Model

TABLE OF CONTENTS

04 Add a Prediction Model

09 Add a Classification Model

14 Add a Clustering Model

18 Add a Survival Analysis

22 Add an Anomaly / Dimensionality Reduction Model

26 Add a Forecasting Model

30 Add a Text Classification Model

34 Add a Text Clustering Model

38 Add an Association Rules Model

Model
QUICK START GUIDE

Analance

AAA Models help users build machine learning analysis through the import, preparation, cleansing and consolidation of data for better deployment and delivery of actionable business insights.

FEATURES

1. Descriptive, predictive, anomaly, and time series algorithms.

KEY BENEFITS

1. Analyze data using multiple algorithms through ensemble.
2. Quickly and easily create models.
3. Discover structure and easily predict categories.
4. Empower even a novice user.
5. Select and apply various modeling techniques and calibrate their parameters to optimal values.

PREREQUISITES

1. Data must be available in any one of the different data connector formats supported by Analance.

RELATIONAL DATABASES

ORACLE, MySQL, PostgreSQL, TERADATA, Microsoft SQL Server, Access, IBM DB2, SQLite, MariaDB

BIG DATA

Hortonworks, IMPALA, MAPR, HIVE, Apache Solr, elasticsearch, hadoop, APACHE DRILL

API

SharePoint, OData Open Data Protocol

FILE

XLS, TXT, CSV, JSON, XML

SOCIAL MEDIA

facebook, twitter

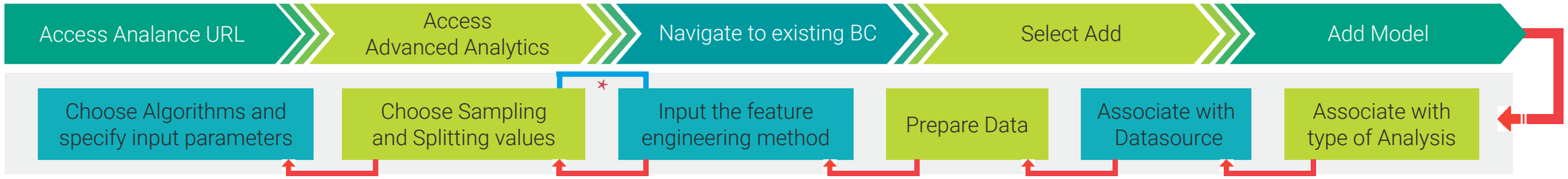
CUBES

Microsoft SQL Server Analysis Services

2. A business challenge must exist.

Model - Workflow (The following is the workflow to add a Model :)

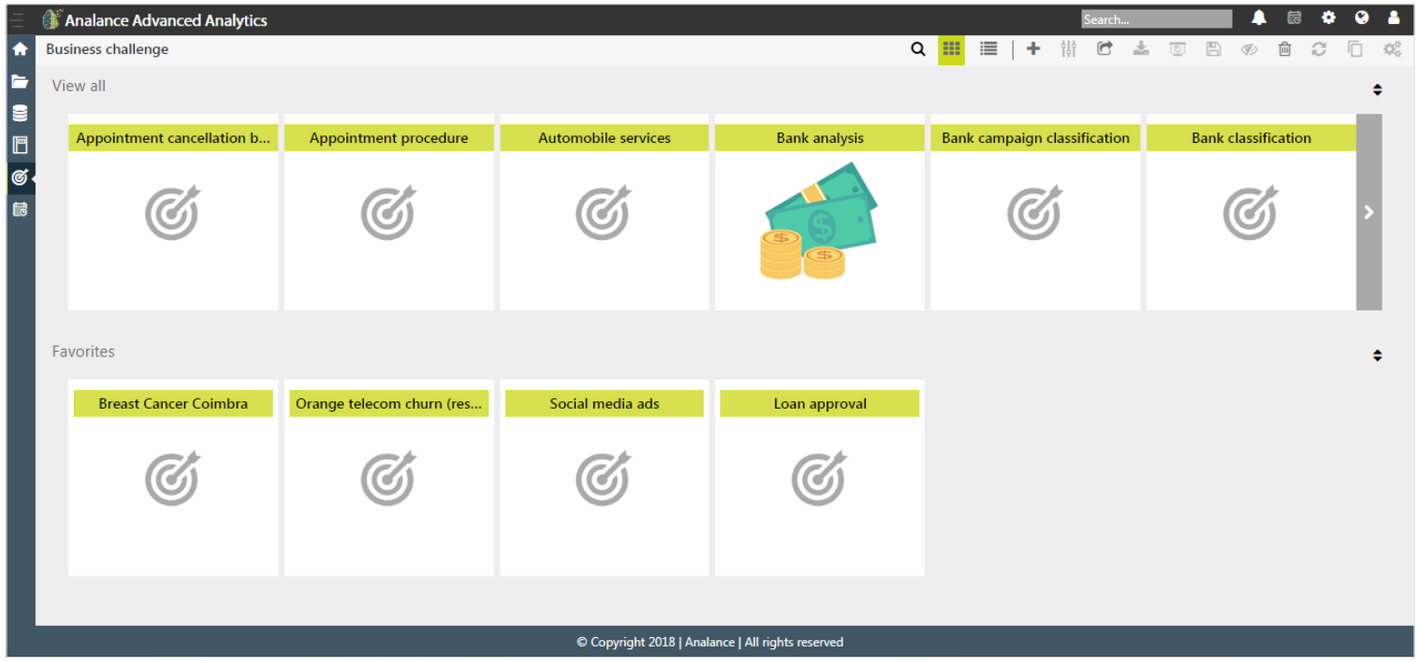
* There are two models: Supervised and Unsupervised.
The feature engineering and sampling and splitting values steps are not required for Unsupervised models.



MENU PATH : >> Business Challenge > Add > Click Model icon

Analance

Add a Prediction Model

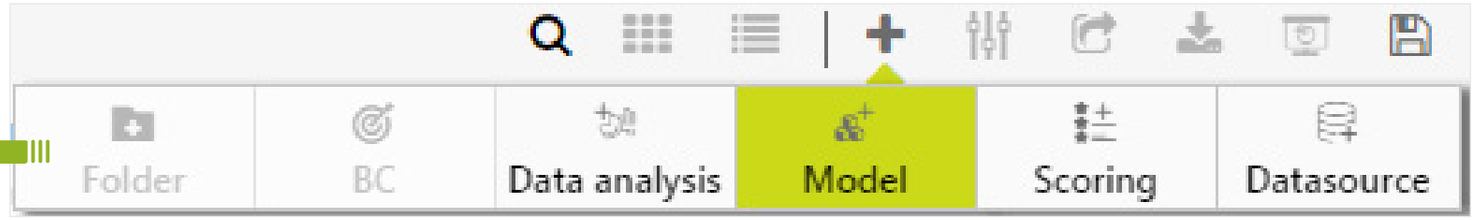


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

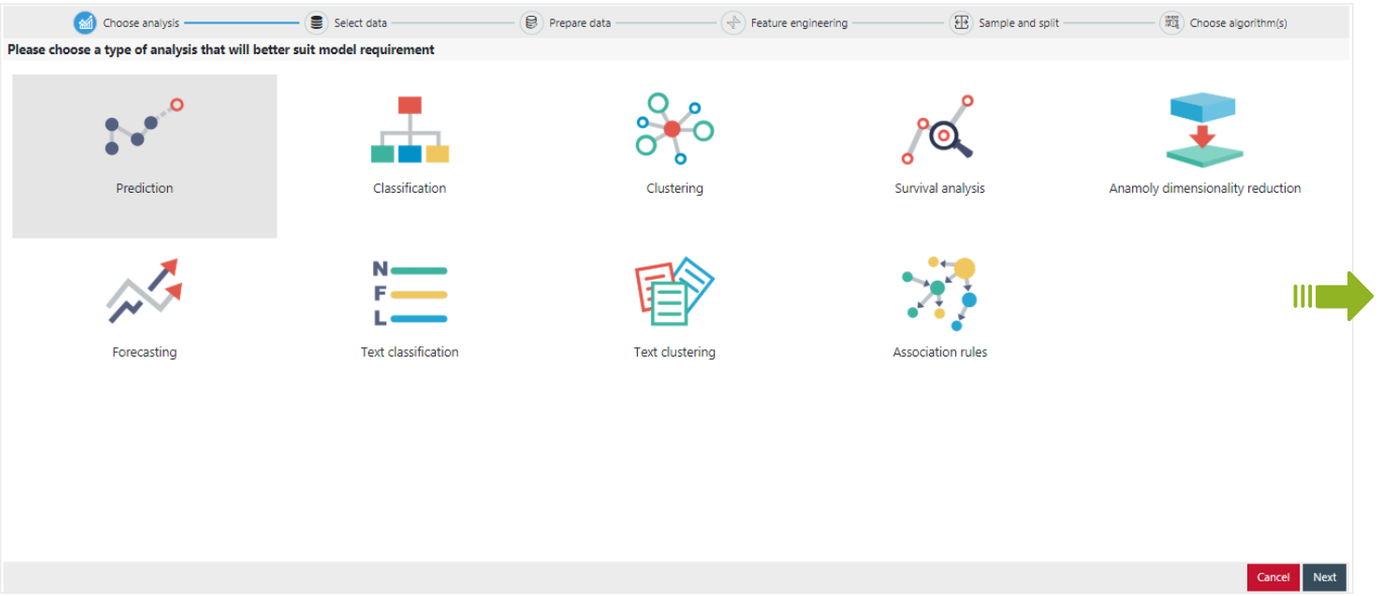
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.



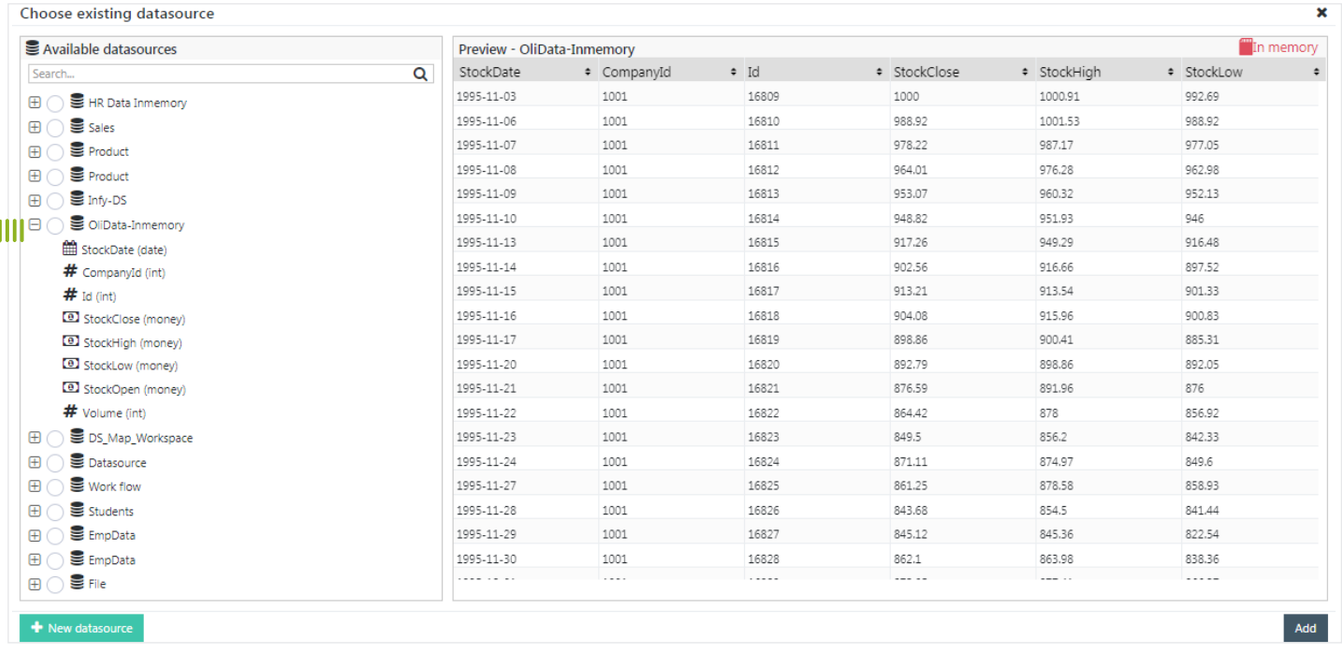
04

Choose Prediction from the available analysis type options and then click Next.

Analance

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

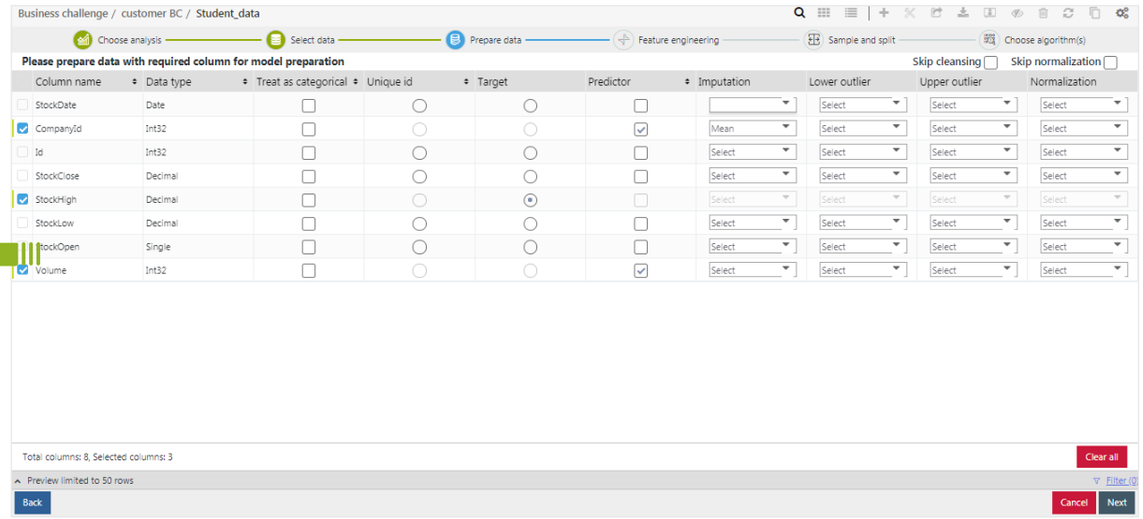


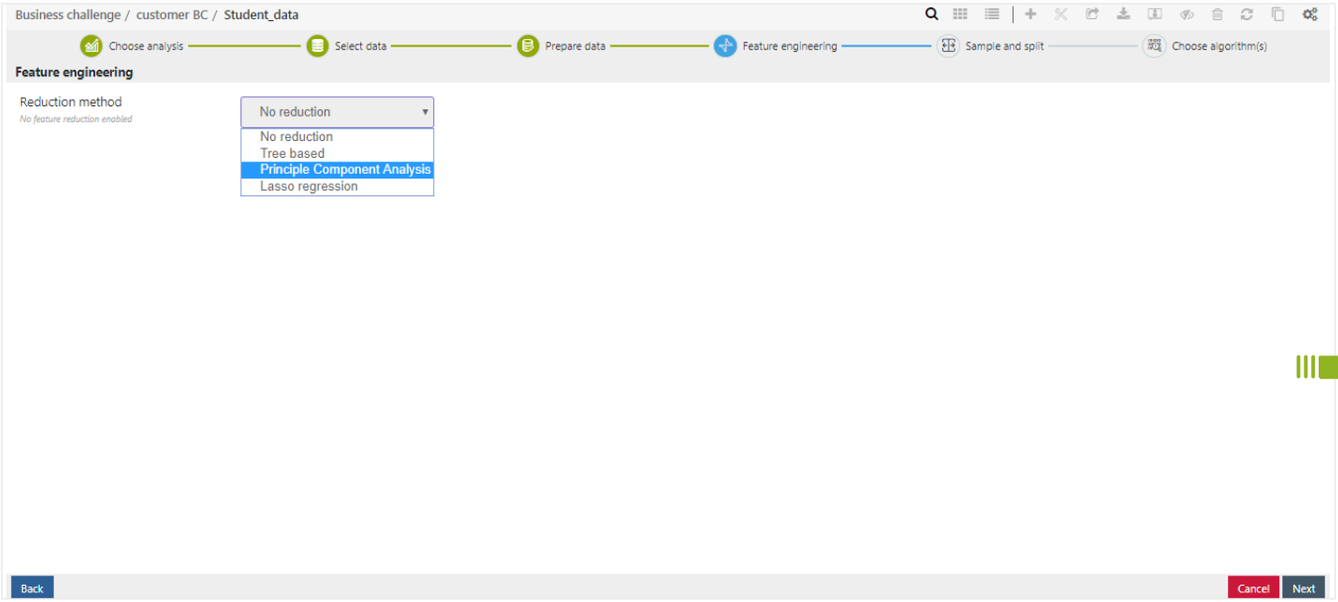
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.

06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

Prepare data	Select to include the variable in the data preparation process.
Treat as categorial	Select if the variable is to be treated as categorical.
Unique id	Select the unique variable if the datasource has a unique value.
Target	This variable is mandatory. Select a variable that you will like to predict that is the target for analysis.
Predictor	This variable is mandatory and independent. Select columns to influence the target variable.
Imputation	Select substitute value in case there are missing values for each variable. The options will vary based on the Data Type. If it is a continuous/numeric value, the options are: Mean, Median, Mode and User Input (Number). If the variable is factor/categorical, the options are: Most frequent, Least frequent and User Input (Text).
Lower outlier	Replace the lower outlier values with one of the options: Percentile 2nd, Percentile 5th and User input.
Upper outlier	Replace the upper outlier values with one of the options: Percentile 95th, Percentile 98th and User input.
Normalization	Select from the options: Z-score, Min Max, Logistic, Log normal and Tanh to normalize data.
Skip Cleansing and Skip Normalization	If available, skip data cleansing and/or normalization.





07 CHOOSE THE FEATURE ENGINEERING OPTIONS :

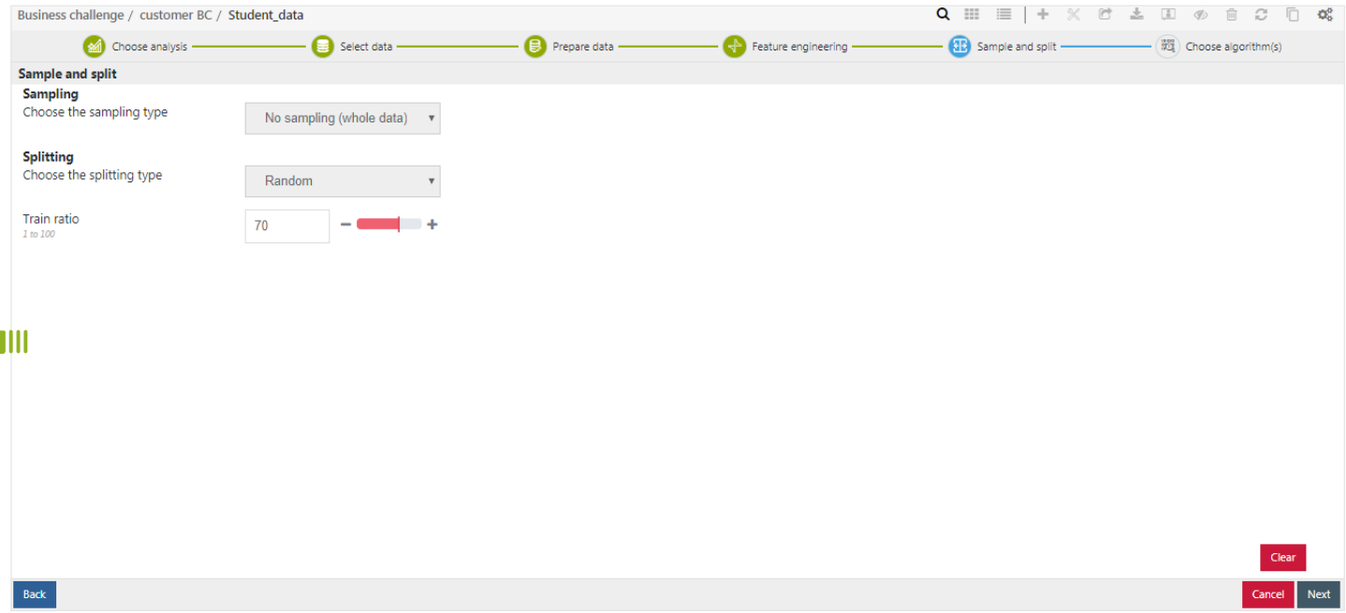
1. In the Reduction method list, choose a reduction method.
2. For the selected method, choose the options:

Three Methods :	
Tree Based	Type the *Number of Trees* and *Depth of Trees*.
Principle Component Analysis	1. From the **Matrix generation based** on options, select **Correlation** or **Covariance**. 2. Type a value or use the slider to increase or reduce the **Proportion of variance to** value as required. 3. Choose a **Normalization** method. Z-score will standardize the values.
Lasso Regression	No options.

