advanced Dataflow Builder interface. The main canvas shows a workflow with several nodes: 'lowercase', 'reordered', 'DU_1', 'HJ_1', 'AG_1', 'VA_1', and 'CB_1'. The workflow is connected by arrows indicating data flow. A 'Python code editor' window is open, showing a function template for processing data. The code includes imports for pandas, numpy, and math, and a function definition for 'FUNCTION(df)' that returns a dataframe 'df'.

Python code editor

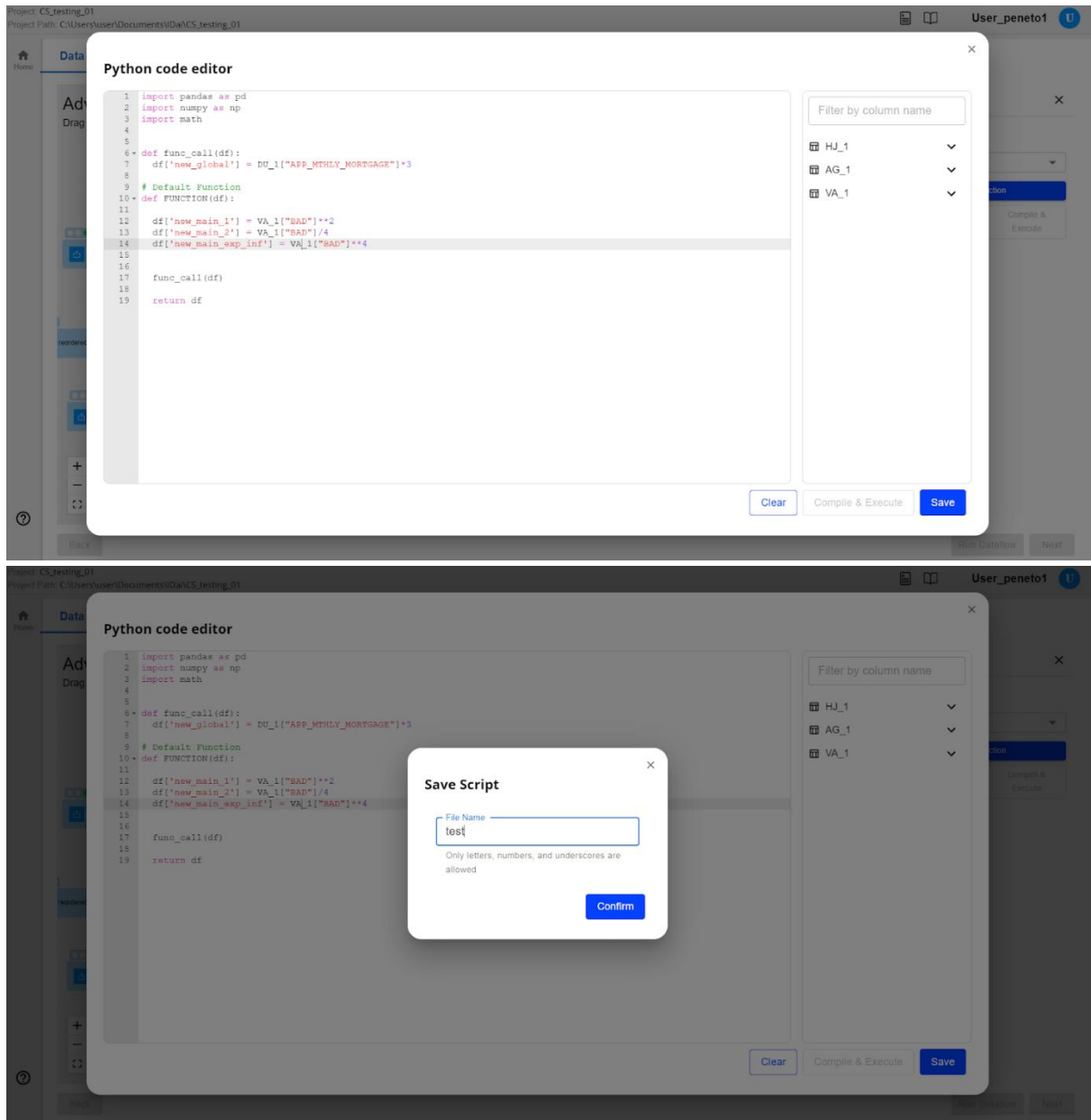
```

1 # Default Imports
2 # Only the following three imports are allowed for use within the program.
3 import pandas as pd
4 import numpy as np
5 import math
6
7 # Default Function
8 def FUNCTION(df):
9     # Please update code to be used here
10    # Output Dataframe variable should always be labelled as 'df'
11
12
13
14
15    return df

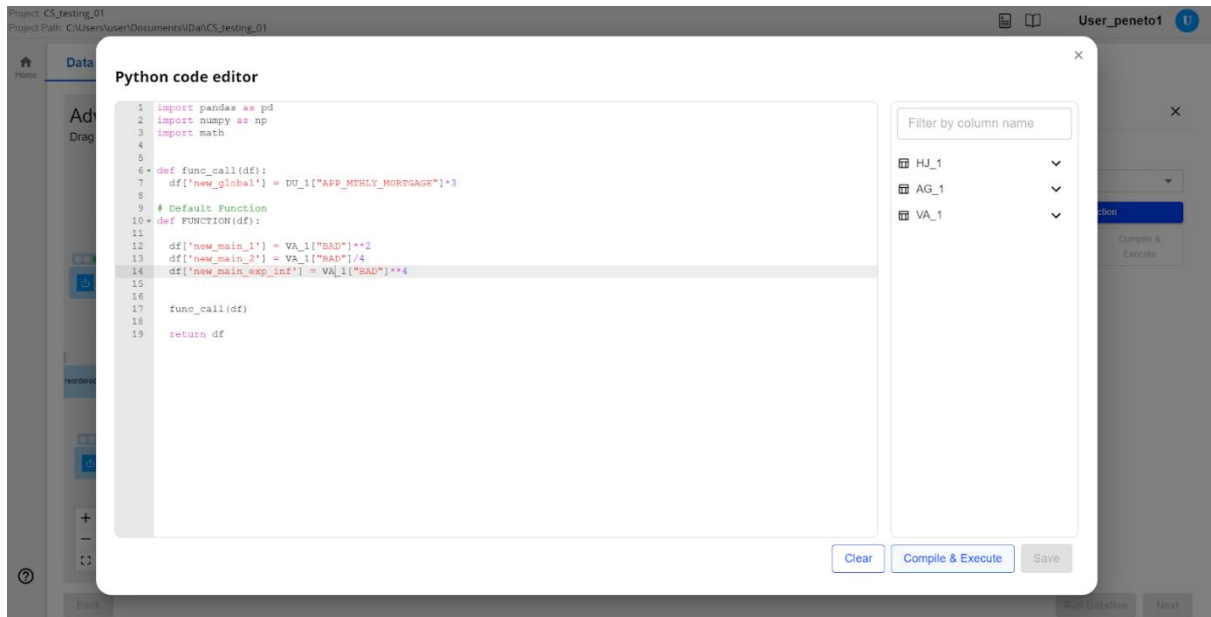
```

Buttons at the bottom of the code editor: Clear, Compile & Execute, Save.

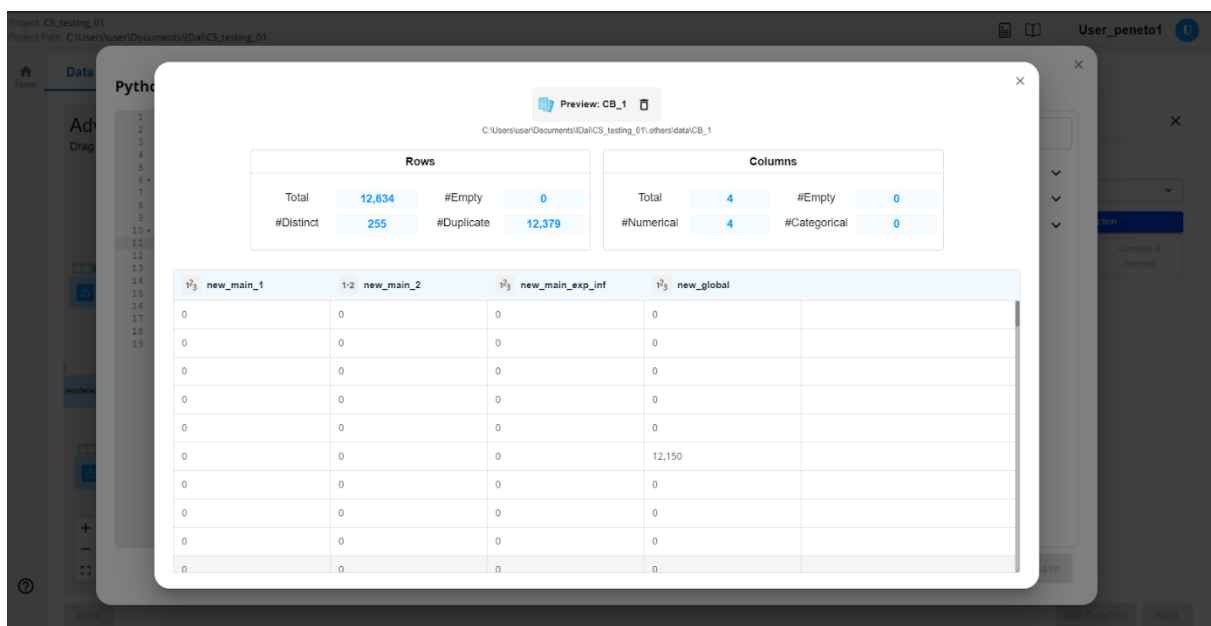
Once the desired code is entered, the function can be saved with an appropriate name for reuse within the workflow.



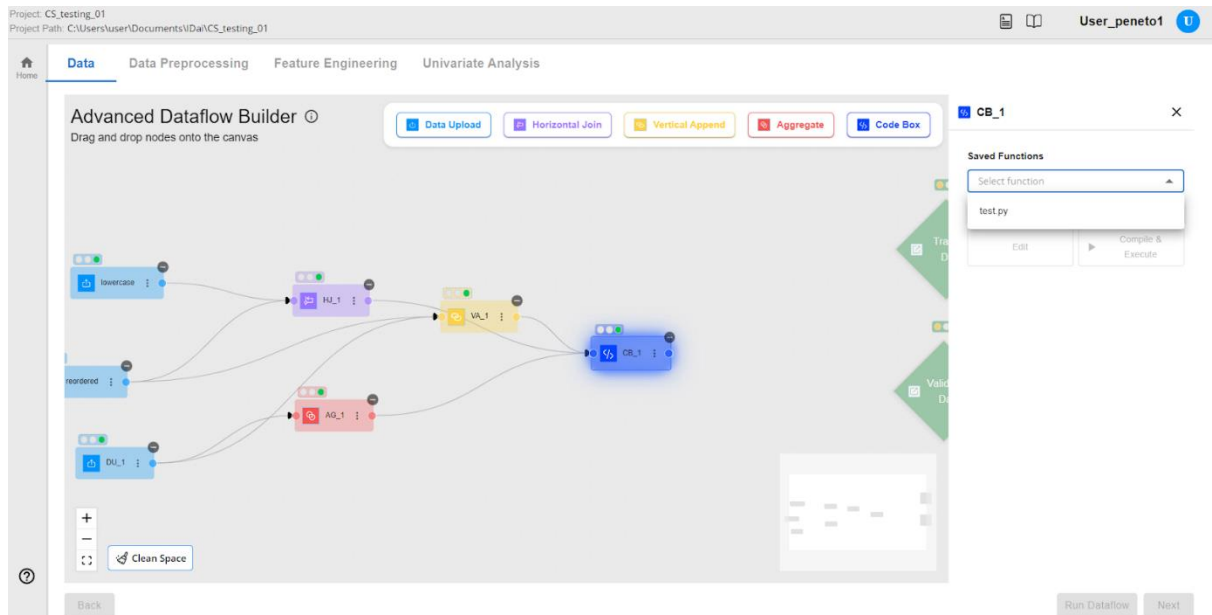
After saving the function, the user can compile and execute the code.



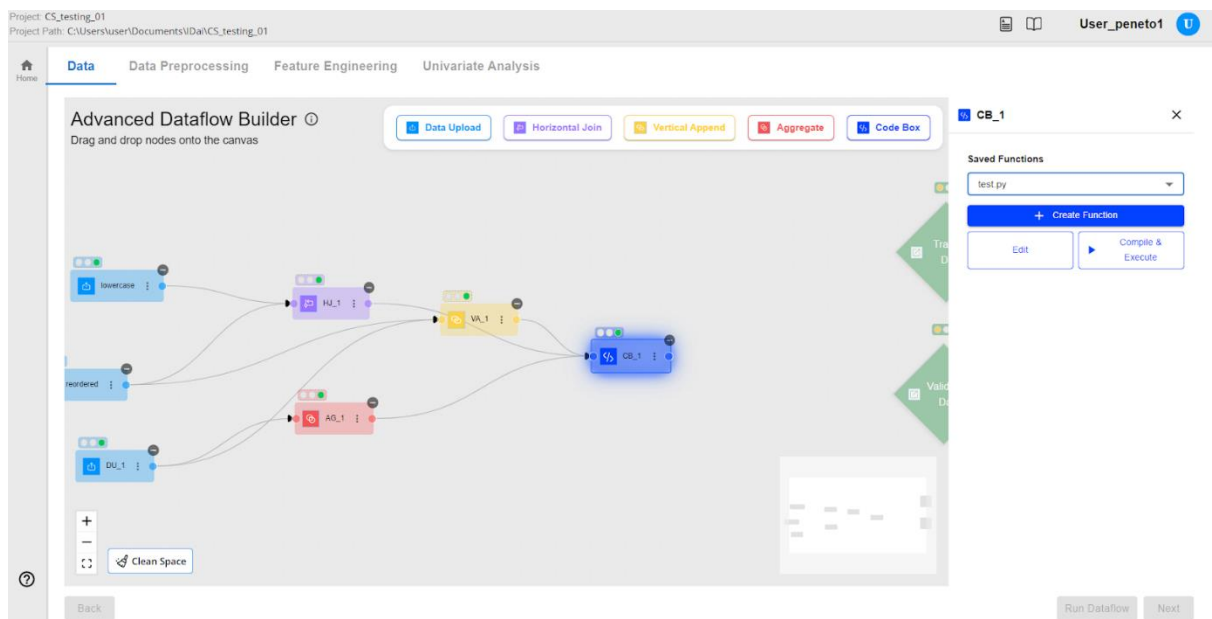
If the code compiles and executes successfully (without errors), an output screen will display the changes made to the dataset.



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



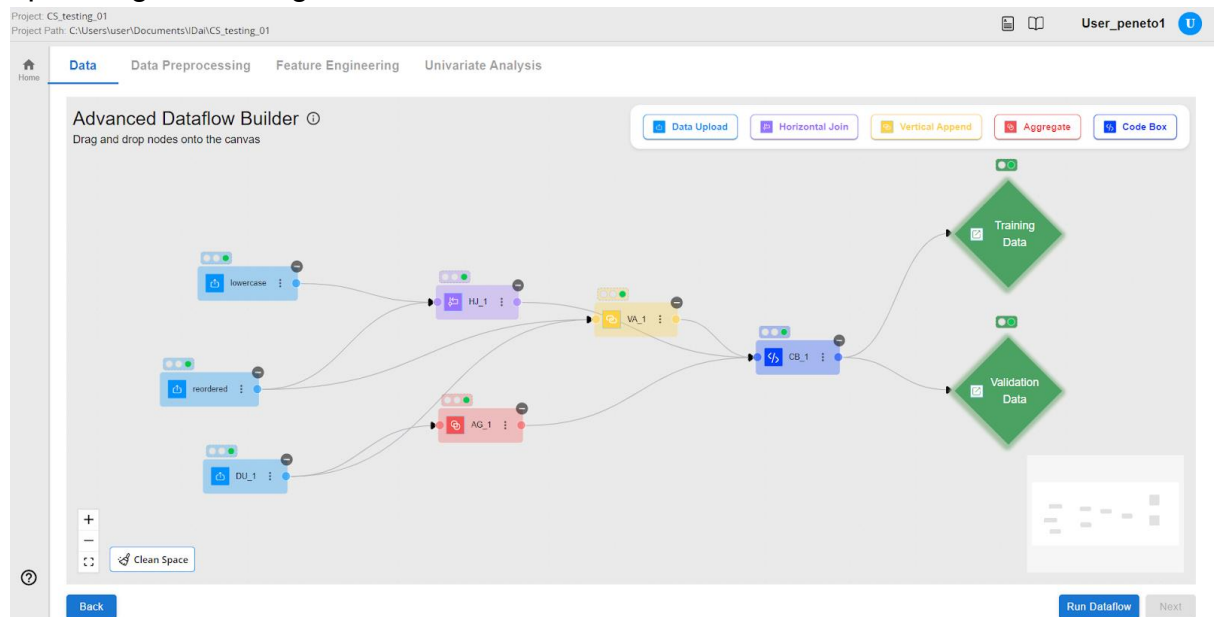
After selecting the desired function, the user can either edit it or click **Compile and Execute** to run it.



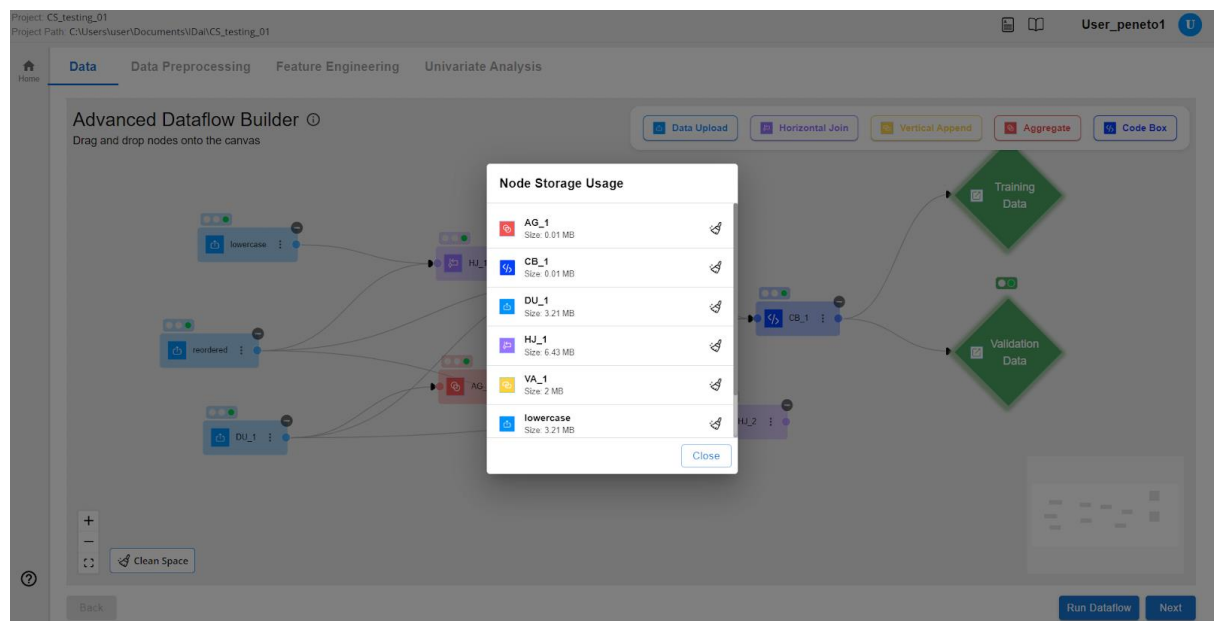
Once the desired workflow is complete, the user can connect its final node to the **Training Data** and **Validation Data** components.

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 screenshot shows the 'Data Engineering' step in the CORESTRAT interface. The left sidebar contains a navigation menu with steps 1 through 6. The main area displays 'Data' statistics and a data table.

Data Statistics:

Rows				Columns			
Total	12,634	#Empty	0	Total	4	#Empty	0
#Distinct	255	#Duplicate	12,379	#Numerical	4	#Categorical	0

Data Table:

	new_main_1	new_main_2	new_main_exp_inf	new_global
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	12,150	
0	0	0	0	
0	0	0	0	

Rows per page: 100 | 1-100 of 1000

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.

The screenshot shows the 'Data Upload' step in the CORESTRAT interface. A dialog box titled 'Validation Dataset Detected' is displayed, asking if the user wants to use an existing validation dataset. The background shows the 'Upload New Data Set' form with fields for 'Add Data Source' and a 'Test Connection' button.

Validation Dataset Detected Dialog:

A validation dataset was previously created using the Advanced Dataflow Builder.
Would you like to use the existing dataset?

Buttons: Cancel, Yes

Background Form: Upload New Data Set

Buttons: Use existing dataset from ADB, Test Connection

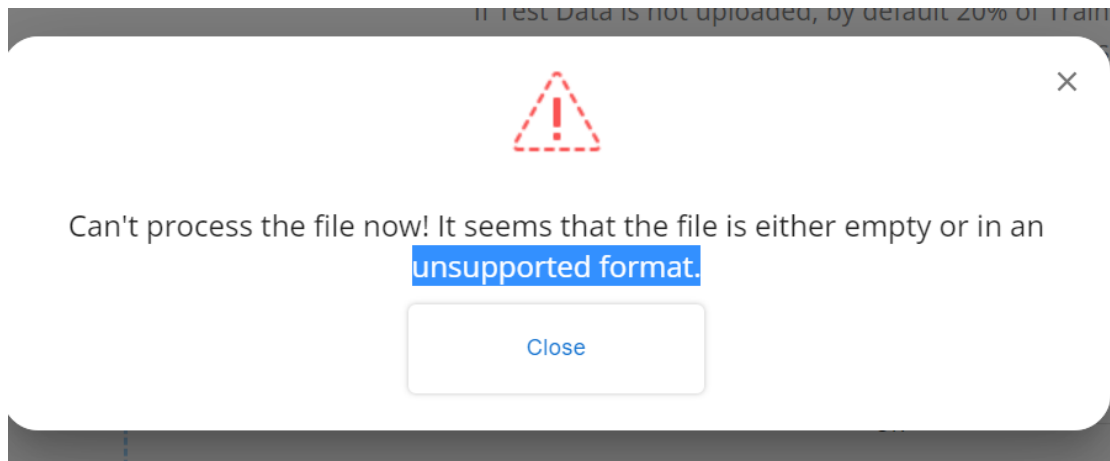
If user selects '**Simple Upload**' they are directed to this page

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.

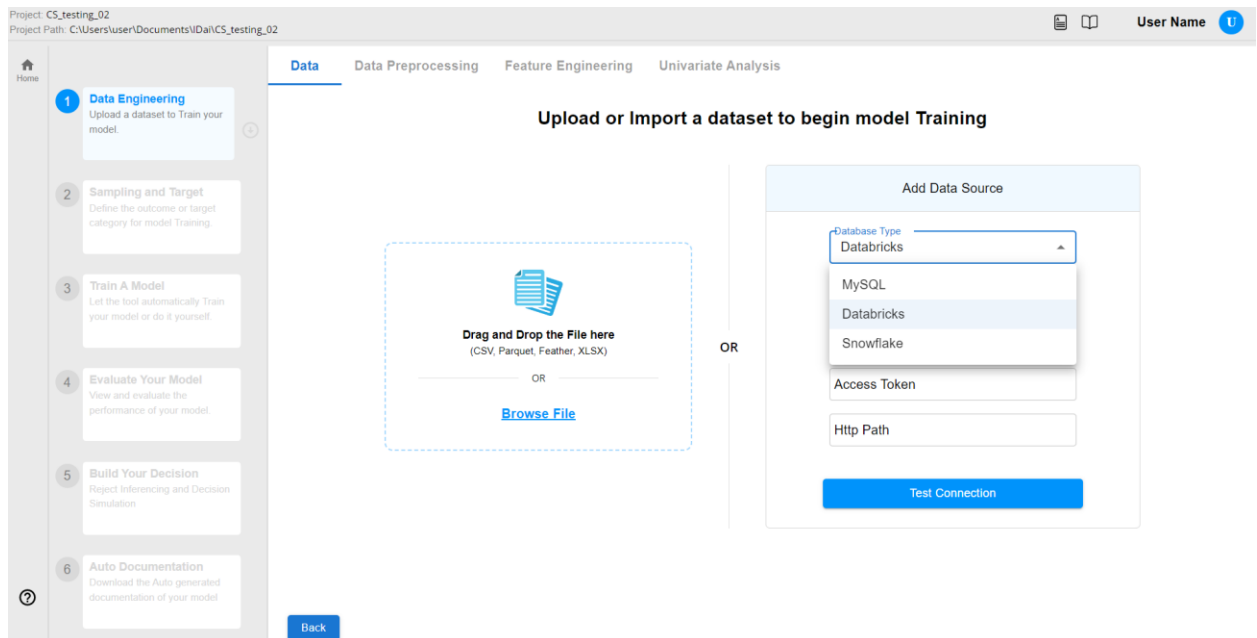


User needs to click on “Close” in the pop-up and go back and upload a dataset in one of the correct formats.

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**.

The screenshot shows the 'Data Engineering' step in the CoreStrat workflow. The main panel displays the 'Add Data Source' form. The form fields are filled with the following values:

- Database Type: MySQL
- Connection Name: MySQL
- Server: los-test-db.mysql.database.azure.com
- Login: csladmin@los-test-db
- Password: (masked with dots)
- Port: 3306
- Schema: (empty)

The 'Test Connection' button is highlighted in blue. A 'Back' button is visible at the bottom left of the main panel.

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'

The screenshot shows the same 'Add Data Source' form, but with an error message displayed below the 'Test Connection' button. The error message is:

⊗ Access denied for user 'csladmin'@'14.195.164.66' (using password: YES)

The 'Test Connection' button is still highlighted in blue.

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

The screenshot shows the 'Data Engineering' step in the workflow. The main heading is 'Upload or Import a dataset to begin model Training'. On the left, a sidebar lists six steps: 1. Data Engineering, 2. Sampling and Target, 3. Train A Model, 4. Evaluate Your Model, 5. Build Your Decision, and 6. Auto Documentation. The central area has a dashed box for 'Drag and Drop the File here' with a 'Browse File' link. To the right, the 'Add Data Source' panel is active, showing a successful connection to a MySQL database. The fields filled are: Connection Name (MySQL), Server (los-test-db.mysql.database.azure.com), Login (csladmin@los-test-db), Password (masked), Port (3306), and Schema (csl_dms_staging_db). A green checkmark indicates 'Connection is successful!'. There are 'CONNECT' and 'RESET' buttons at the bottom.

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

This screenshot shows the same interface after clicking 'Reset'. The 'Add Data Source' panel is now configured for Databricks. The fields are: Database Type (Databricks), Connection Name (Databricks), Host URL, Access Token, and Http Path. A 'Test Connection' button is at the bottom. The 'Drag and Drop' area and sidebar remain the same. A 'Back' button is visible at the bottom left of the main content area.

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.

Intelligent Decision AI
Project: Klop4
Project Path: C:\Users\HP\Documents\U...

Shubham

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.

Upload Data | Data Preprocessing | Feature Engineering | Data Univariate Analysis

Filter by database or table name

#Database: default
#Table: file_1

default
file_1
file_2
file_3

1 select * from default.file_1

Run Query

#Rows: 500,000 #Columns: 102 Source: Query - Custom SQL

col_1	col_2	col_3	col_4	col_5	col_6	col_7	col_8	col_9	col_10	col_11
0.99946...	0.06460...	0.90771...	0.21277...	0.40986...	0.53817...	0.30413...	0.21628...	0.38245...	0.13450...	0.4823...
0.93744...	0.10923...	0.77536...	0.84812...	0.12274...	0.40596...	0.71404...	0.74477...	0.50088...	0.81734...	0.5114...
0.62927...	0.13444...	0.85831...	0.14038...	0.47274...	0.39320...	0.98963...	0.24714...	0.82609...	0.01486...	0.0054...
0.85956...	0.80275...	0.34445...	0.53060...	0.14468...	0.59651...	0.65979...	0.29203...	0.89580...	0.19540...	0.3494...
0.46120...	0.64926...	0.72889...	0.39686...	0.21816...	0.67459...	0.59148...	0.00269...	0.73154...	0.65718...	0.9754...
0.73672...	0.00697...	0.65515...	0.65325...	0.48792...	0.58092...	0.05919...	0.86104...	0.46202...	0.17708...	0.0254...
0.64914...	0.90575...	0.05024...	0.54468...	0.48655...	0.11911...	0.78796...	0.96573...	0.16899...	0.17004...	0.6363...
0.11964...	0.86160...	0.81780...	0.86293...	0.52607...	0.32170...	0.46252...	0.28794...	0.43351...	0.33853...	0.7345...
0.64995...	0.93910...	0.02157...	0.80484...	0.00409...	0.34489...	0.72760...	0.93604...	0.77375...	0.34432...	0.6364...
0.26317...	0.89728...	0.08494...	0.27189...	0.85820...	0.99436...	0.58512...	0.35099...	0.75534...	0.59543...	0.3901...
0.89426...	0.71626...	0.52904...	0.37688...	0.96837...	0.81504...	0.51145...	0.22797...	0.27544...	0.91512...	0.8195...
0.30663...	0.72081...	0.16889...	0.67515...	0.72831...	0.56327...	0.22113...	0.44085...	0.72360...	0.04495...	0.6584...
0.59807...	0.22714...	0.45133...	0.34070...	0.77342...	0.66122...	0.50086...	0.55308...	0.64416...	0.04921...	0.3894...
0.09903...	0.32091...	0.76948...	0.37866...	0.54120...	0.12570...	0.06692...	0.83935...	0.55956...	0.42378...	0.0894...
0.10681...	0.81690...	0.74996...	0.05381...	0.21980...	0.40845...	0.52306...	0.10032...	0.73297...	0.77631...	0.5943...
0.39056...	0.73795...	0.83005...	0.90699...	0.04985...	0.60071...	0.25166...	0.48854...	0.65861...	0.54611...	0.6223...

