Quick Start Reference Card - MODEL

MODEL

With AAA Model, build a machine learning analysis that allows to import, prepare & cleanse and consolidate data for better deployment and deliver an actionable business insights

FEATURES

1. Descriptive, Predictive, Anomaly, and Time series algorithms.

BENEFITS

1. Analyze data using multiple algorithms through ensemble.
2. Quickly and easily create models.
3. Discover structure and easily predict categories.
4. Capability to empower even a novice user.
5. Various modeling techniques are selected and applied and their parameters are calibrated to optimal values

PREREQUISITE

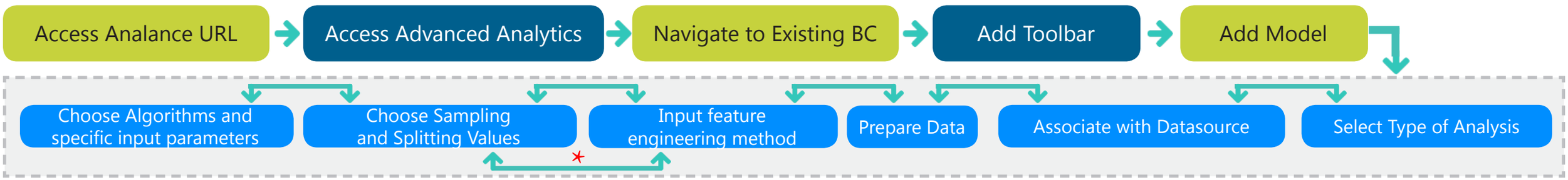
1. Data should be available in any one of different data connector format supported in Analance.

RELATIONAL DATABASES	BIG DATA	API	FILE	SOCIAL MEDIA
ORACLE	hadoop	OData	CSV	facebook
MySQL, PostgreSQL	HIVE, Solr	SharePoint	TXT	twitter
TERADATA	Hortonworks	Google Analytics	XLS	
Microsoft SQL Server, Access	Apache Spark, HDP		XML	
IBM DB2	cloudera IMPALA, MAPR, cloudera			

MODEL - WORK FLOW

WORK FLOW TO ADD MODEL IS SHOWN BELOW:

MENU PATH: >> Add > Click Model icon

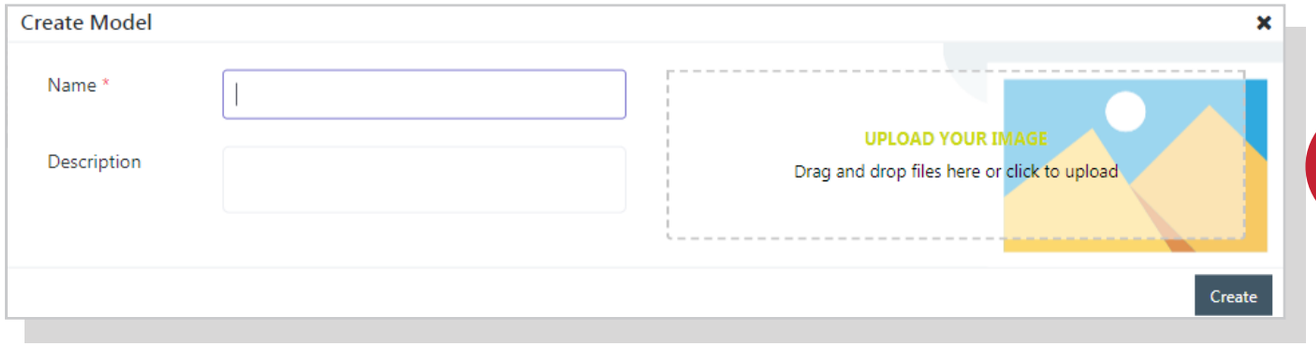


1. Log into Analance using Analance URL
2. Select Advanced Analytics
3. Navigate to existing BC.
4. Select Add Toolbar- Navigate to the Model path to create Model. Refer to the Path section for more details.
5. Add a Model

* There are two models: Supervised and Unsupervised. Both the steps, feature engineering method and sampling and splitting values step is not required for unsupervised models.