3. Click Next.

08 CHOOSE THE SAMPLE AND SPLIT TYPES :

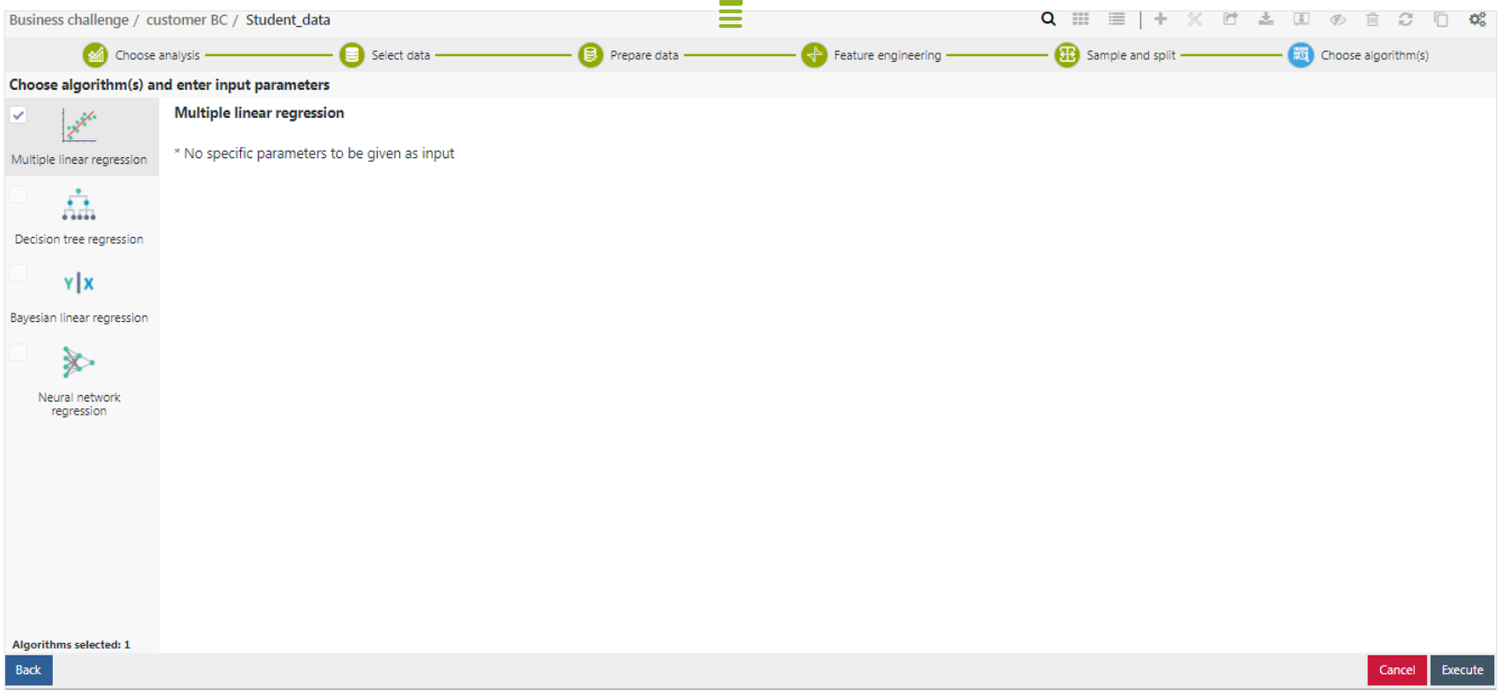
1. In the Sampling list, choose a sampling type.
2. Depending on the chosen sampling type, type the No of Samples and choose the Column Information.
3. In the Splitting list, choose the splitting type.
4. Type a value or use the slider to increase or reduce the Training ratio value as required.



09 TO CHOOSE THE ALGORITHM(S) :

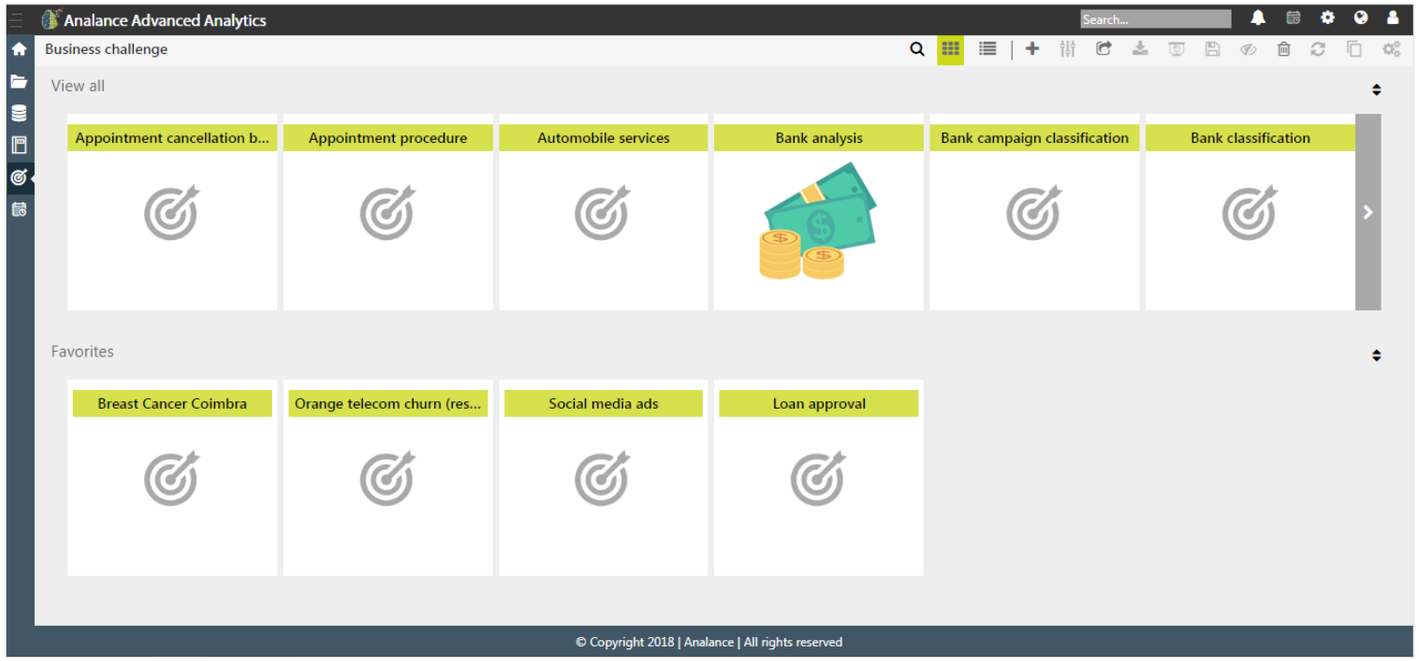
1. Choose the algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
Multiple linear regression	No input parameter
Decision Tree regression	Minimum split, Minimum bucket, Maximum Depth
Bayesian linear regression	No input parameter
Neural network regression	Sensitive value
Random forest regression	Maximum depth, Number of trees

Add a Classification Model

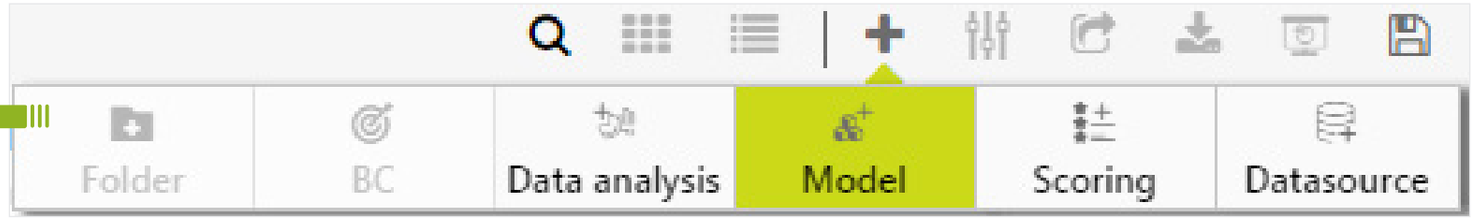


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

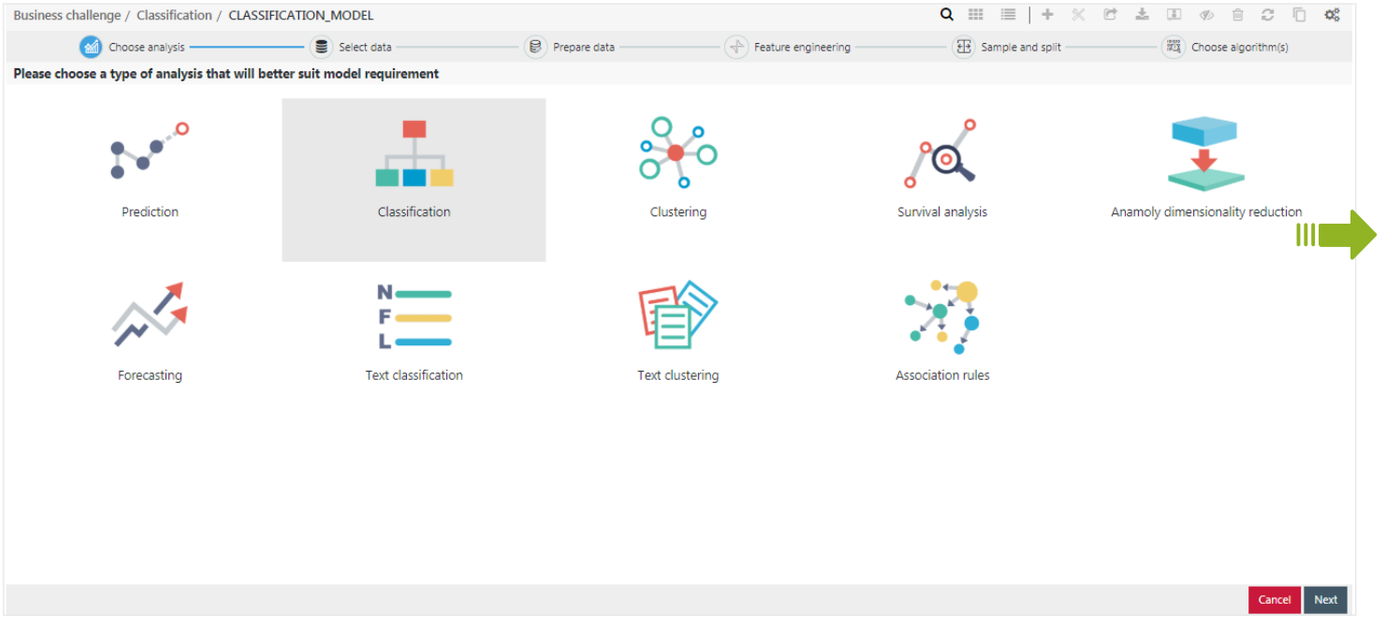
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.



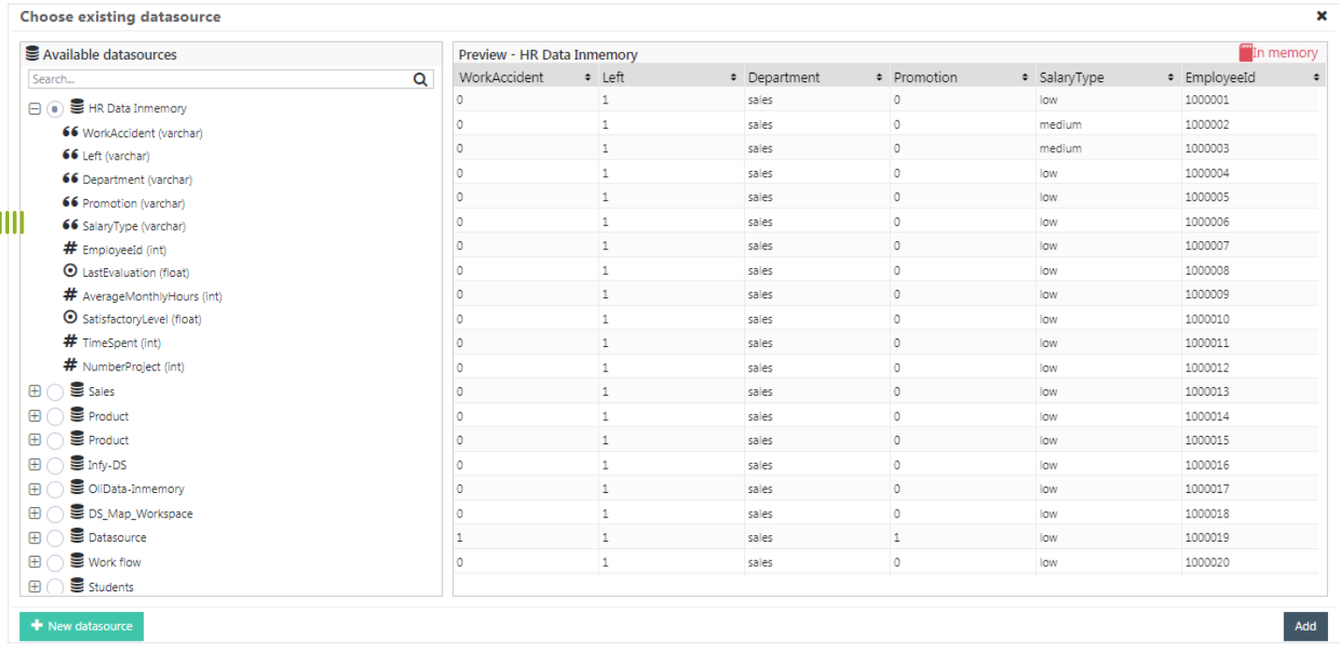
04

Choose Classification from the available analysis type options and then click Next.

Analance

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

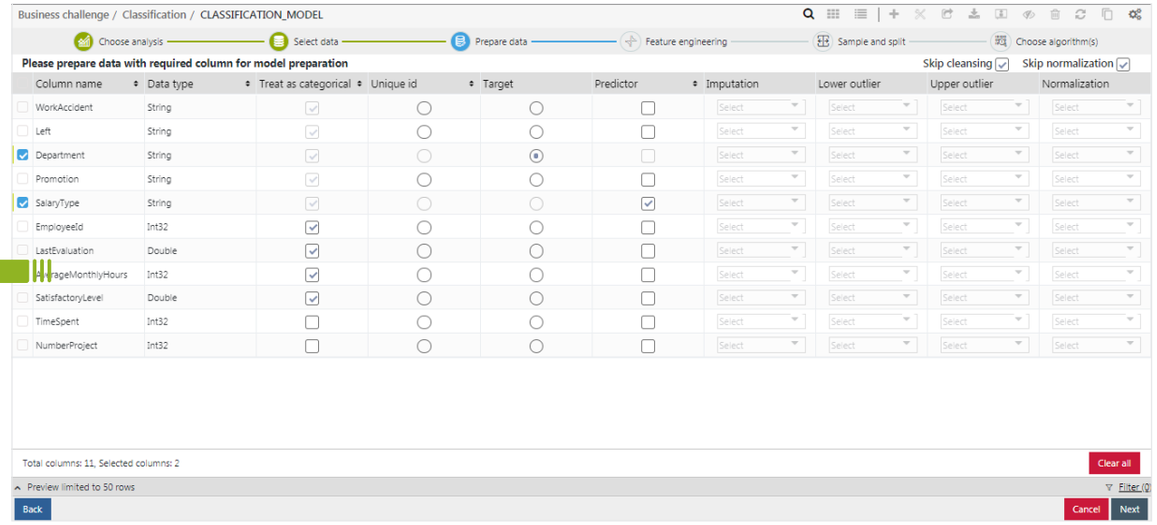


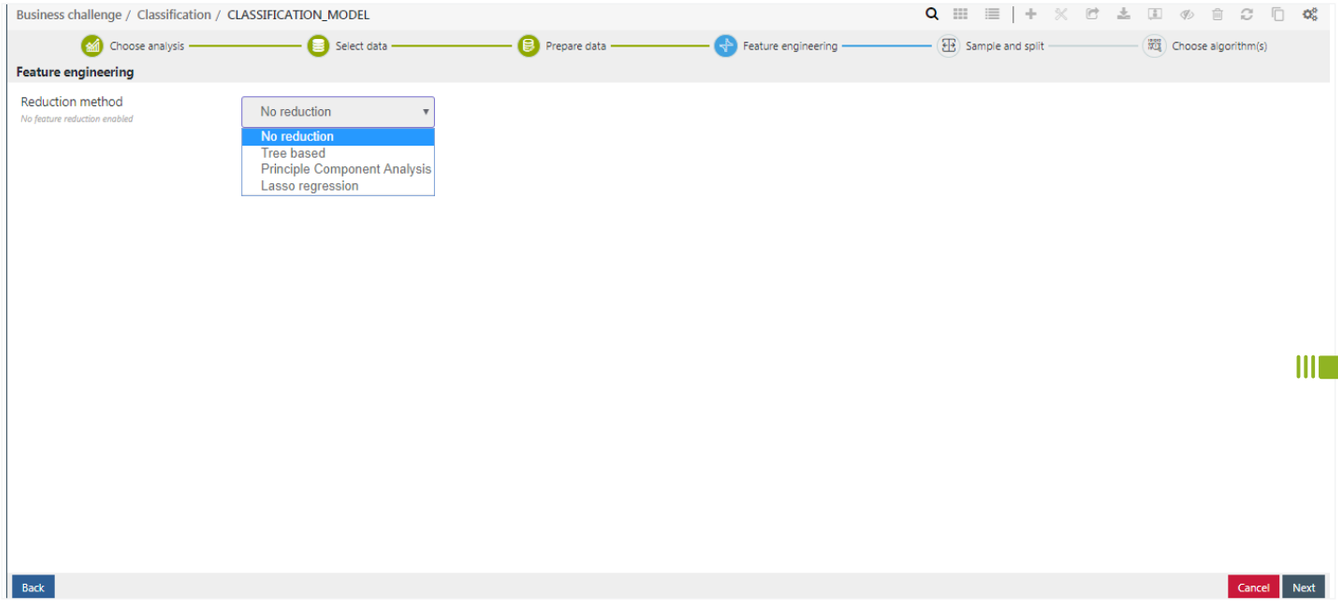
> *Note : You can also add a new datasource by clicking **New datasource** and following the step-by-step wizard.*

06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

Prepare data	Select to include the variable in the data preparation process.
Treat as categorial	Select if the variable is to be treated as categorical.
Unique id	Select the unique variable if the datasource has a unique value.
Target	This variable is mandatory. Select a variable that you will like to predict that is the target for analysis.
Predictor	This variable is mandatory and independent. Select columns to influence the target variable.
Imputation	Select substitute value in case there are missing values for each variable. The options will vary based on the Data Type. If it is a continuous/numeric value, the options are: Mean, Median, Mode and User Input (Number). If the variable is factor/categorical, the options are: Most frequent, Least frequent and User Input (Text).
Lower outlier	Replace the lower outlier values with one of the options: Percentile 2nd, Percentile 5th and User input.
Upper outlier	Replace the upper outlier values with one of the options: Percentile 95th, Percentile 98th and User input.
Normalization	Select from the options: Z-score, Min Max, Logistic, Log normal and Tanh to normalize data.
Skip Cleansing and Skip Normalization	If available, skip data cleansing and/or normalization.