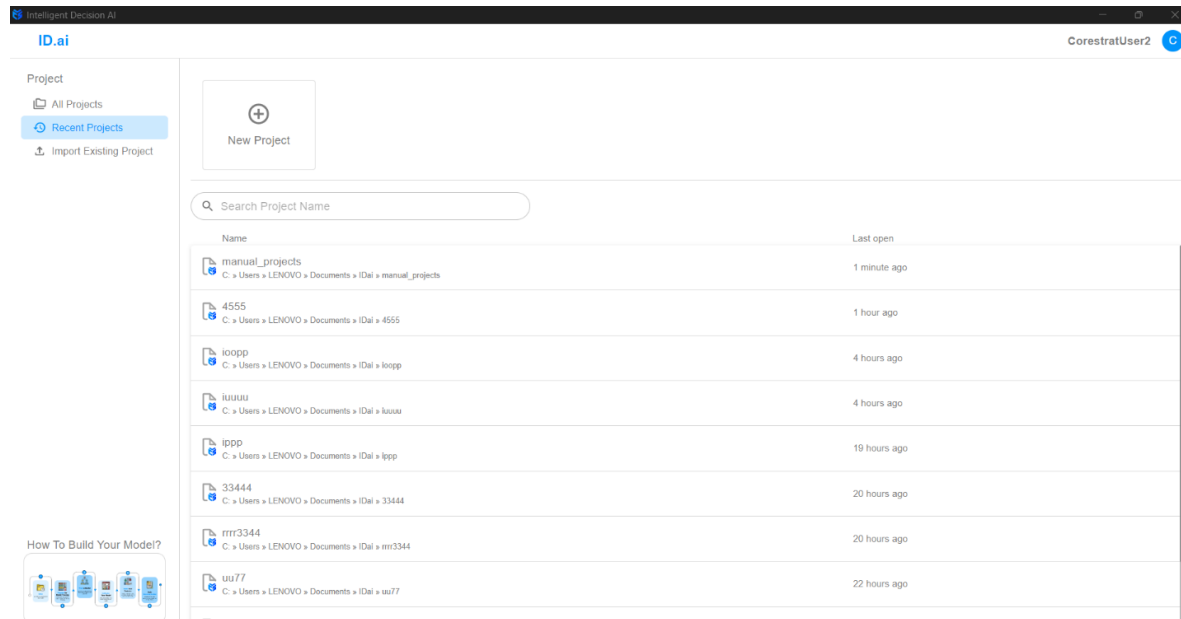
1-100 of 500000 < >

Back Next

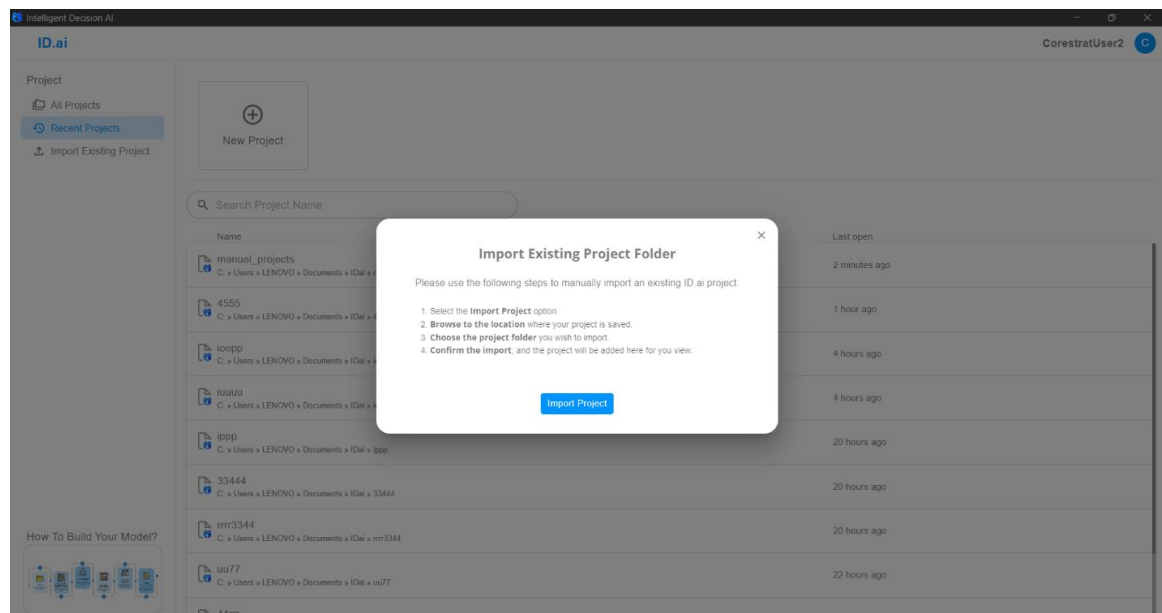
User needs to click on “Close” in the pop-up and go back and upload a dataset in one of the correct formats.

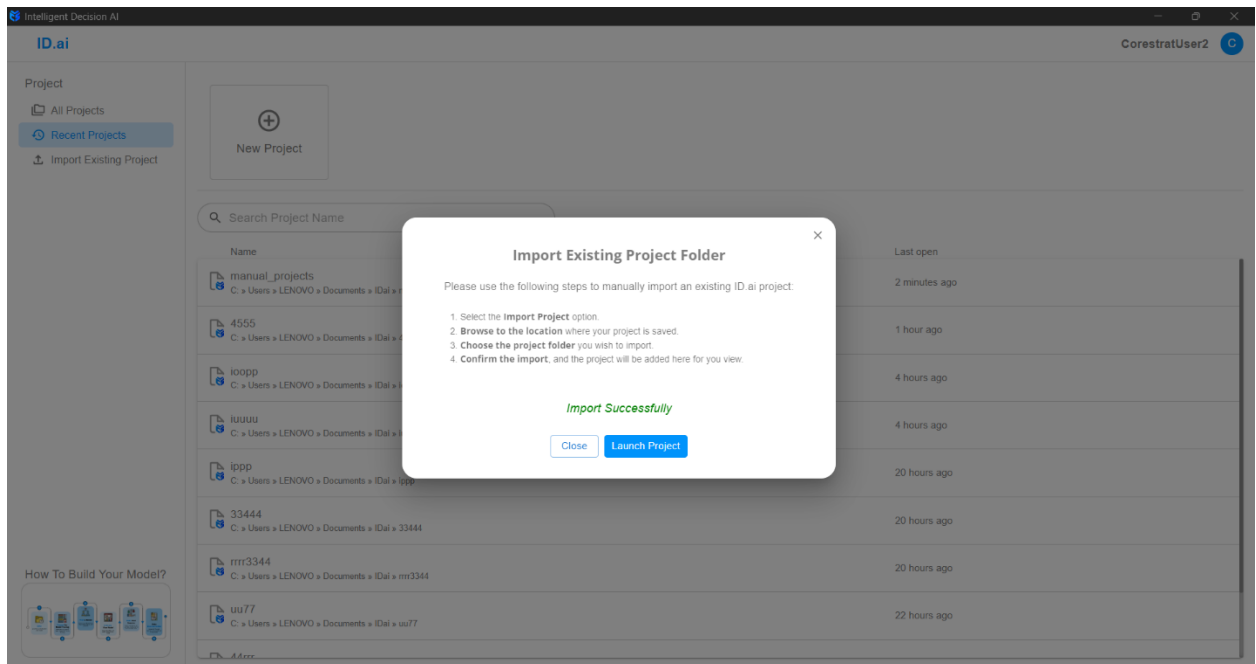
Existing Projects

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

Project: demo11
Project Path: C:\Users\LENOVO\Documents\...

Upload Data | Data Preprocessing | Feature Engineering | Data Univariate Analysis

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

train.csv

Rows				Columns			
Total	15,000	#Empty	0	Total	50	#Empty	0
#Distinct	15,000	#Duplicate	0	#Numerical	49	#Categorical	1

APP_PROD_CODE	APP_APPLICATION_KEY	APP_MTHLY_MORTGAGE	MONTHS_SINCE_OLDEST_TRADE	AVG_CURRENT_BALANCE	AVG_MONT
REGULAR	1	0	151	396	54
REGULAR	2	0	161	3,317	51
REGULAR	3	0	123	1,651	78
REGULAR	4	0	226	2,379	102
REGULAR	5	0	185	3,149	79
REGULAR	6	4,050	297	8,831	125
REGULAR	7	0	602	494	226
REGULAR	8	0	62	1,000	38


Rows per page: 100 | 1-100 of 1000

[Next](#)


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.

Project: CS_testing_02
Project Path: C:\Users\User\Documents\IDa\CS_testing_02



User Name 

Data Data Preprocessing Feature Engineering Univariate Analysis

train_EVERYTHING_1.feather 

Rows				Columns			
Total	15,006	#Empty	4	Total	119	#Empty	1
#Distinct	15,000	#Duplicate	2	#Numerical	113	#Categorical	4


APP_PROD_CODE	ALL_SINGLE	ALL_UNIQUE	EMPTY COL	DATE	TOP_3_FEATURE_AR
REGULAR	1	02b347d8155d73c15c328c9f...	--	28/5/2020	OTH_HME_AI
REGULAR	1	8af642af14ac909a353973e4b...	--	24/3/2023	OTH_HME_AI
REGULAR	1	940dfbcea633bcf78b120bf93...	--	25/2/2020	GSL_CV_GL
REGULAR	1	96a4d69280f3a255628964f2b...	--	31/8/2023	OTH_HME_AI
REGULAR	1	96a4d69280f3a255628964f2b...	--	31/8/2023	OTH_HME_AI
REGULAR	1	9de2c83009ca37a21684dd14...	--	13/12/2023	CV_GSL_DB
REGULAR	1	a3d114cdd72c7b7f3655b42d...	--	6/4/2023	CV_GSL_GSE
REGULAR	1	a44ecb7c6176e1b9bb89715f...	--	2/1/2019	SH_OTH_HME

Rows per page: 100 1-100 of 999  


Next

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**'.

Project: CS_testing_02
Project Path: C:\Users\User\Documents\IDa\CS_testing_02


User Name 

Data Data Preprocessing Feature Engineering Univariate Analysis



Let AI Do It

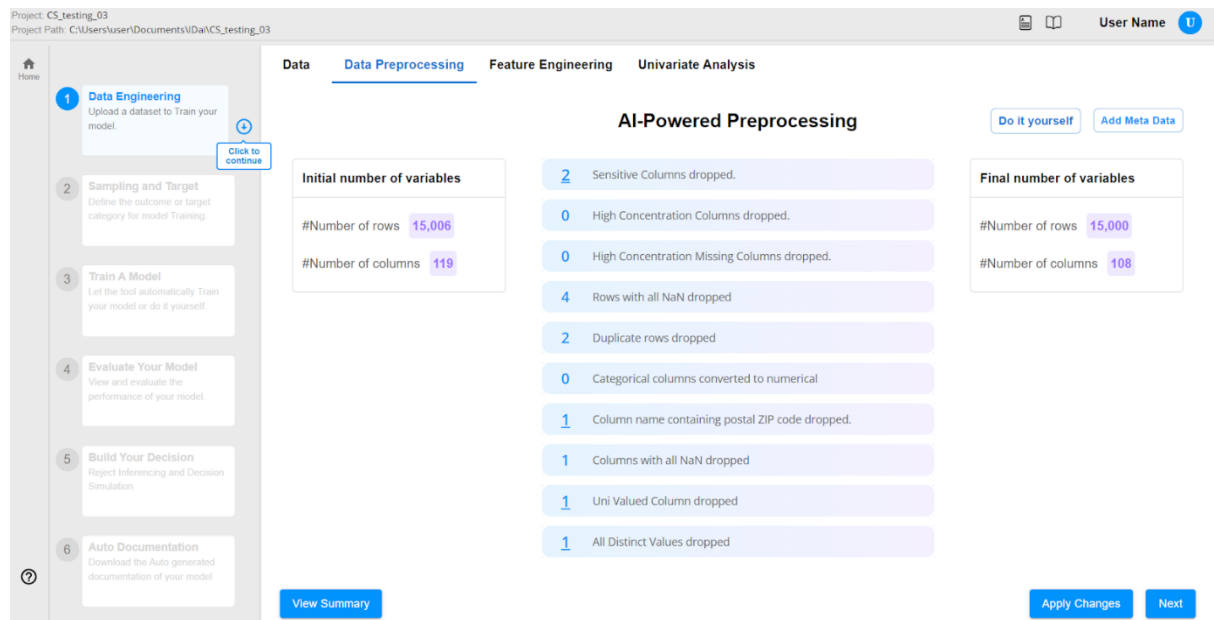
Our advanced AI will automatically preprocess your data. Use this intuitive tool to clean, transform, and analyze your data.



Do It Yourself

Greater control and accuracy, but time-consuming and requires expertise

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | **Data Preprocessing** | Feature Engineering | Univariate Analysis

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.

Rows and Columns | Column Data Type | Variable Treatment

Duplicate Rows 2
☒ Keep unique Rows

Rows with all NaN 4
☒ Drop all NaN rows

Columns with all NaN 0
☒ Drop all NaN Columns

Columns name contains postal ZIP code 1
☒ Drop all ZIP columns

Date Time Column Drop
DATE ☒
D_SNAPSHOT ☐
AVG_CURRENT_BALANCE ☒

Uni Valued Column Drop
ALL_SINGLE ☒

All Distinct Values Drop
ALL_UNIQUE ☒

Duplicate Columns Retain
MTHS_SNC_LAST_30P_DAYS ☒
MTHS_SNC_LAST_DLQ ☐
NBR_TRDLNS_30P_DPD_EVER ☒
NBR_TRDLNS_30P_DPD_EVER_1 ☐
TTL_BALANCE_ON_USUAL_REVOLV... ☒
TTL_REVOLVING_BALANCES ☐

[Let AI do it](#) [Add Meta Data](#)

[View Summary](#) [Apply Changes](#) [Next](#)

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | **Data** | Data Preprocessing | Feature Engineering | Univariate Analysis

Rows and Columns | **Column Data Type** | Variable Treatment

Let AI do it | Add Meta Data

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.

Numerical Columns (Convert to Categorical)

Columns with ≤20 unique values

All ☐

☒ ALL_SINGLE ☒

EMPTY COL ☒

1-2 EFFICIENT_MAX_DLQ_023_MONTHS__REVOLVING ☒

1-2 EFFICIENT_MAX_DLQ_EVER ☒

Remaining Columns

☒ APP_APPLICATION_KEY

☒ DOBER

☒ APP_MTHLY_INCOME

☒ APP_MTHLY_MORTGAGE

1-2 MONTHS_SINCE_OLDEST_TRADE

[View Summary](#)

Categorical Columns (Convert to numerical if required)

Potential Columns Identified

! None

Remaining Columns

☒ APP_PROD_CODE

☒ TOP_3_FEATURE_AREAS_V1

1-2 EFFICIENT_MAX_DLQ_023_MONTHS__REVOLVING

1-2 EFFICIENT_MAX_DLQ_EVER

1-2 NBR_BANK_OR_NATL_REVOLVING_TRDLNS_75PRC_OF_AMOUNT

[Apply Changes](#) [Next](#)

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | **Data Preprocessing** | Feature Engineering | Univariate Analysis

Rows and Columns | Column Data Type | **Variable Treatment**

Let AI do it | Add Meta Data

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

Sensitive Variables 2

Search Variable...

☒ Select All

☒ AGE

☒ GENDER

Single Value Concentration 37

Search Variable... Set Threshold 50 %

☐ Select All

TOO_NEW_OR_STALE_FILE_INDI...	97.27 %
BAD	95.50 %
APP_MTHLY_MORTGAGE	88.65 %
BANKRUPTCY_INDICATOR	88.54 %
NBR_TRDLNS_60P_DPD_023_MO...	85.85 %
NBR_BANK_OR_NATL_REVOLVIN...	85.55 %
NBR_RETAIL_TRDLNS_OVERLIMIT	85.27 %
NBR_TRDLNS_OPENED_023_MO...	85.03 %
NBR_ADVERSE_PUBLIC_RECORDS	84.29 %
MTHS_SNC_MOST_RECENT_ADV...	84.29 %

Missing Value Concentration 0

Search Variable... Set Threshold 10 %

No Variable Found!

[View Summary](#) [Apply Changes](#) [Next](#)

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | **Data Preprocessing** | Feature Engineering | Univariate Analysis

Rows and Columns | Column Data Type | **Variable Treatment**

Let AI do it | Add Meta Data

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

Sensitive Variables 2

Search Variable...

☒ Select All

☒ AGE

☒ GENDER

Single Value Concentration 37

Search Variable... Set Threshold 50 %

☐ Select All

TOO_NEW_OR_STALE_FILE_INDI...	97.27 %
BAD	95.50 %
APP_MTHLY_MORTGAGE	88.65 %
BANKRUPTCY_INDICATOR	88.54 %
NBR_TRDLNS_60P_DPD_023_MO...	85.85 %
NBR_BANK_OR_NATL_REVOLVIN...	85.55 %
NBR_RETAIL_TRDLNS_OVERLIMIT	85.27 %
NBR_TRDLNS_OPENED_023_MO...	85.03 %
NBR_ADVERSE_PUBLIC_RECORDS	84.29 %
MTHS_SNC_MOST_RECENT_ADV...	84.29 %

Missing Value Concentration 0

Search Variable... Set Threshold 10 %

No Variable Found!

[View Summary](#) [Apply Changes](#) [Next](#)

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDah\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | Category Encoding

Code-It-Yourself

Create and modify features using custom Python code.

```

1 # Default Imports
2 # Only the following three imports
3 # are allowed for use within the program.
4 import pandas as pd
5 import numpy as np
6 import math
7
8 # Default Function
9 def FUNCTION(df):
10     # Please update code to be used here
11     # Code will be applied to both train and test dataset
12     # Dataframe variable should always be labelled as 'df'
13
14     df['new1'] = df['APP_MTHLY_INCOME']*20
15
16
17     return df

```

Columns in your Data

Search Variable

- AMT_CR_LIMIT
- APP_APPLICATION_KEY
- APP_MTHLY_INCOME
- APP_MTHLY_MORTGAGE
- APP_PROD_CODE
- AVG_CURRENT_BALANCE
- AVG_MONTHS_IN_FILE
- AVG_MONTHS_IN_FILE_R...
- AVG_RETAIL_BALANCE

List of Engineered Features

Next →

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 Engineered Features

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDah\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | Category Encoding

Code-It-Yourself

Create and modify features using custom Python code.

```

1 # Default Imports
2 # Only the following three imports
3 # are allowed for use within the program.
4 import pandas as pd
5 import numpy as np
6 import math
7
8 # Default Function
9 def FUNCTION(df):
10     # Please update code to be used here
11     # Code will be applied to both train and test dataset
12     # Dataframe variable should always be labelled as 'df'
13
14     df['new1'] = df['APP_MTHLY_INCOME']*20
15
16
17     return df

```

Columns in your Data

Search Variable

- AMT_CR_LIMIT
- APP_APPLICATION_KEY
- APP_MTHLY_INCOME
- APP_MTHLY_MORTGAGE
- APP_PROD_CODE
- AVG_CURRENT_BALANCE
- AVG_MONTHS_IN_FILE
- AVG_MONTHS_IN_FILE_R...
- AVG_RETAIL_BALANCE

List of Engineered Features

1. new1

Compiled Successfully

Compile Execute

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | Category Encoding

Code-It-Yourself ⓘ
Create and modify features using custom Python code.

```

1 # Default Imports
2 # Only the following three imports
3 # are allowed for use within the program.
4 import pandas as pd
5 import numpy as np
6 import math
7
8 # Default Function
9 def FUNCTION(df):
10     # Please update code to be used here
11     # Code will be applied to both train and test dataset
12     # Dataframe variable should always be labelled as 'df'
13
14     df['new1'] = df['APP_MONTHLY_INCOME']*20
15
16
17     return df

```

Columns in your Data ⓘ

Search Variable

- APP_AMT_CR_LIMIT
- APP_APPLICATION_KEY
- APP_MONTHLY_INCOME
- APP_MONTHLY_MORTGAGE
- APP_PROD_CODE
- AVG_CURRENT_BALANCE
- AVG_MONTHS_IN_FILE
- AVG_MONTHS_IN_FILE_R
- AVG_RETAIL_BALANCE

List of Engineered Features

- new1

Code Executed Successfully

Compile Execute

Next →

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | **Two-Way Interaction** | Variable Scaling and Transformation | Category Encoding

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.

List of Categorical Variables

Search Variable

- ☒ APP_PROD_CODE
- ☒ EFFICIENT_MAX_DLQ_023_MO...
- ☒ EFFICIENT_MAX_DLQ_EVER
- ☐ NBR_BANK_OR_NATL_REVOLV...

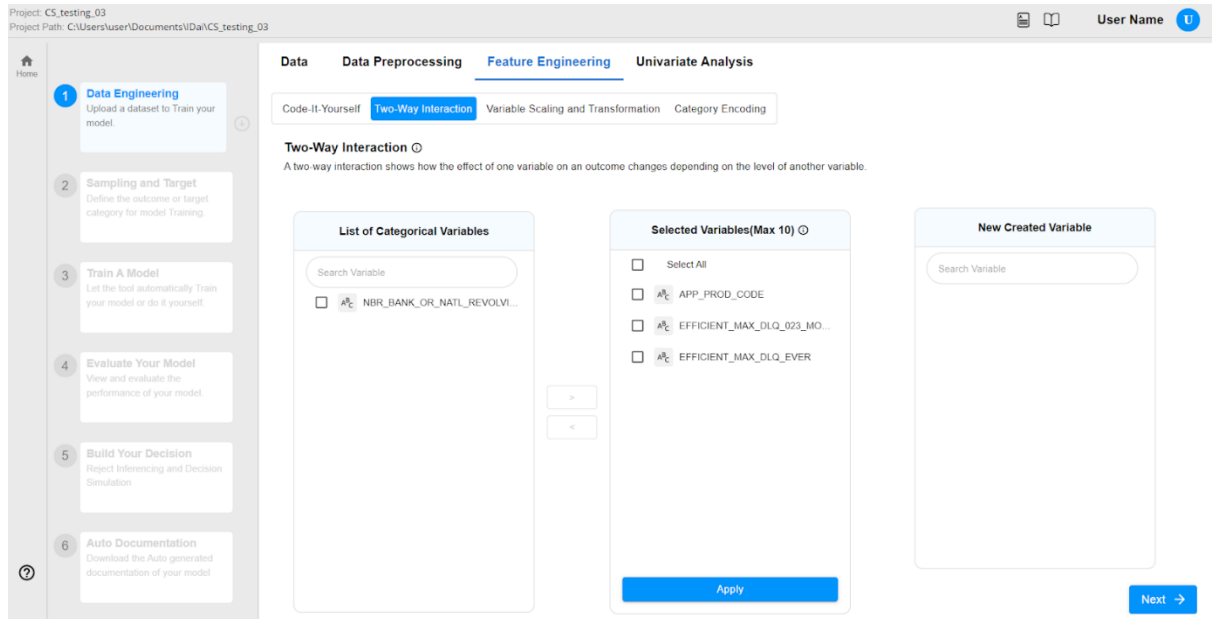
> <

Selected Variables(Max 10)

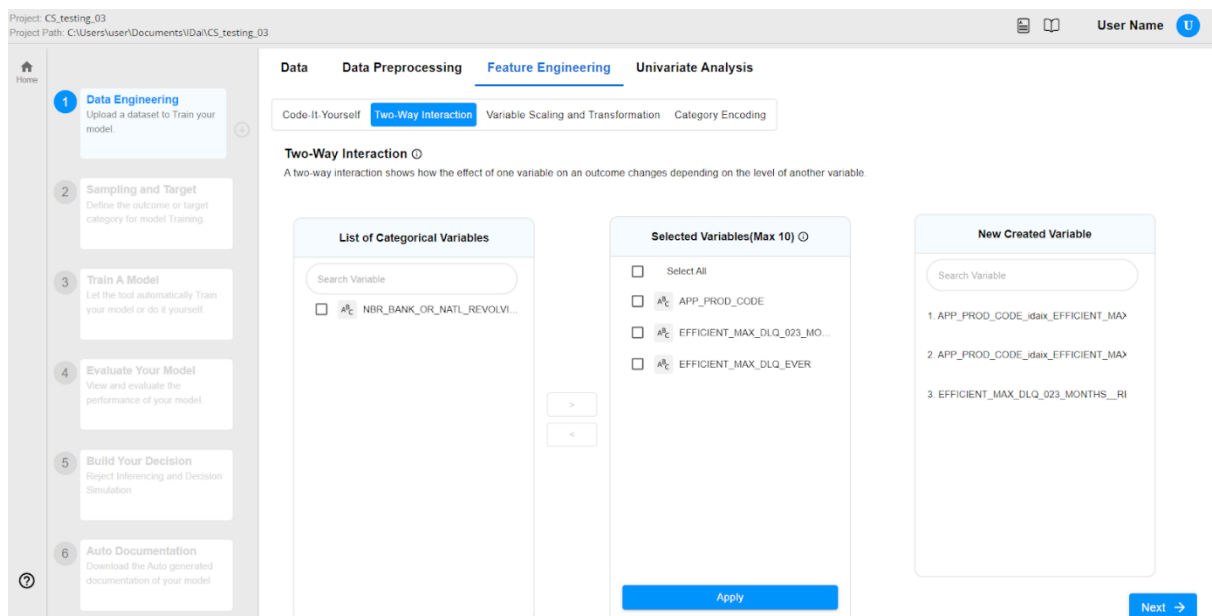
New Created Variable

Search Variable

Next →



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**’

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | 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



Let AI Do It

Our advanced AI will automatically select the transformation technique that minimizes variability by reducing standard deviation



Do it Yourself

Greater control and accuracy, but time-consuming and requires expertise

Next →

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

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | 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

Do it yourself

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	Auto	Loading..	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATI_RE...	45.42	Auto	Loading..	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATI_RE...	37.88	Auto	Loading..	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Auto	Loading..	<input type="checkbox"/>
2-3 APP_MTHLY_MORTGAGE	15.59	Auto	Loading..	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Auto	Loading..	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Auto	Loading..	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Auto	Loading..	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Auto	Loading..	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Auto	Loading..	<input type="checkbox"/>

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\Da\CS_testing_03

Home | 1 Data Engineering | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	None	--	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	None	--	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.59	None	--	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	None	--	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	None	--	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	None	--	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	None	--	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	None	--	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104 | < >

Let AI do it

After Transformation

#Variables: 104
#Transformed Variables: 0

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 0
#Medium: 1 0
#High: 102 0

Next →

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Home | 1 Data Engineering | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	Log	Loading	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	None	--	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	None	--	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.59	None	--	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	None	--	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	None	--	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	None	--	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	None	--	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	None	--	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104 | < >

Let AI do it

After Transformation

#Variables: 104
#Transformed Variables: 0

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 0
#Medium: 1 0
#High: 102 0

Next →

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

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Home | User Name | U

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	Auto	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	Log	--	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Square Root	--	<input type="checkbox"/>
2-3 APP_MTHLY_MORTGAGE	15.59	Standardization	--	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Min-Max Scaler	--	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Power (Box Cox)	--	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Power (Yeo Johnson)	--	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Robust 95-5	--	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Quantile (Normal)	--	<input type="checkbox"/>
		Quantile (Uniform)	--	<input type="checkbox"/>

100 1-100 of 104

After Transformation

#Variables: 104
#Transformed Variables: 0

Relative Variability Profile ⓘ
Coefficient of Variation (CV)

	#Low	#Medium	#High
1	1	1	102
0	0	0	0

Let AI do it

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	None	--	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	None	--	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	None	--	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.58	None	--	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	None	--	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	None	--	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	None	--	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	None	--	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	None	--	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104

Let AI do it

After Transformation

#Variables: 104
#Transformed Variables: 0

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 (green), 0 (green)
#Medium: 1 (yellow), 0 (yellow)
#High: 102 (orange), 0 (orange)

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_60P_DP...	278	Auto	Loading..	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	Auto	Loading..	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	Auto	Loading..	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Auto	Loading..	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.58	Auto	Loading..	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Auto	Loading..	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Auto	Loading..	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Auto	Loading..	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Auto	Loading..	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Auto	Loading..	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104

Let AI do it

After Transformation

#Variables: 104
#Transformed Variables: 104

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 (green), 63 (green)
#Medium: 1 (yellow), 13 (yellow)
#High: 102 (orange), 28 (orange)

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Home | 1 Data Engineering | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_90P_DP...	278	Log (Auto)	0.01	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	Log (Auto)	0.08	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	Log (Auto)	0.08	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Log (Auto)	0.01	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.59	Power (Box Cox) (Auto)	0.21	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Log (Auto)	0.04	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Log (Auto)	0.07	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Log (Auto)	0.01	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104 | < >

Let AI do It

After Transformation

#Variables: 104
#Transformed Variables: 104

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 (70)
#Medium: 1 (2)
#High: 102 (32)

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Home | 1 Data Engineering | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_90P_DP...	278	Log (Auto)	0.01	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	Log (Auto)	0.08	<input checked="" type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	Log (Auto)	0.08	<input type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Log (Auto)	0.01	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.59	Power (Box Cox) (Auto)	0.21	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Log (Auto)	0.04	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Log (Auto)	0.07	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Log (Auto)	0.01	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104 | < >

Let AI do It

After Transformation

#Variables: 104
#Transformed Variables: 104

Relative Variability Profile
Coefficient of Variation (CV)

#Low: 1 (70)
#Medium: 1 (2)
#High: 102 (32)

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | 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

Variables	CV (Before)	Transformation Type	CV (After)	Retain Original Variable
1-2 NBR_REVL_OR_OPEN_TRDLNS_90P_DP...	278	Log (Auto)	0.01	<input checked="" type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	45.42	Log (Auto)	0.08	<input type="checkbox"/>
1-2 HIGHEST_UTIL_ON_BANK_OR_NATL_RE...	37.88	Log (Auto)	0.08	<input checked="" type="checkbox"/>
1-2 AVG_CURRENT_BALANCE	17.14	Log (Auto)	0.01	<input type="checkbox"/>
1-2 APP_MTHLY_MORTGAGE	15.58	Power (Box Cox) (Auto)	0.21	<input type="checkbox"/>
1-2 PRC_RETAIL_TRDLNS_WITH_BALANCE...	15.04	Log (Auto)	0.04	<input type="checkbox"/>
1-2 NBR_REVL_OR_OPN_TRDLNS_50PRC...	12.65	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NBR_REVOLVING_TRDLNS_50PRC_OF...	12.48	Log (Auto)	0.03	<input type="checkbox"/>
1-2 NET_FRACTION_REVOLVING_BURDEN...	11.81	Log (Auto)	0.07	<input type="checkbox"/>
1-2 TTL_REVOLVING_OR_OPEN_BALANCES	11.31	Log (Auto)	0.01	<input type="checkbox"/>

Rows per page: 100 | 1-100 of 104 | < >

Let AI do It

After Transformation

#Variables: 107
#Transformed Variables: 104

Relative Variability Profile ⓘ
Coefficient of Variation (CV)

#Low: 1 (70)
#Medium: 1 (2)
#High: 102 (32)

Next →

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

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | **Category Encoding**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	2	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	None	0	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	12	None	0	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Next →

Before Encoding

#Variables: 4

After Encoding

#Variables: 4
#Created Variables: 0

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | **Category Encoding**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	One Hot Encoding	0	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	12	Frequency Encoding	0	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Next →

Before Encoding

#Variables: 4

After Encoding

#Variables: 4
#Created Variables: 0

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | **Category Encoding**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	One Hot Encoding	Leading	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	None	0	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	12	None	0	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Next →

Before Encoding

#Variables: 4

After Encoding

#Variables: 4
#Created Variables: 0

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | **Category Encoding**

Category Encoding

Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	None	0	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	17	None	0	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4

Next →

Before Encoding

#Variables: 4

After Encoding

#Variables: 4

#Created Variables: 0

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDA\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

1 Upload a dataset to Train your model.
2 Define the outcome or target category for model Training.
3 Let the tool automatically Train your model or do it yourself.
4 View and evaluate the performance of your model.
5 Reject Inferencing and Decision Simulation.
6 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**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	None	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	One Hot Encoding	0	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	Frequency Encoding	0	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	12	None	0	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Next →

Before Encoding

#Variables 4

After Encoding

#Variables 4

#Created Variables 0

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDA\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

1 Upload a dataset to Train your model.
2 Define the outcome or target category for model Training.
3 Let the tool automatically Train your model or do it yourself.
4 View and evaluate the performance of your model.
5 Reject Inferencing and Decision Simulation.
6 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**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	One Hot Encoding	Loading..	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	One Hot Encoding	Loading..	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	One Hot Encoding	Loading..	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	12	One Hot Encoding	Loading..	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Next →

Before Encoding

#Variables 4

After Encoding

#Variables 4

#Created Variables 0

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	One Hot Encoding	3	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	One Hot Encoding	13	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	One Hot Encoding	11	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	17	One Hot Encoding	17	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Before Encoding

#Variables: 4

After Encoding

#Variables: 44
#Created Variables: 44

Next →

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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**

Category Encoding ⓘ
Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables ⓘ	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	One Hot Encoding	Loading...	<input checked="" type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS...	13	One Hot Encoding	13	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	One Hot Encoding	11	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	17	One Hot Encoding	17	<input type="checkbox"/>

Rows per page: 100 | 1-4 of 4 | < >

Before Encoding

#Variables: 4

After Encoding

#Variables: 44
#Created Variables: 44

Next →

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Code-It-Yourself | Two-Way Interaction | Variable Scaling and Transformation | **Category Encoding**

Category Encoding

Encoding is performed to represent categorical variables as binary vectors.