Quick Start Reference Card - MODEL

PROCEDURE TO ADD PREDICTION MODEL

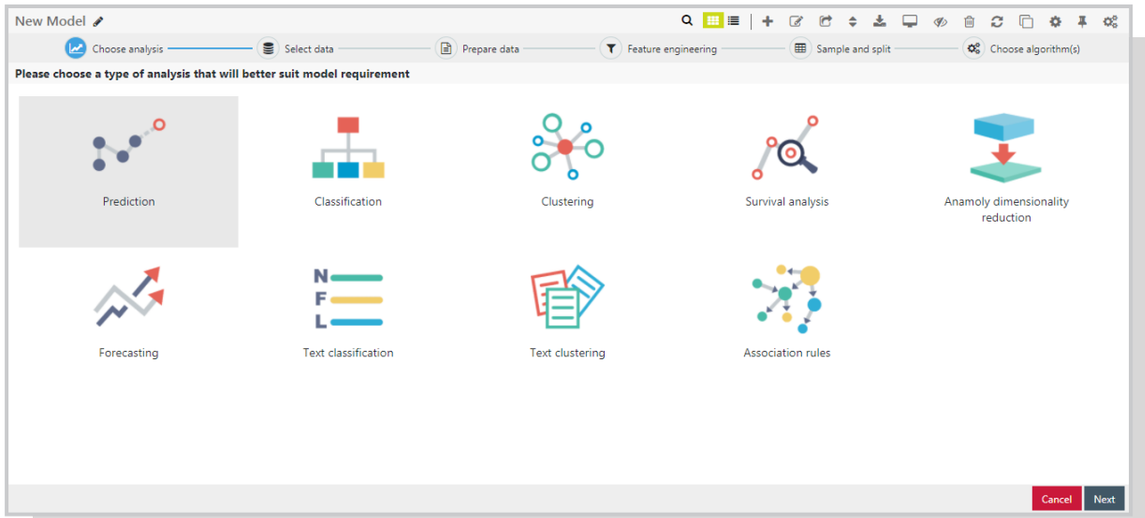


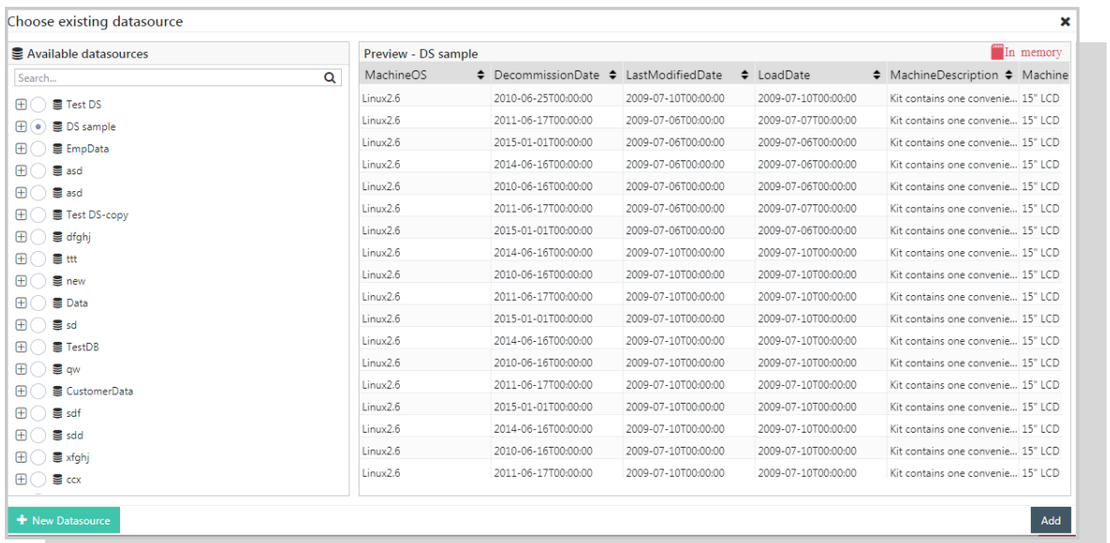
1

Create Model :
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

Choose Analysis :
Select Prediction analysis from the list of options.