07 CHOOSE THE FEATURE ENGINEERING OPTIONS :

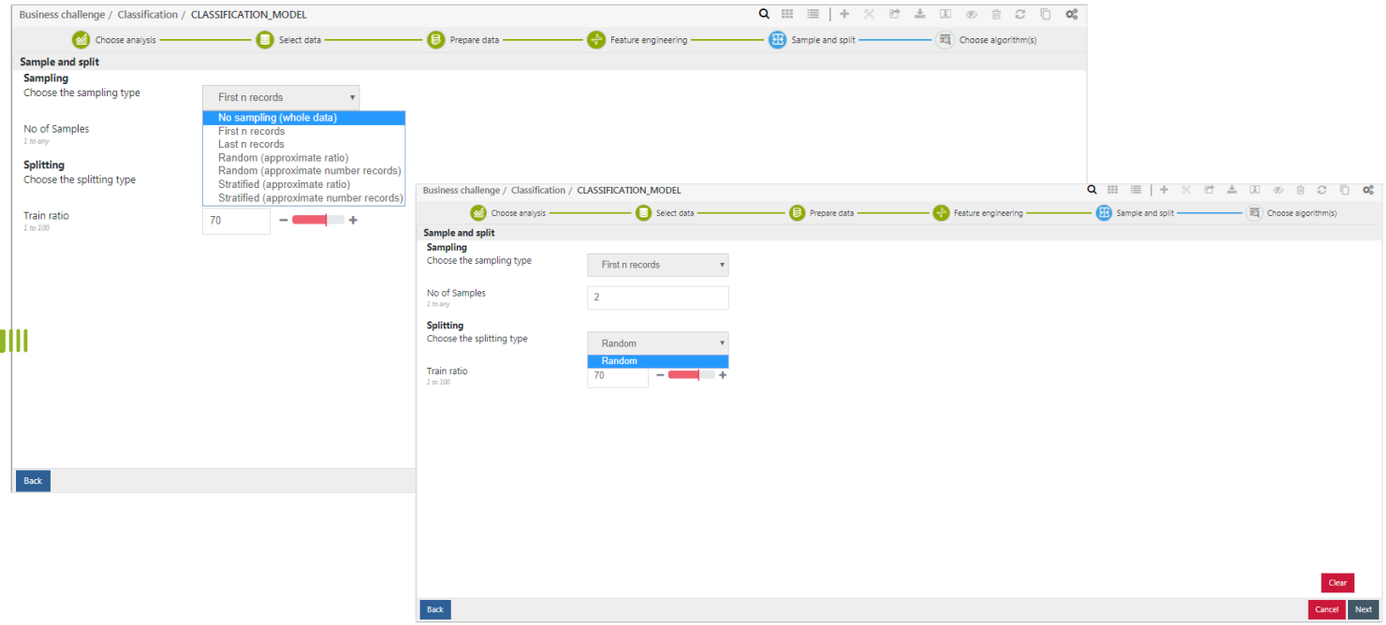
1. In the Reduction method list, choose a reduction method.
2. For the selected method, choose the options:

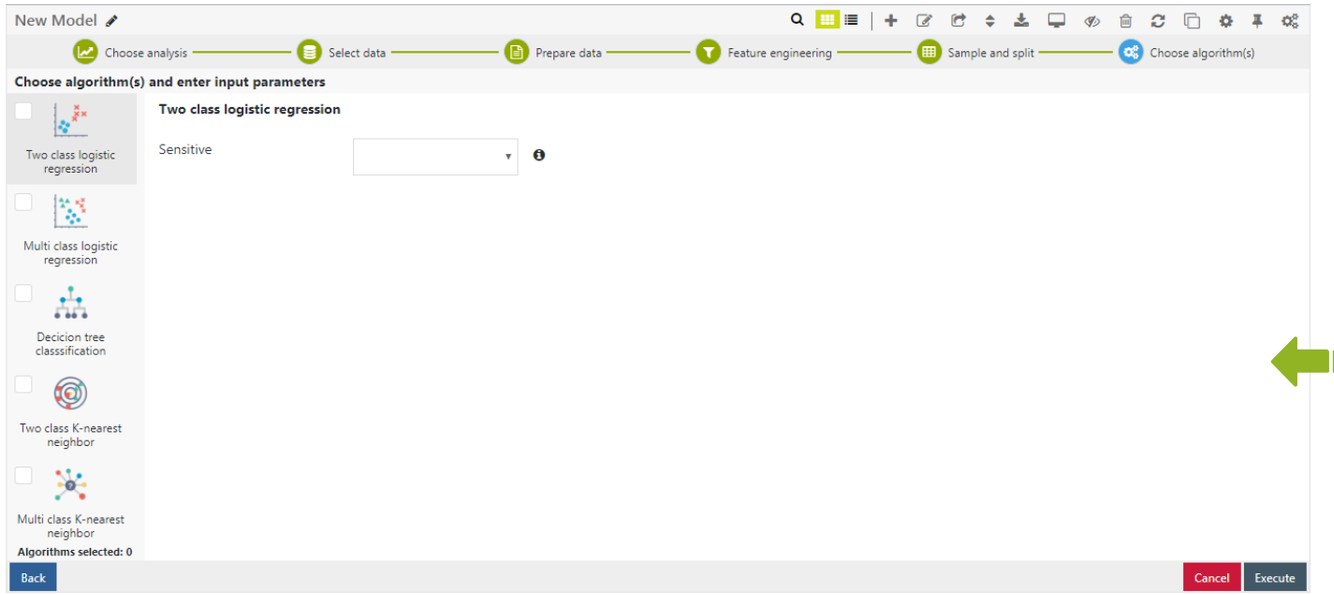
Three Methods :	
Tree Based	Type the *Number of Trees* and *Depth of Trees*.
Principle Component Analysis	1. From the **Matrix generation based** on options, select **Correlation** or **Covariance**. 2. Type a value or use the slider to increase or reduce the **Proportion of variance to** value as required. 3. Choose a **Normalization** method. Z-score will standardize the values.
Lasso Regression	No options.

3. Click Next.

08 CHOOSE THE SAMPLE AND SPLIT TYPES :

1. In the Sampling list, choose a sampling type.
2. Depending on the chosen sampling type, type the No of Samples and choose the Column Information.
3. In the Splitting list, choose the splitting type.
4. Type a value or use the slider to increase or reduce the Training ratio value as required.





09 TO CHOOSE THE ALGORITHM(S) :

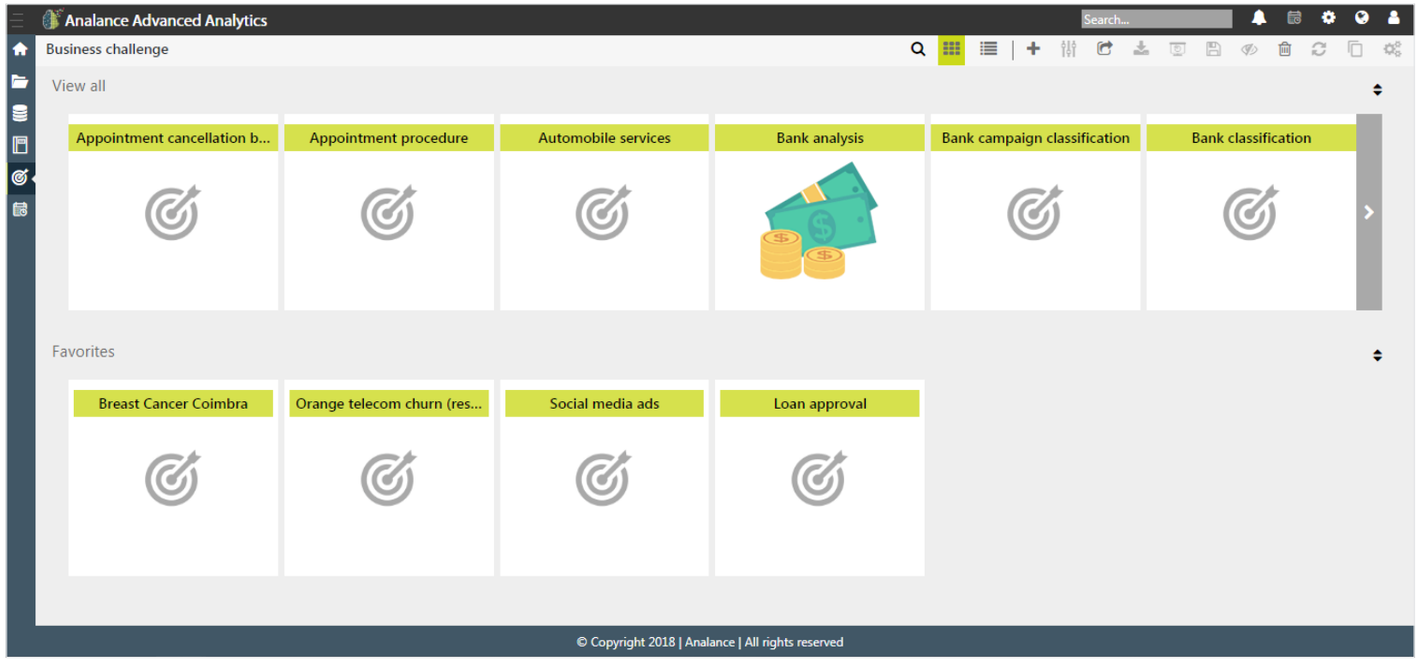
1. Choose the algorithm(s) for calculating the values for the model and specify the input parameters. Based on the input, the algorithms displayed can be either two class or multi class.
2. Click Execute.

The model will be created in the background. When the model is ready, a notification will appear in the Notification area.

Algorithms	Input Parameters
Two class logistic regression	Sensitive value
Multi class logistic regression	No input parameter
Two class decision tree classification	Sensitive value
Two class K-nearest neighbor	Sensitive value
Multi class k-nearest neighbor	No input parameter
Two class random forest	Maximum depth, number of tress, sensitive, minimum split, minimum bucket, maximum leaf node
Two class neural network	Sensitive value

Algorithms	Input Parameters
Two class support vector machine	Sensitive, Cost function C, Gamma value, Kernel
Two class gradient boosting machine	Minimum split, Maximum depth, Sensitive, Bagging size, Estimators, Feature strategy, Learning rate, Loss function
Ada boosting	Minimum split, Maximum depth, Sensitive, Criterion, Estimators, Feature strategy, Learning rate, Type
Multi class random forest	Minimum split, Minimum bucket, Maximum depth, Number of trees, Maximum leaf node
Multi class Ordinal regression	No input parameter
Two class naïve bayes classification	Sensitive class
Multi class naïve bayes classification	No input parameter

Add a Clustering Model

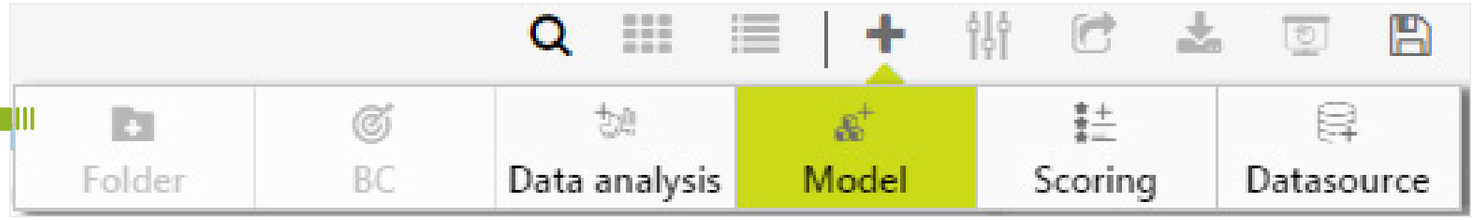


01 SELECT A BUSINESS CHALLENGE :

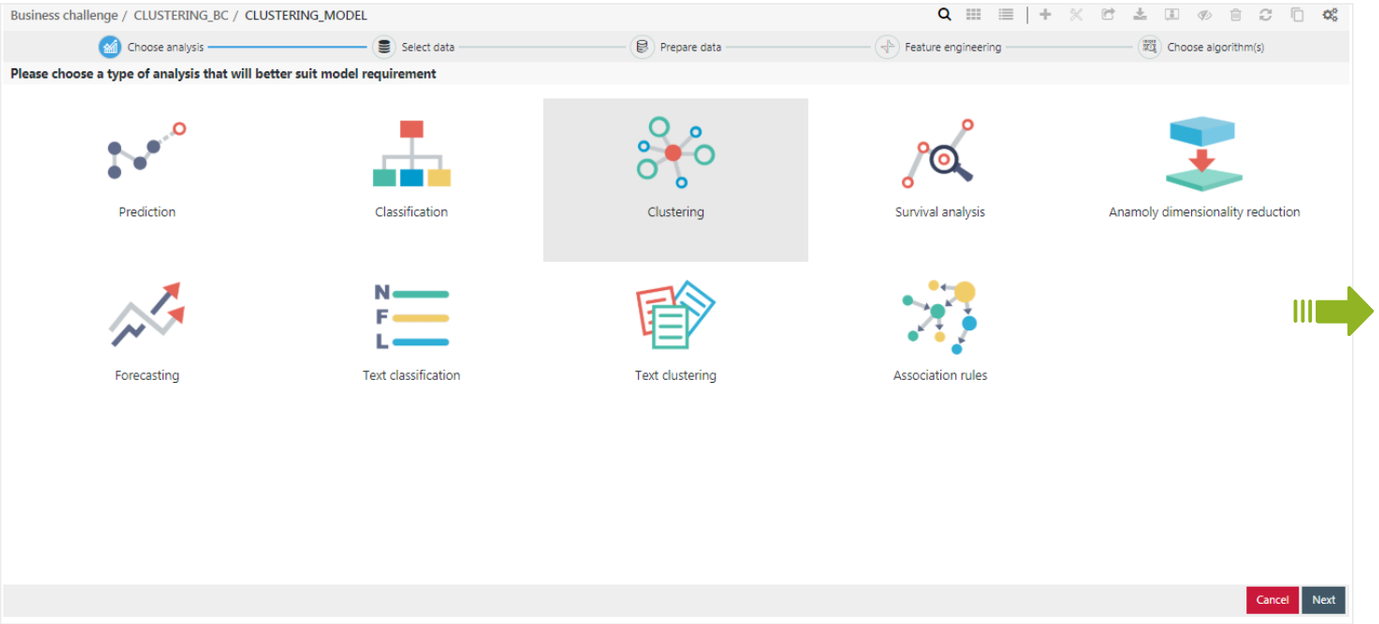
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

02

Click Add (+) > Model.



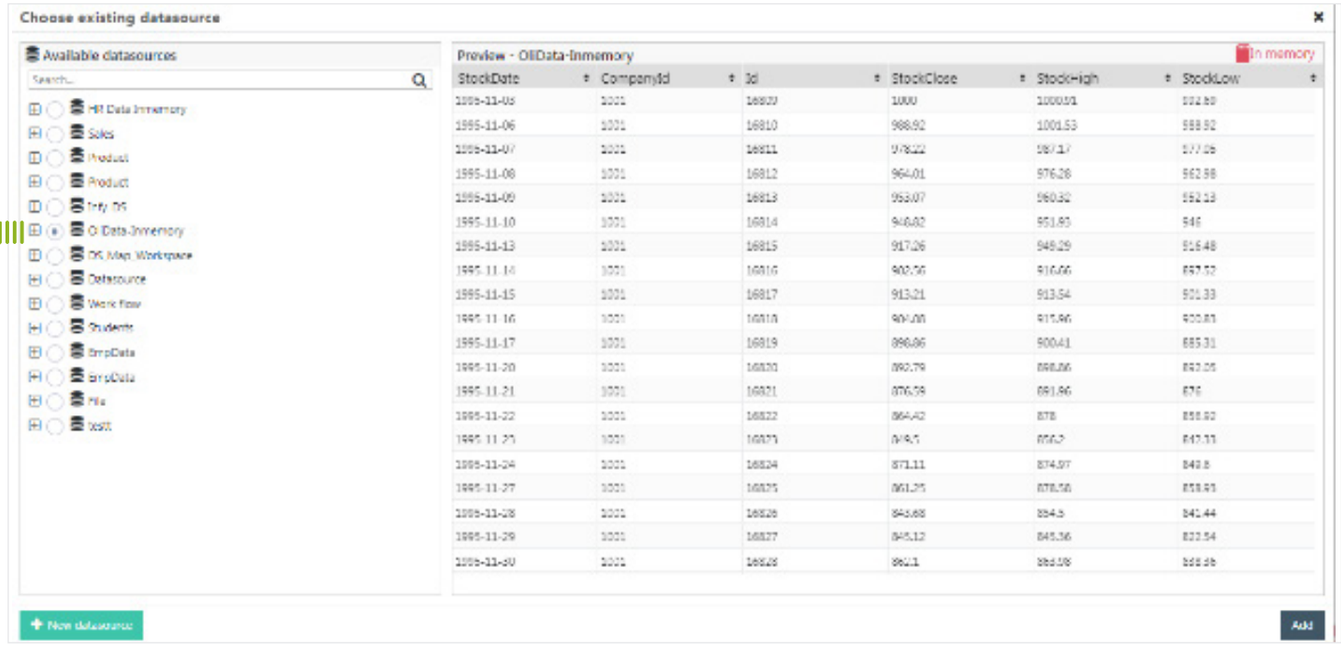
03 Type the Name and Description of the model and then click Create.



04 Choose Clustering from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

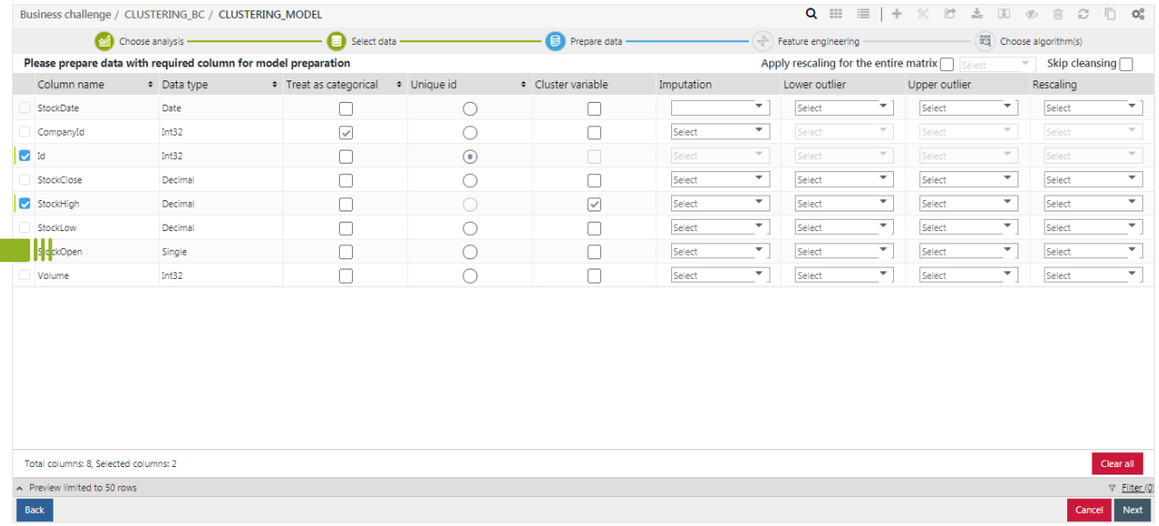


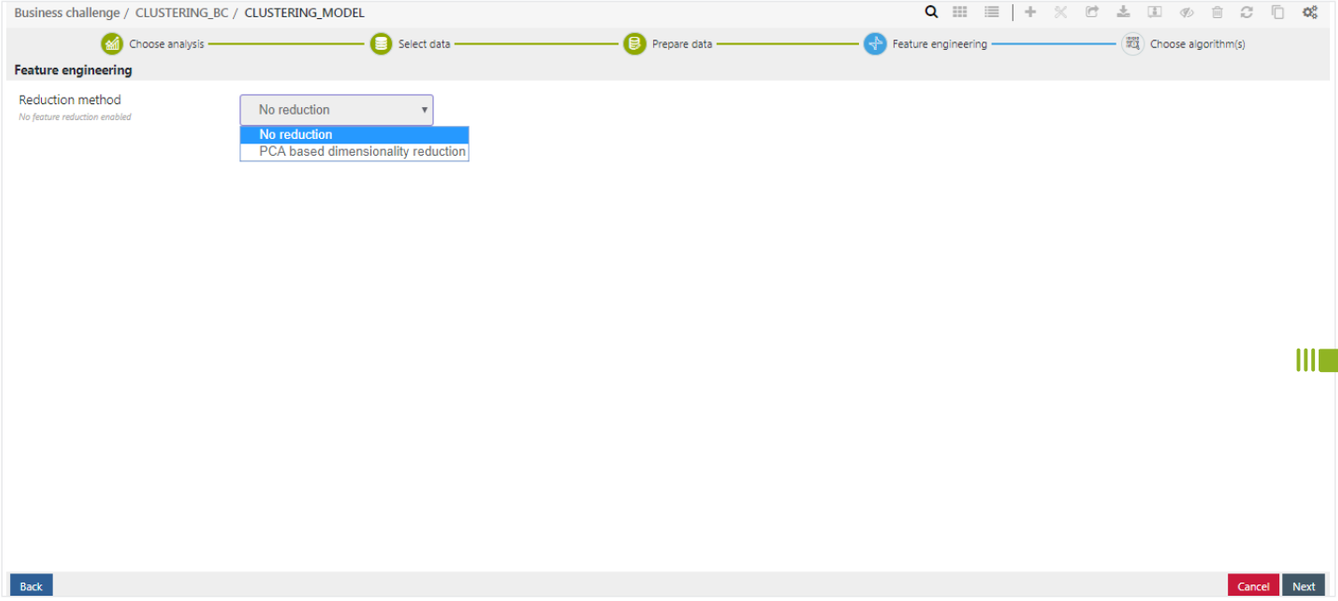
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.