Search Variable

Candidate Variables	#Unique Categories	Encoding Type	#Binary Columns Created	Retain Original Variable
APP_PROD_CODE	3	One Hot Encoding	3	<input checked="" type="checkbox"/>
EFFICIENT_MAX_DLQ_023_MONTHS_...	13	One Hot Encoding	13	<input type="checkbox"/>
EFFICIENT_MAX_DLQ_EVER	11	One Hot Encoding	11	<input type="checkbox"/>
NBR_BANK_OR_NATL_REVOLVING_TR...	17	One Hot Encoding	17	<input type="checkbox"/>

Before Encoding

#Variables: 4

After Encoding

#Variables: 45
#Created Variables: 44

Rows per page: 100 | 1-4 of 4 | < | >

Next →

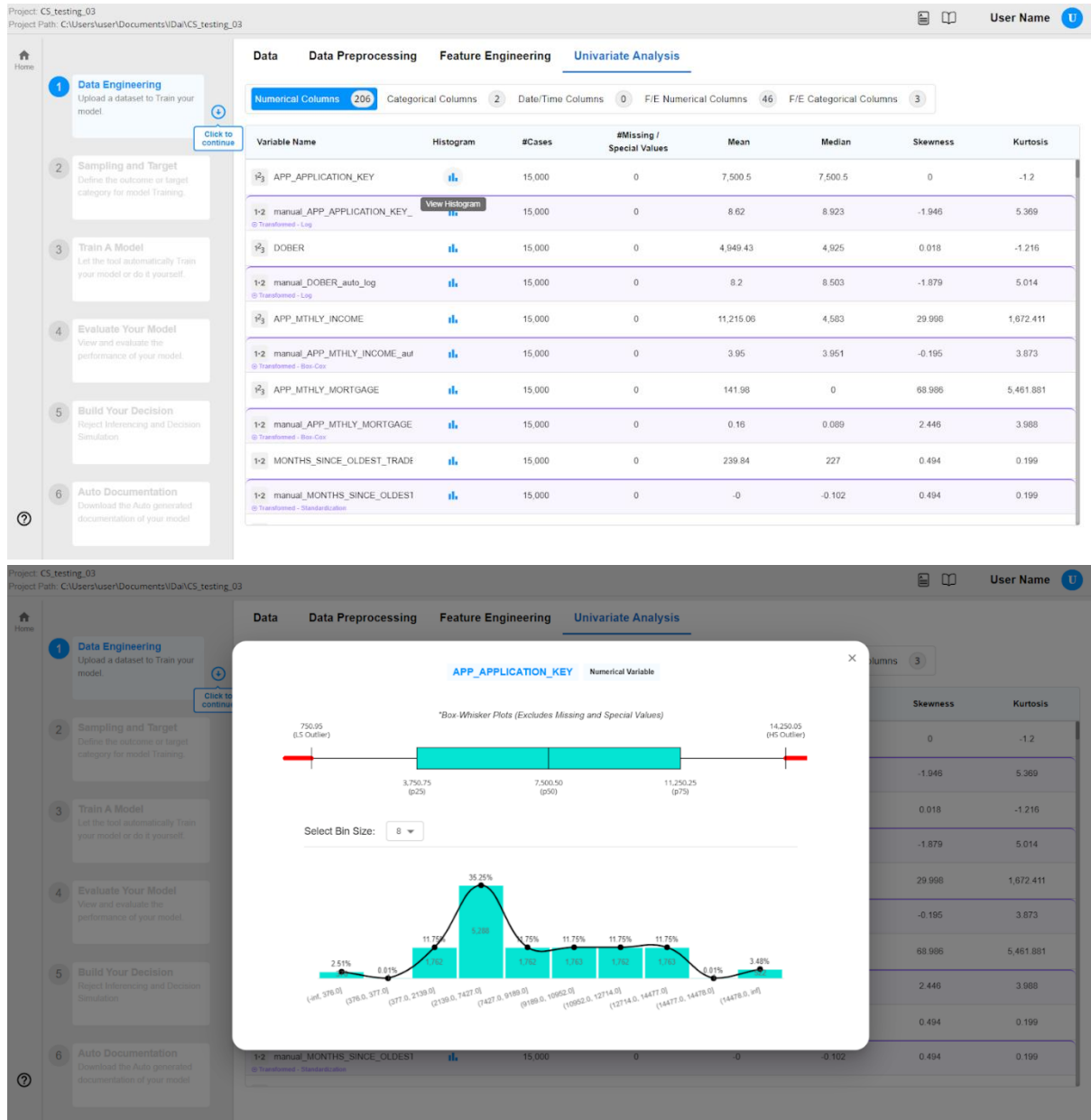
Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home | Data | Data Preprocessing | **Feature Engineering** | Univariate Analysis

Numerical Columns: 209 | Categorical Columns: 2 | Date/Time Columns: 0 | F/E Numerical Columns: 46 | F/E Categorical Columns: 3

Variable Name	Histogram	#Cases	#Missing / Special Values	Mean	Median	Skewness	Kurtosis
APP_APPLICATION_KEY		15,000	0	7,500.5	7,500.5	0	-1.2
manual_APP_APPLICATION_KEY_... <small>@ Transformed - Log</small>		15,000	0	8.62	8.923	-1.946	5.369
DOBER		15,000	0	4,949.43	4,925	0.018	-1.216
manual_DOBER_auto_log <small>@ Transformed - Log</small>		15,000	0	8.2	8.503	-1.879	5.014
APP_MTHLY_INCOME		15,000	0	11,215.06	4,583	29.996	1,672.411
manual_APP_MTHLY_INCOME_aut <small>@ Transformed - Box-Cox</small>		15,000	0	3.95	3.951	-0.195	3.873
APP_MTHLY_MORTGAGE		15,000	0	141.98	0	68.986	5,461.881
manual_APP_MTHLY_MORTGAGE <small>@ Transformed - Box-Cox</small>		15,000	0	0.16	0.089	2.446	3.988
MONTHS_SINCE_OLDEST_TRADE		15,000	0	239.84	227	0.494	0.199
manual_MONTHS_SINCE_OLDEST <small>@ Transformed - Standardization</small>		15,000	0	-0	-0.102	0.494	0.199

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



Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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		15,000	550	0	oth_hme_ai	5,619

[View Histogram](#)

Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDa\CS_testing_03

Home | **1 Data Engineering** | 2 Sampling and Target | 3 Train A Model | 4 Evaluate Your Model | 5 Build Your Decision | 6 Auto Documentation

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		15,000	550	0	oth_hme_ai	5,619

APP_PROD_CODE Categorical Variable

Category	Count	Percentage
premier	5,148	34.29%
regular	6,317	42.11%
student	3,535	23.59%

Project CS_testing_02

Project Path: C:\Users\user\Documents\IDa\CS_testing_02

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

← Data Preprocessing

Upload Meta Data

Uploading meta data is optional. However, it is helpful when:

- You need to assign specific roles to variables in the training dataset
- You have list of default or special values for certain variables(list of Special Values must be separated by comma in the Meta File)

Drag and Drop the File here

(CSV,Parquet,Feather,XLSX)

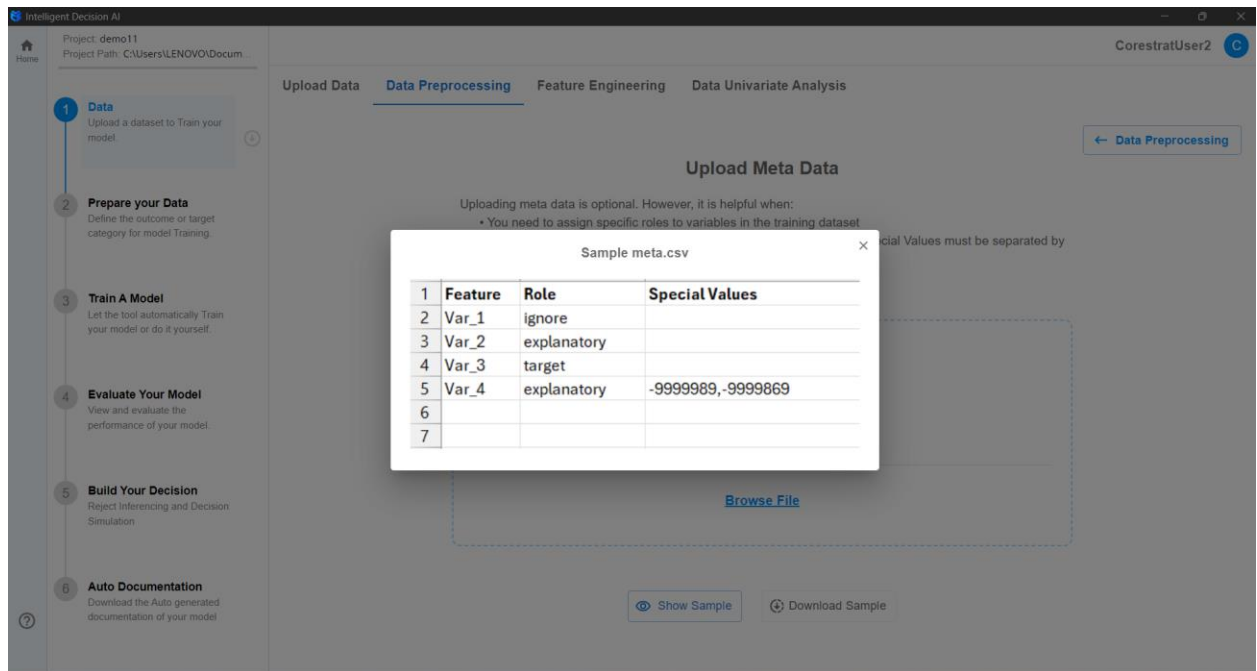
OR

[Browse File](#)

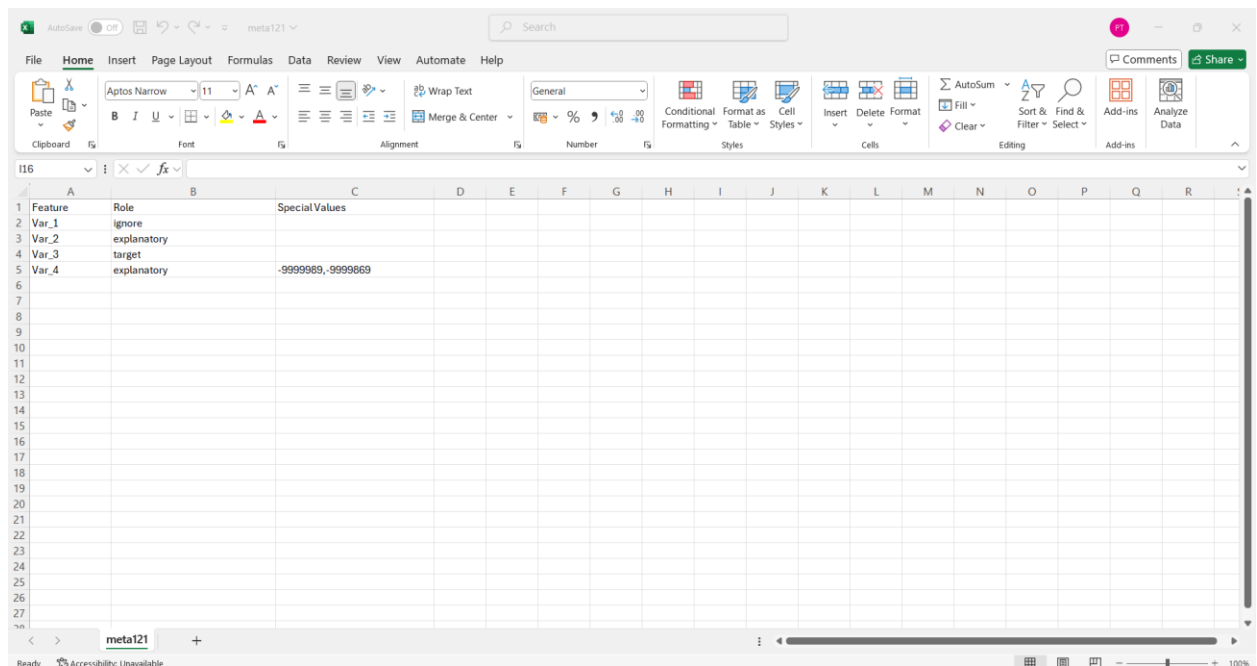
Show Sample

Download Sample

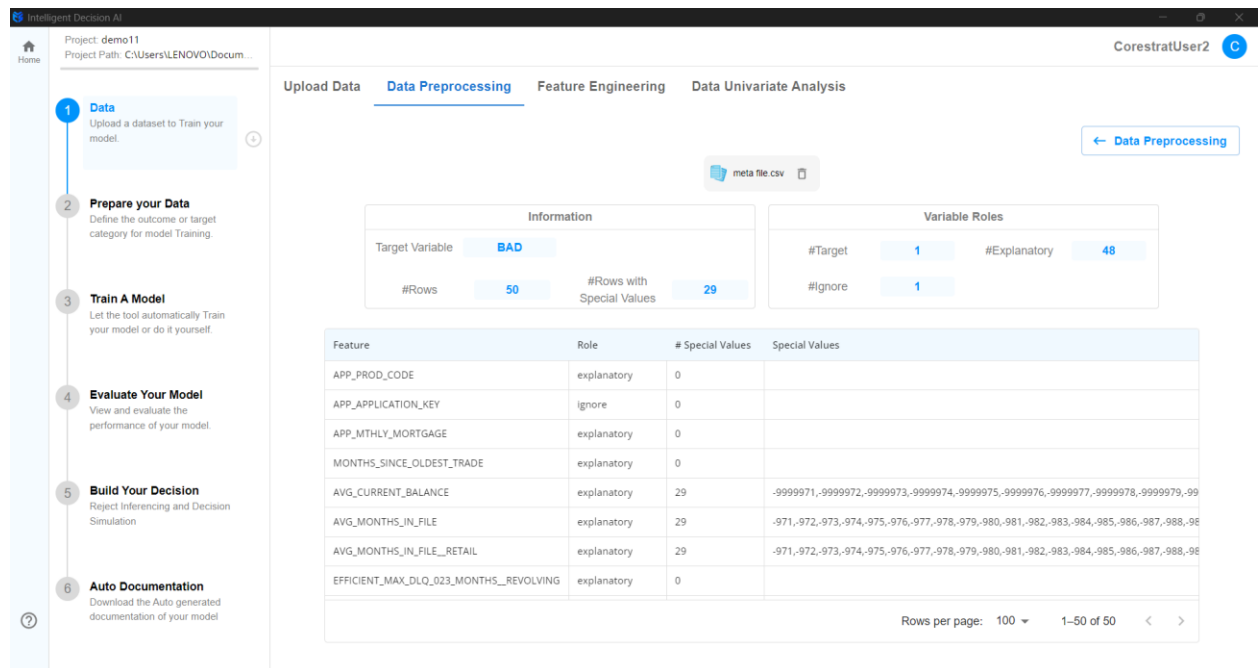
56 | Page



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.



Project: demo11
Project Path: C:\Users\LENOVO\Docum...

CorestratUser2

Upload Data **Data Preprocessing** Feature Engineering Data Univariate Analysis

← Data Preprocessing

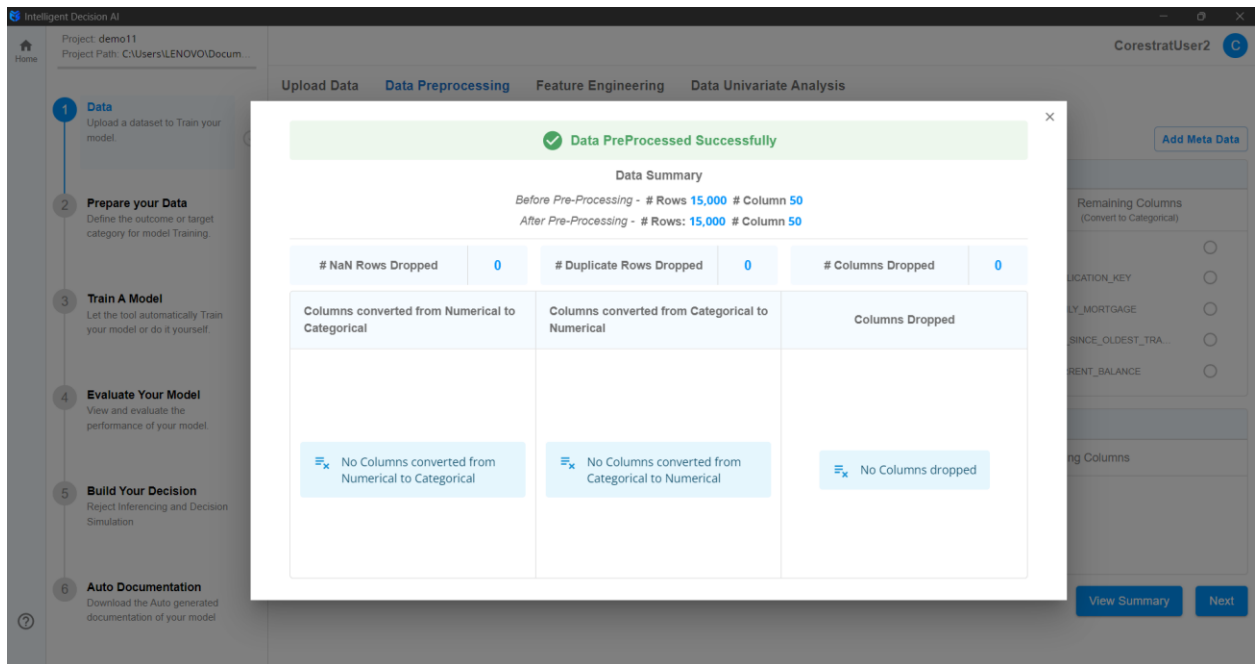
meta file.csv

Information		Variable Roles	
Target Variable	BAD	#Target	1
#Rows	50	#Explanatory	48
	#Rows with Special Values	#Ignore	1
	29		

Feature	Role	# Special Values	Special Values
APP_PROD_CODE	explanatory	0	
APP_APPLICATION_KEY	ignore	0	
APP_MTHLY_MORTGAGE	explanatory	0	
MONTHS_SINCE_OLDEST_TRADE	explanatory	0	
AVG_CURRENT_BALANCE	explanatory	29	-.9999971,-.9999972,-.9999973,-.9999974,-.9999975,-.9999976,-.9999977,-.9999978,-.9999979,-.9999980,-.9999981,-.9999982,-.9999983,-.9999984,-.9999985,-.9999986,-.9999987,-.9999988,-.9999989,-.9999990,-.9999991,-.9999992,-.9999993,-.9999994,-.9999995,-.9999996,-.9999997,-.9999998,-.9999999
AVG_MONTHS_IN_FILE	explanatory	29	-.971,-.972,-.973,-.974,-.975,-.976,-.977,-.978,-.979,-.980,-.981,-.982,-.983,-.984,-.985,-.986,-.987,-.988,-.989,-.990,-.991,-.992,-.993,-.994,-.995,-.996,-.997,-.998,-.999
AVG_MONTHS_IN_FILE_RETAIL	explanatory	29	-.971,-.972,-.973,-.974,-.975,-.976,-.977,-.978,-.979,-.980,-.981,-.982,-.983,-.984,-.985,-.986,-.987,-.988,-.989,-.990,-.991,-.992,-.993,-.994,-.995,-.996,-.997,-.998,-.999
EFFICIENT_MAX_DLQ_023_MONTHS_REVOLVING	explanatory	0	

Rows per page: 100 1-50 of 50

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)

Project: Demo22
Project Path: C:/Users/User/Documents/IDai/Demo22

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

Step 1 **Select Target Variable**
In progress

Step 2 **Define Target Categories**
Pending

Step 3 **Split Train Test**
Pending

Step 4 **Variable(s) to be Ignored**
Pending

Step 5 **Variable x Target Insights**
Pending

Select Target Variable
The chosen variable will serve as the decision tree's root node. Only variables with **distinct values up to 10** will be shown.

Candidate Target Variables

Search Variable

- ☐ NBR_BANK_OR_NATL_REVOLVING_TRDLNS_O...
- ☐ NBR_FINANCE_COMPANY_TRDLNS_ACTIVE_O...
- ☐ NBR_FINANCE_COMPANY_TRDLNS_OPENED_O...
- ☐ BANKRUPTCY_INDICATOR
- ☐ NBR_RETAIL_TRDLNS_OVERLIMIT
- ☐ NBR_TRDLNS_OPENED_023_MONTHS_30P_DP...
- ☐ TOO_NEW_OR_STALE_FILE_INDICATOR
- ☒ BAD

Selected Variable for Target

Next

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.

Project: Demo22
Project Path: C:/Users/User/Documents/IDai/Demo22

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

Step 1 **Select Target Variable**
Completed

Step 2 **Define Target Categories**
In progress

Step 3 **Split Train Test**
Pending

Step 4 **Variable(s) to be Ignored**
Pending

Step 5 **Variable x Target Insights**
Pending

Select the categories that you want to train the model for BAD

Selected Categories will be treated as "Target"

Distinct Values/Categories

Category	Count	Percentage
<input type="checkbox"/> 0	14,328	(95.52%)

Selected Values/Categories

Category	Count	Percentage
<input checked="" type="checkbox"/> 1	672	(4.48%)

Back

Next

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.

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.

5 Build Your Decision

Reject Inferencing and Decision Simulation

6 Auto Documentation

Download the Auto generated documentation of your model

Step 1

Select Target Variable

Completed

Step 2

Define Target Categories

Completed

Step 3

Split Train Test

In Progress

Step 4

Variable(s) to be Ignored

Pending

Step 5

Variable x Target Insights

Pending

Split Train Test

BAD

This screen allows you to choose variables to draw a stratified sample from the uploaded dataset.

Selected Variables for Strata (Max 4)

☐ BAD

Train 70 + -

Test 30

Seed Value 404

Redraw the test

Candidate Strata Variables

Search Candidate Variable

☐ APP_PROD_CODE
☐ EFFICIENT_MAX_DLO_EVER
☐ NBR BANK OR NATL REVOLVIN...

	Train	Test
Total After Split	10,500	4,500
Target	470 (4.48%)	202 (4.49%)
Non Target	10,030 (95.52%)	4,298 (95.51%)

Back

Next

Project: Demo22
Project Path: C:/Users/user/Documents/IDai/Demo22

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

Step 1
Select Target Variable
Completed

Step 2
Define Target Categories
Completed

Step 3
Split Train Test
Completed

Step 4
Variable(s) to be Ignored
In Progress

Step 5
Variable x Target Insights
Pending

Variable(s) to be Ignored
Examples include special ID, serial number, and personal data.

List of All Variables

Search Variable

- ☐ APP_PROD_CODE
- ☐ TOP_3_FEATURE_AREAS_W1
- ☐ DOBER
- ☐ APP_MTHLY_INCOME
- ☐ APP_MTHLY_MORTGAGE
- ☐ MONTHS_SINCE_OLDEST_TRADE
- ☐ AVG_CURRENT_BALANCE
- ☐ AVG_MONTHS_IN_FILE

> <

Variable(s) to be ignored

No potential variables were identified.

Back Next

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Processing information value for all 154 variables. Please hold tight as it progresses.

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
APP_PROD_CODE	Loading..	Loading..	--	--
TOP_3_FEATURE_ARE...	Loading..	Loading..	--	--
HIGHEST_UTIL_ON_BA...	Loading..	Loading..	--	--
NBR_REVL_OR_OPEN...	Loading..	Loading..	--	--
APP_PROD_CODE_idai...	Loading..	Loading..	--	--
APP_PROD_CODE_idai...	Loading..	Loading..	--	--
EFFICIENT_MAX_DLQ...	Loading..	Loading..	--	--
EFFICIENT_MAX_DLQ...	Loading..	Loading..	--	--
EFFICIENT_MAX_DLQ...	Loading..	Loading..	--	--

Rows per page: 100 | 1-100 of 154

Inferred Relationship

- Monotonic Increase: 0
- Monotonic Decrease: 0
- U-Shape: 0
- Bell-Shape: 0
- Categorical: 0
- Manual: 0
- Fine Classing: 0

IV - Fine Classing

- Suspicious IV (IV > 2.0): 0
- Strong IV (0.5 < IV < 2.0): 0
- Moderate IV (0.1 < IV < 0.5): 0
- Weak IV (0.02 < IV < 0.1): 0
- Very Low IV (IV < 0.02): 0

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

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Search...

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
manual_TOT_FIN_CHR...	27.84%	2.3051	Loading..	Loading..
manual_AMT_CR_LIMIT...	20.88%	1.179	Loading..	Loading..
manual_NET_FRACTIO...	18.81%	0.9	Loading..	Loading..
manual_NET_FRACTIO...	18.09%	0.8863	Loading..	Loading..
HIGHEST_UTIL_ON_BA...	17.01%	0.8758	Loading..	Loading..
manual_NET_FRACTIO...	18.27%	0.8492	Loading..	Loading..
manual_HIGHEST_UTIL...	17.19%	0.7618	Loading..	Loading..
manual_HIGHEST_UTIL...	17.41%	0.7462	Loading..	Loading..
manual_HIGHEST_UTIL...	16.86%	0.7344	Loading..	Loading..

Rows per page: 100 | 1-100 of 154

Inferred Relationship

- Monotonic Increase: 0
- Monotonic Decrease: 0
- U-Shape: 0
- Bell-Shape: 0
- Categorical: 0
- Manual: 0
- Fine Classing: 0

IV - Fine Classing

- Suspicious IV (IV > 2.0): 1
- Strong IV (0.5 < IV < 2.0): 18
- Moderate IV (0.1 < IV < 0.5): 45
- Weak IV (0.02 < IV < 0.1): 29
- Very Low IV (IV < 0.02): 60

Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDa\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.
[Click to continue](#)
- 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

Step 1 **Select Target Variable** Completed

Step 2 **Define Target Categories** Completed

Step 3 **Split Train Test** Completed

Step 4 **Variable(s) to be Ignored** Completed

Step 5 **Variable x Target Insights** In Progress

Variable Classing vs. Target Smart Variables-Grouping Multi-Collinearity Insights Variable Selection Walk

Search

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
1-2 manual_TOT_FIN_CHR...	27.84%	2.3051	Bell-Shape (IV Optimal)	2.3051
1-2 manual_AMT_CR_LIMIT...	20.88%	1.179	Monotonic Decrease (IV ...)	1.179
1-2 manual_NET_FRACTIO...	18.81%	0.9	U-Shape (IV Optimal)	0.5848
1-2 manual_NET_FRACTIO...	18.69%	0.8663	U-Shape (IV Optimal)	0.5755
1-2 HIGHEST_UTIL_ON_BA...	17.01%	0.8758	U-Shape (IV Optimal)	0.8758
1-2 manual_NET_FRACTIO...	18.27%	0.8492	U-Shape (IV Optimal)	0.8492
1-2 manual_HIGHEST_UTIL...	17.19%	0.7618	U-Shape (IV Optimal)	0.6858
1-2 manual_HIGHEST_UTIL...	17.41%	0.7462	U-Shape (IV Optimal)	0.5958
1-2 manual_HIGHEST_UTIL...	16.86%	0.7344	U-Shape (IV Optimal)	0.6644

Rows per page: 100 1-100 of 154

[Back](#)

Inferred Relationship 164

- Monotonic Increase 16
- Monotonic Decrease 18
- U-Shape 19
- Bell-Shape 20
- Categorical 5
- Manual 0
- Fine Classing 76

IV - Fine Classing 153

- Suspicious IV (IV > 2.0) 1
- Strong IV (0.5 ≤ IV < 2.0) 18
- Moderate IV (0.1 ≤ IV < 0.5) 45
- Weak IV (0.02 ≤ IV < 0.1) 29
- Very Low IV (IV < 0.02) 60

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)

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Search...