2

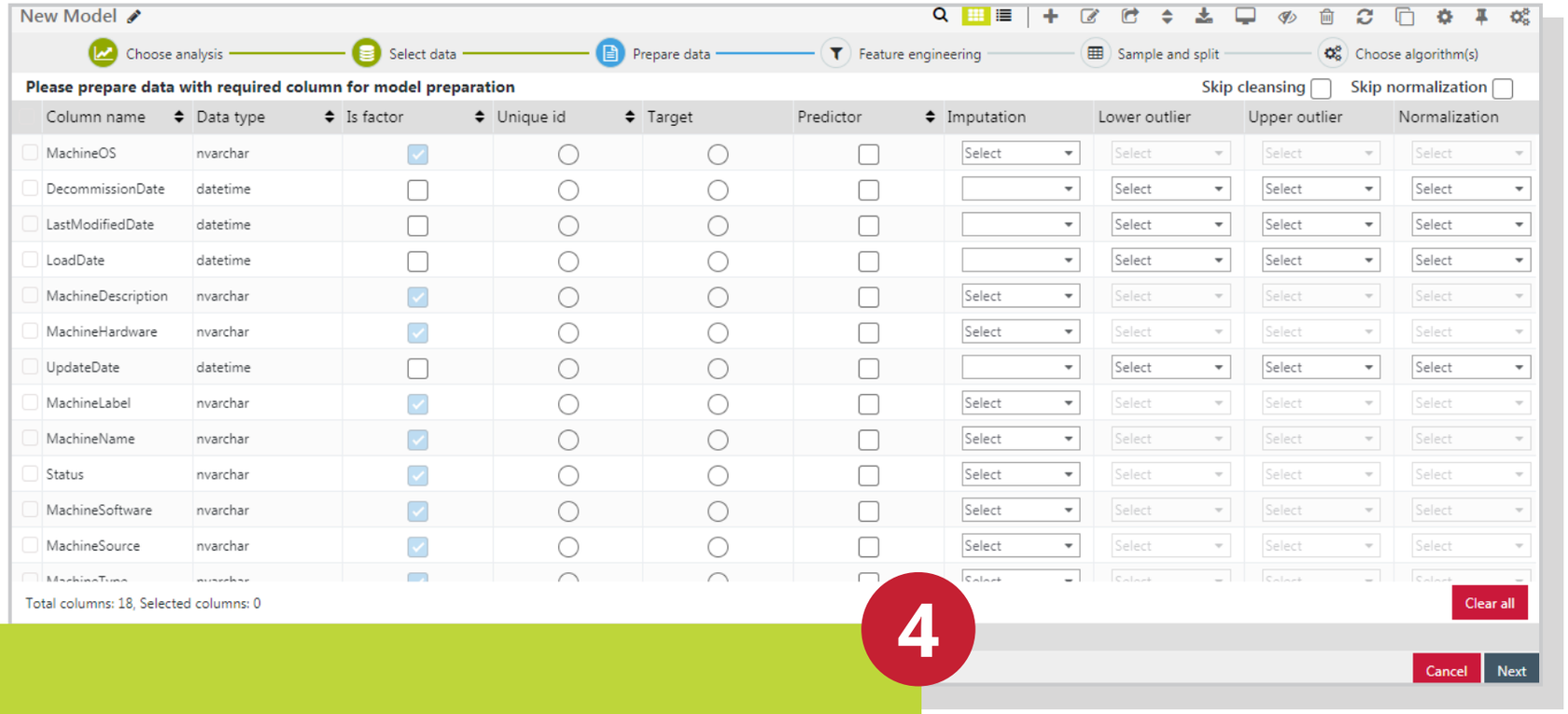




3

Choose existing Datasource :
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

Quick Start Reference Card - MODEL



Prepare data : Prepare the data with the required columns.