06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the **Prepare data** table and then click **Next**.

Prepare data	Select to include the variable in the data preparation process.
Treat as categorial	Select if the variable is to be treated as categorical.
Unique id	Select the unique variable if the datasource has a unique value.
Target	This variable is mandatory. Select a variable that you will like to predict that is the target for analysis.
Predictor	This variable is mandatory and independent. Select columns to influence the target variable.
Imputation	Select substitute value in case there are missing values for each variable. The options will vary based on the Data Type. If it is a continuous/numeric value, the options are: Mean, Median, Mode and User Input (Number). If the variable is factor/categorical, the options are: Most frequent, Least frequent and User Input (Text).
Lower outlier	Replace the lower outlier values with one of the options: Percentile 2nd, Percentile 5th and User input.
Upper outlier	Replace the upper outlier values with one of the options: Percentile 95th, Percentile 98th and User input.
Normalization	Select from the options: Z-score, Min Max, Logistic, Log normal and Tanh to normalize data.
Skip Cleansing and Skip Normalization	If available, skip data cleansing and/or normalization.





07 CHOOSE THE FEATURE ENGINEERING OPTIONS :

1. In the Reduction method list, choose a reduction method.
2. If you choose PCA based dimensionality reduction :

2a	Type the *Number of Trees* and *Depth of Trees*.
2b	For **Matrix generation based on**, select **Correlation** or **Covariance**. For **Proportion of variance to keep**, enter a number between 0 and 100.

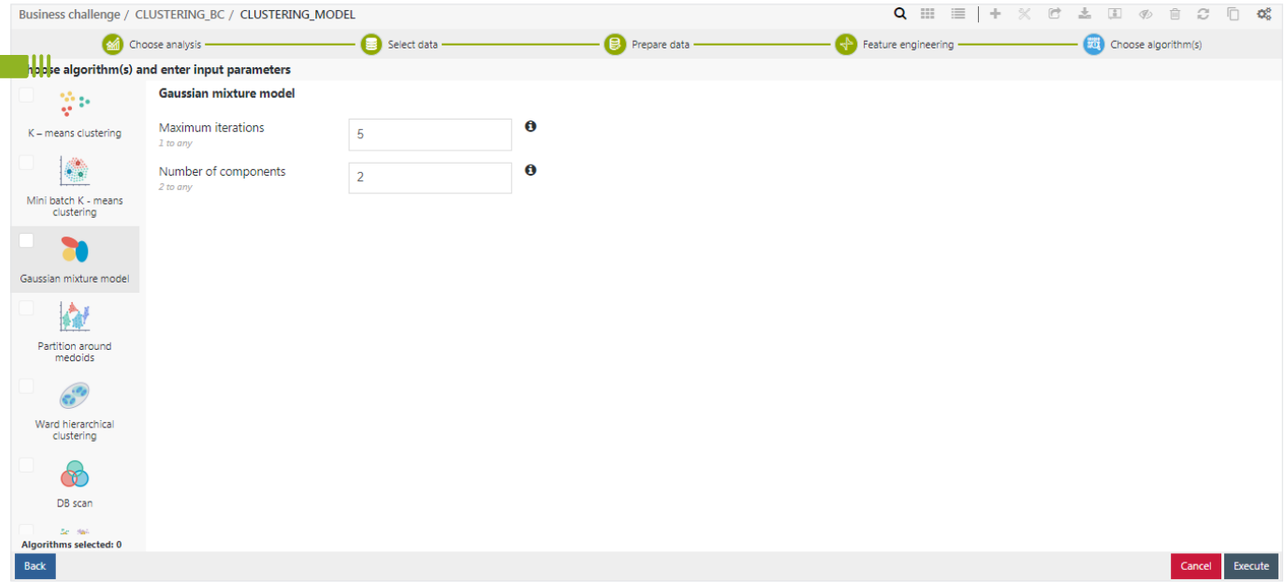
3. Click Next.

08 TO CHOOSE THE ALGORITHM(S) :

1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

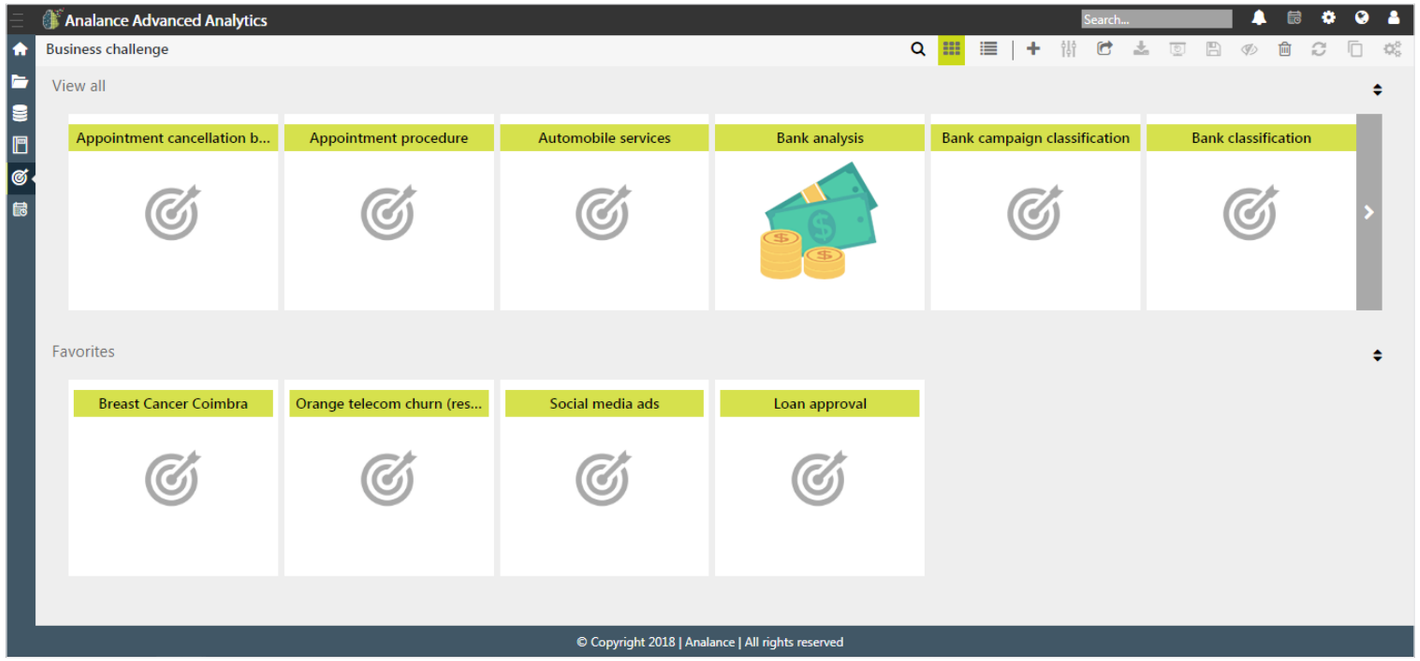
The model will be created in the background. When the model is ready, a notification will appear in the Notification area.

Algorithms	Input Parameters
K-means clustering	Exact K
Mini batch K – means clustering	Exact K, Batch size, Maximum iterations, Number of initialization, Initialization fraction, Initialization
Gaussian mixture model	Minimum iterations, Number of components
Partition around mediods	Exact K, Distance metric
Ward hierarchical clustering	Exact K, Distance metric, Method
DB Scan	Optimal epsilon(eps), Minimum points
Adaptive K-means clustering	Maximum iterations, Distance metric, Minimum K, Maximum K, Threshold value, N Start



Analance

Add a Survival Analysis Model

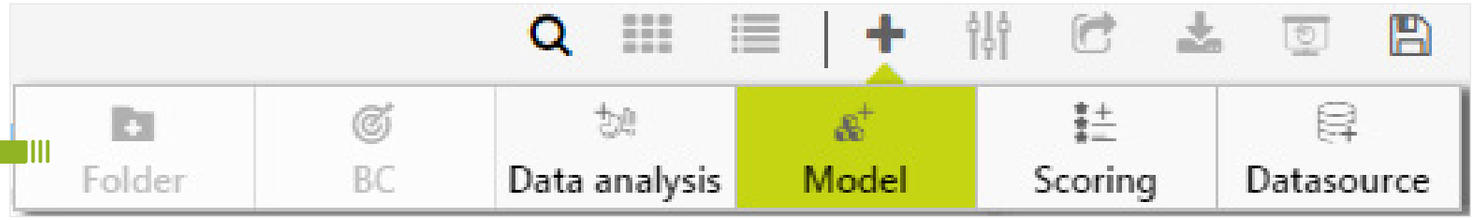


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

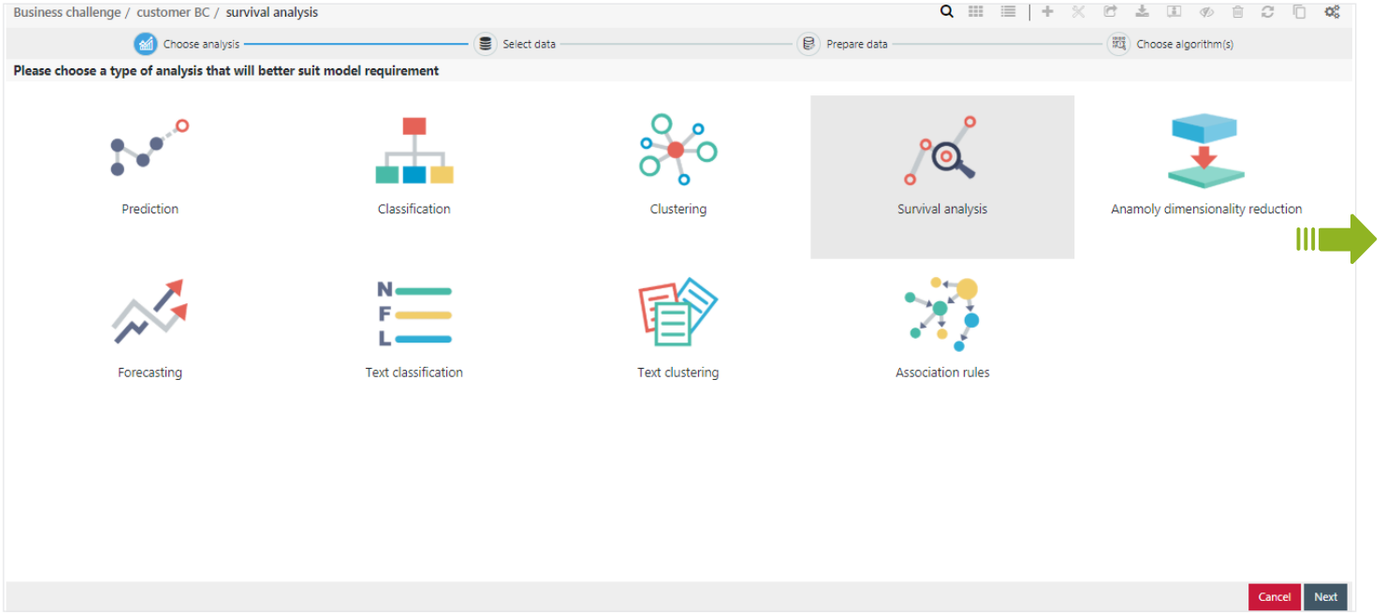
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.



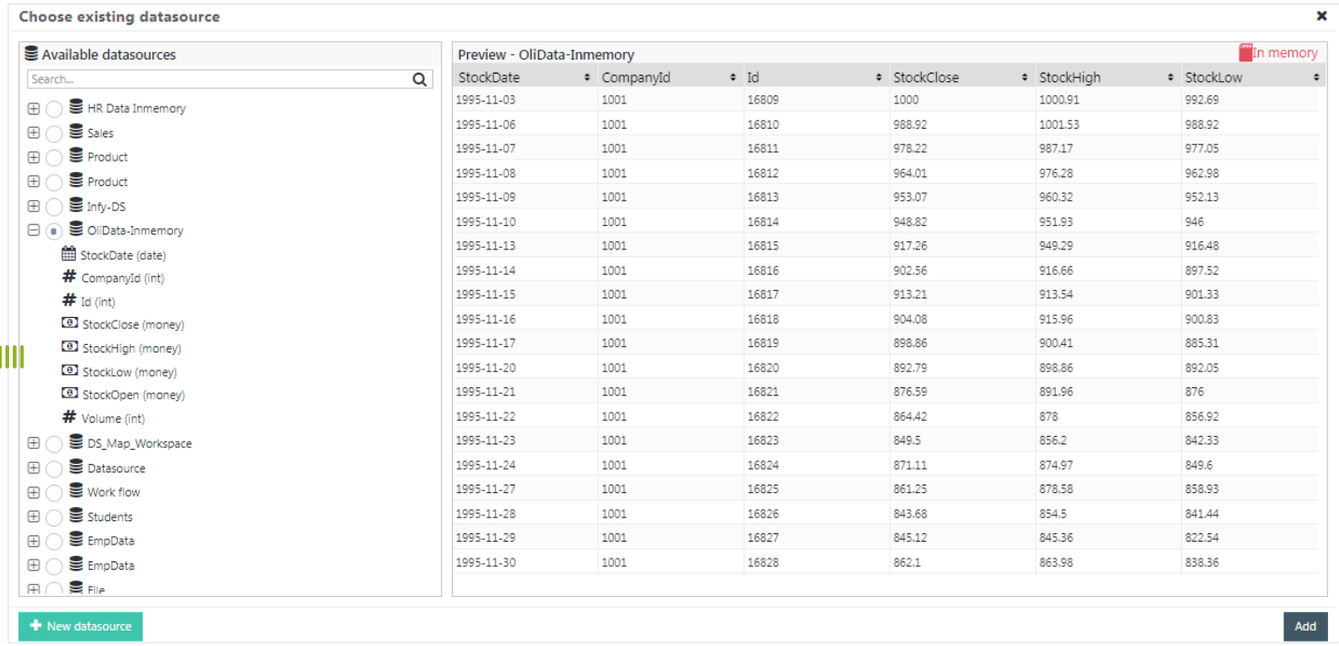
04

Choose Survival analysis from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

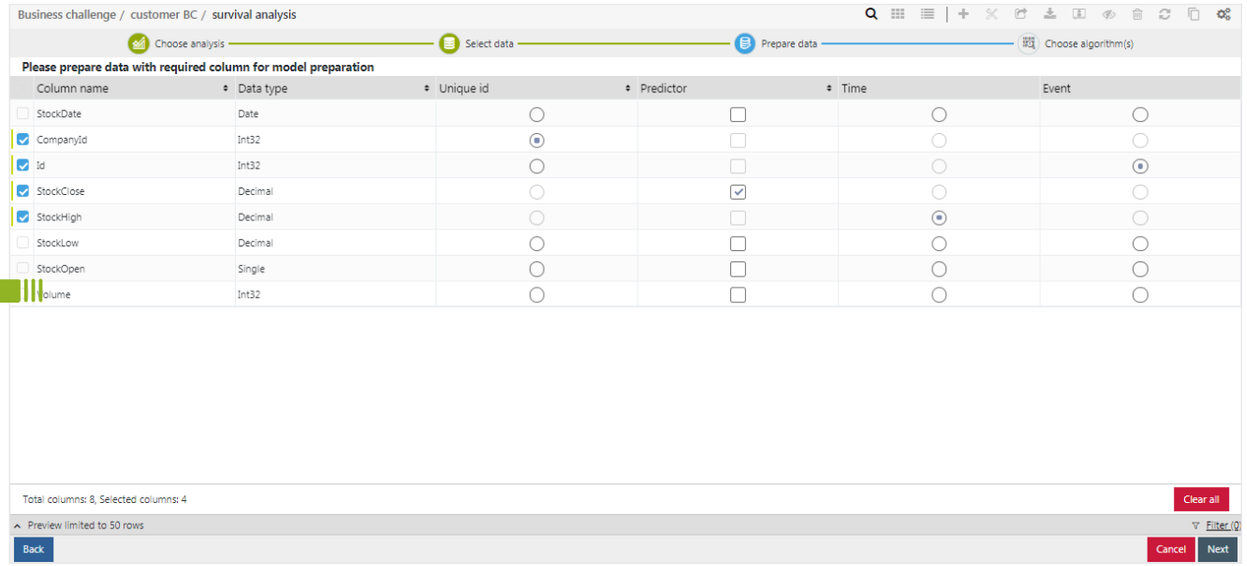
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.