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
1-2 manual_NET_FRACTIO...	17.99%	0.6775	Monotonic Increase (IV 0...)	0.3082
1-2 manual_HIGHEST_UTIL...	17.03%	0.629	Monotonic Increase (IV 0...)	0.2913
1-2 manual_NBR_DEROG_T...	12.35%	0.3392	Monotonic Increase (IV 0...)	0.3391
1-2 manual_PRC_TRDLNS_I...	11.96%	0.2786	Monotonic Increase (IV 0...)	0
1-2 manual_NBR_COLLECT...	11.96%	0.2557	Monotonic Increase (IV 0...)	0.2557
1-2 manual_NBR_INQUIRI...	9.63%	0.1998	Monotonic Increase (IV 0...)	0.1998
1-2 manual_NBR_TRDLNS_...	9.48%	0.1834	Monotonic Increase (IV 0...)	0.1833
1-2 manual_NBR_INQUIRI...	8.81%	0.1806	Monotonic Increase (IV 0...)	0.1806
1-2 manual_NBR_TRDLNS_...	8.9%	0.1726	Monotonic Increase (IV 0...)	0.1726
1-2 manual_NBR_INQUIRI...	8.81%	0.1806	Monotonic Increase (IV 0...)	0.1806
1-2 manual_NBR_TRDLNS_...	8.81%	0.1649	Monotonic Increase (IV 0...)	0.1649
1-2 manual_NBR_INSTALLM...	8.63%	0.1532	Monotonic Increase (IV 0...)	0.1532

Rows per page: 100 1-16 of 16

Back

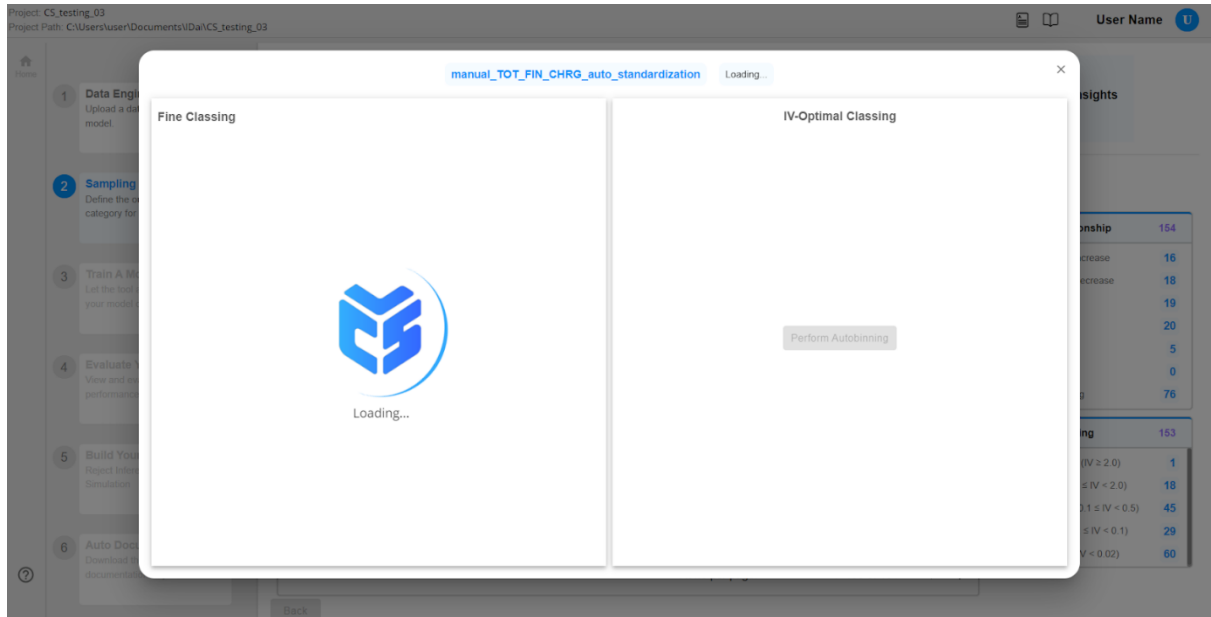
Inferred Relationship 154

- Monotonic Increase 16
- Monotonic Decrease 18
- U-Shape 19
- Bell-Shape 19
- Categorical 5
- Manual 1
- Fine Classing 76

IV - Fine Classing 153

- Suspicious IV (IV ≥ 2.0) 1
- Strong IV (0.5 ≤ IV < 2.0) 18
- Moderate IV (0.1 ≤ IV < 0.5) 45
- Weak IV (0.02 ≤ IV < 0.1) 29
- Very Low IV (IV < 0.02) 60

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'** text box. 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.

Project: CS_testing_03
Project Path: C:\Users\user\Documents\IDa\CS_testing_03

User Name

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Search...

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
1-2 manual_TOT_FIN_CHR...	27.84%	2.3051	Manual	2.2374
1-2 manual_AMT_CR_LIMIT...	20.88%	1.179	Monotonic Decrease (IV ...)	1.1720
1-2 manual_NET_FRACTIO...	18.81%	0.9	U-Shape (IV Optimal)	0.5848
1-2 manual_NET_FRACTIO...	18.69%	0.8863	U-Shape (IV Optimal)	0.5755
1-2 HIGHEST_UTIL_ON_BA...	17.01%	0.8758	U-Shape (IV Optimal)	0.8758
1-2 manual_NET_FRACTIO...	18.27%	0.8492	U-Shape (IV Optimal)	0.8492
1-2 manual_HIGHEST_UTIL...	17.19%	0.7618	U-Shape (IV Optimal)	0.6858
1-2 manual_HIGHEST_UTIL...	17.41%	0.7462	U-Shape (IV Optimal)	0.5858
1-2 manual_HIGHEST_UTIL...	16.86%	0.7344	U-Shape (IV Optimal)	0.6644

Rows per page: 100 | 1-100 of 154

Back

Inferred Relationship 154

- Monotonic Increase 16
- Monotonic Decrease 18
- U-Shape 19
- Bell-Shape 19
- Categorical 5
- Manual 1
- Fine Classing 76

IV - Fine Classing 153

- Suspicious IV ($IV \geq 2.0$) 1
- Strong IV ($0.5 \leq IV < 2.0$) 18
- Moderate IV ($0.1 \leq IV < 0.5$) 45
- Weak IV ($0.02 \leq IV < 0.1$) 29
- Very Low IV ($IV < 0.02$) 60

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.

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.

5 Build Your Decision
Reject Inferencing and Decision Simulation

6 Auto Documentation
Download the Auto generated documentation of your model

Step 1 Select Target Variable **Completed**

Step 2 Define Target Categories **Completed**

Step 3 Split Train Test **Completed**

Step 4 Variable(s) to be ignored **Completed**

Step 5 Variable x Target Insights **In Progress**

Variable Classing vs. Target Smart Variables-Grouping Multi-Collinearity Insights Variable Selection Walk

Processing information value for all variables. Please hold tight as it progresses.

Search...

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
Sort by ASC	27.8%	2.4103	Bell-Shape (IV Optimal)	2.4103
Sort by DESC	21.04%	1.2449	Monotonic Decrease (IV Optimal)	1.2449
Filter	18.03%	0.2694	Monotonic Increase (IV Optimal)	0.7433
	17.95%	0.0501	Monotonic Increase (IV Optimal)	0.7324
NET_FRACTION_REVOLV...	17.22%	0.9755	Monotonic Increase (IV Optimal)	0.4389
HIGHEST_UTIL_ON_REVOL...	16.62%	0.8917	U-Shape (IV Optimal)	0.8917
HIGHEST_UTIL_ON_BANK...	17.01%	0.8758	U-Shape (IV Optimal)	0.8758
HIGHEST_UTIL_ON_REVOL...	16.33%	0.8725	U-Shape (IV Optimal)	0.8725
HIGHEST_UTIL_ON_BANK...	16.71%	0.8464	U-Shape (IV Optimal)	0.8464

Rows per page: 100 1-100 of 106

Back

Inferred Relationship 106

- Monotonic Increase 21
- Monotonic Decrease 18
- U-Shape 27
- Bell-Shape 28
- Categorical 2
- Manual 0
- Fine Classing 10

IV - Fine Classing 106

- Suspicious IV (IV ≥ 2.0) 1
- Strong IV (0.5 \leq IV < 2.0) 25
- Moderate IV (0.1 \leq IV < 0.5) 56
- Weak IV (0.02 \leq IV < 0.1) 20
- Very Low IV (IV < 0.02) 4

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

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

Step 1 Select Target Variable **Completed**

Step 2 Define Target Categories **Completed**

Step 3 Split Train Test **Completed**

Step 4 Variable(s) to be ignored **Completed**

Step 5 Variable x Target Insights **In Progress**

Variable Classing vs. Target Smart Variables-Grouping Multi-Collinearity Insights Variable Selection Walk

Search...

Variables	Correlation	IV - Fine Classing	Inferred Relationship	IV - Final Classing
Columns Operator Value			Manual	2.2374
X Variables contains Filter value			Monotonic Decrease (IV ...)	1.1129
1-2 manual_NET_FRACTIO...	18.81%	0.9	U-Shape (IV Optimal)	0.5848
1-2 manual_NET_FRACTIO...	18.69%	0.8863	U-Shape (IV Optimal)	0.5755
1-2 HIGHEST_UTIL_ON_BA...	17.01%	0.8758	U-Shape (IV Optimal)	0.8758
1-2 manual_NET_FRACTIO...	18.27%	0.8492	U-Shape (IV Optimal)	0.8492
1-2 manual_HIGHEST_UTIL...	17.19%	0.7618	U-Shape (IV Optimal)	0.6858
1-2 manual_HIGHEST_UTIL...	17.41%	0.7462	U-Shape (IV Optimal)	0.5858
1-2 manual_HIGHEST_UTIL...	16.86%	0.7344	U-Shape (IV Optimal)	0.6644

Rows per page: 100 1-100 of 154

Back

Inferred Relationship 154

- Monotonic Increase 16
- Monotonic Decrease 18
- U-Shape 19
- Bell-Shape 19
- Categorical 5
- Manual 1
- Fine Classing 76

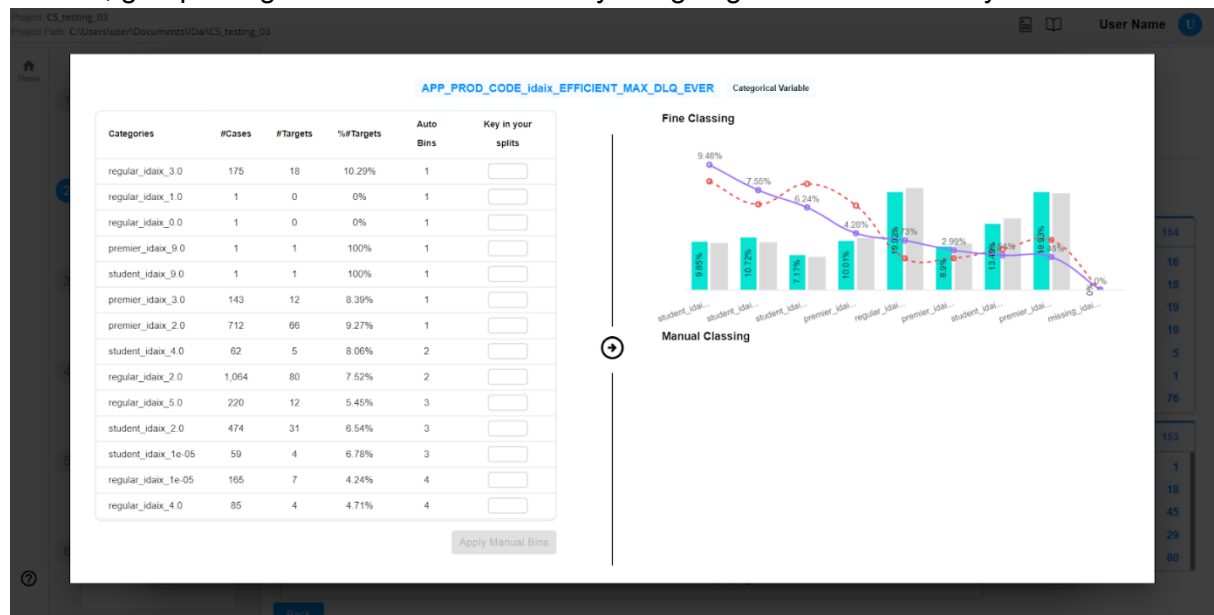
IV - Fine Classing 153

- Suspicious IV (IV ≥ 2.0) 1
- Strong IV (0.5 \leq IV < 2.0) 18
- Moderate IV (0.1 \leq IV < 0.5) 45
- Weak IV (0.02 \leq IV < 0.1) 29
- Very Low IV (IV < 0.02) 60

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.



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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | **Smart Variables-Grouping** | Multi-Collinearity Insights | Variable Selection Walk

Smart Variables-Grouping

Number of Clusters: 70

Variance Explained (%): 0.02

>= IV: 1

<= IV: 1

☐ Use Binned Variables

☒ Use Original Variables

Run Algorithm

Final Clusters

ID	#Features	Proportions
Cluster1	12	95.61%
Cluster2	6	89.62%
Cluster3	6	99.91%
Cluster4	3	95.87%
Cluster5	3	88.56%
Cluster6	8	99.28%
Cluster7	2	80.98%
Cluster8	2	98.93%
Cluster9	3	100.00%

Final Set of Variables

1-2 manual_NBR_BANK_OR...	Cluster1
1-2 manual_NBR_TRDLSNS_9...	Cluster2
1-2 manual_NBR_RETAIL_TR...	Cluster3
1-2 manual_NBR_REVL_OR...	Cluster4
1-2 manual_TTL_TRADE_LIN...	Cluster5
1-2 manual_NET_FRACTION...	Cluster6
1-2 manual_NBR_MORTGAG...	Cluster7
1-2 manual_MTHS_SNC_OLD...	Cluster8
1-2 manual_TTL_INSTALLME...	Cluster9
1-2 manual_NBR_MORTGAG...	Cluster10
1-2 manual_NBR_TRDLSNS_O...	Cluster11
1-2 manual_NBR_BANK_OR...	Cluster12

Results

#Final Clusters: 35

#Final Set of Variables: 35

Variance explained by ... 40.23%

Skip Save

Back

If user selects 'Use Binned Variables':

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | **Smart Variables-Grouping** | Multi-Collinearity Insights | Variable Selection Walk

Smart Variables-Grouping

Number of Clusters: 70

Variance Explained (%): 0.02

>= IV: 1

<= IV: 1

☒ Use Binned Variables

☐ Use Original Variables

Run Algorithm

Skip Clustering

Back

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Generating Cluster(s)...

Project: CS_testing_03
Project Path: C:\Users\User\Documents\Da\CS_testing_03

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Smart Variables-Grouping

Number of Clusters: 70
Variance Explained (%): 70
>= IV: 0.02
<= IV: 1

☒ Use Binned Variables
☐ Use Original Variables

Run Algorithm

Final Clusters

ID	#Features	Proportions
Cluster1	3	82.30%
Cluster2	4	72.53%
Cluster3	4	99.44%
Cluster4	7	81.66%
Cluster5	1	100.00%
Cluster6	4	91.16%
Cluster7	2	88.03%
Cluster8	2	94.71%
Cluster9	5	87.39%

Final Set of Variables

Variable	Cluster
manual_AVG_MONTHS_I...	Cluster1
manual_NBR_TRDLNS_3...	Cluster2
manual_TTL_BANK_OR...	Cluster3
manual_NBR_REVL_OR...	Cluster4
manual_DOBER_auto_b...	Cluster5
manual_NBR_RETAIL_TR...	Cluster6
manual_NBR_MORTGAG...	Cluster7
manual_NET_FRACTION...	Cluster8
manual_NBR_BANK_OR...	Cluster9
manual_TTL_AMOUNT_N...	Cluster10
manual_NBR_REVL_OR...	Cluster11
manual_NBR_INQUIRES...	Cluster12

Results

#Final Clusters: 35
#Final Set of Variables: 35
Variance explained by ... 39.33%

Skip Save

Back

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

User Name: U

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

Save Clustering?

Only the features in final set of Variables list explaining variance of 38.46% in the data will be used for further journey. Click Save for confirmation.

Clustering data saved successfully!

Close Save

Variables: Cluster1, Cluster2, Cluster3, Cluster4, Cluster5, Cluster6, Cluster7, Cluster8, Cluster9, Cluster10, Cluster11, Cluster12

Results: #Final Clusters: 35, #Final Set of Variables: 35, Variance explained by ... 38.46%

Buttons: Skip, Save, Revert Changes

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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

User Name: U

Step 1: Select Target Variable (Completed)
Step 2: Define Target Categories (Completed)
Step 3: Split Train Test (Completed)
Step 4: Variable(s) to be Ignored (Completed)
Step 5: Variable x Target Insights (In Progress)

Variable Classing vs. Target | Smart Variables-Grouping | Multi-Collinearity Insights | Variable Selection Walk

IV Lower Threshold: 0.02 | IV Upper Threshold: 1 | Correlation Threshold: 0.95 | VIF Threshold: 10 | Generate VIF

Select Variable: All Variables

Correlated Variable Pairs (> Correlation Threshold)

VIF Range	Count
Very High Multicollinearity (VIF > 10)	0
High Multicollinearity (5 ≤ VIF < 10)	0
Moderate Multicollinearity (1.5 ≤ VIF < 5)	0
Low or No Multicollinearity (VIF < 1.5)	35

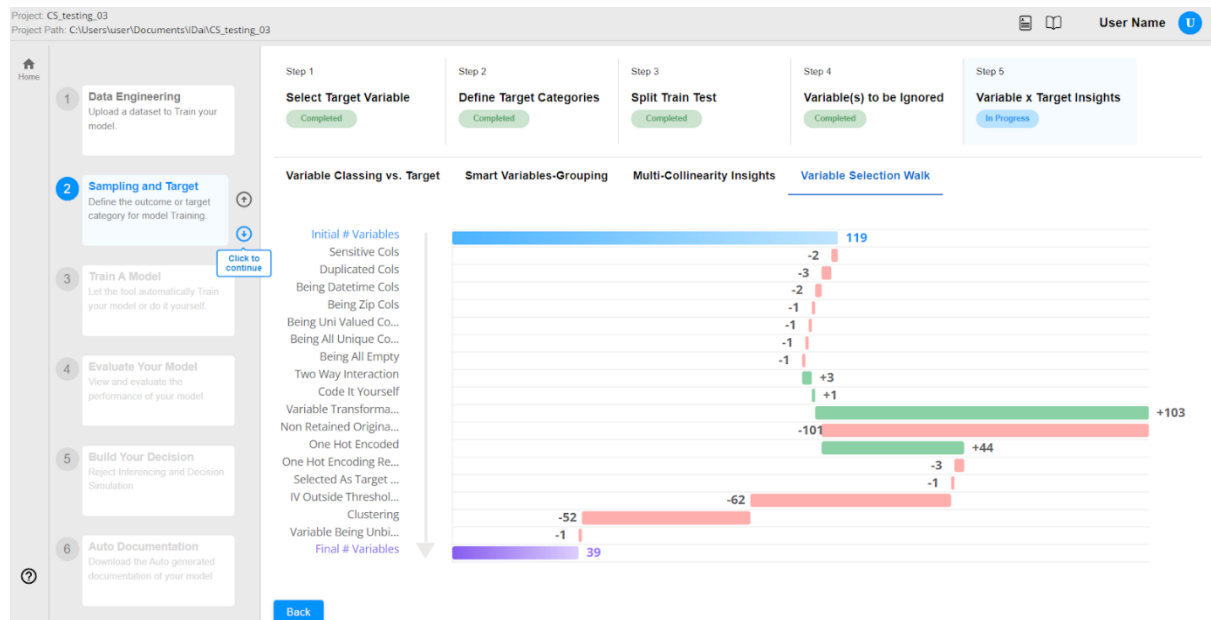
Buttons: Back

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 screenshot displays the CORESTRAT IntelliDecision.ai interface. On the left, a sidebar lists six steps: 1. Data Engineering, 2. Sampling and Target, 3. Train A Model, 4. Evaluate Your Model, 5. Build Your Decision, and 6. Auto Documentation. The main workspace shows a workflow with five steps: Step 1 (Select Target Variable, Completed), Step 2 (Define Target Categories, Completed), Step 3 (Split Train Test, Completed), Step 4 (Variable(s) to be Ignored, Completed), and Step 5 (Variable x Target Insights, In Progress). Below these steps, there are tabs for Variable Classing vs. Target, Smart Variables-Grouping, Multi-Collinearity Insights, and Variable Selection Walk. A modal window is open in the center, titled "Variable Pairs with High Correlation (> Correlation Threshold)". It contains a table with columns: Variable 1 (Retained), Variable 2 (Dropped), and Correlation between Variable 1 & Variable 2. The table is empty, and a message states: "No highly correlated variable pairs identified." In the background, a bar chart is visible, and a table on the right shows VIF analysis results: VIF (VIF ≥ 10) with 0 results, VIF (VIF < 10) with 0 results, Moderate Multicollinearity (1.5 ≤ VIF < 5) with 0 results, and Low or No Multicollinearity (VIF < 1.5) with 39 results.

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

Project: Demo22
Project Path: C:/Users/User/Documents/IDai/Demo22

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

Performance Interpretation Guide

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 01

Selected

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Variable(s) in Final Tree

Variable(s) in Final Model

Please enter remarks...

Model Ensembling

Ensemble

Use as Final Model

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Train: -- Test: --

Variable(s) in Final Tree

Variable(s) in Final Model

Please enter remarks...

Summary Model 01

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

Project: Demo22
Project Path: C:/Users/User/Documents/IDai/Demo22

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

Performance Interpretation Guide

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 01

Selected

Train: 66 Test: 62

Train: 60 Test: 80

Train: 83 Test: 0.29

Train: 99 Test: 85

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)

Select Method(s) of Ensembling

☒ Logistic Regression

Apply

Variable(s) in Final Model

Please enter remarks...

Model Ensembling Successful!

Close

Summary Model 01

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

Project: IDai/CS_testing_03
Project Path: C:/Users/user/Documents/IDai/CS_testing_03

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
Inspect Inference and Decision Simulation
- 6 Auto Documentation
Download the Auto generated documentation of your model

Target Variable
APP_PROD_CODE

Change Target Variable Auto Grow Tree

Click on nodes for options to grow/de-grow the tree.

ID 0	Train	Test
Total	178	77
Target	24	11
Target %	13.48%	14.29%

Summary Model 01

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\IDa\CS_testing_01

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](#)

	Train	Test
total	178	77
get	24	11
et %	13.48%	14.29%

Summary **Model 01**

Project: CS_testing_01
Project Path: C:\Users\User\Documents\IDa\CS_testing_01

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

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)

Number of Random Forest Trees (≥ 10 and ≤ 500)

404 100

Max Depth of Each Tree (≥ 0 and ≤ 10)

Min Split Loss Reduction (≥ 0 and ≤ 5)

6 0

Min Sample # to Split a Node (≥ 1 and ≤ 500)

Max Delta Step Size (≥ 0 and ≤ 100)

Fraction of Features to Consider (≥ 0.5 and ≤ 1)

20 20 0.8

Regularization Parameters

L1 Regularization (Sparsity) (≥ 0 and ≤ 5)

L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)

0 1

XG Boost

	Train	Test
al	178	77
et	24	11
et %	13.48%	14.29%

Project: IDai/CS_testing_03
Project Path: C:/Users/user/Documents/IDai/CS_testing_03

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)

Max Depth of Each Tree (≥ 0 and ≤ 10)

Min Sample # to Split a Node (≥ 1 and ≤ 500)

Max Number of Leaves (≥ 0 and ≤ 10)

404 0 20 6

Feature Importance Type

Gain

Min Split Loss Reduction (≥ 0 and ≤ 5)

Fraction of Features to Consider (≥ 0.5 and ≤ 1)

0 0.8

Regularization Parameters

L1 Regularization (Sparsity) (≥ 0 and ≤ 5)

L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)

0 1

Learning Parameters

Step Size for Each Iteration (≥ 0.01 and ≤ 0.3)

Number of Boosting Rounds (Tree) (≥ 100 and ≤ 500)

0.1 100

Evaluation Parameters

Evaluation Metrics for Validation Data

negative log-likelihood

	Train	Test
al	178	77
et	24	11
et %	13.48%	14.29%

Project: CS_testing_01
Project Path: C:\Users\user\Documents\IDa\CS_testing_01

User: User_peneto1

Train A Model
Let the tool automatically Train your model or do it yourself.

XG Boost

Neural Network

Neural Structure

Compiler Parameters

Loss: Binary Cross Entropy
Optimizer: adam

Metrics: accuracy

Epochs: 100
Batch Size: 32

Early Stopping Flag: ☒ Yes ☐ No

Patience (> 2 and < 20): 10
Min Delta (> 0.0001 and < 1): 0.01

Train Test

Train	Test
178	77
24	11
3.48%	14.29%

Save Settings

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.

Project: IDa\CS_testing_03
Project Path: C:\Users\user\Documents\IDa\CS_testing_03

User: User Name

Train A Model
Let the tool automatically Train your model or do it yourself.

Auto Grow Tree

Click on nodes for options to grow/de-grow the tree

ID No: 0
Variable(s) to be used in Model Building: 98

IntelliDecision.ai does it all for you
This option builds a decision tree or analytical model through intelligent calculations.

Auto Grow (Decision tree) OR **Auto ML Model** (Build classification model)

Do-it-Yourself (DIY)
This option allows you to do things manually as per your preferences and needs.

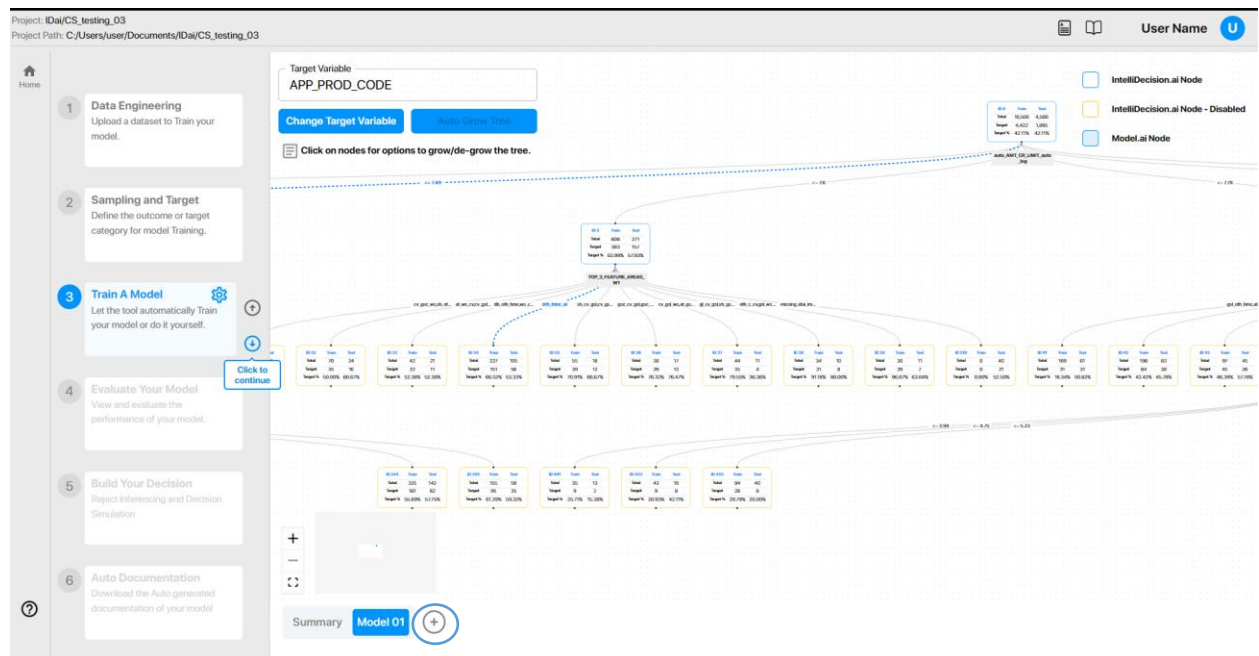
Collapse Node **Add Your Splits** **Manual Grow**

Summary **Model 01**

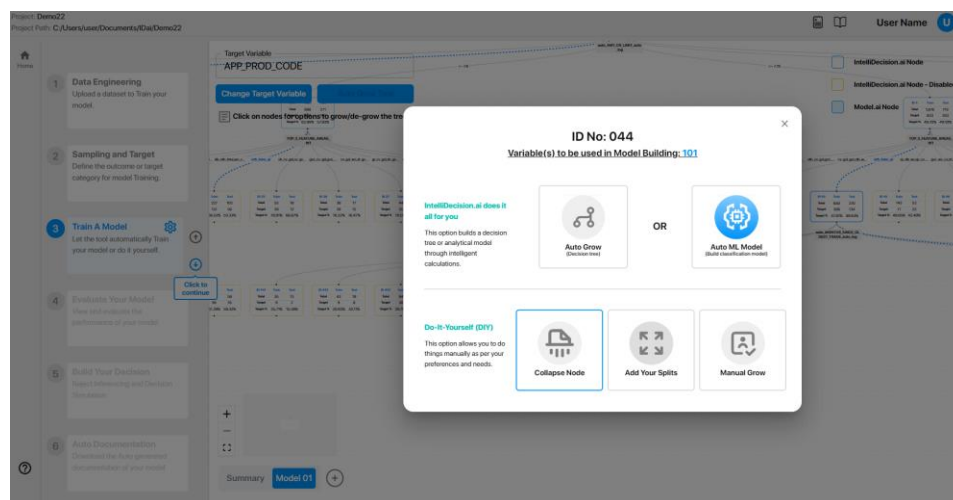
Auto Grow

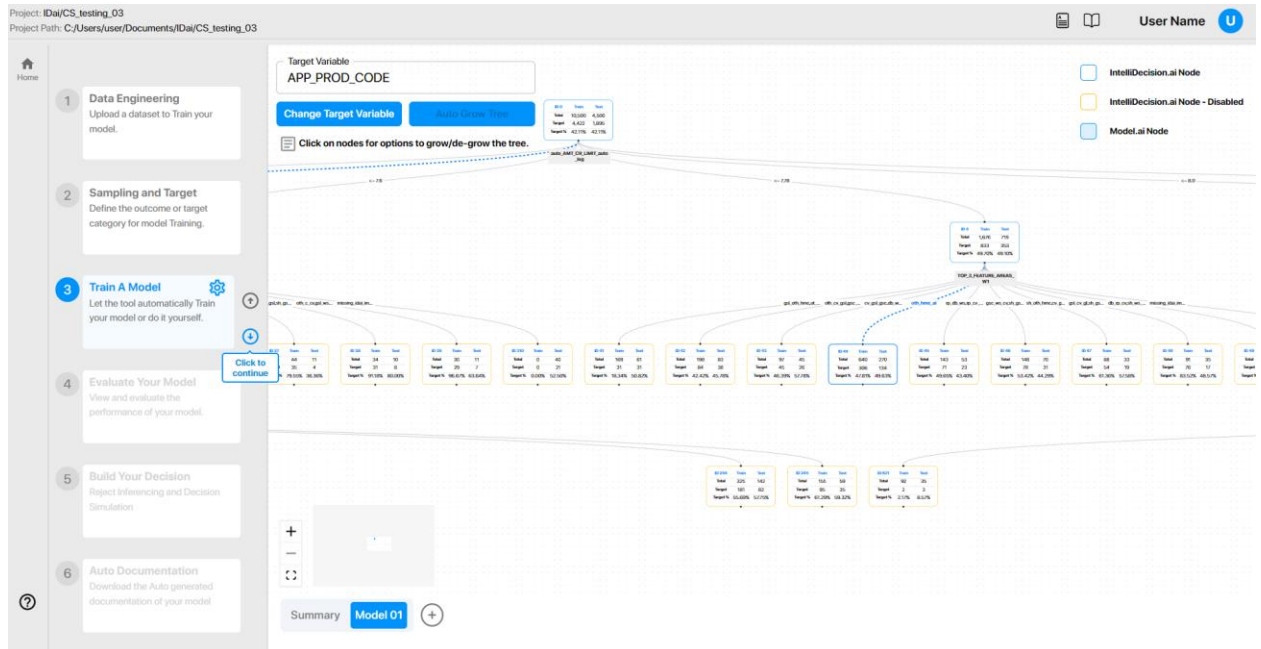
When the “**Auto Grow**” option is selected, a tree is developed starting from the root node and progressing to subsequent levels based on robust separation within each level and across different levels.