06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

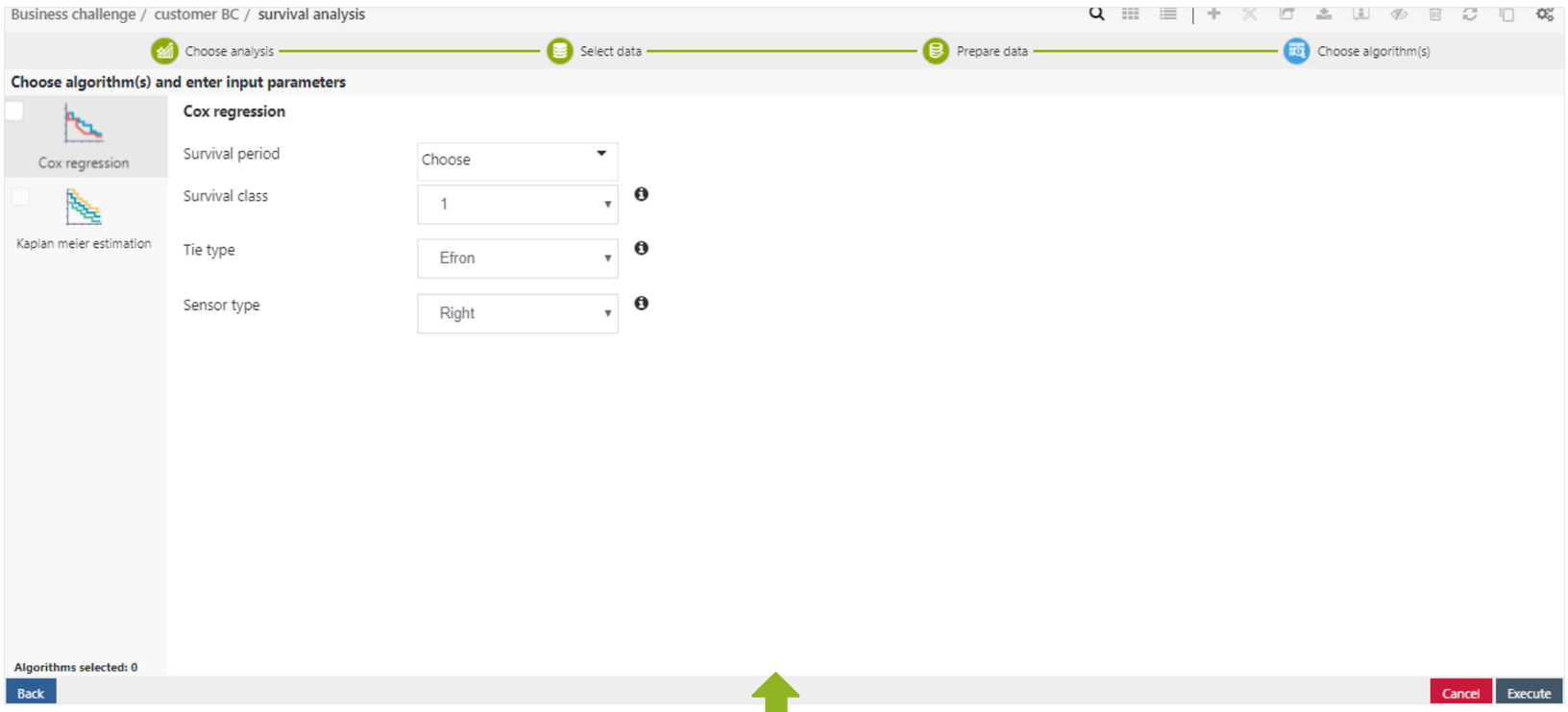
Prepare data	Select to include the variable in the data preparation process.
Unique id	Select the unique variable if the datasource has a unique value.
Predictor	This column value is mandatory and independent. Select columns to influence the target variable. On selecting Predictor for a particular variable, all the other parameters are disabled.
Time	This column value is mandatory. On selecting Time for a particular variable, all the other parameters are disabled.
Event	This column is mandatory. On selecting an event for a particular variable, all other parameters for the variable are disabled.



07 TO CHOOSE THE ALGORITHM(S) :

1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

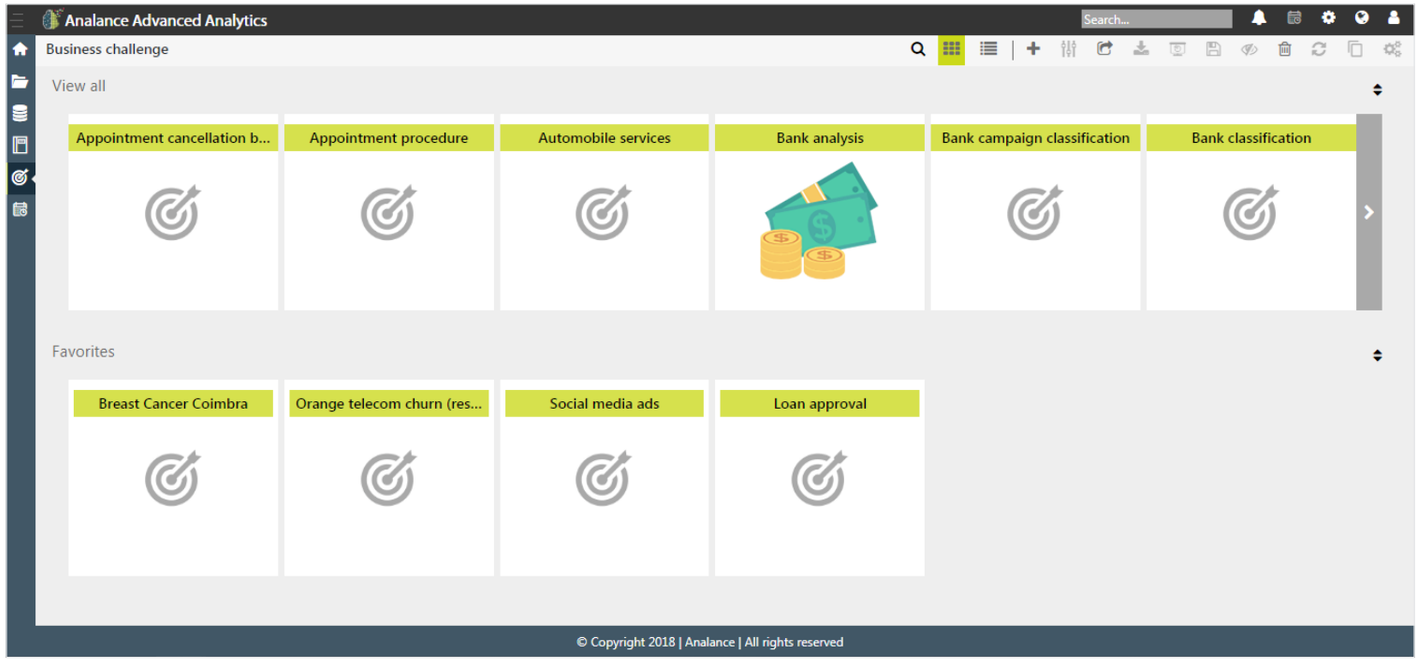
The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
Cox Regression	Survival period, Survival class, Tie Type, Sensor Type
Kaplar Meier Estimation	Censor Type, Variable Type, Confidence Interval. The UniqueID and Predictor column information is not applicable for this algorithm.

Analance

Add an Anomaly / Dimensionality Reduction Model

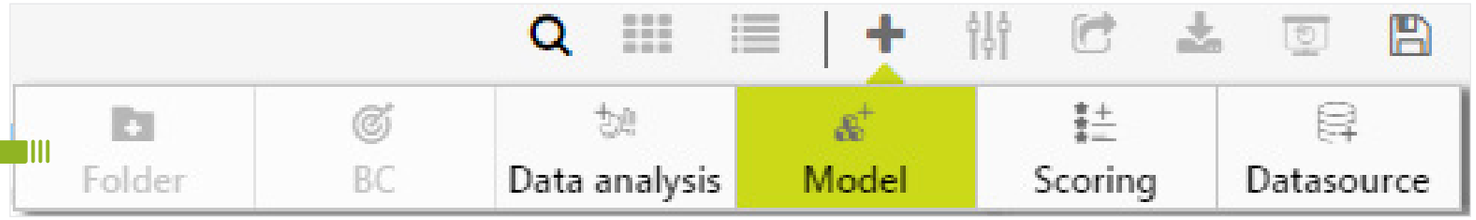


01 SELECT A BUSINESS CHALLENGE :

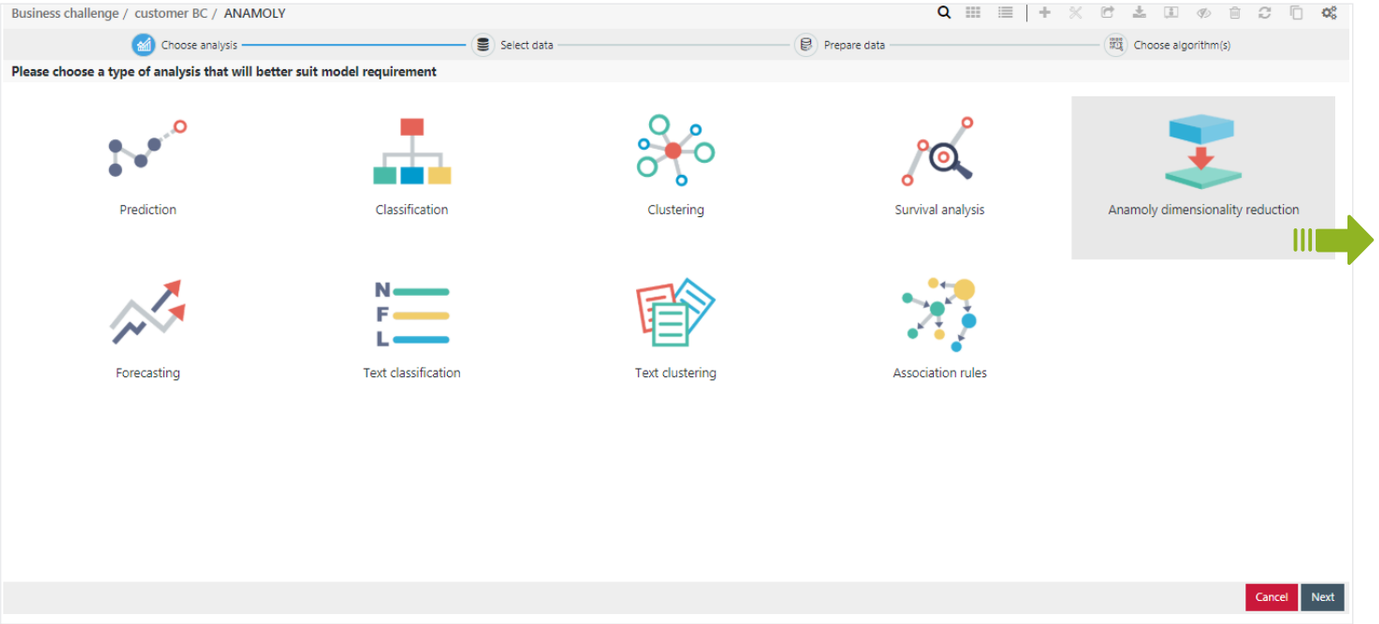
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

02

Click Add (+) > Model.



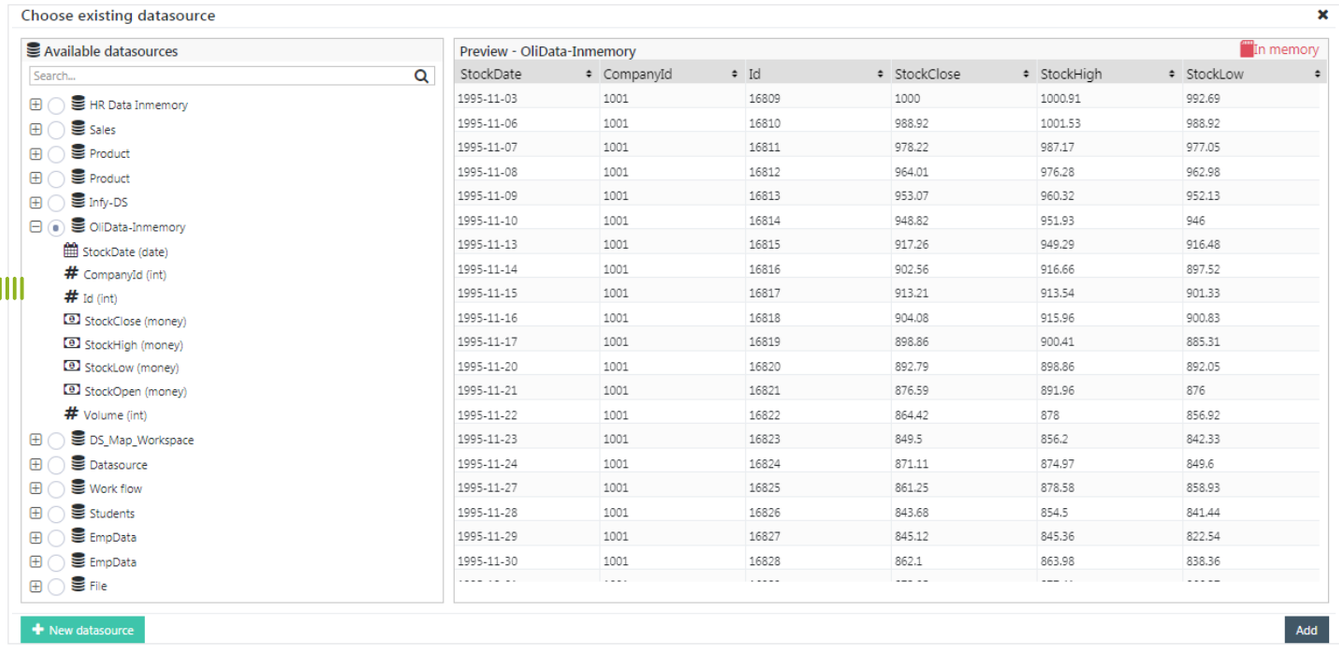
03 Type the Name and Description of the model and then click Create.



04 Choose Anomaly dimensionality reduction from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

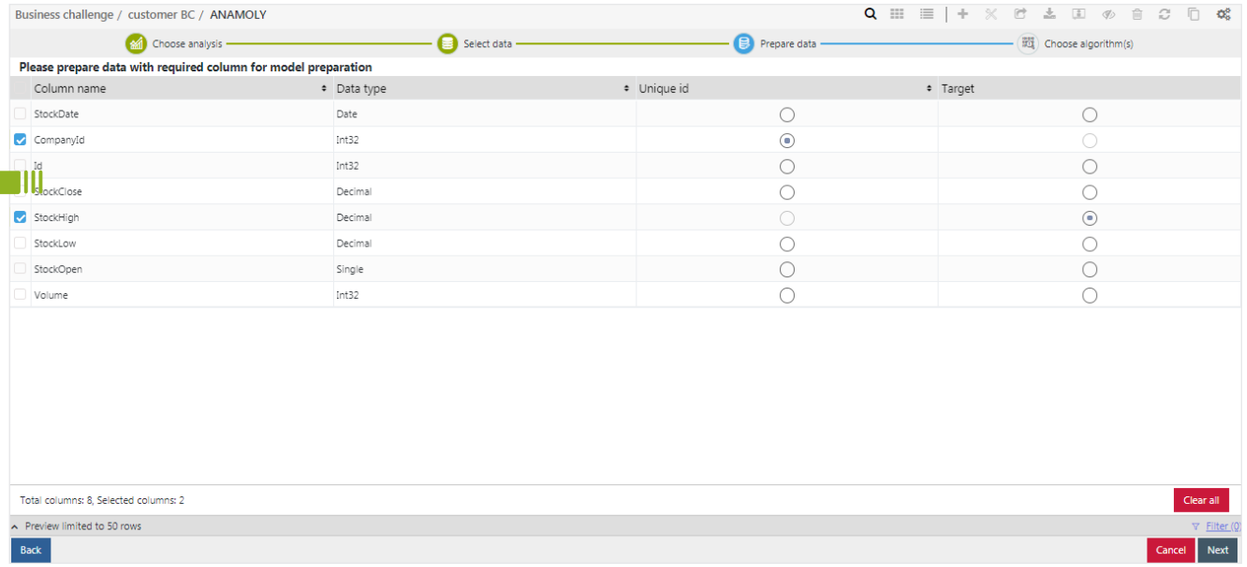


Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.

06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

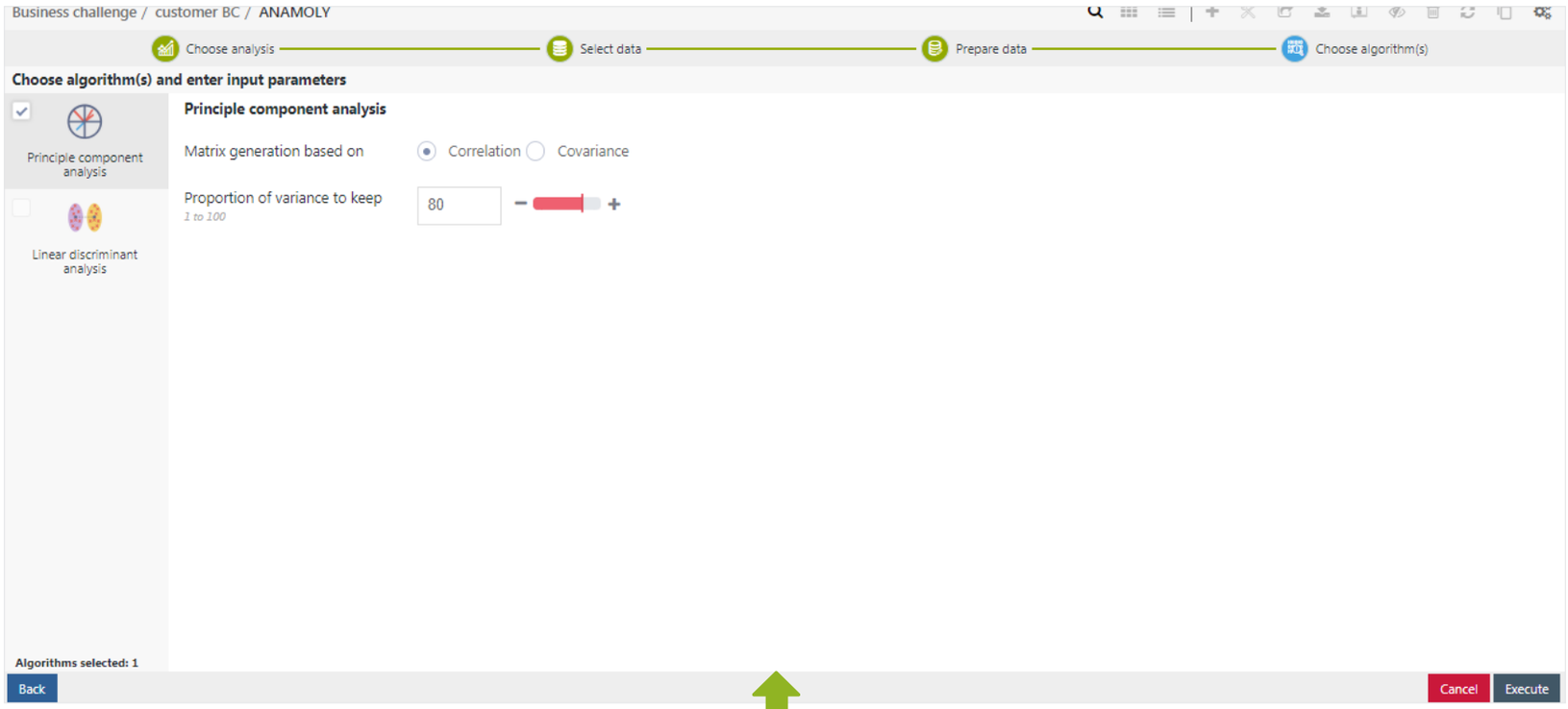
Prepare data	Select to include the variable in the data preparation process.
Unique id	Select the unique variable if the datasource has a unique value. On selecting the **Unique id** column for a variable **Target** column is disabled for the variable.
Target	This variable value is mandatory. Select a variable that you will like to predict that is dependent for the analysis. On selecting the **Target** column for a variable **Unique id** column is disabled for the variable.



07 TO CHOOSE THE ALGORITHM(S) :

1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

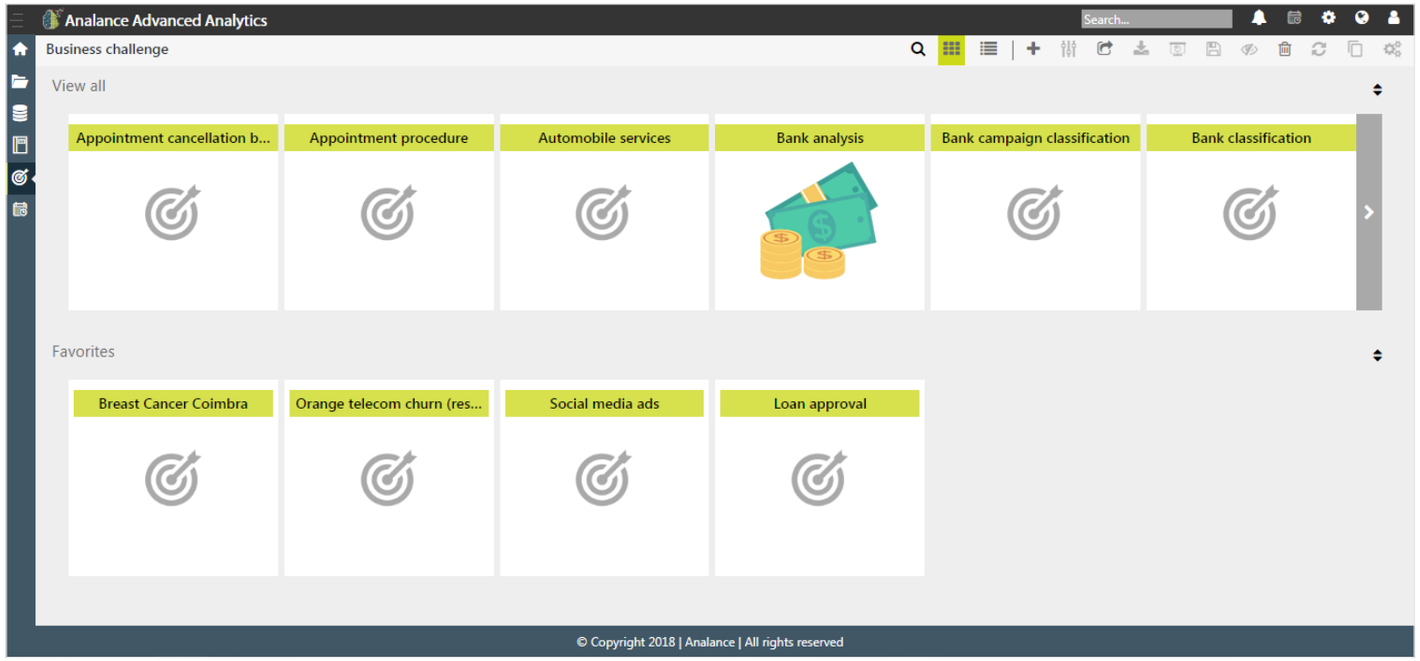
The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
Principle Component Analysis	Matrix generated based on: Correlation / Covariance Proportion of variance to keep: Increase or decrease the slider
Linear Discriminant Analysis	No input parameter

Analance

Add a Forecasting Model

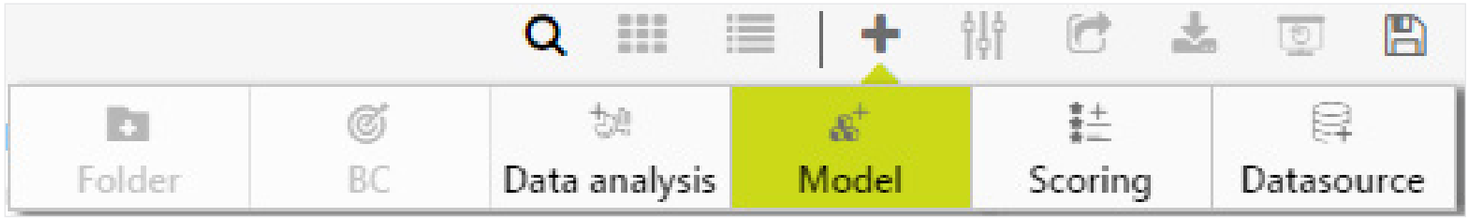


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

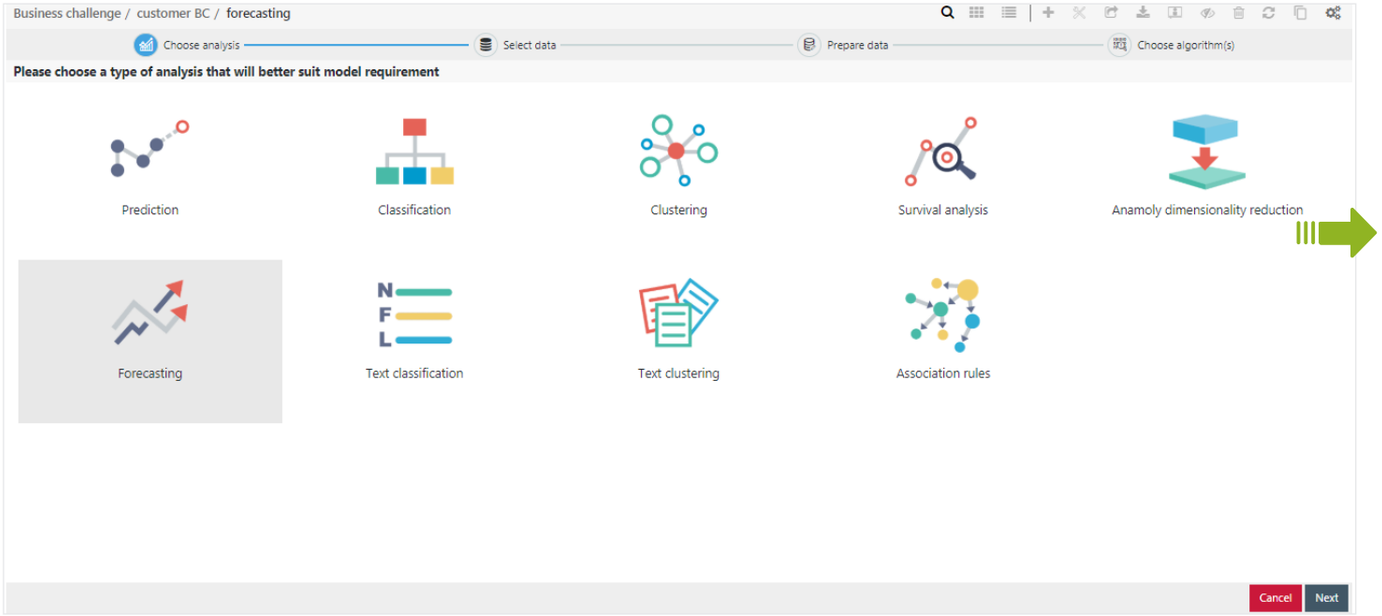
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.



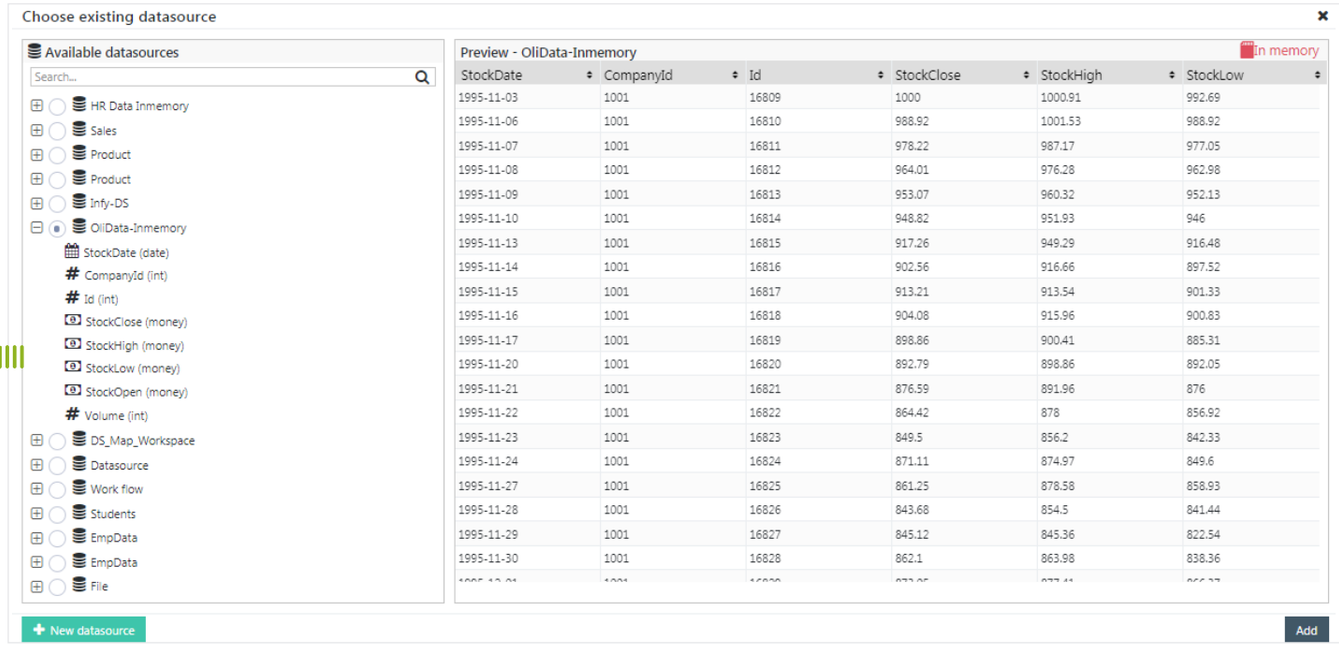
04

Choose Forecasting from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

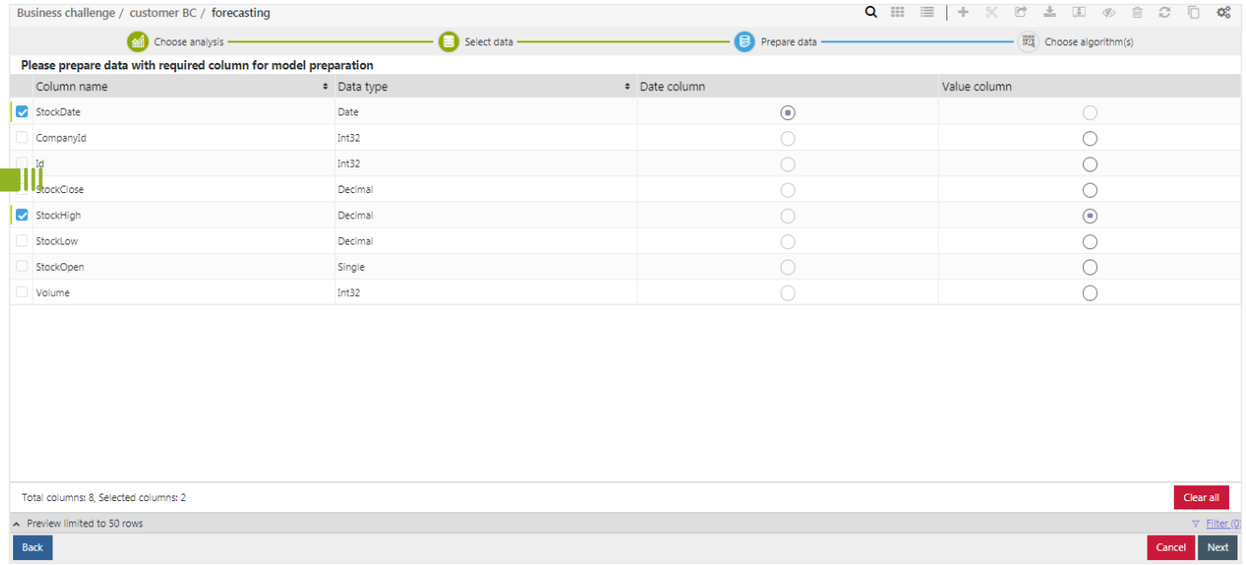
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.



06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

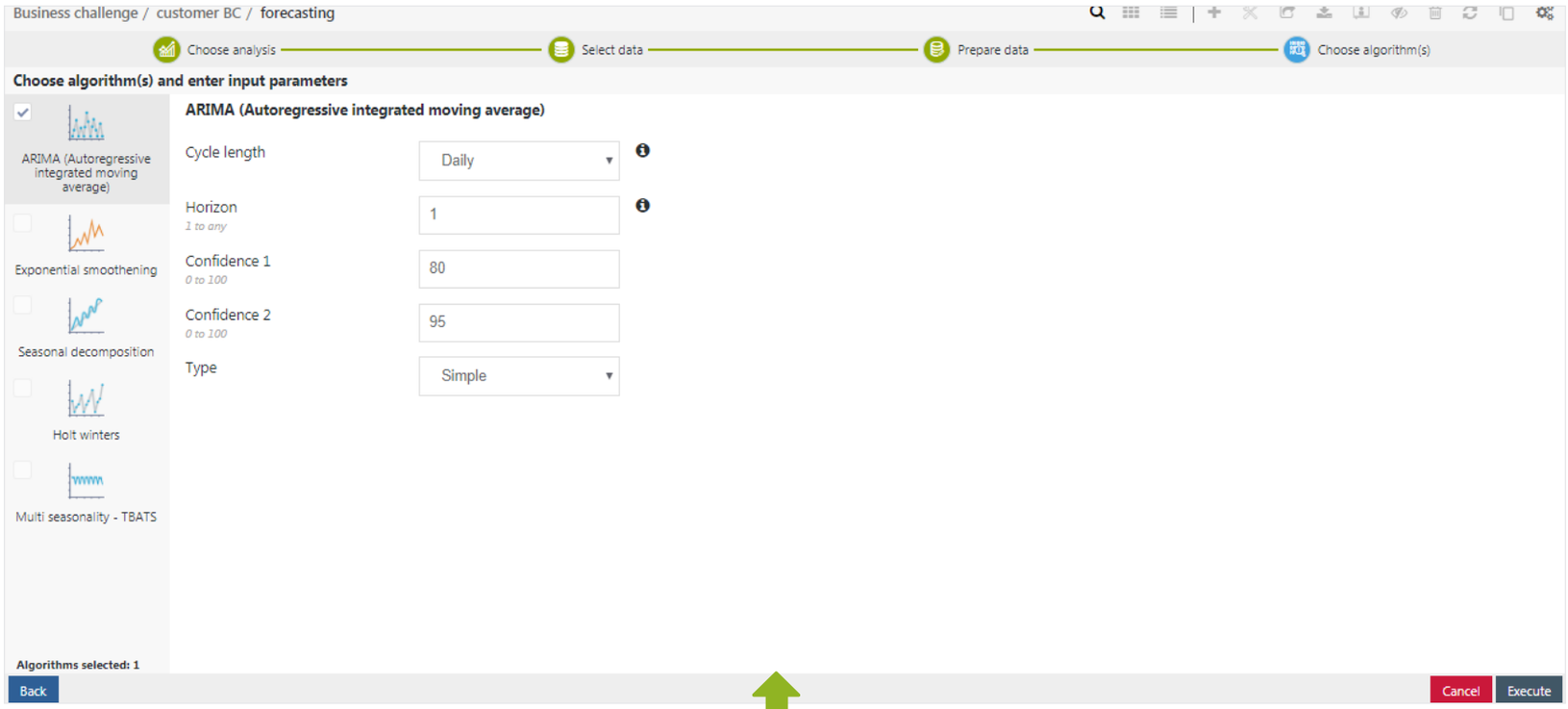
Prepare data	Select to include the variable in the data preparation process.
Data column	This is mandatory column. Select the date variable that will be the base for forecasting the value. On selecting the Date column for a variable, the **Value** column is disabled for the variable.
Value column	This is also a mandatory column. Select the desired forecasted value. On selecting the Value column for a variable **Date** column is disabled for the variable.



07 TO CHOOSE THE ALGORITHM(S) :

1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

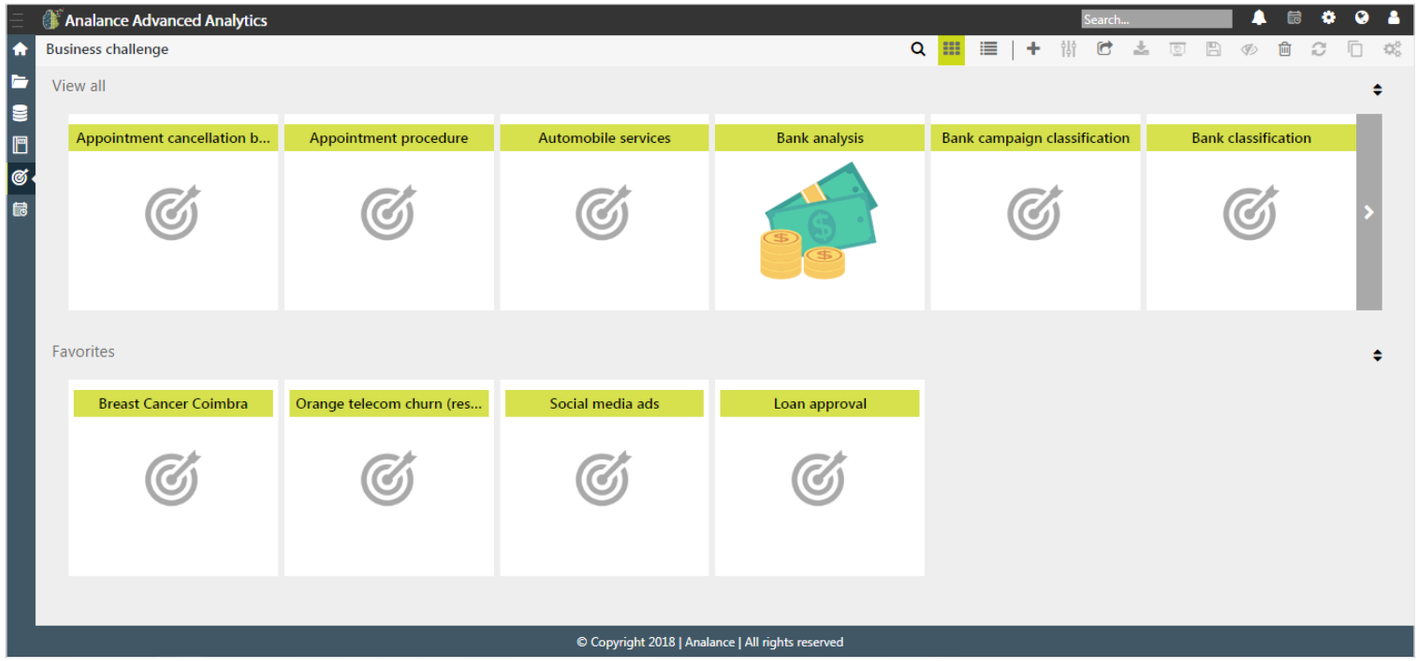
The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
ARIMA (Auto Regressive Integrated Moving Average)	Cycle length, Horizon, Confidence 1, Confidence 2, Type
Exponential smoothening	Cycle length, Horizon, Confidence 1, Confidence 2
Seasonal decomposition	Cycle length, Horizon, Confidence 1, Confidence 2
Holt winters	Cycle length, Horizon, Confidence 1, Confidence 2, Type
Multi seasonality (TBATS)	Cycle length, Horizon, Confidence 1, Confidence 2

Analance

Add a Text Classification Model

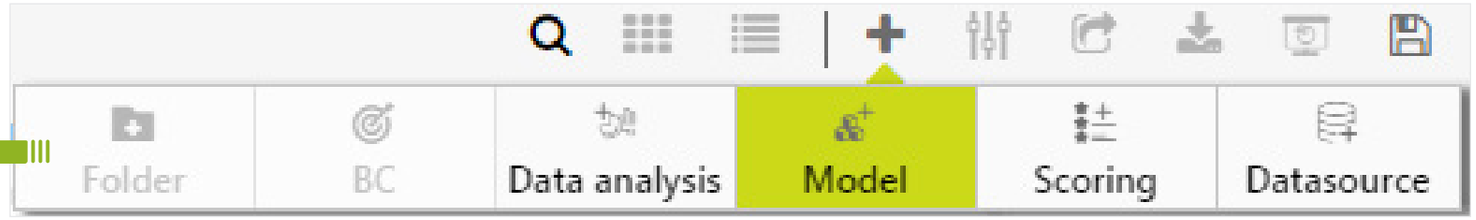


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

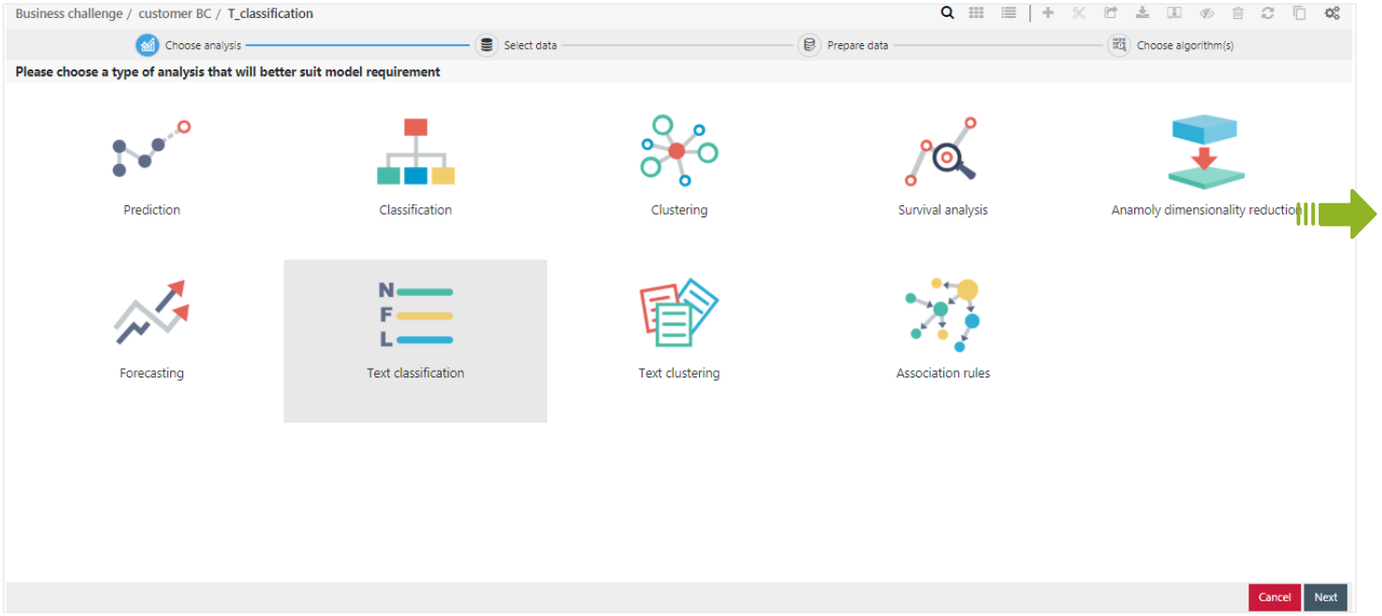
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.