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

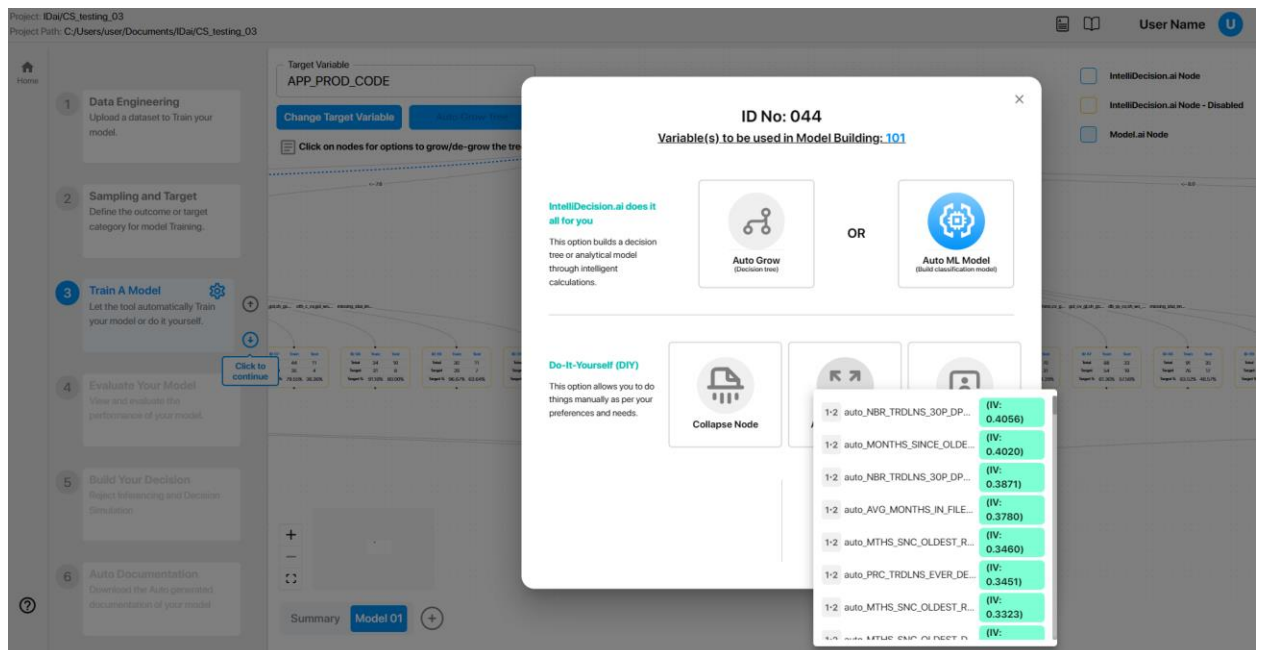


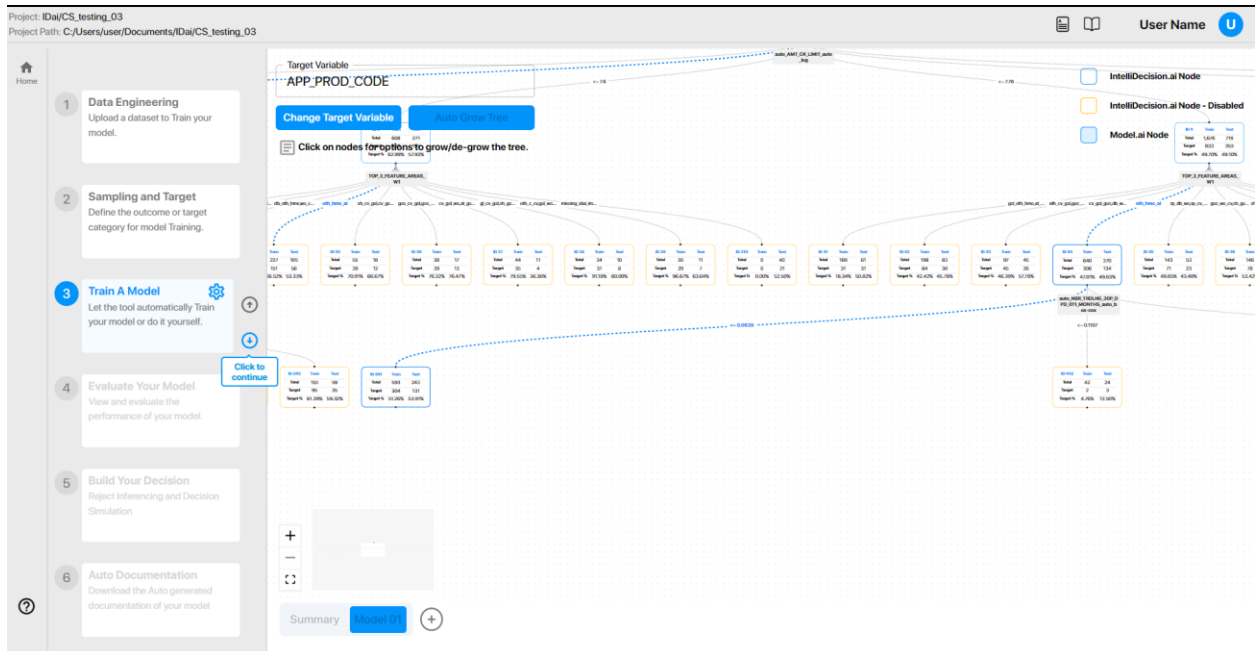
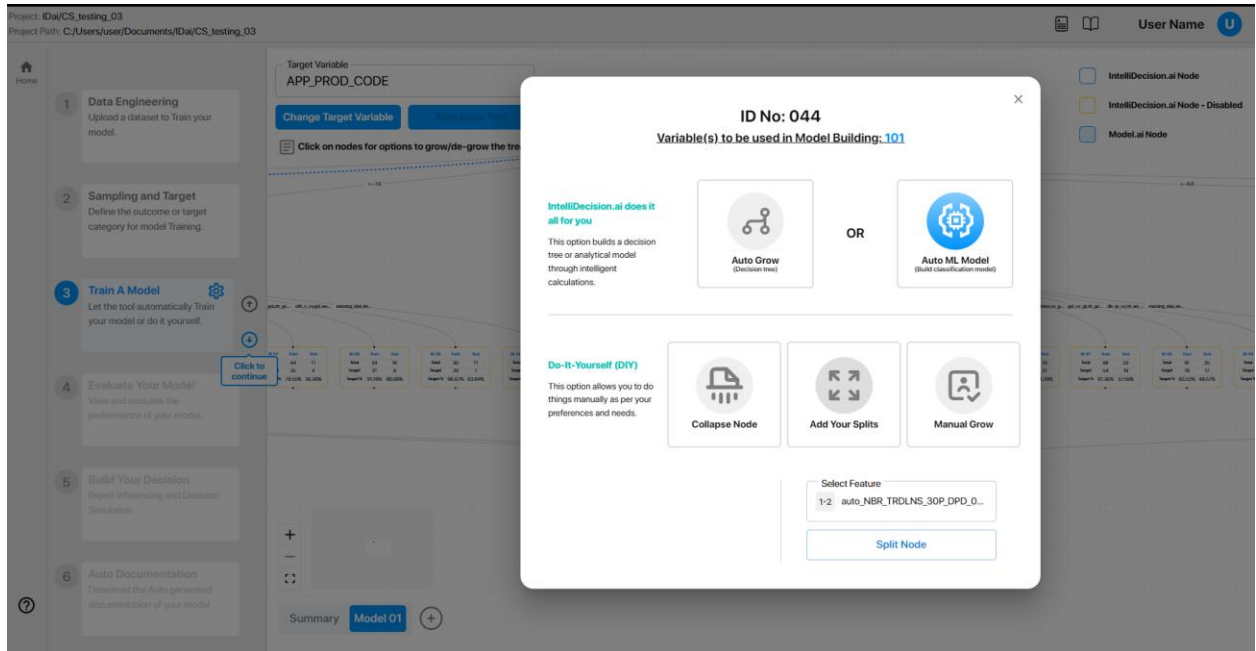
If user does not wish to split a particular node into sub-classes, they can go back to that node and click on “**Collapse Node**” option.





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.



Expand ID No: 064 (49.73 - 164.44)

Enter the desired number of bins

		Test			Train			Grouping	
#	Range	#Cases	#Target	%#Target	#Cases	#Target	%#Target	Select categories to be grouped together	Create your own splits
1	49.73 - 61.20	44	1	2.27%	112	2	1.79%	<input type="text"/>	<p>Specify the split point by comma separated</p> <input type="text"/> <p>Submit</p>
2	61.20 - 72.67	41	0	0.00%	116	0	0.00%	<input type="text"/>	
3	72.67 - 84.14	51	0	0.00%	98	1	1.02%	<input type="text"/>	
4	84.14 - 95.61	36	0	0.00%	109	2	1.83%	<input type="text"/>	
5	95.61 - 107.08	34	0	0.00%	81	0	0.00%	<input type="text"/>	
6	107.08 - 118.56	29	0	0.00%	86	2	2.33%	<input type="text"/>	
7	118.56 - 130.03	26	0	0.00%	68	1	1.47%	<input type="text"/>	
8	130.03 - 141.50	24	0	0.00%	62	0	0.00%	<input type="text"/>	
9	141.50 - 152.97	26	0	0.00%	88	1	1.14%	<input type="text"/>	
10	152.97 - 164.44	33	0	0.00%	60	1	1.67%	<input type="text"/>	

[Start a New Project](#)



Expand ID No: 064 (49.73 - 164.44)

Enter the desired number of bins

		Test			Train			Grouping	
#	Range	#Cases	#Target	%#Target	#Cases	#Target	%#Target	Select categories to be grouped together	Create your own splits
1	49.73 - 61.20	44	1	2.27%	112	2	1.79%	<input type="text" value="1"/>	<p>Specify the split point by comma separated</p> <input type="text"/> <p>Submit</p>
2	61.20 - 72.67	41	0	0.00%	116	0	0.00%	<input type="text" value="1"/>	
3	72.67 - 84.14	51	0	0.00%	98	1	1.02%	<input type="text" value="1"/>	
4	84.14 - 95.61	36	0	0.00%	109	2	1.83%	<input type="text" value="1"/>	
5	95.61 - 107.08	34	0	0.00%	81	0	0.00%	<input type="text" value="2"/>	
6	107.08 - 118.56	29	0	0.00%	86	2	2.33%	<input type="text" value="2"/>	
7	118.56 - 130.03	26	0	0.00%	68	1	1.47%	<input type="text"/>	
8	130.03 - 141.50	24	0	0.00%	62	0	0.00%	<input type="text"/>	
9	141.50 - 152.97	26	0	0.00%	88	1	1.14%	<input type="text"/>	
10	152.97 - 164.44	33	0	0.00%	60	1	1.67%	<input type="text"/>	

Expand ID: 064

	Test	Train
Cases	244	881
Target	1	10
Target%	0.20%	1.14%

Grouping 1

	Test	Train
Cases	172	435
Target	1	0
Target%	0.88%	1.18%

Grouping 2

	Test	Train
Cases	83	187
Target	0	2
Target%	0.00%	1.20%

Grouping 3

	Test	Train
Cases	109	278
Target	0	3
Target%	0.00%	1.08%

[Expand Node](#)

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

1 Data

Upload a dataset to Train your model.

Settings

User Name

U

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%	<input type="text"/>	<div>Specify the split point by comma separated</div> <div>1800,2000</div> <div>Submit</div>
2	1940.00 - 2080.00	514	8	1.56%	1174	15	1.28%	<input type="text"/>	
3	2080.00 - 2220.00	181	1	0.55%	448	7	1.56%	<input type="text"/>	
4	2220.00 - 2360.00	34	1	2.94%	47	3	6.38%	<input type="text"/>	
5	2360.00 - 2500.00	230	7	3.04%	510	7	1.37%	<input type="text"/>	
6	2500.00 - 2640.00	5	0	0.00%	26	0	0.00%	<input type="text"/>	
7	2640.00 - 2780.00	32	0	0.00%	100	1	1.00%	<input type="text"/>	
8	2780.00 - 2920.00	18	1	5.56%	45	1	2.22%	<input type="text"/>	
9	2920.00 - 3060.00	228	1	0.44%	506	3	0.59%	<input type="text"/>	
10	3060.00 - 3200.00	15	0	0.00%	64	3	4.69%	<input type="text"/>	

Start a New Project

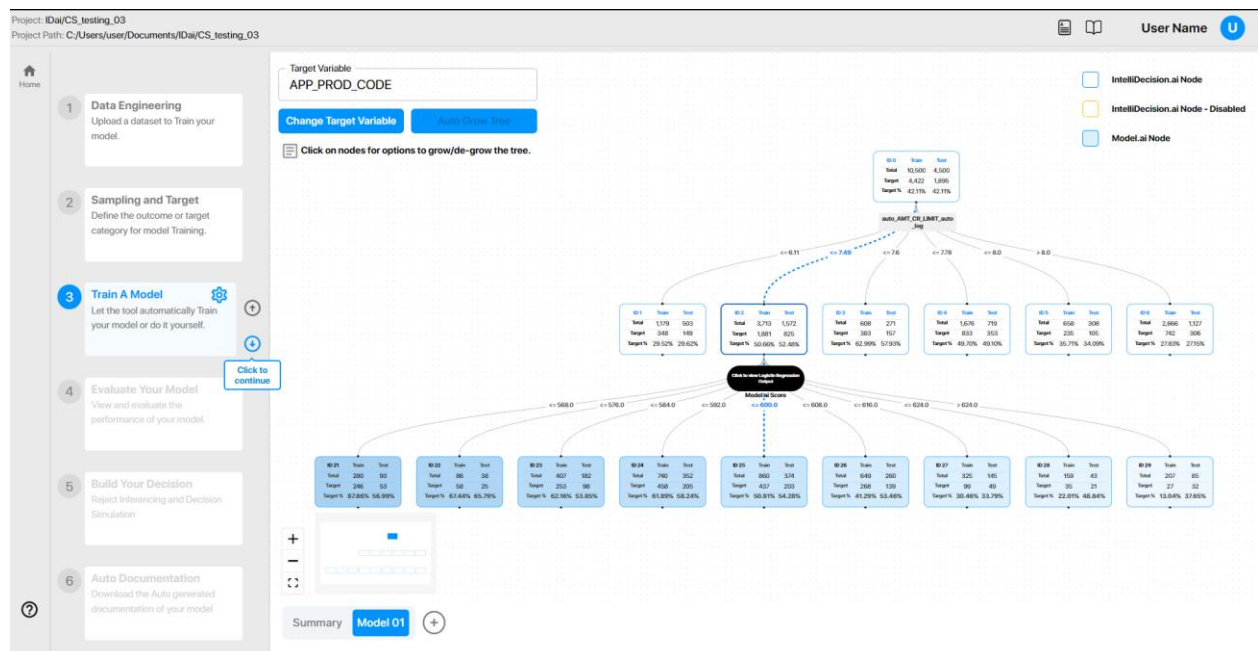
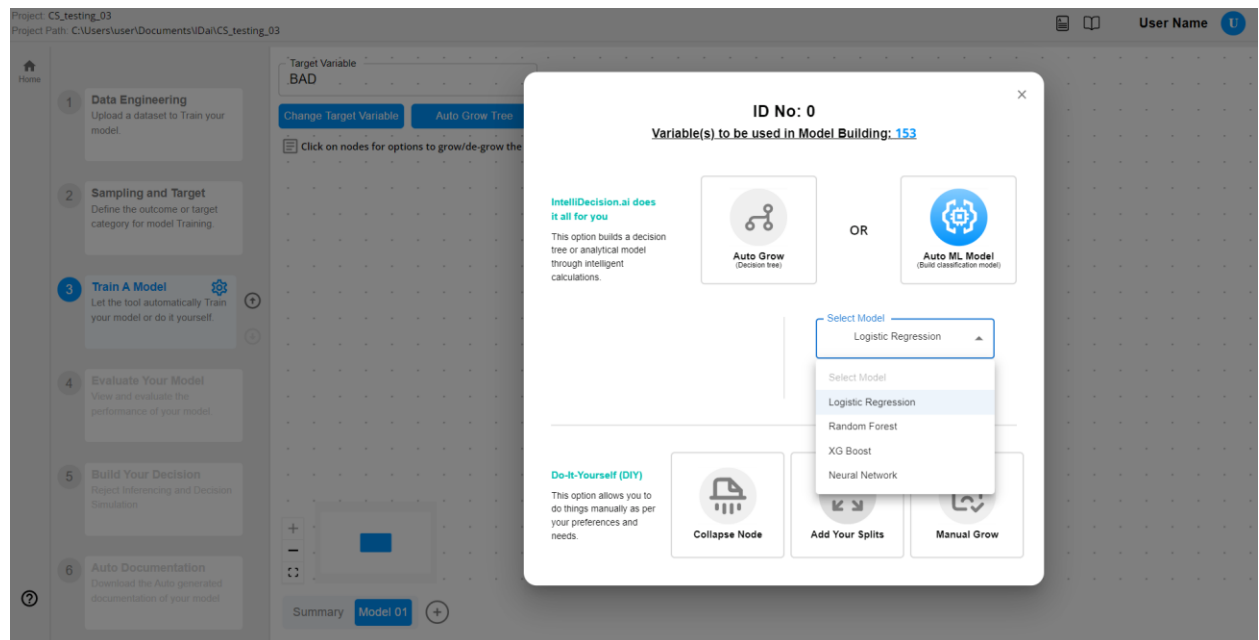
Refresh

Cancel

Save

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.

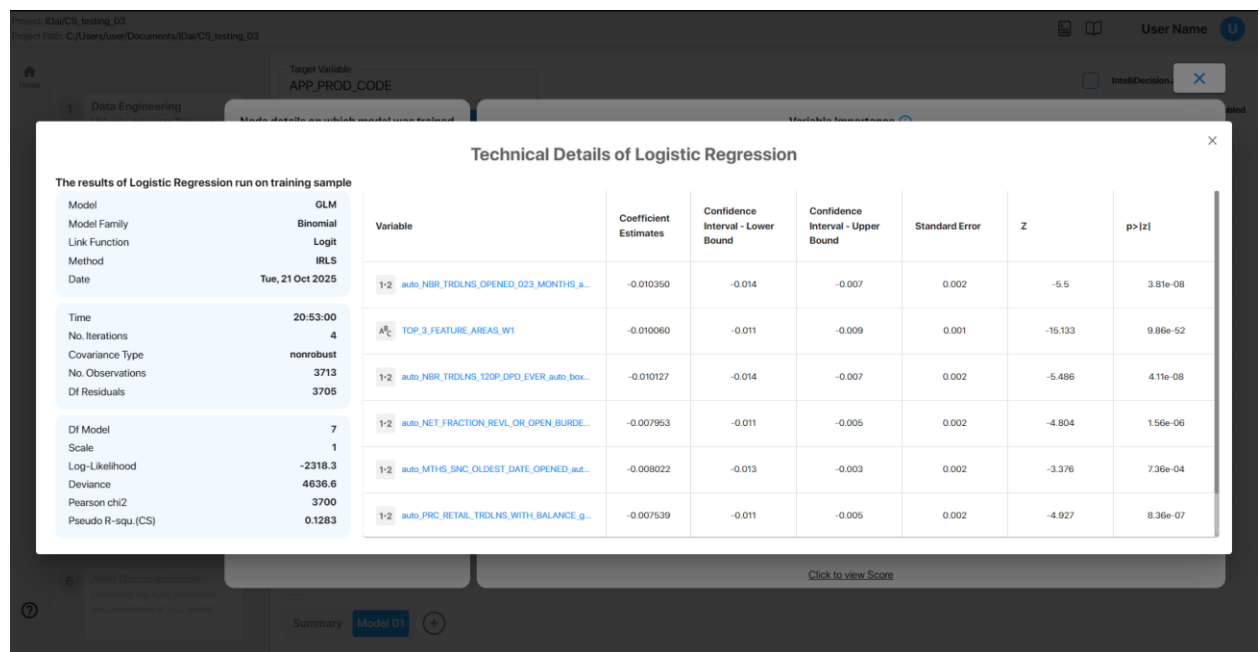


User can click on “**Click to view Logistic Regression Output**” to see the following results for the node on which model.ai was run:

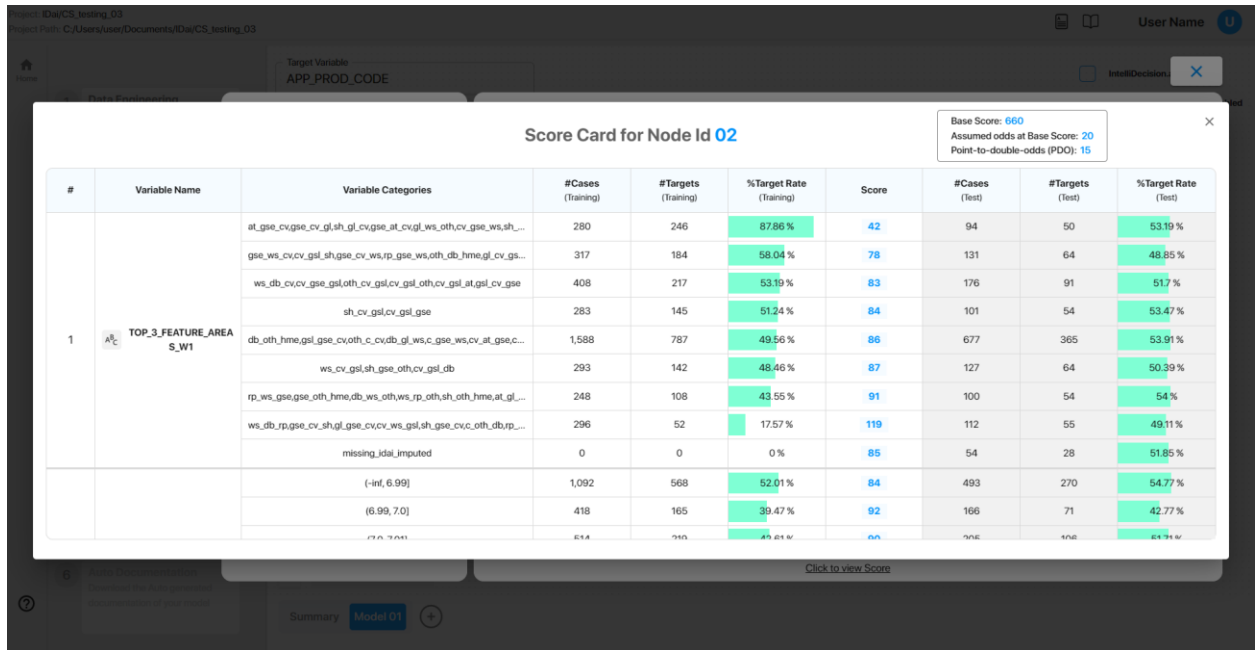
1. Node details, variable importance, model performance metrics & target rate by score.



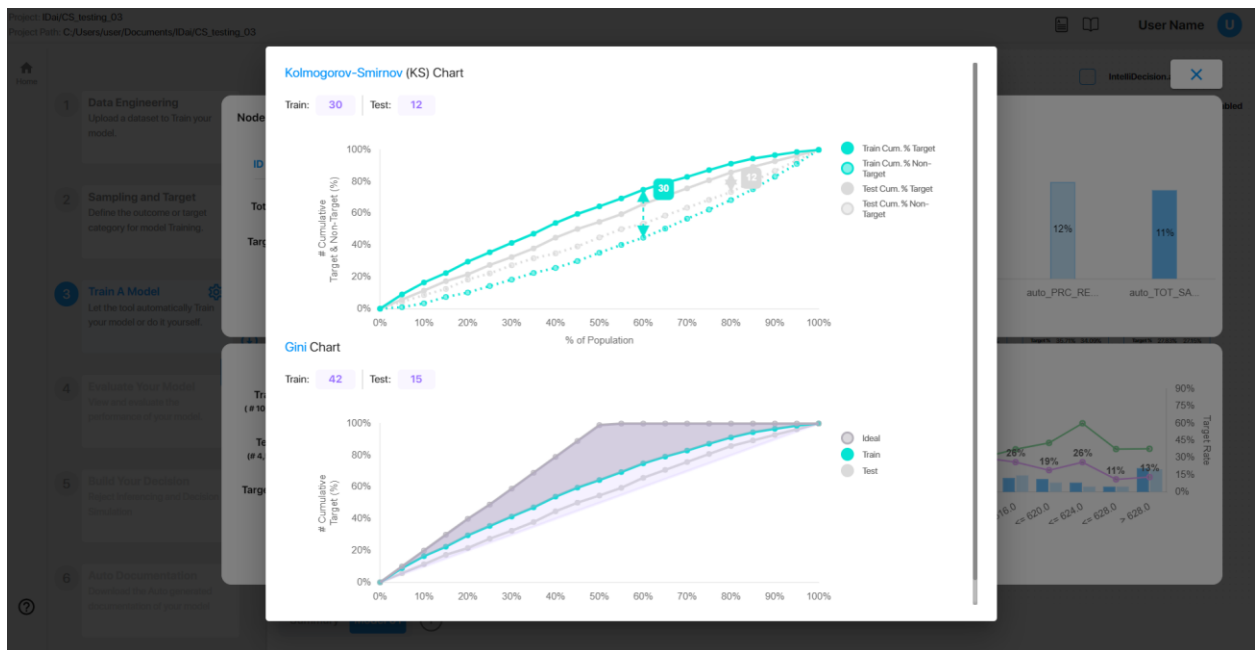
2. Logistic regression technical details



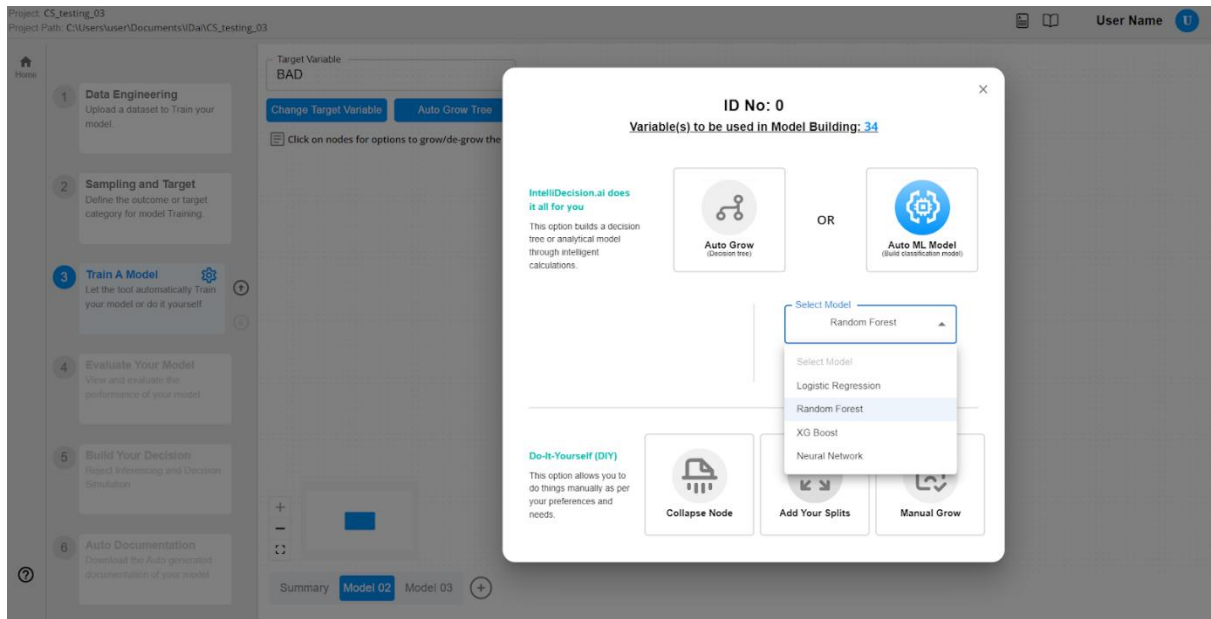
3. Scorecard details for that node



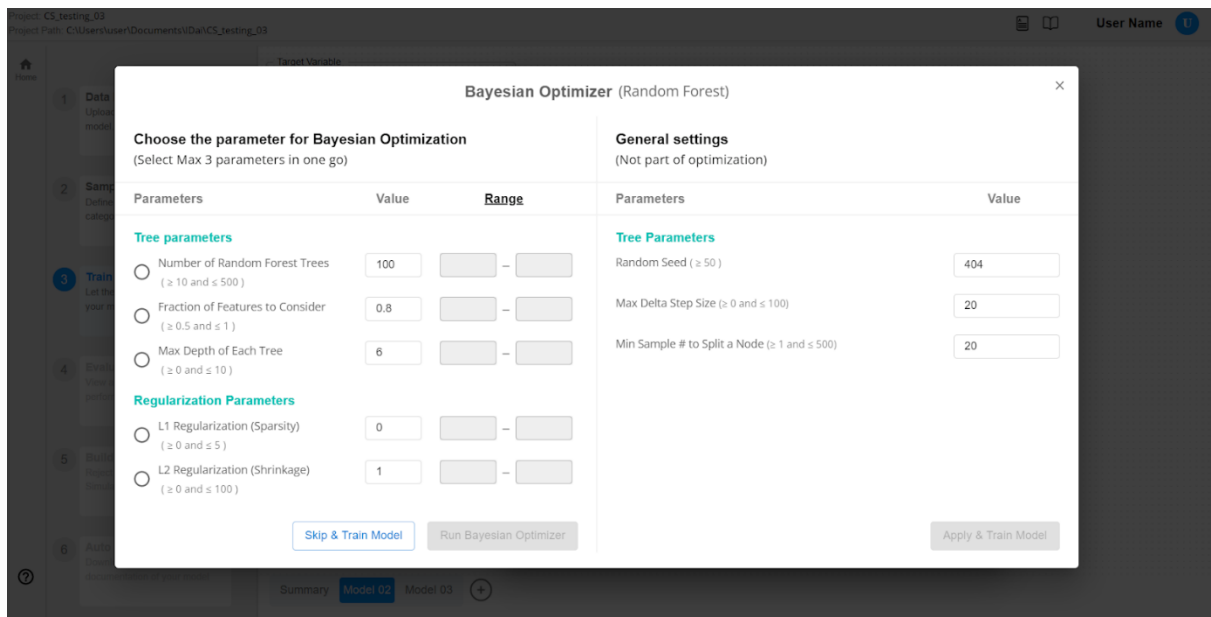
4. User can click on “**Click to see the graph**” in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.