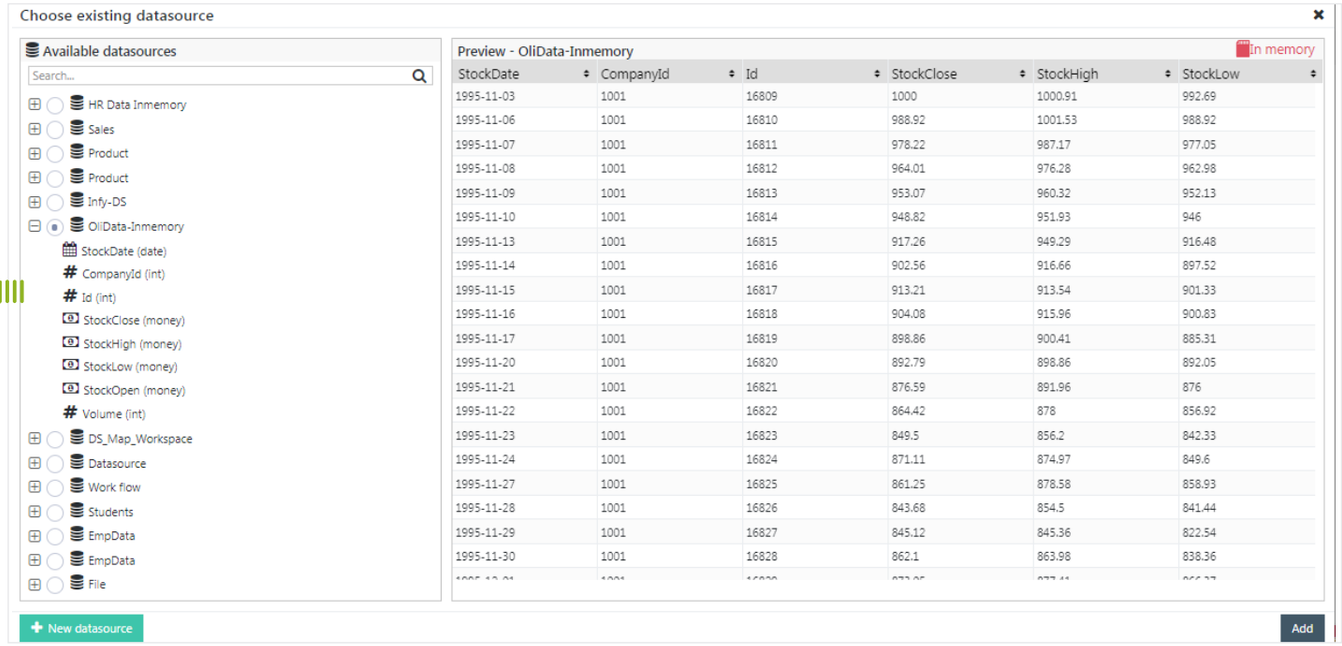
04

Choose Text classification from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

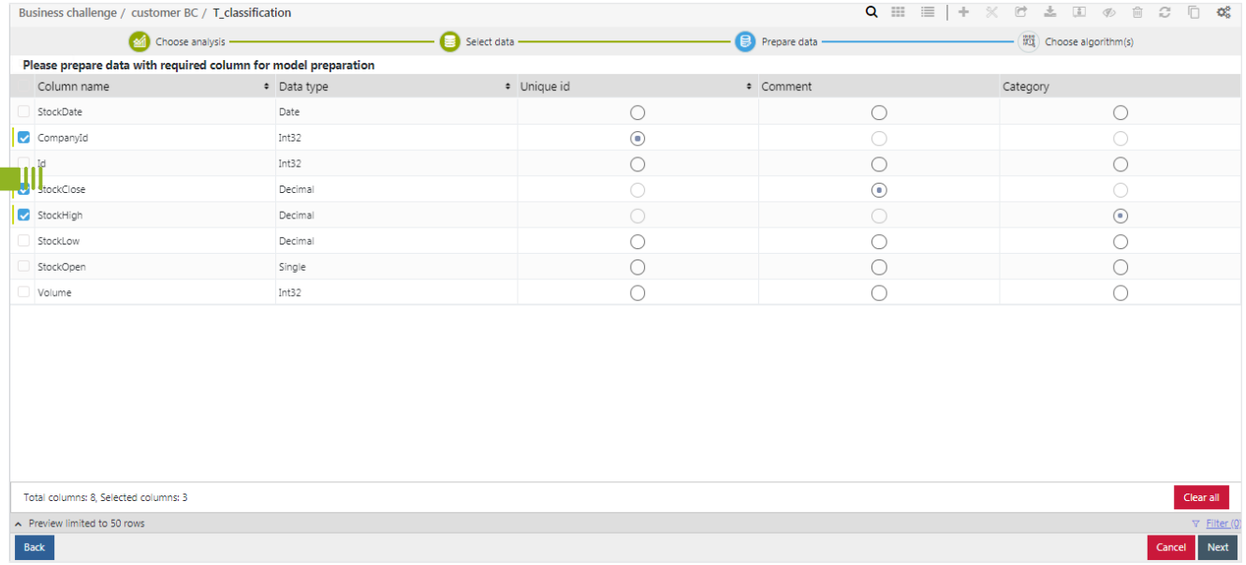
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.



06 PREPARE DATA :

To prepare the data for use in the model, select from the following options in the Prepare data table and then click Next.

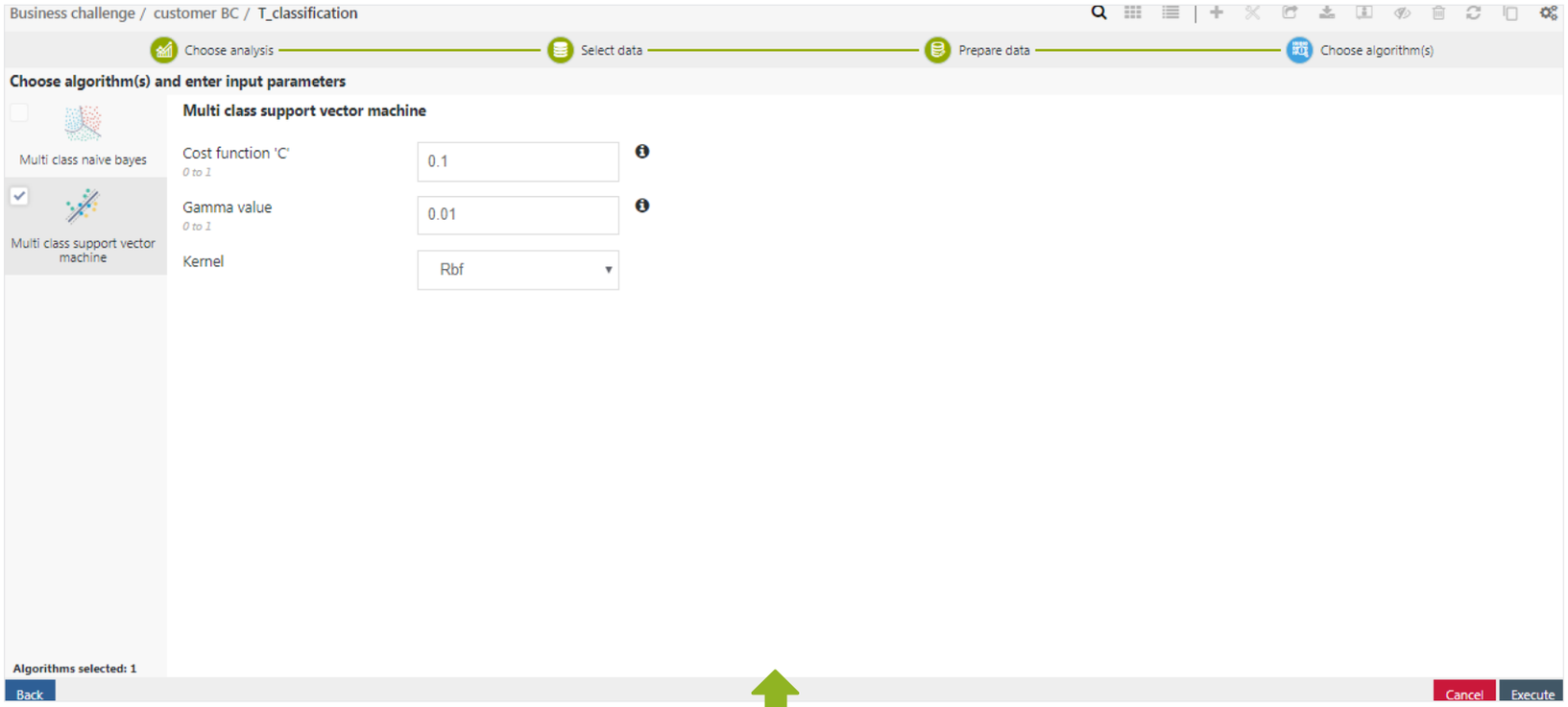
Unique id	This column is mandatory. Select the unique label to identify the variable as a unique comment.
Comment	This column is mandatory. Select the column that has text comments. If Comment is selected for a variable, then **Unique id** and **Category** for that variable are disabled. You cannot define more than one criteria for a variable.
Category	This column is mandatory. Select this column to define the variable as a Category.



07 TO CHOOSE THE ALGORITHM(S) :

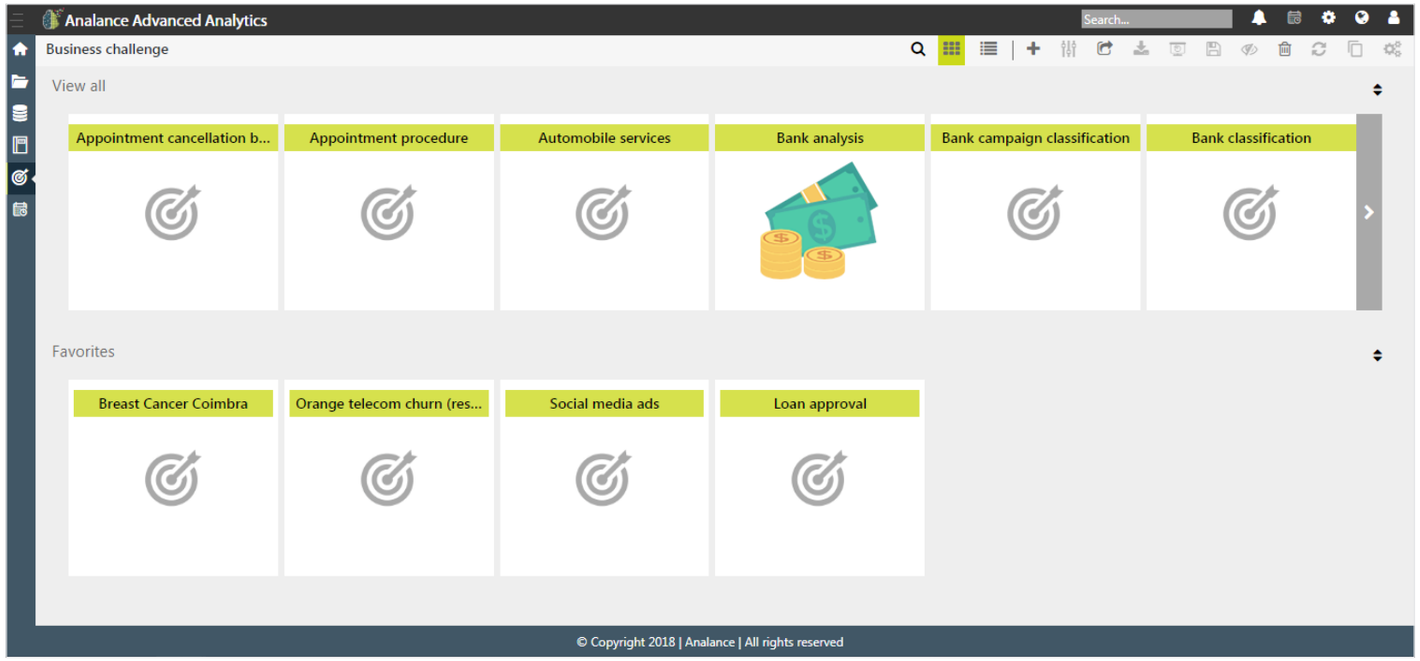
1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters. Based on the input, the algorithms displayed can be either two class or multi class.
2. Click Execute.

The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
Two class naïve bayes	Sensitive class
Two class K-nearest neighbor	Sensitive class

Add a Text Clustering Model

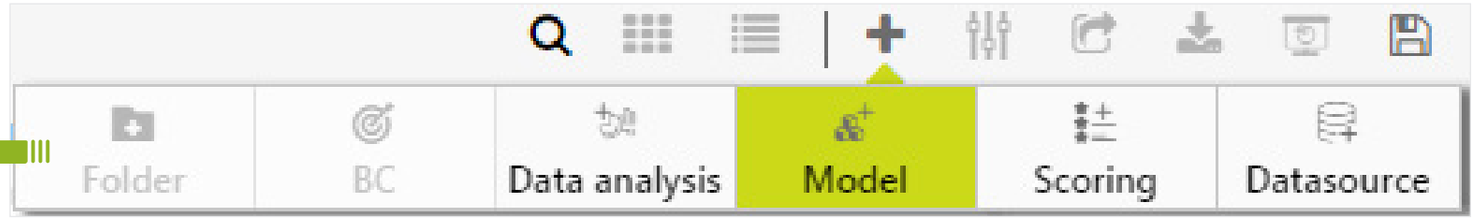


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

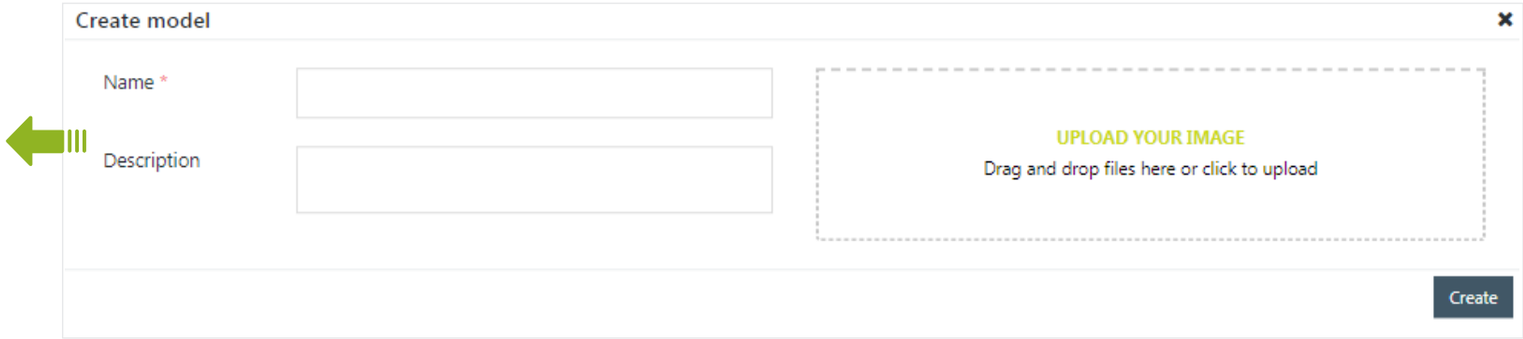
02

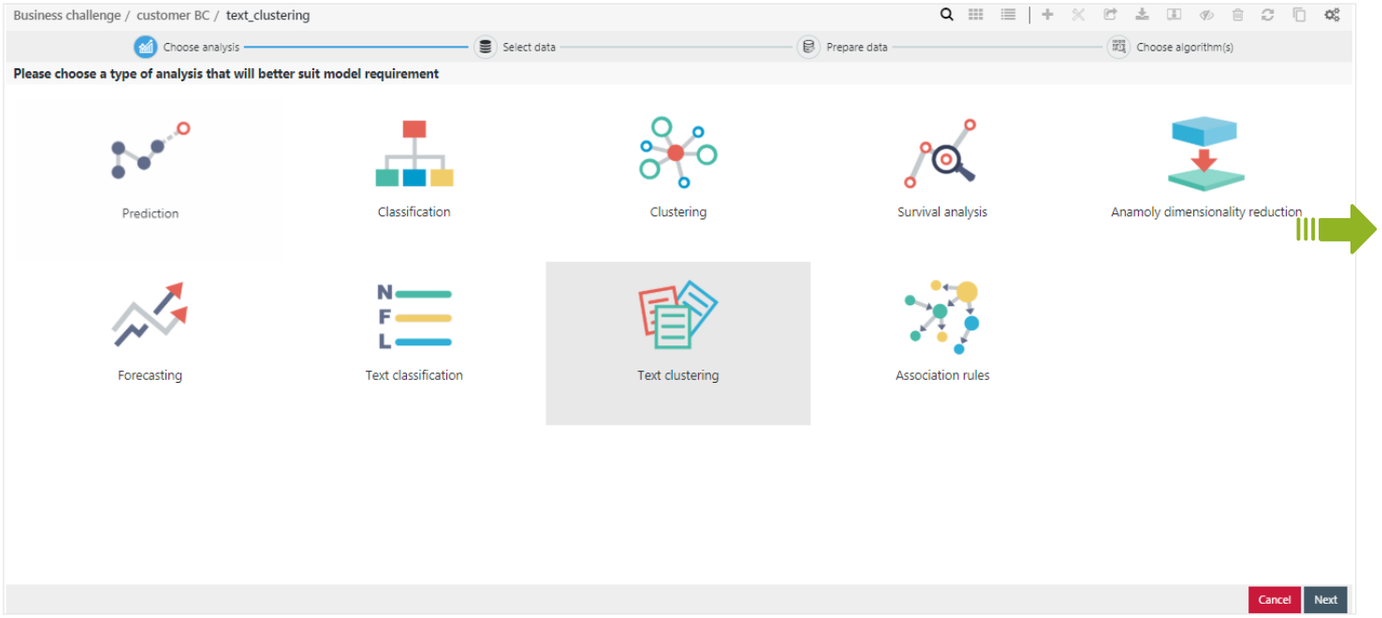
Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.