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.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Target Variable

Bayesian Optimizer (Random Forest)

Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)

Parameters	Value	Range
Tree parameters		
<input checked="" type="radio"/> Number of Random Forest Trees (≥ 10 and ≤ 500)	100	10 - 500
<input checked="" type="radio"/> Fraction of Features to Consider (≥ 0.5 and ≤ 1)	0.8	0.5 - 1
<input type="radio"/> Max Depth of Each Tree (≥ 0 and ≤ 10)	6	-
Regularization Parameters		
<input checked="" type="radio"/> L1 Regularization (Sparsity) (≥ 0 and ≤ 5)	0	0 - 5
<input type="radio"/> L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)	1	-

Skip & Train Model Run Bayesian Optimizer

General settings
(Not part of optimization)

Parameters	Value
Tree Parameters	
Random Seed (≥ 50)	404
Max Delta Step Size (≥ 0 and ≤ 100)	20
Min Sample # to Split a Node (≥ 1 and ≤ 500)	20

Apply & Train Model

Summary Model 02 Model 03 +

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Target Variable

Bayesian Optimizer (Random Forest)

Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)

Parameters	Value	Range
Tree parameters		
<input checked="" type="radio"/> Number of Random Forest Trees (≥ 10 and ≤ 500)	100	10 - 500
<input checked="" type="radio"/> Fraction of Features to Consider (≥ 0.5 and ≤ 1)	0.8	0.5 - 1
<input type="radio"/> Max Depth of Each Tree (≥ 0 and ≤ 10)	6	-
Regularization Parameters		
<input type="radio"/> L1 Regularization (Sparsity) (≥ 0 and ≤ 5)	0	-
<input checked="" type="radio"/> L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)	1	0 - 100

Skip & Train Model Please wait...

General settings
(Not part of optimization)

Parameters	Value
Tree Parameters	
Random Seed (≥ 50)	404
Max Delta Step Size (≥ 0 and ≤ 100)	20
Min Sample # to Split a Node (≥ 1 and ≤ 500)	20

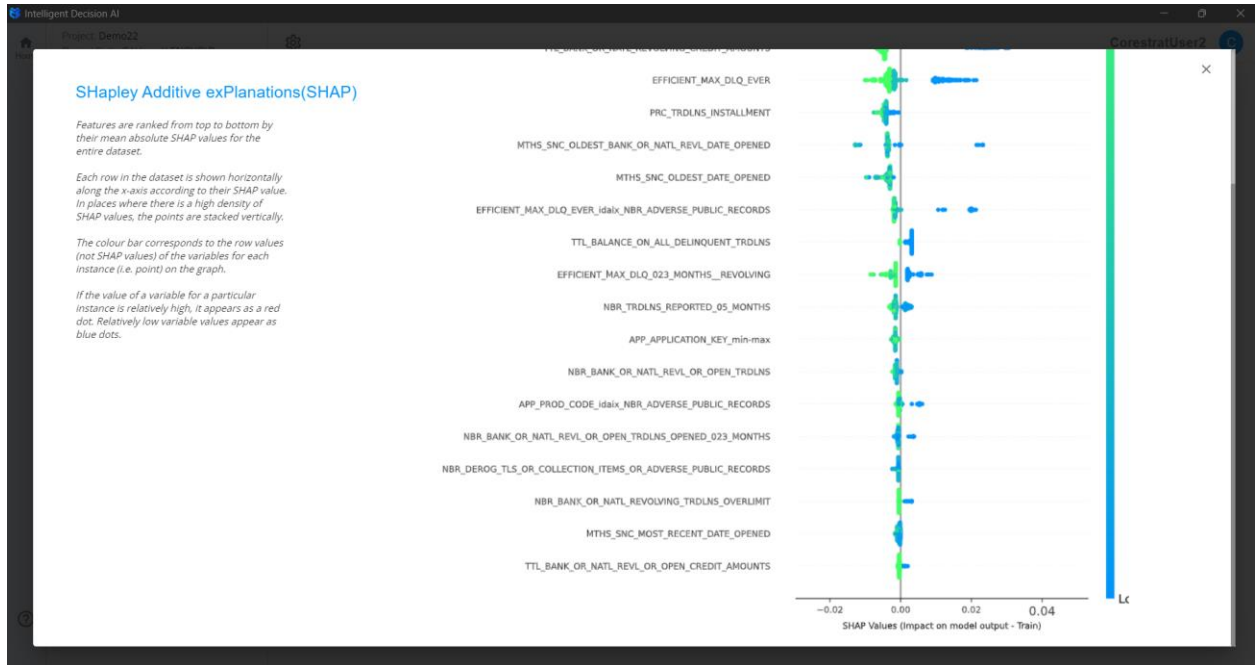
Apply & Train Model

Summary Model 02 Model 03 +

[illegible]

- 94 | Page

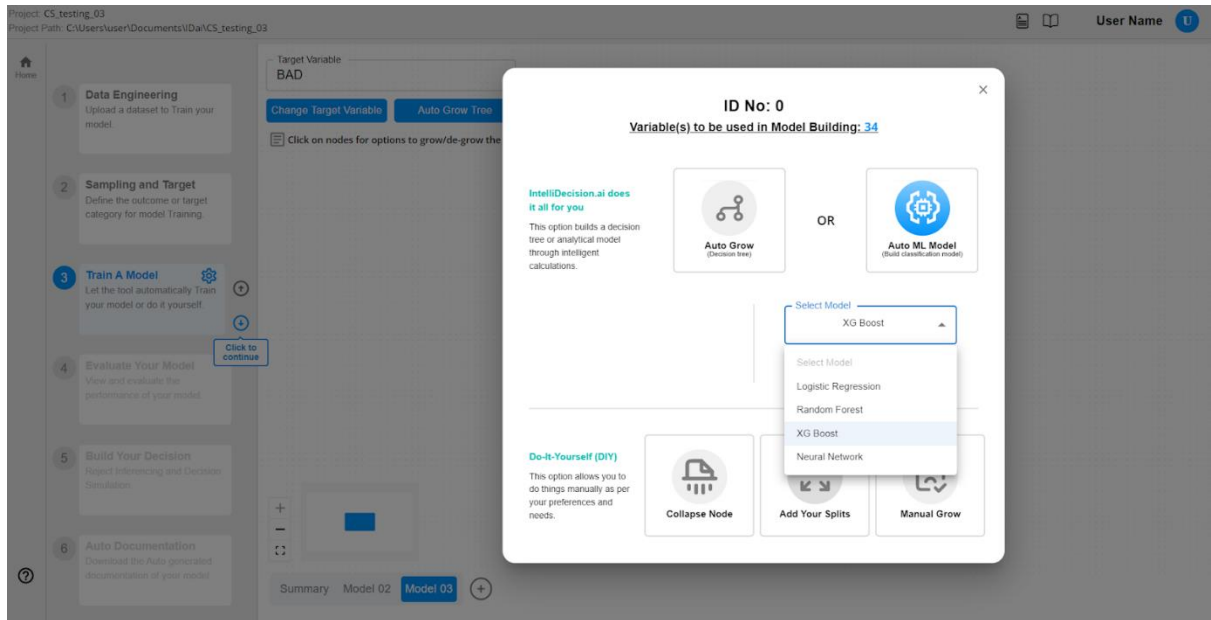




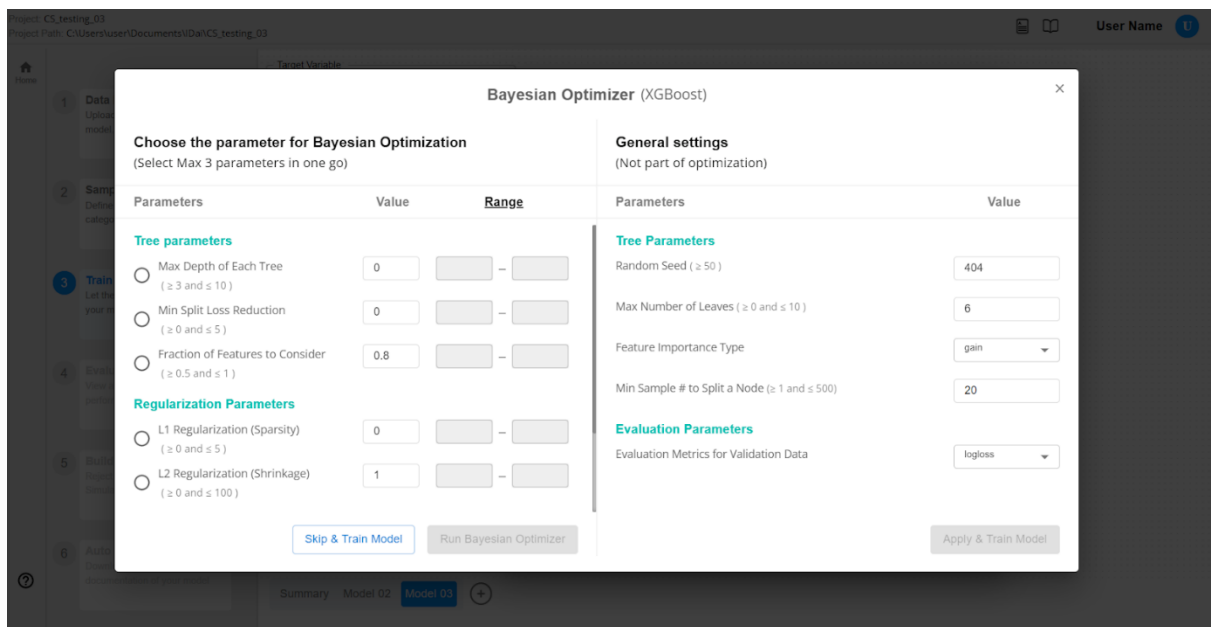
- User can click on **“Click to view SHAP chart”** to see the SHAP chart, s(he) can also select **“Click to see the graph”** in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.



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.

Project: CS_testing_03
Project Path: C:\Users\user\Documents\DA\CS_testing_03

Target Variable

Bayesian Optimizer (XGBoost)

Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)

Parameters	Value	Range
Tree parameters		
<input checked="" type="radio"/> Max Depth of Each Tree (≥ 3 and ≤ 10)	0	3 - 10
<input type="radio"/> Min Split Loss Reduction (≥ 0 and ≤ 5)	0	-
<input checked="" type="radio"/> Fraction of Features to Consider (≥ 0.5 and ≤ 1)	0.8	0.5 - 1
Regularization Parameters		
<input type="radio"/> L1 Regularization (Sparsity) (≥ 0 and ≤ 5)	0	-
<input checked="" type="radio"/> L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)	1	0 - 100

Skip & Train Model Run Bayesian Optimizer

General settings
(Not part of optimization)

Parameters	Value
Tree Parameters	
Random Seed (≥ 50)	400
Max Number of Leaves (≥ 0 and ≤ 10)	6
Feature Importance Type	gain
Min Sample # to Split a Node (≥ 1 and ≤ 500)	20
Evaluation Parameters	
Evaluation Metrics for Validation Data	logloss

Apply & Train Model

Project: CS_testing_03
Project Path: C:\Users\user\Documents\DA\CS_testing_03

Target Variable

Bayesian Optimizer (XGBoost)

Choose the parameter for Bayesian Optimization
(Select Max 3 parameters in one go)

Parameters	Value	Range
Tree parameters		
<input checked="" type="radio"/> Max Depth of Each Tree (≥ 3 and ≤ 10)	0	3 - 10
<input type="radio"/> Min Split Loss Reduction (≥ 0 and ≤ 5)	0	-
<input checked="" type="radio"/> Fraction of Features to Consider (≥ 0.5 and ≤ 1)	0.8	0.5 - 1
Regularization Parameters		
<input type="radio"/> L1 Regularization (Sparsity) (≥ 0 and ≤ 5)	0	-
<input checked="" type="radio"/> L2 Regularization (Shrinkage) (≥ 0 and ≤ 100)	1	0 - 100

Skip & Train Model Please wait...

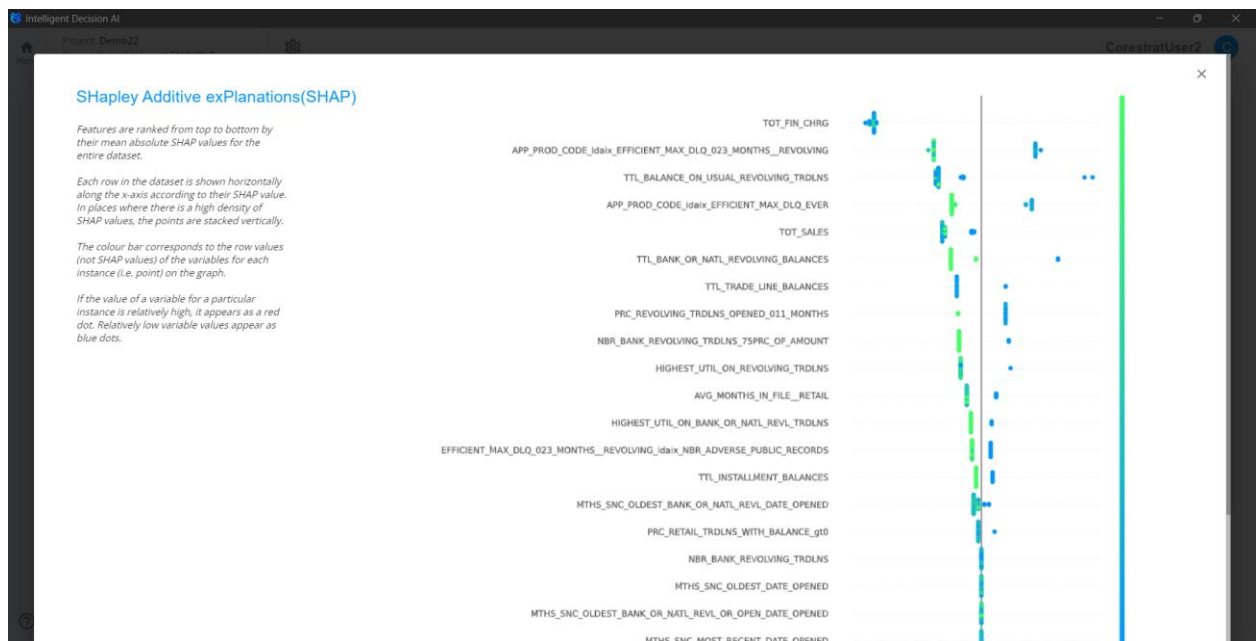
General settings
(Not part of optimization)

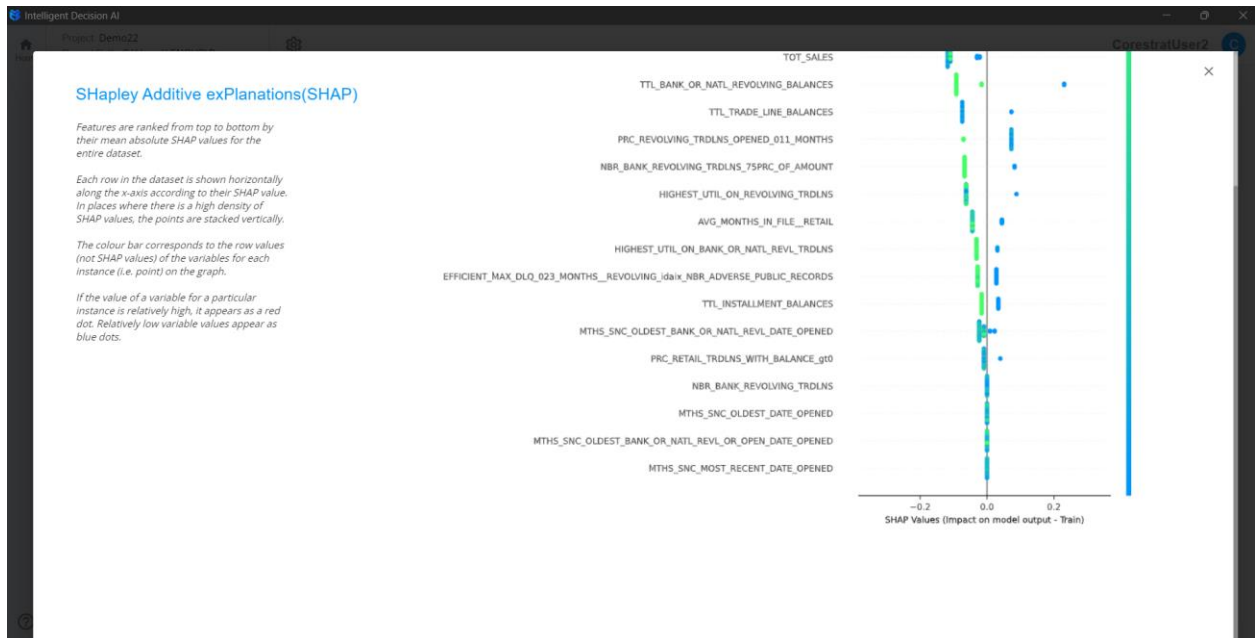
Parameters	Value
Tree Parameters	
Random Seed (≥ 50)	400
Max Number of Leaves (≥ 0 and ≤ 10)	6
Feature Importance Type	gain
Min Sample # to Split a Node (≥ 1 and ≤ 500)	20
Evaluation Parameters	
Evaluation Metrics for Validation Data	logloss

Apply & Train Model

[illegible]

www.corestrat.ai 99 | Page

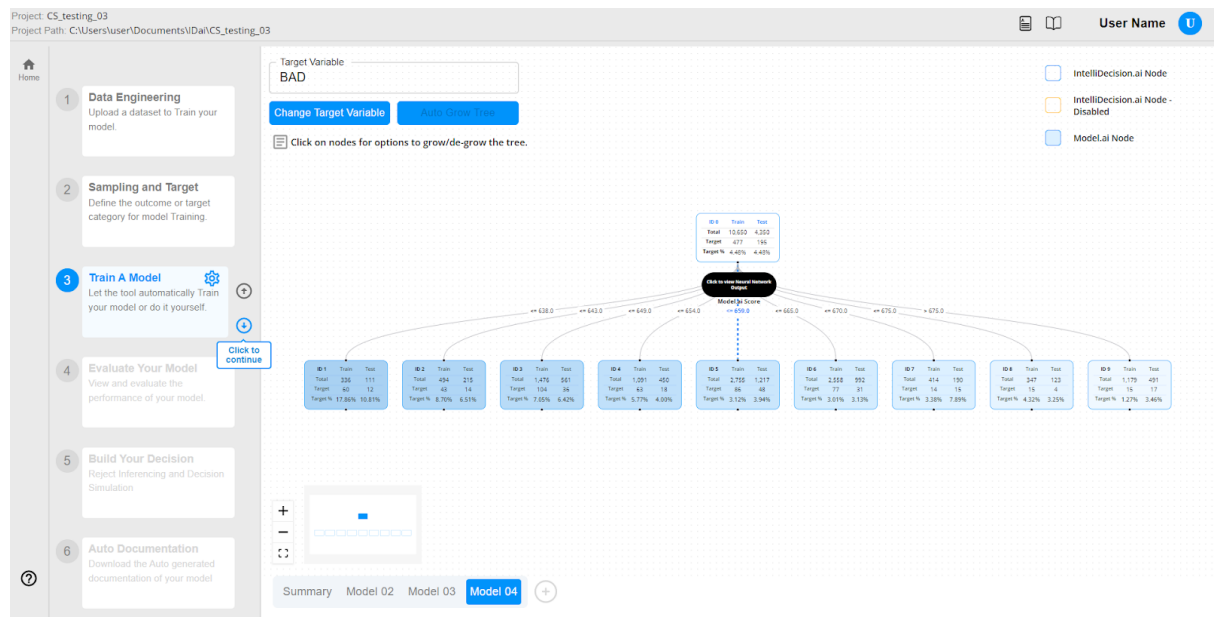
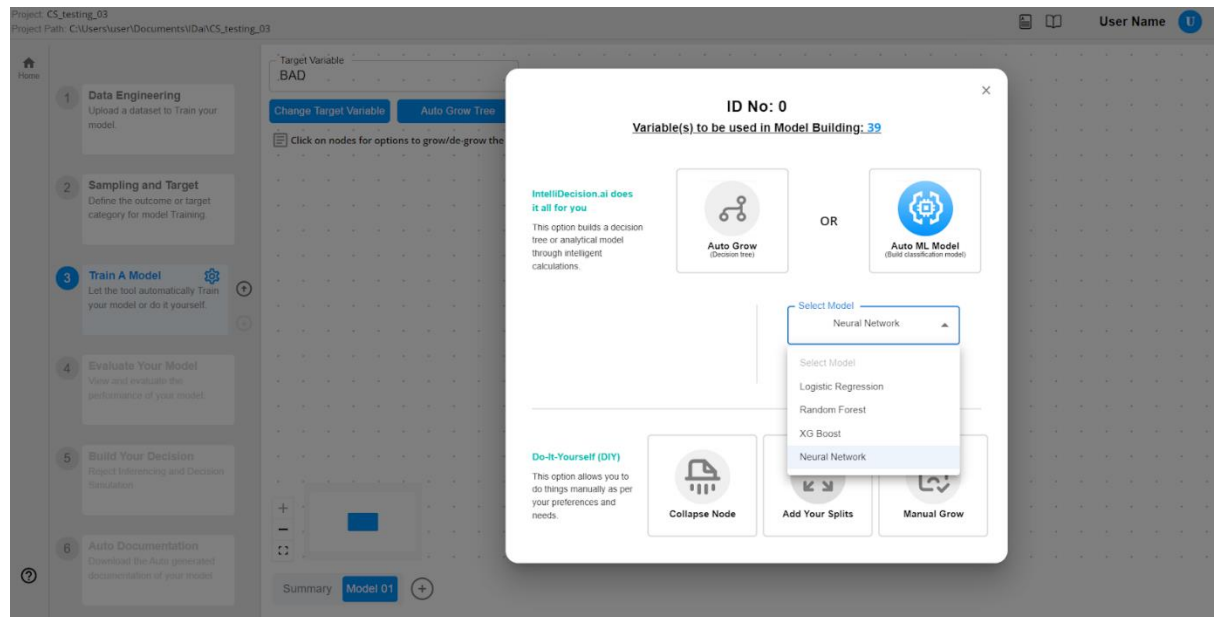




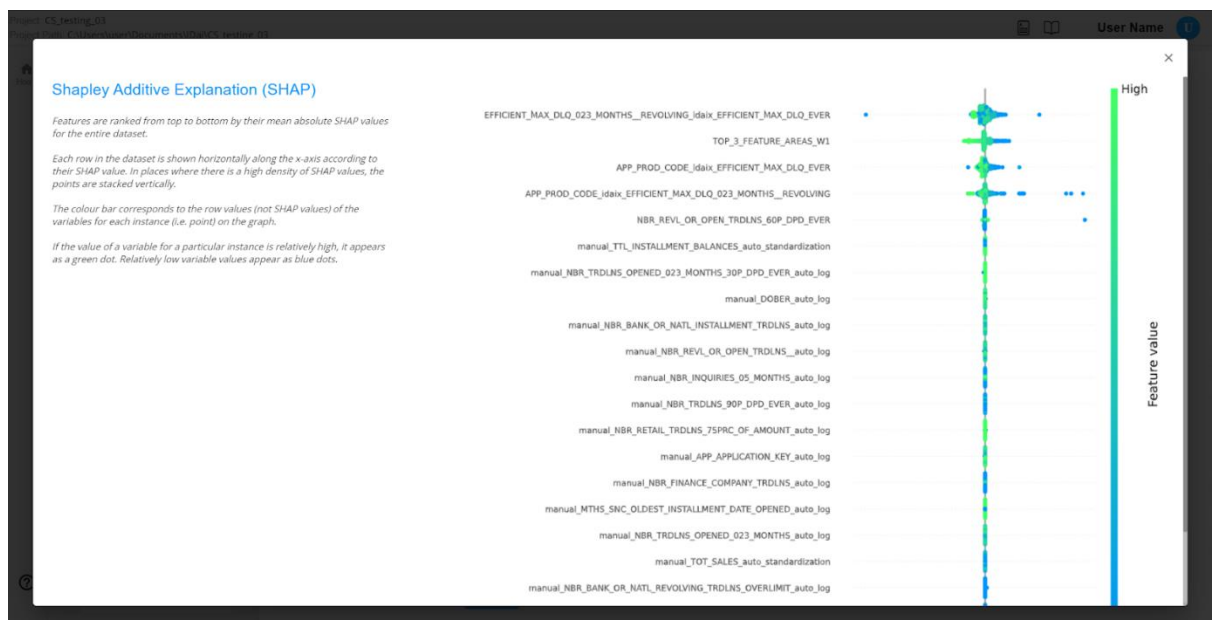
4. User can click on **“Click to view SHAP chart”** to see the SHAP chart, s(he) can also select **“Click to see the graph”** in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test samples.



1. Neural Networks Model

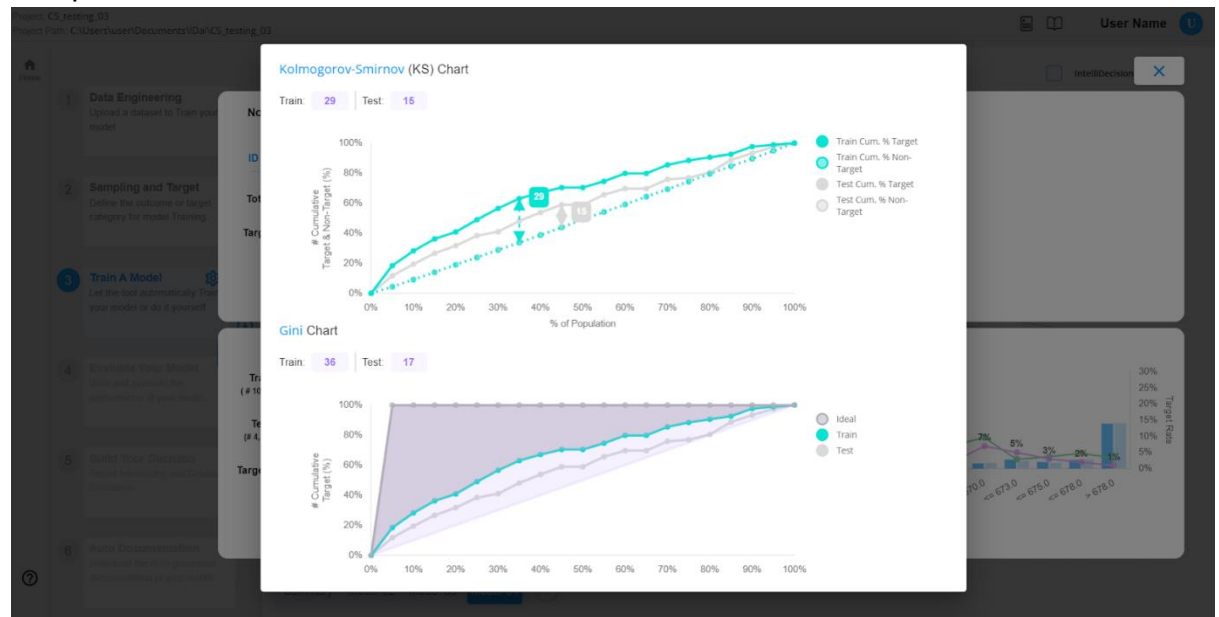


2. Neural Network provides a bar graph of variable importance and a SHAP chart along with KS and GINI charts. which can be viewed by clicking on '**Click to view Neural Network Output**'.

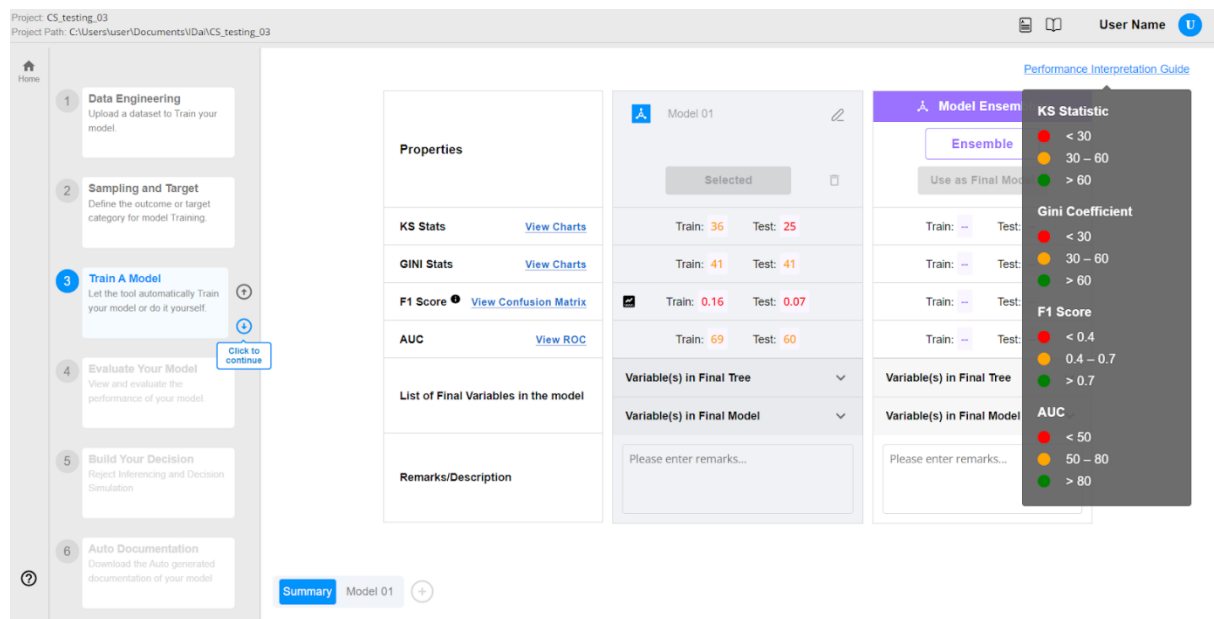


3. User can click on “**Click to view SHAP chart**” to see the SHAP chart, s(he) can also select “**Click to see the graph**” in the model performance metrics tab to get a visual representation of KS and Gini values for the train and test

samples.



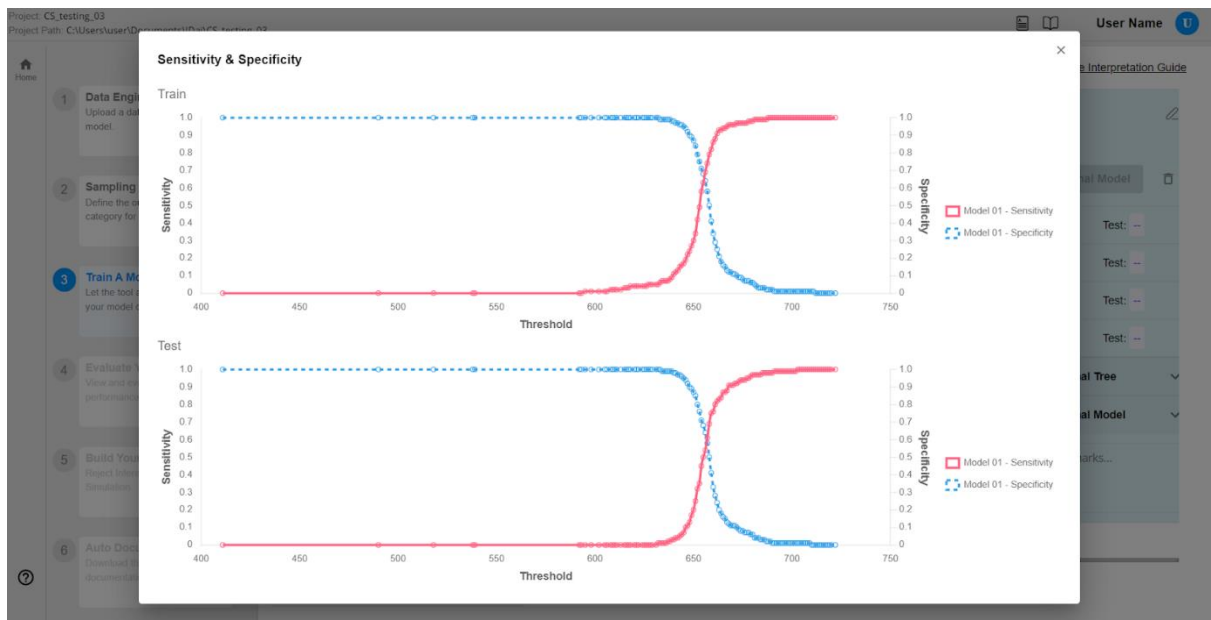
On the ‘**Summary**’ screen, there is a ‘**Performance Interpretation Guide**’ provided to enable the users to make sense of the color depictions used on this screen



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.



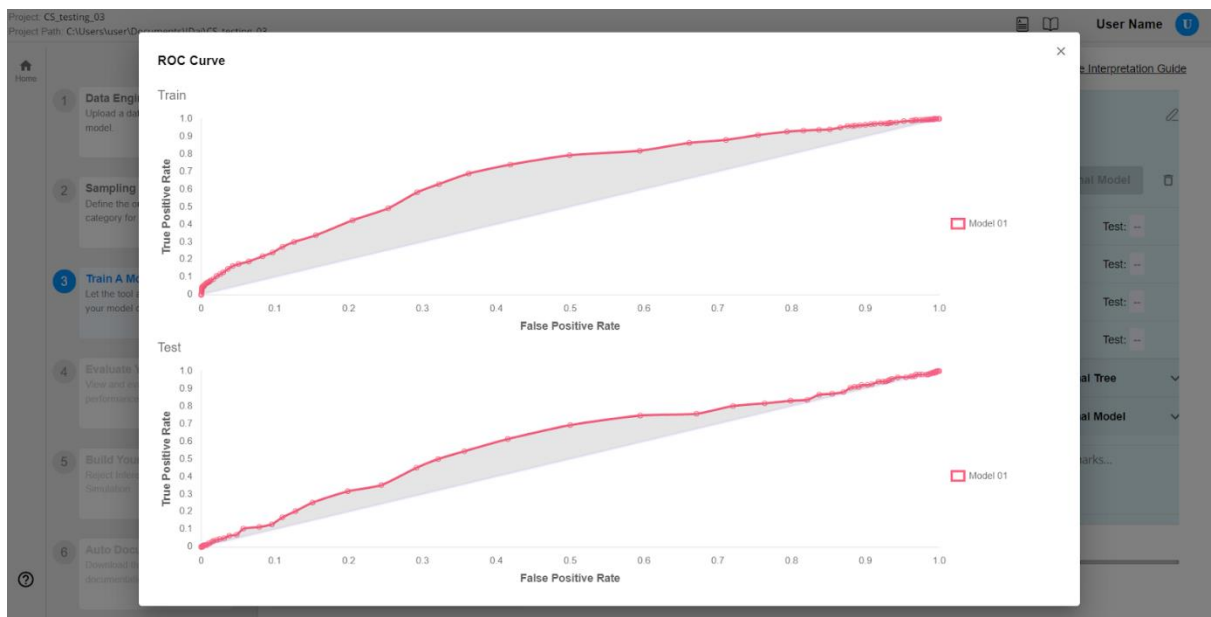
This page helps users visualize how the model's **sensitivity** (true positive rate) and **specificity** (true negative rate) change with different threshold values.

- 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\IDa\CS_testing_03

Home

- Data Engineering**
Upload a dataset to Train your model.
- Sampling and Target**
Define the outcome or target category for model Training.
- Train A Model**
Let the tool automatically Train your model or do it yourself.
[Click to continue](#)
- Evaluate Your Model**
View and evaluate the performance of your model.
- Build Your Decision**
Reject Inferencing and Decision Simulation
- Auto Documentation**
Download the Auto generated documentation of your model

[Summary](#) Model 02 Model 03 +

Properties

[View Charts](#)

[View Charts](#)

[View Confusion Matrix](#)

[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...

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Home

- Data Engineering**
Upload a dataset to Train your model.
- Sampling and Target**
Define the outcome or target category for model Training.
- Train A Model**
Let the tool automatically Train your model or do it yourself.
[Click to continue](#)
- Evaluate Your Model**
View and evaluate the performance of your model.
- Build Your Decision**
Reject Inferencing and Decision Simulation
- Auto Documentation**
Download the Auto generated documentation of your model

[Summary](#) Model 02 Model 03 +

Properties

[View Charts](#)

[View Charts](#)

[View Confusion Matrix](#)

[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_MTH5_SNC_OLDEST_REV...

☐ APP_PROD_CODE_premier

☐ NBR_BANK_OR_NATI_REVOLVIN...

☐ manual_HIGHEST_UTIL_ON_BAN...

Select Method(s) of Ensembling

[Apply](#)

Please enter remarks...

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

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

Performance Interpretation Guide

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)

Select Method(s) of Ensembling

☐ Average

☒ Best

☐ Worst

☐ Logistic Regression

Apply

Summary Model 02 Model 03

Clicking 'Apply' will initiate the ensembling process.

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

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

Performance Interpretation Guide

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)

Select Method(s) of Ensembling

☐ Average

☒ Best

☐ Worst

☐ Logistic Regression

Please Wait...

Summary Model 02 Model 03

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Performance Interpretation Guide

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...

Model 02

Selected

Train: 55 Test: 49

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 Successful!

Close

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

Summary Model 02 Model 03

Project: CS_testing_03
Project Path: C:\Users\User\Documents\IDa\CS_testing_03

Performance Interpretation Guide

Model Ensembling

Ensemble

Use as Final Model

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 02

Selected

Train: 55 Test: 49

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...

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

Summary Model 02 Model 03

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.

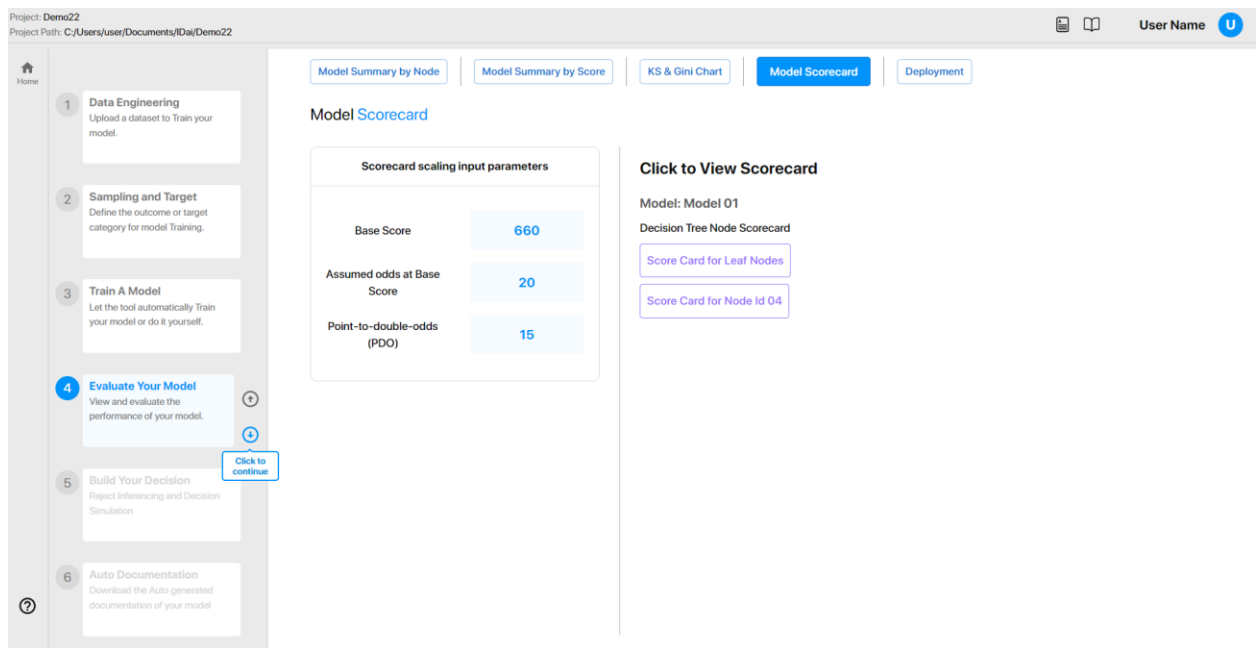
- Model summary by Node – Information on all nodes used in the model, the counts and bad rates in both Train and Test samples.



- 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

Project: Demo22
Project Path: C:/Users/user/Documents/IDa/Demo22

Model Summary by Node | Model Summary by Score | KS & Gini Chart | **Model Scorecard** | Deployment

Score Card for Node Id 04

Base Score: 660
Assumed odds at Base Score: 20
Point-to-double-odds (PDO): 15

#	Variable Name	Variable Categories	#Cases (Training)	#Targets (Training)	%Target Rate (Training)	Score	#Cases (Test)	#Targets (Test)	%Target Rate (Test)
1	TOP_3_FEATURE_AREA_S_W1	db_rp_cv_sh_ws_db_sh_rp_gsl_sh_rp_db_gsl_cv_ws_sh_cv_gsl_gsl_rp...	104	88	84.62 %	44	17	9	52.94 %
		gsl_cv_gsl_sh_gsl_cv_sh_ws_cv_cv_gsl_rp_ws_rp_oth_cv_gse_s_at_ga...	91	76	83.52 %	46	35	17	48.57 %
		sh_oth_hme_cv_gsl_gsl_cv_gsl_rp_ws_gse_cv_gse_gsl	88	54	61.36 %	74	33	19	57.58 %
		gse_ws_cv_sh_gse_oth_db_oth_hme_ws_cv_gsl	146	78	53.42 %	81	70	31	44.29 %
		rp_db_ws_rp_cv_gsl_rp_gse_oth_rp_gse_ws_gsl_ws_cv_gse_at_cv_g...	143	71	49.65 %	85	53	23	43.4 %
		oth_hme_ai	640	306	47.81 %	87	270	134	49.63 %
		cv_gsl_gse_db_ws_oth	97	45	46.39 %	88	45	26	57.78 %
		oth_cv_gsl_gse_cv_ws_cv_gsl_oth_gsl_cv_oth_cv_sh_gsl_gsl_gse_cv...	198	84	42.42 %	92	83	38	45.78 %
		gsl_oth_hme_at_cv_gsl_ws_db_sh_rp_gsl_cv_sh_gsl_oth_rp_sh_ws...	169	31	18.34 %	120	61	31	50.82 %
		missing_idai_imputed	0	0	0 %	85	52	25	48.08 %
		(-inf, 6.99]	41	18	43.9 %	90	20	9	45 %

6 Auto Documentation
Download the Auto generated documentation of your model

- 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

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 Inference 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

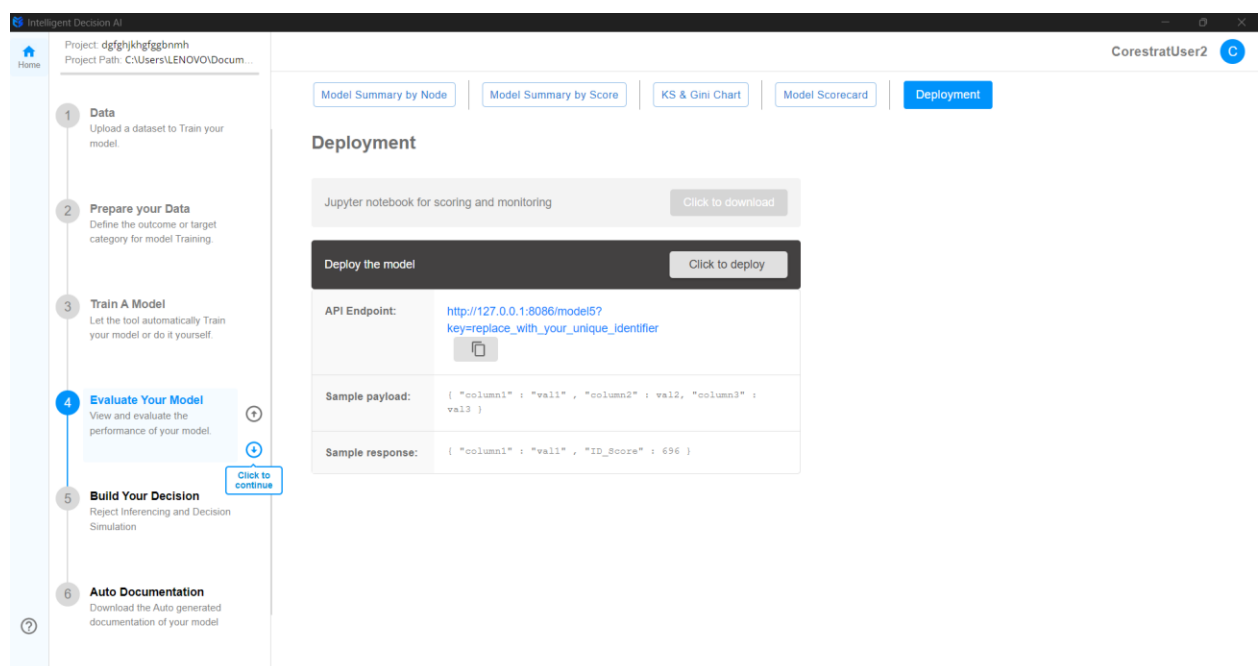
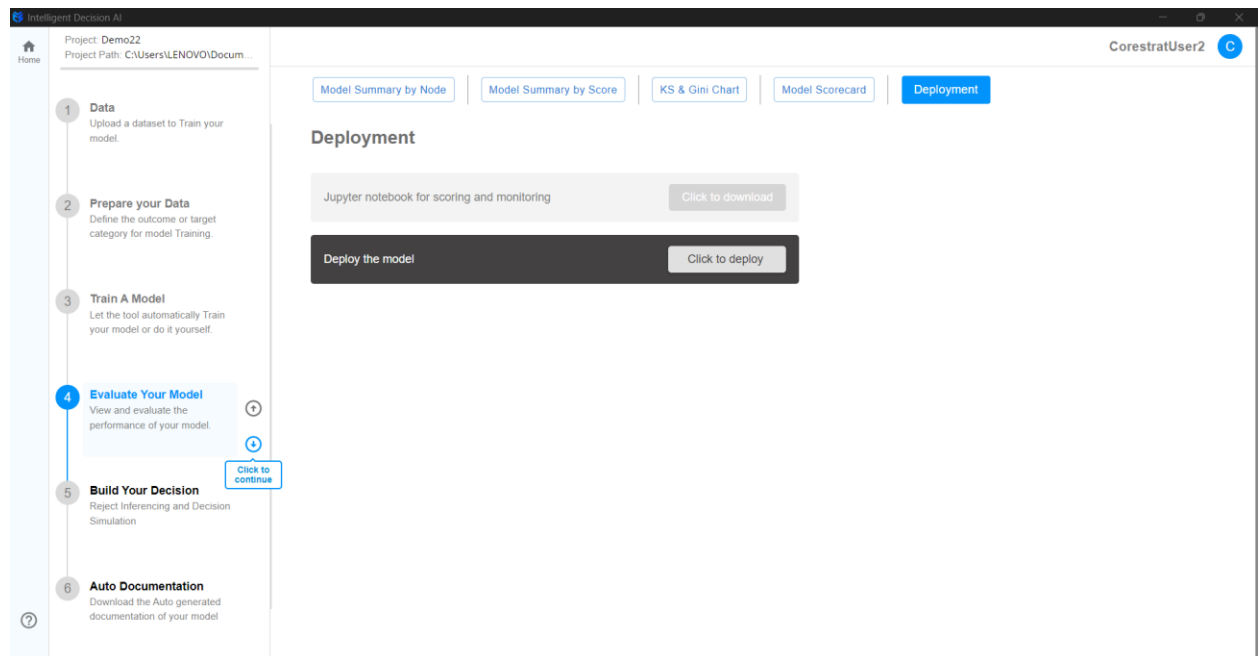
Model: Model 01
Decision Tree Node Scorecard

[Score Card for Leaf Nodes](#)

[Shapley values for Node Id 04](#)



- **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.



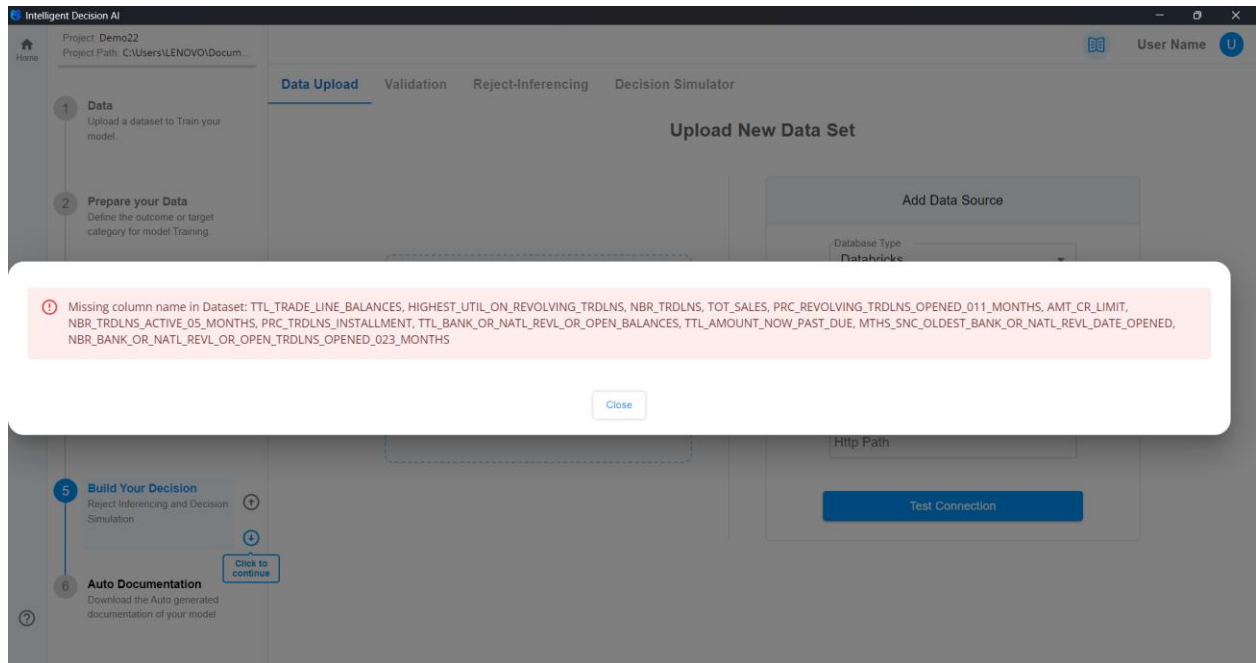
Build Your Decision

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)

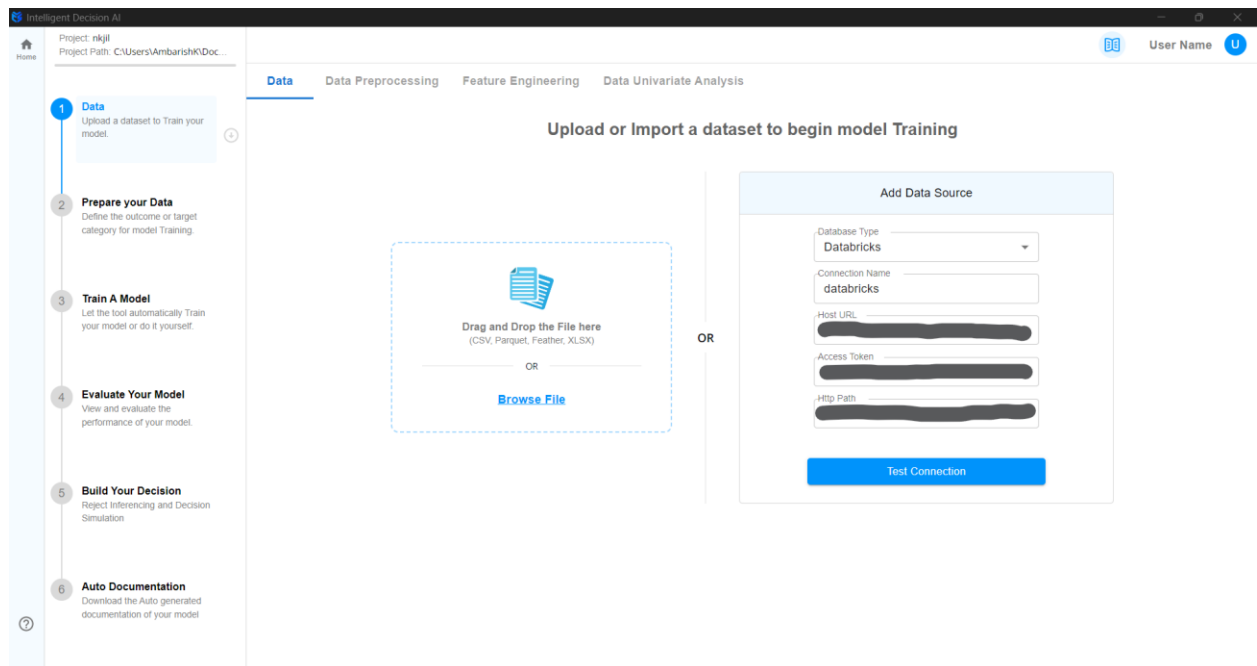
The screenshot displays the CORESTRAT IntelliDecision.ai web application interface. At the top, the project name 'Demo22' and path 'C:/Users/user/Documents/IDai/Demo22' are shown. The user's name and a profile icon are in the top right. The main navigation bar includes 'Data Upload', 'Validation', 'Reject-Inferencing', and 'Decision Simulator'. The 'Data Upload' tab is active, showing the 'Upload New Data Set' section. This section has two options: 'Drag and Drop the File here (CSV, Parquet, Feather, XLSX)' with a 'Browse File' link, and 'Add Data Source' which includes fields for 'Database Type' (set to Databricks), 'Connection Name' (Databricks), 'Host URL', 'Access Token', and 'Http Path', followed by a 'Test Connection' button. A sidebar on the left shows a progress bar with six steps: 1. Data Engineering, 2. Sampling and Target, 3. Train A Model, 4. Evaluate Your Model, 5. Build Your Decision (current step), and 6. Auto Documentation. A 'Click to continue' button is next to step 5. A 'Next' button is at the bottom right.

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.

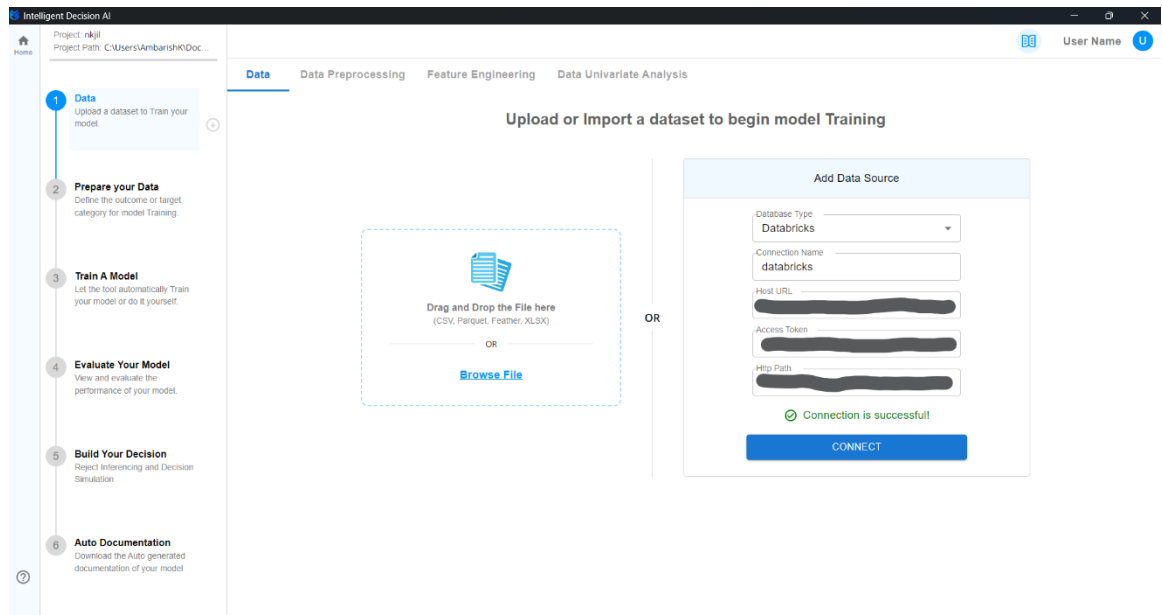


Data can also be added using Databricks. User needs to follow the below steps:

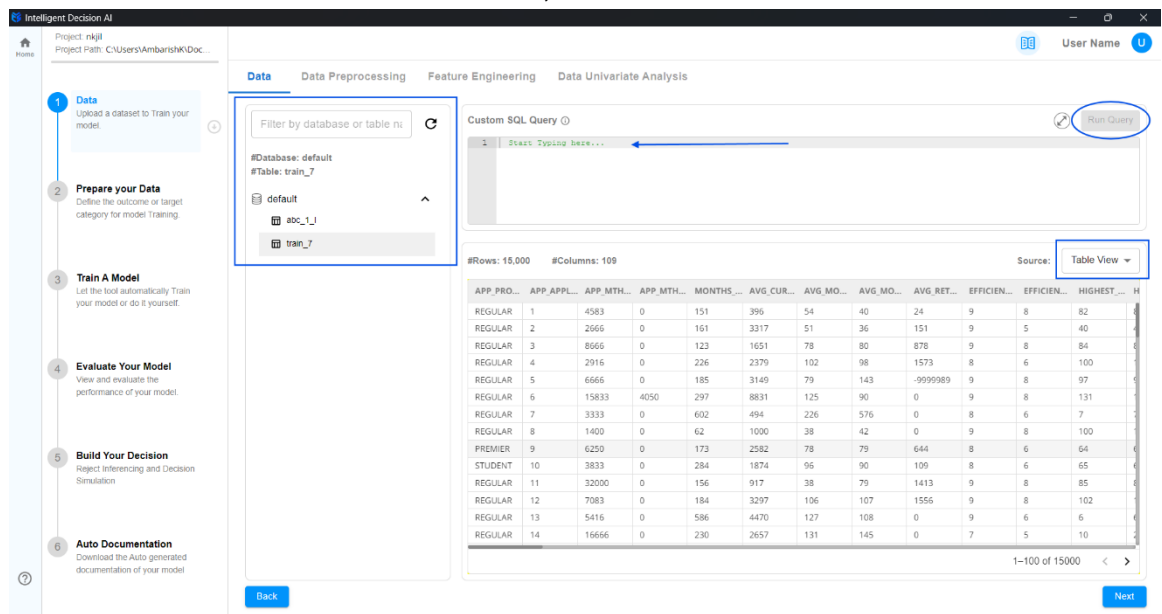
1. Enter the Host URL, Access Token and the Http Path in the allocated fields.
2. Click on Test Connection



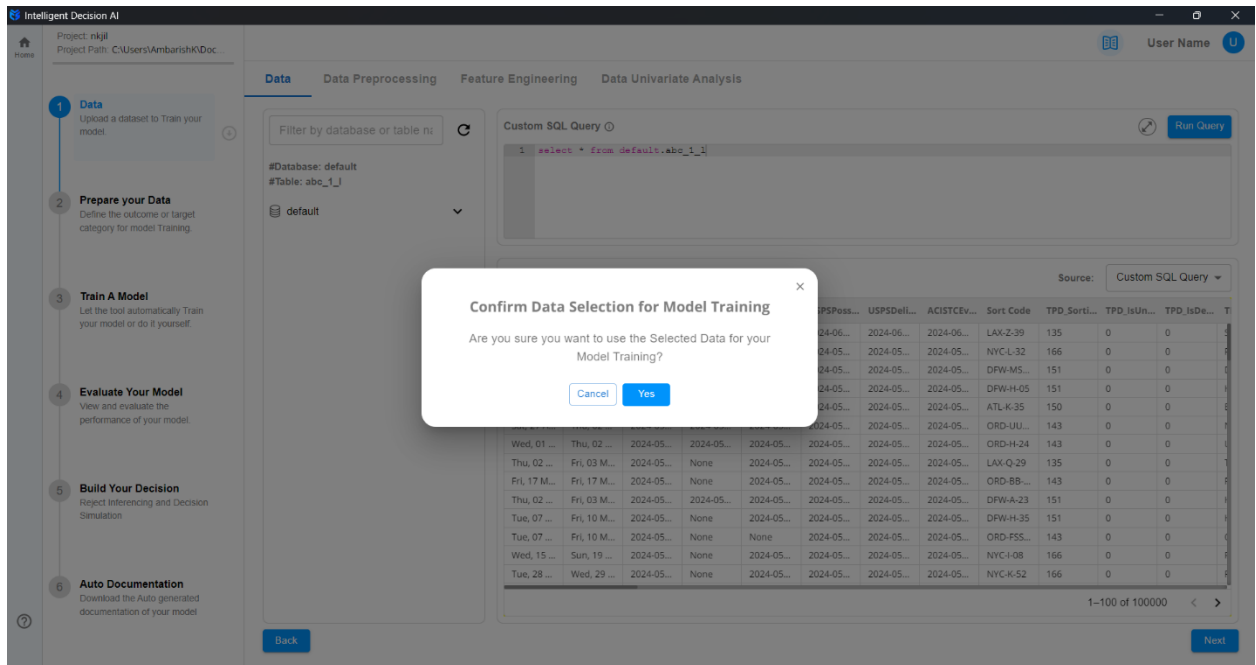
- Once it shows “**Connection successful**”, click on “**Connect**”.