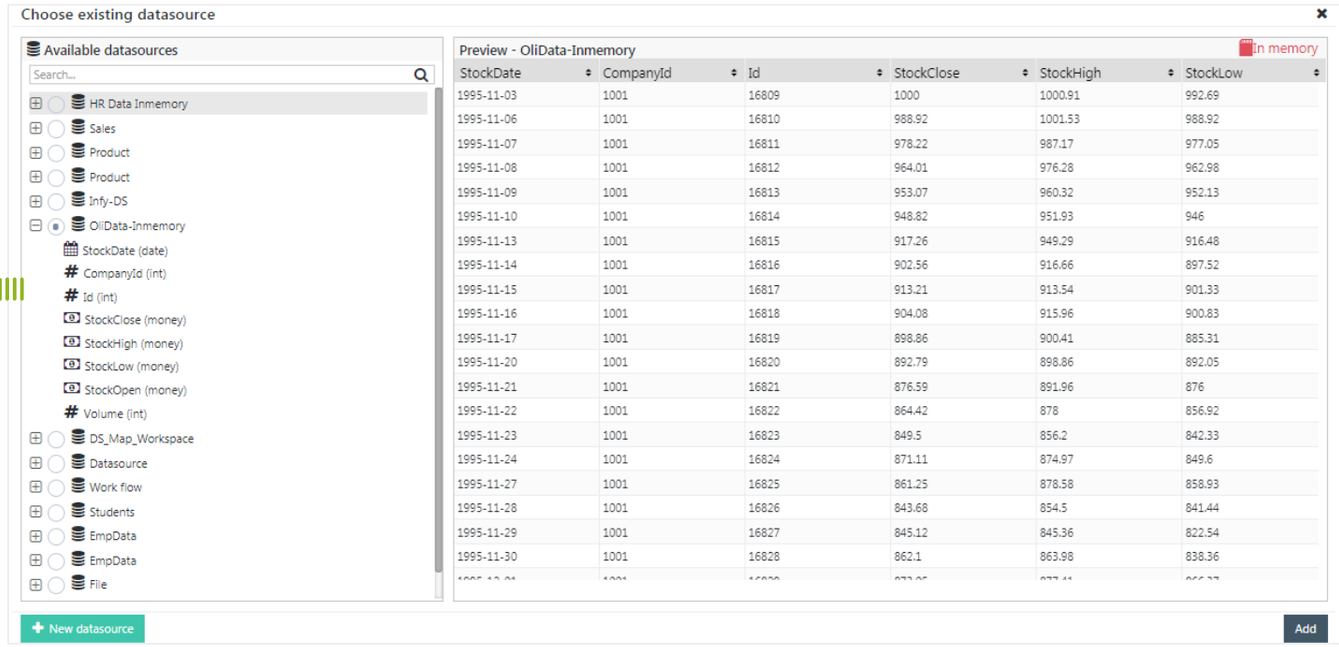
04

Choose Text clustering from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.

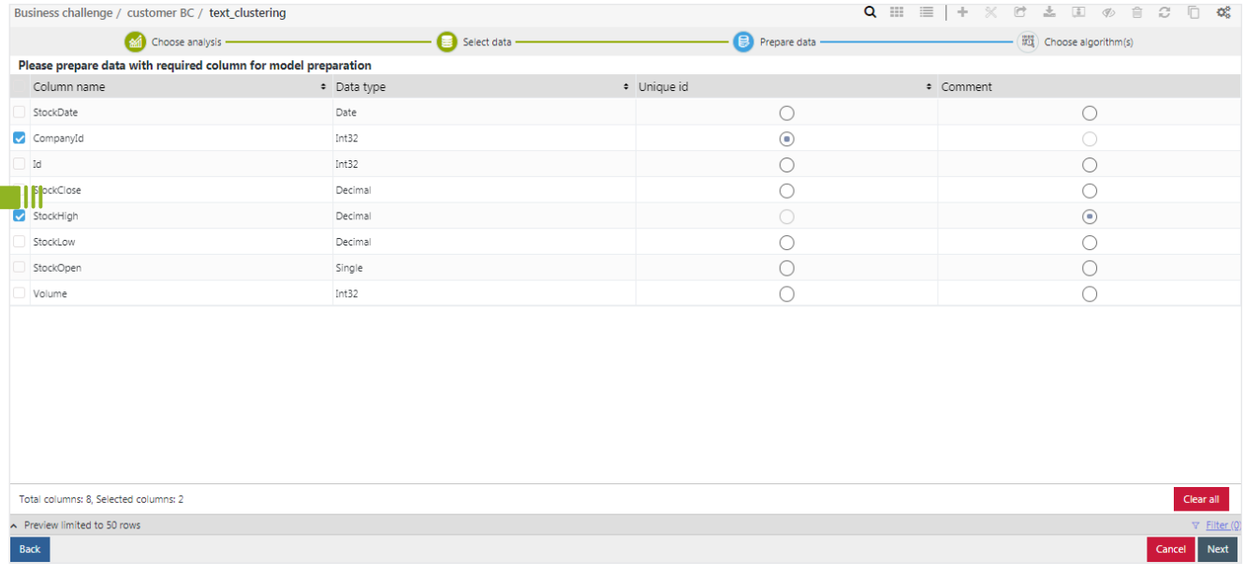
Note : You can also add a new datasource by clicking New datasource and following the step-by-step wizard.



06 PREPARE DATA :

Select to include the variable in the data preparation process.

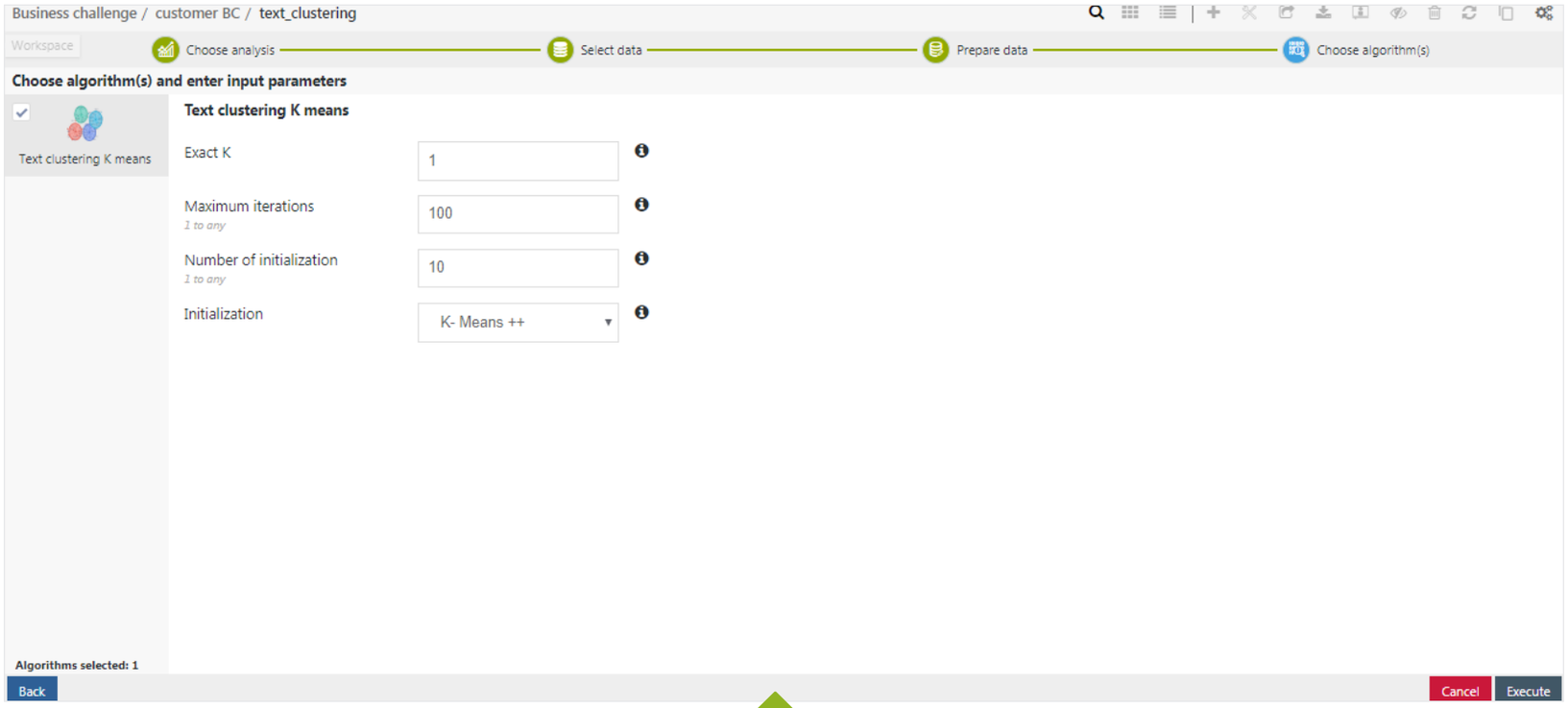
Prepare Data	Prepare the data with the required columns.
Unique id	This column is mandatory. Select the unique label to identify the variable as a unique comment.
Comment	This column is mandatory. Select **Comment** for the variable that has text comments.



07 TO CHOOSE THE ALGORITHM(S) :

1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

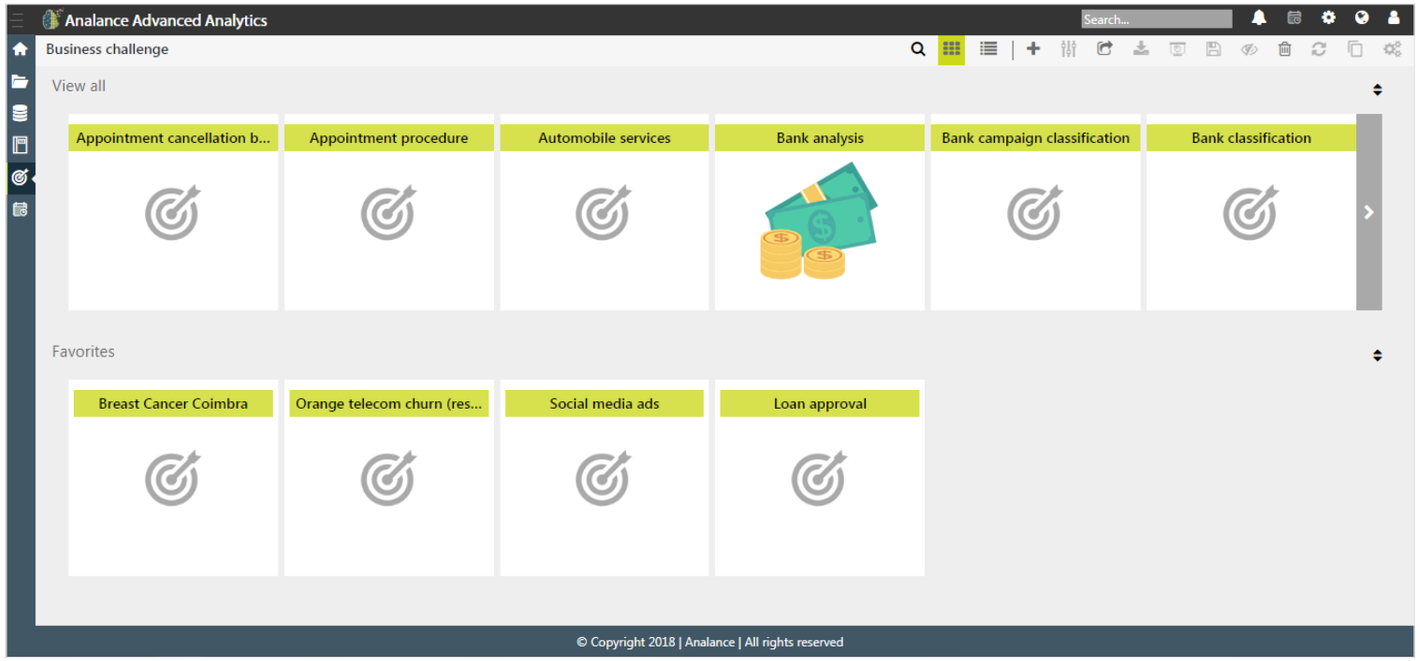
The model will be created in the background. When the model is ready, a notification will appear in the Notification area.



Algorithms	Input Parameters
TextClustering K Means	Exact K, Maximum number of iterations, Number of initialization, Initialization

Analance

Add an Association Rules Model

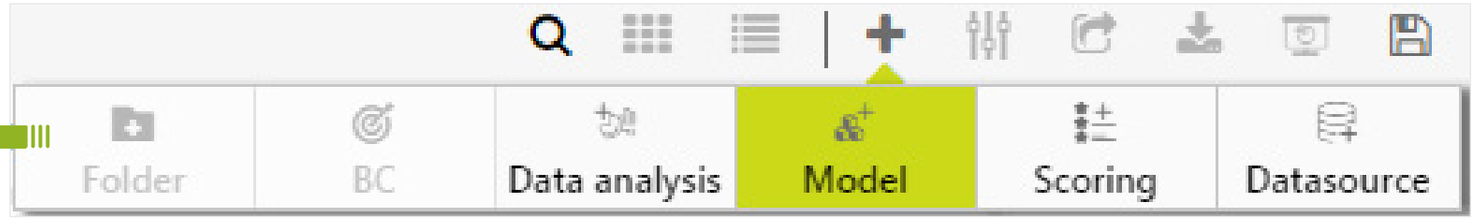


01 SELECT A BUSINESS CHALLENGE :

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.
3. Under Business challenge, select a Business challenge.

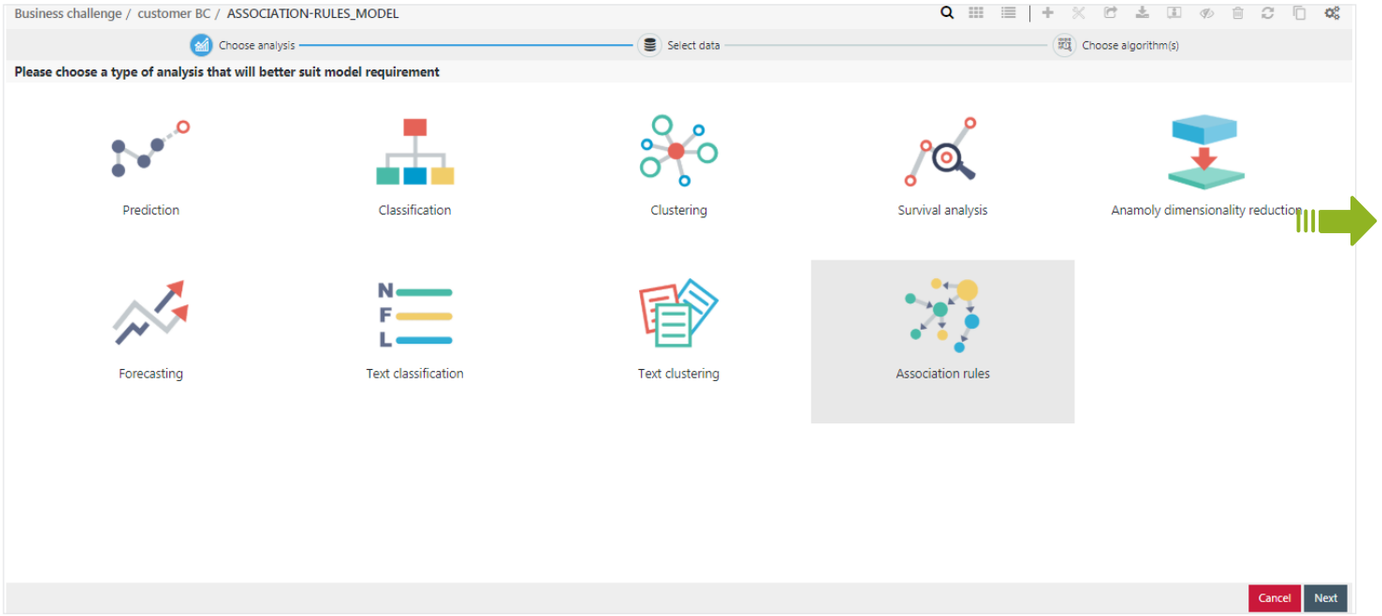
02

Click Add (+) > Model.



03

Type the Name and Description of the model and then click Create.

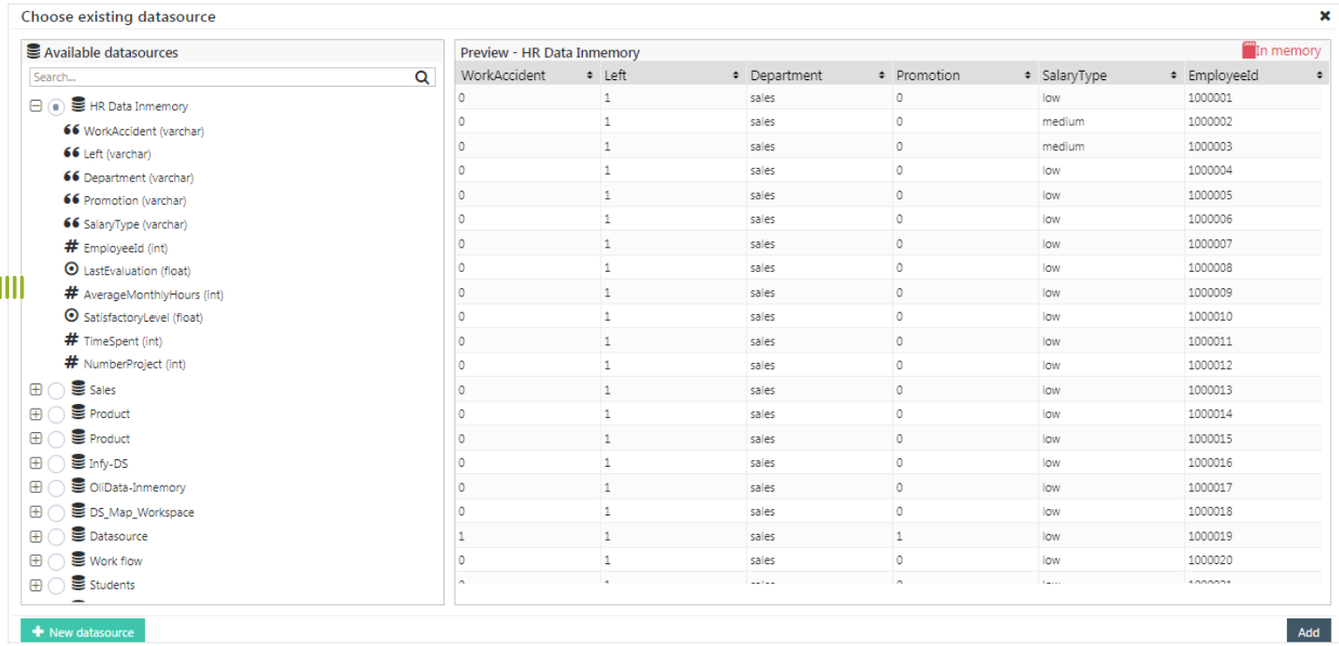


04

Choose Association rules from the available analysis type options and then click Next.

05 TO SELECT THE DATA :

1. Click Select a datasource for Data Analysis.
2. Under Available datasources, select an existing datasource.
   A preview will be shown.
3. Click Add.



Note : You can also add a new datasource by clicking ***New datasource*** *and following the step-by-step wizard.*

06 TO CHOOSE THE ALGORITHM(S) :

1. Choose the algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click Execute.

The model will be created in the background. When the model is ready, a notification will appear in the Notification area.

Algorithms	Input Parameters
Market_Basket_Analysis	Criterion, Value