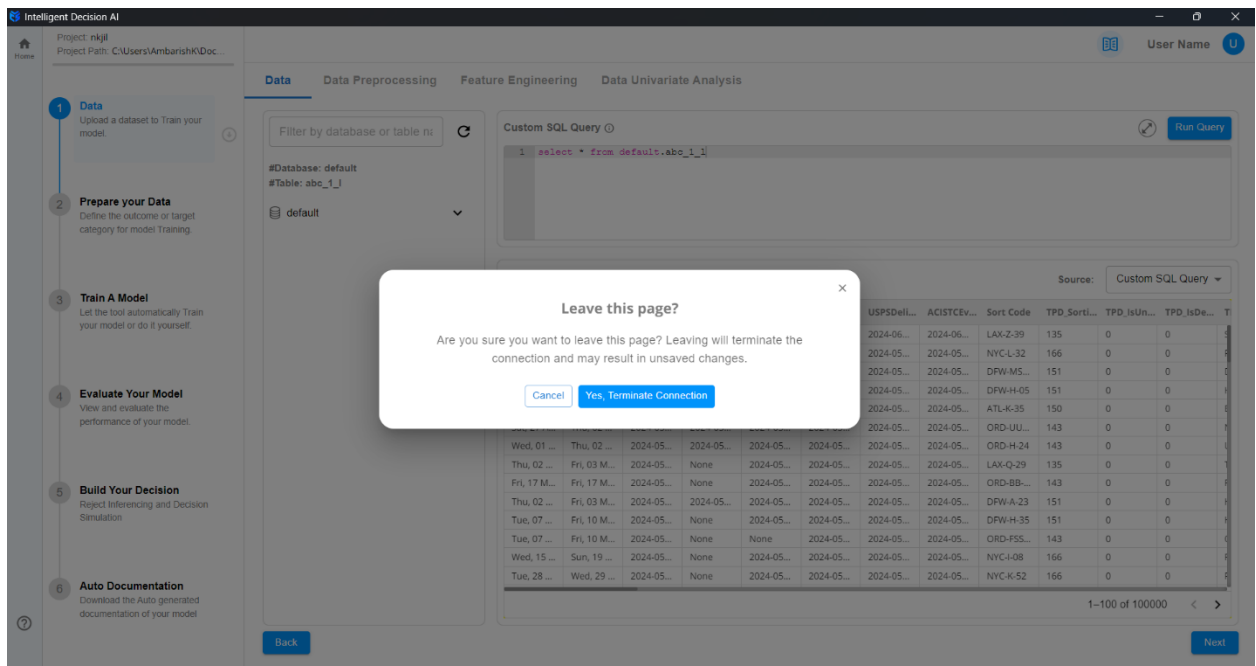
- User can choose the table from the left side from the database of need (refer image).
- Write a custom SQL Query in the editor provided to get data from the table based on user criteria. Once done, click on “**Run Query**” (refer image). This is an optional step.
- The source of the data shown can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



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



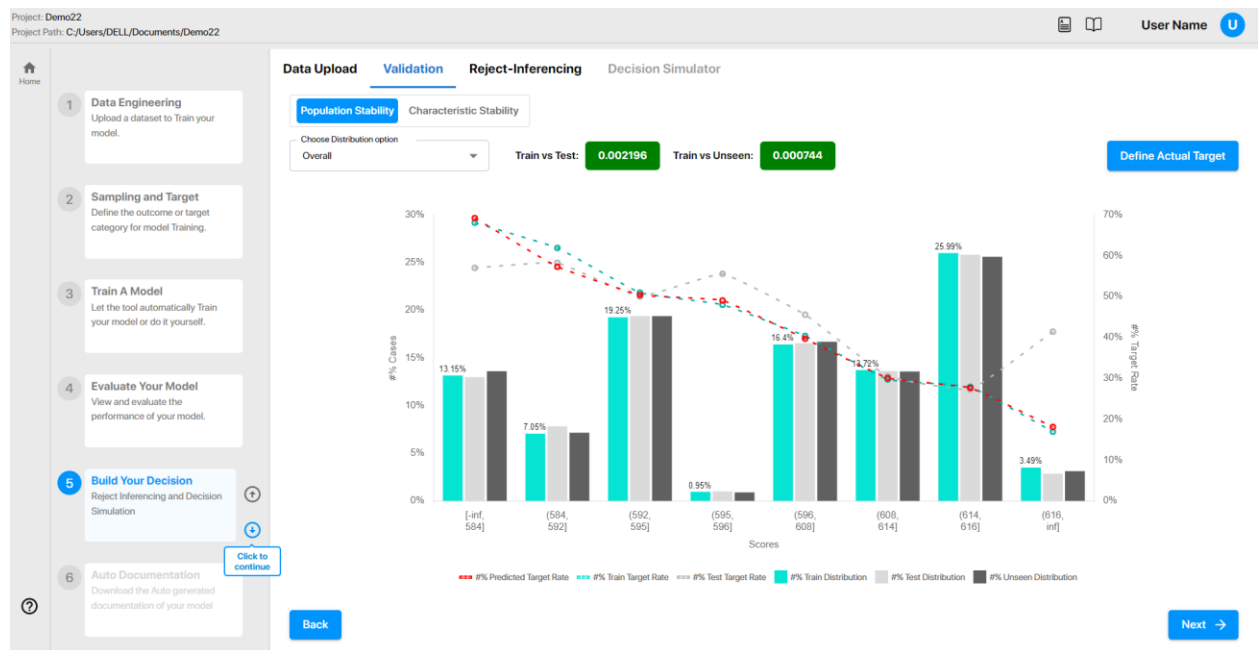
If you want to terminate the connection, click on the **“Back”** button as shown in the image and click on **“Yes Terminate Connection”** when prompted.



The next screen provides the population stability and characteristic stability.

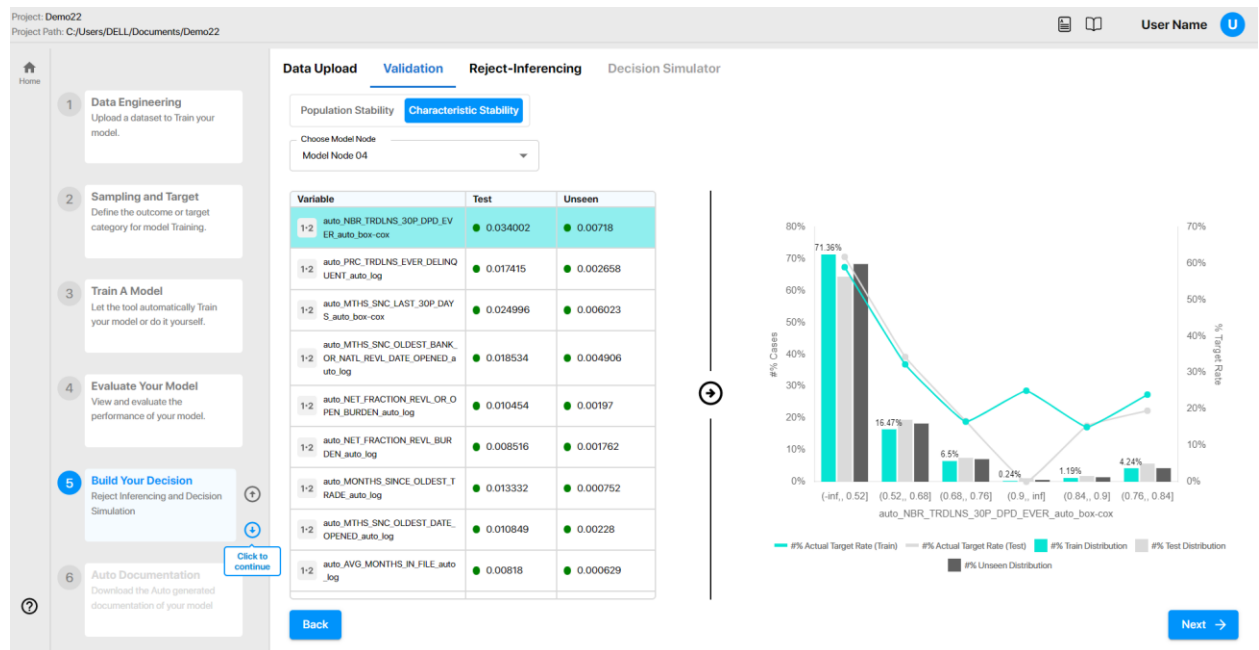
The Population Stability Index (PSI) measures how much the data used to train the model differs from new, unseen data. It helps check if the model's predictions are still reliable over time. A high PSI means the data has changed a lot, which could affect the model's performance.

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.

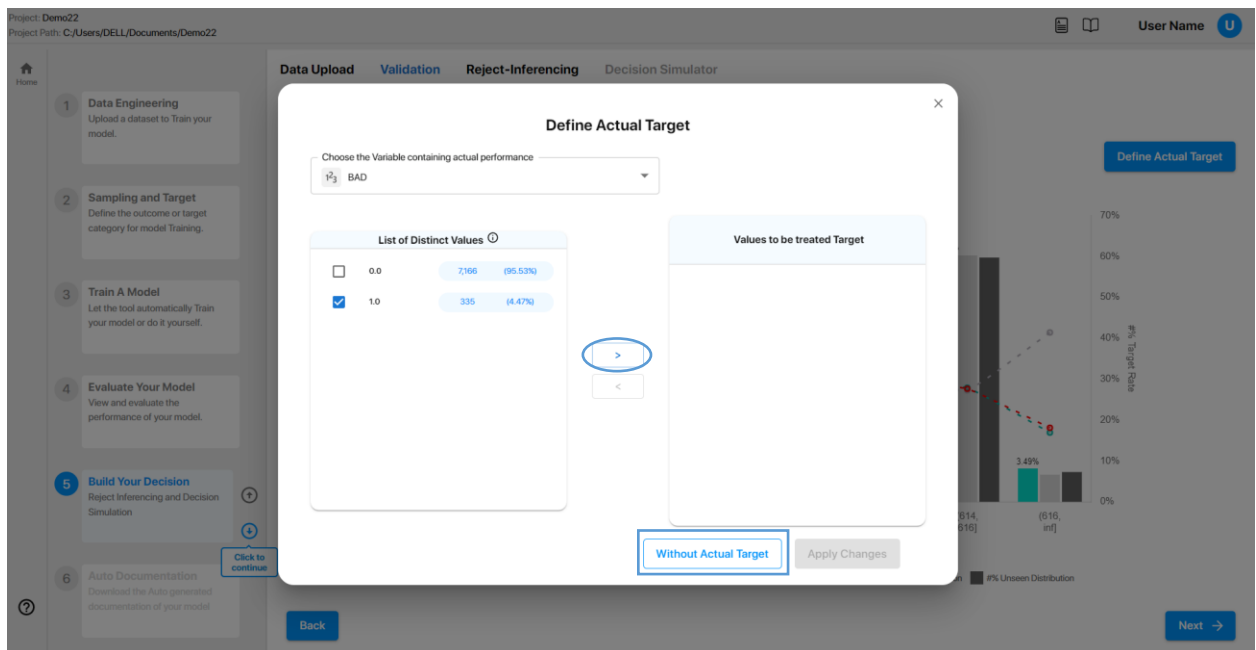


“Define Actual Target” button in Population stability section

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).

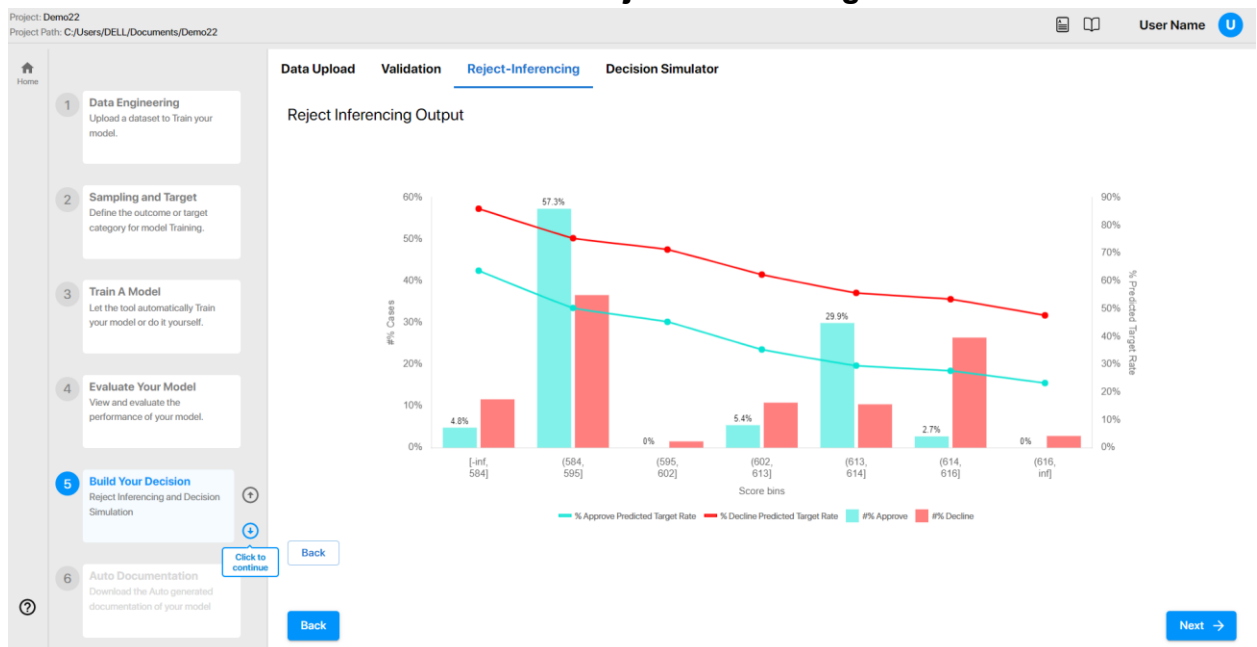
OR

- 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.

The screenshot shows the 'Reject-Inferencing' configuration screen. On the left is a sidebar with a progress bar showing six steps: 1. Data Engineering, 2. Sampling and Target, 3. Train A Model, 4. Evaluate Your Model, 5. Build Your Decision (current step), and 6. Auto Documentation. The main area has tabs for 'Data Upload', 'Validation', 'Reject-Inferencing' (active), and 'Decision Simulator'. Under 'Perform Reject Inferencing', there's a dropdown to 'Choose the Variable containing past decisions' set to 'BAD'. Below this is a 'List of Distinct Values' table with two rows: 0.0 (7,166 cases, 95.53%) and 1.0 (335 cases, 4.47%). The 1.0 row is selected. To the right of the table is a 'Values to be treated as pass decision = "Approved"' box. Further right, under 'Select Methods', the 'Reject Inferencing Method' is set to 'Parceling' and the 'Reject Inferencing Factor' is set to 3. At the bottom right, there are buttons for 'Without Reject Inferencing', 'Apply Changes', and 'Next'. A 'Back' button is located at the bottom left of the main area.

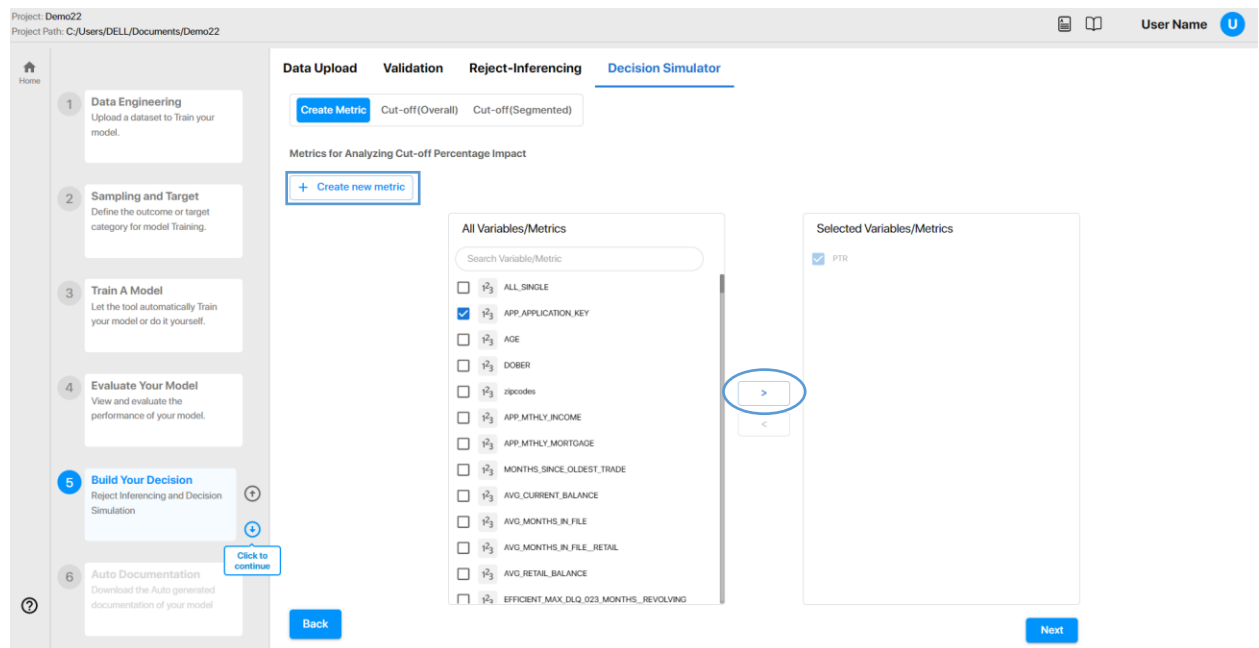
It would display the output as represented in below image. Click on the "**Back**" button on bottom left corner to return to "**Reject Inferencing**" screen.



Decision Simulator

Create Metric

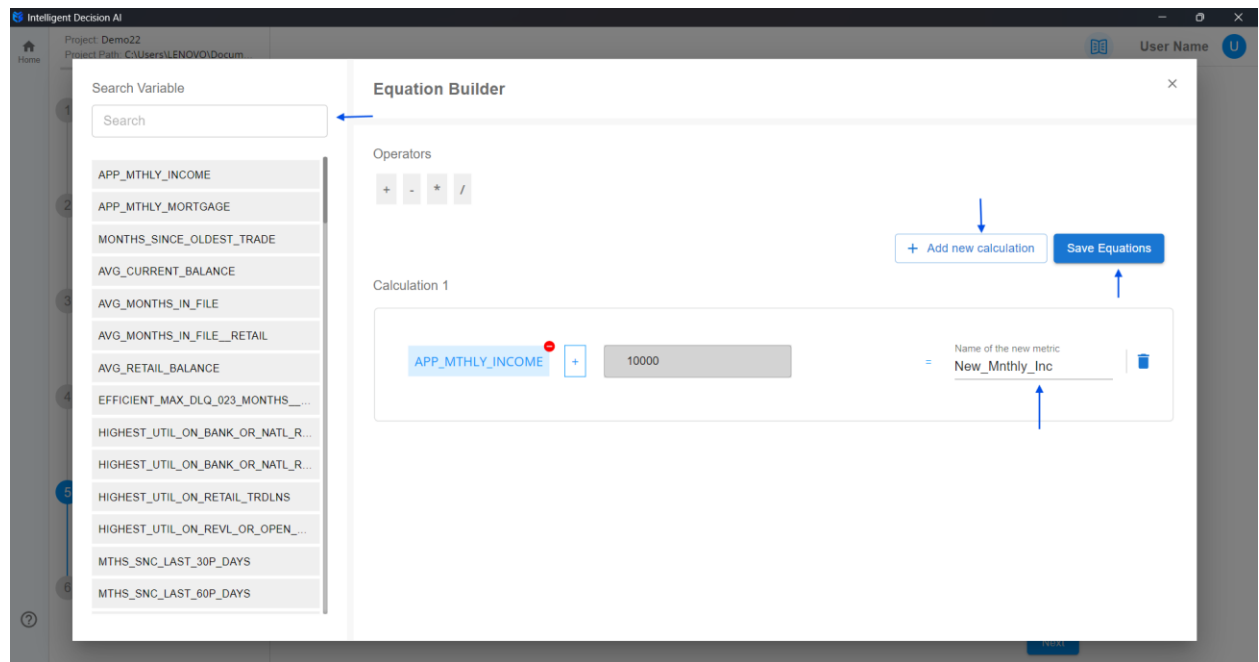
- User can select the metrics from the list of “**All Variables/Metrics**” from their uploaded dataset (refer image, blue oval)
- AND
- User can click on “**+ Create new metric**” to create new metrics of their own (refer image, blue rectangle)
- OR
- Click on “**Next**” button to skip this and proceed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

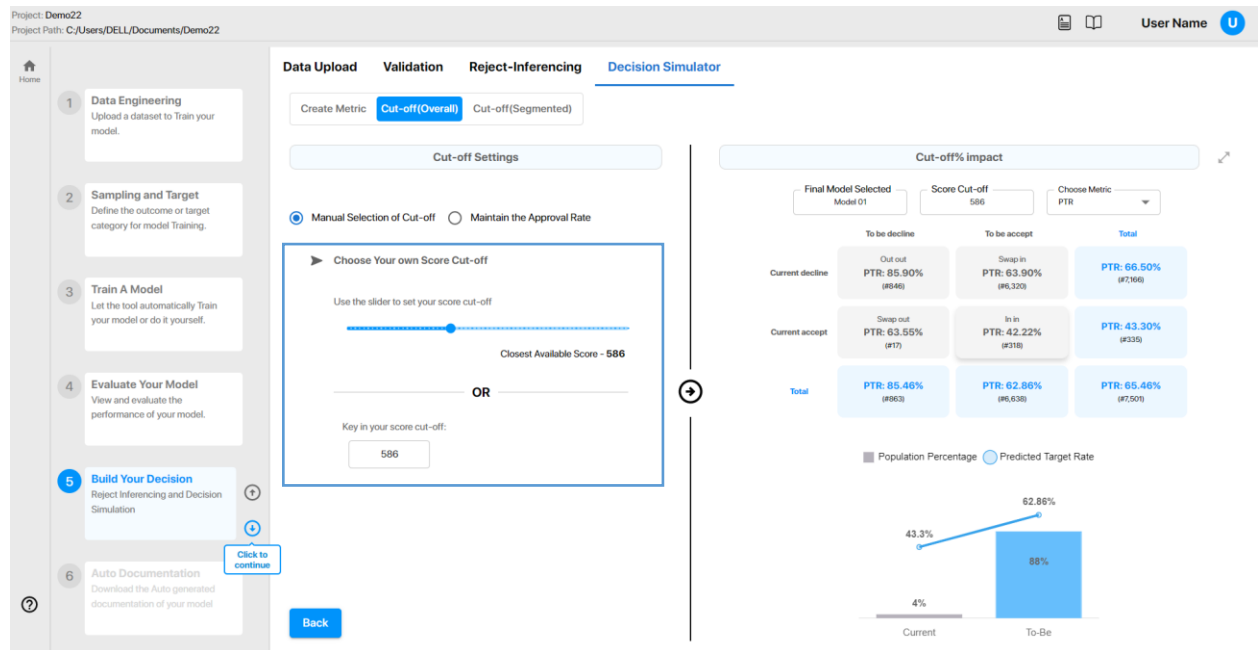
Users can also choose actual score point values to analyse the impact of this on target and approval rates.

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.

Project: Demo22
Project Path: C:/Users/DELL/Documents/Demo22

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

Click to continue

Back

Data Upload **Validation** **Reject-Inferencing** **Decision Simulator**

Create Metric Cut-off(Overall) **Cut-off(Segmented)**

Input Score splits separated by comma
600,650,700,750 **Apply**

Score	Total
[-inf, 584] (PTR: 85.55%)	844
(584, 595] (PTR: 73.65%)	2,813
(595, 602] (PTR: 71.28%)	106
(602, 613] (PTR: 61.65%)	790
(613, 614] (PTR: 52.59%)	844
(614, 616] (PTR: 53.29%)	1,904
(616, inf] (PTR: 47.61%)	200

Choose Segmentation Variable

- APP_PROD_CODE
- ALL_SINGLE
- GENDER
- NBR_FINANCE_COMPANY_TRDLNS_ACTIVE_05_MONTHS
- NBR_FINANCE_COMPANY_TRDLNS_OPENED_011_MONTHS
- NBR_MORTGAGE_TRDLNS_WITH_BALANCE_gt0
- BANKRUPTCY_INDICATOR
- NBR_RETAIL_TRDLNS_OVERLIMIT
- NBR_TRDLNS_OPENED_023_MONTHS_30P_DPD_EVER
- TOO_NEW_OR_STALE_FILE_INDICATOR
- BAD
- auto_BANKRUPTCY_INDICATOR_auto_box-cox
- auto_NBR_TRDLNS_60P_DPD_023_MONTHS_auto_box-cox
- auto_NBR_FINANCE_COMPANY_TRDLNS_auto_box-cox

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).

Project: Demo22
Project Path: C:/Users/DELL/Documents/Demo22

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

Click to continue

Back

Data Upload **Validation** **Reject-Inferencing** **Decision Simulator**

Create Metric Cut-off(Overall) **Cut-off(Segmented)**

*Click on the cells to allocate new accept/reject decision.

Final Model Selected: Model 01

Choose Metric: PTR

Segmentation Variable: APP_PROD_CODE


Range / Binning (Click to edit)	Total	PREMIER	REGULAR	STUDENT
[-inf, 584] (PTR: 85.55% (#844))	83.16% (#140)	86.25% (#583)	84.95% (#121)	
(584, 595] (PTR: 73.65% (#2,813))	73.36% (#831)	73.53% (#1,476)	74.48% (#506)	
(595, 602] (PTR: 71.28% (#106))	71.45% (#17)	71.23% (#61)	71.27% (#28)	
(602, 613] (PTR: 61.65% (#790))	61.54% (#297)	61.59% (#281)	61.9% (#212)	
(613, 614] (PTR: 52.59% (#844))	52.16% (#245)	52.54% (#248)	52.92% (#350)	
(614, 616] (PTR: 53.29% (#1,904))	53.3% (#948)	53.25% (#516)	53.3% (#441)	
(616, inf] (PTR: 47.61% (#200))	47.61% (#96)	47.89% (#16)	47.02% (#96)	
Total (#7,501)	62.22% (#2,565)	69.7% (#3,162)	62.5% (#1,754)	
To-Be Decline (#3,763)	74.72% (#988)	76.96% (#2,129)	76.28% (#655)	
To-Be Accept (#3,738)	54.39% (#1,577)	55.2% (#1,062)	54.29% (#1,099)	

Apply

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.

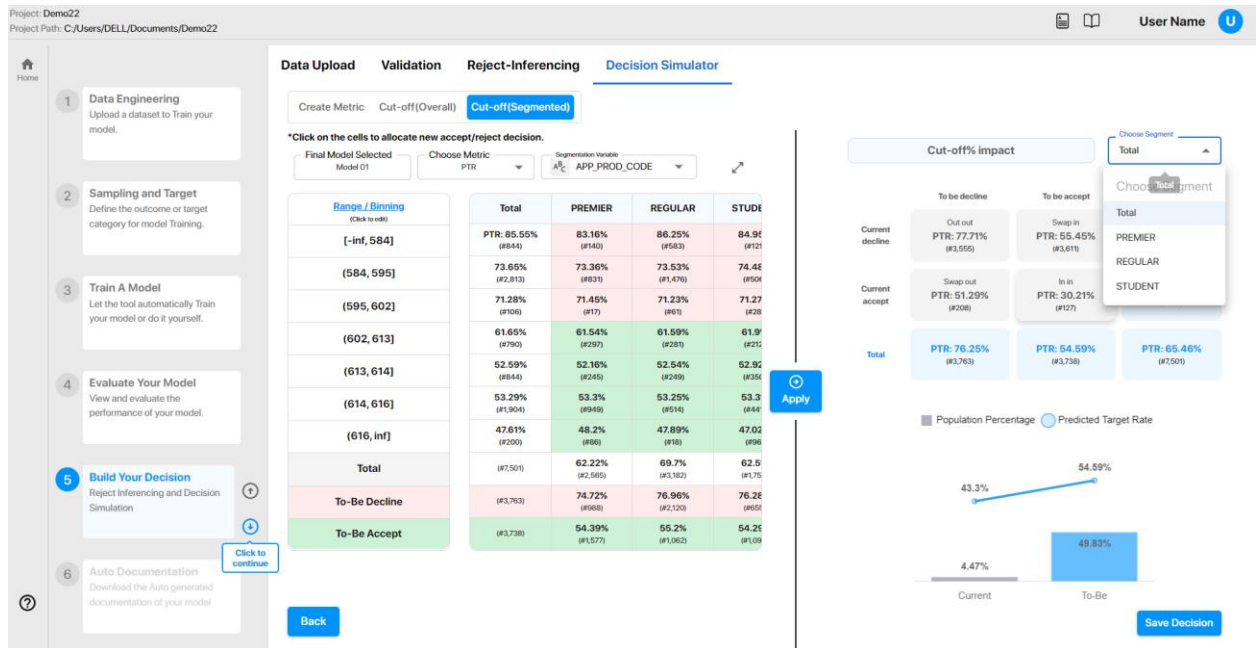
The  option can be used to toggle between expanded and compact view.

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**”

- 129 | Page



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


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

1 / 122

170%

User Name 

 **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\ldai\<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\ldai\<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”