Is factor: 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 value is mandatory. Select a variable that you will like to predict that is dependent for the analysis.

Predictor: This column value 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 then the options are: Mean, Median, Mode and User Input (Number). If the variable is factor / categorical then the Options: Most frequent, Least frequent, User Input (Text).

Lower outlier: Replace the lower outlier values with any one of the options: Percentile 2nd, Percentile 5th, User input.

Upper outlier: Replace the upper outlier values with any one of the options: Percentile 95th, Percentile 98th, User input.

Normalization: Select from the options Z-score, Min Max, Logistic, Log normal, Tanh to normalize data.

Skip Cleansing and Skip Normalization: if available data is cleansed and normalized or user wants to view raw data to be analyzed.

Quick Start Reference Card - MODEL

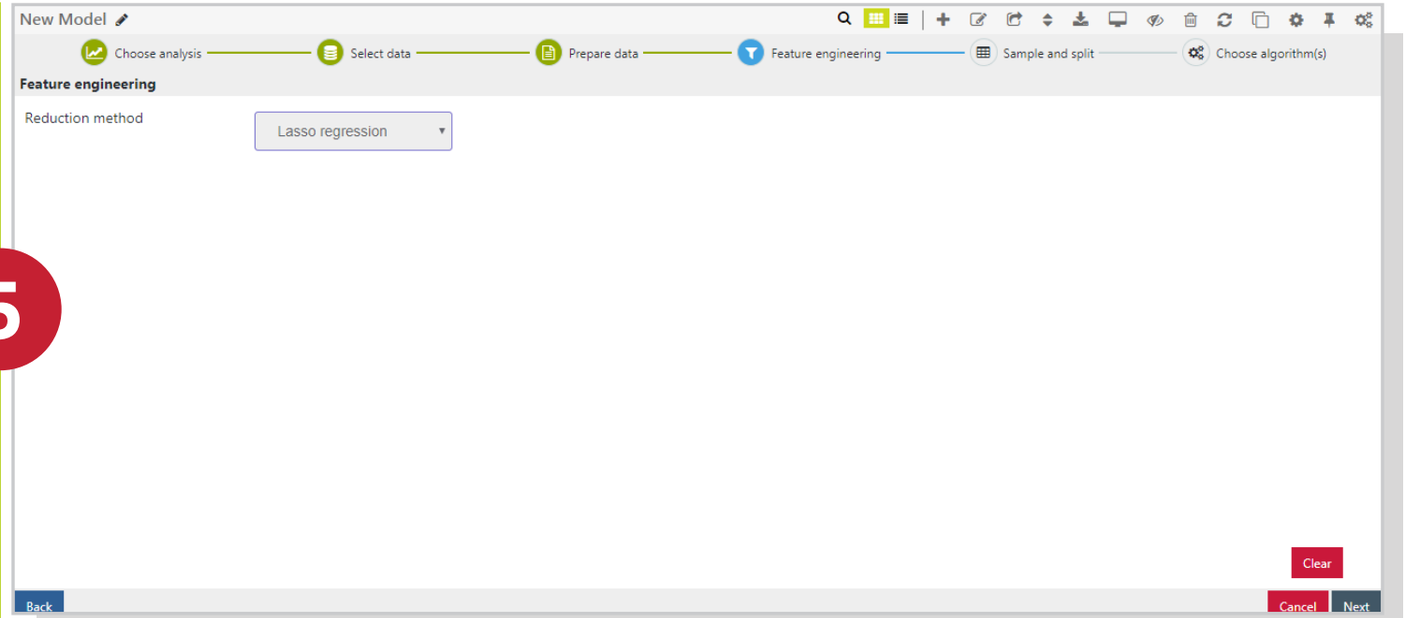
Feature engineering :
Select the reduction method. Tree based, Principle component analysis, Lasso Regression.

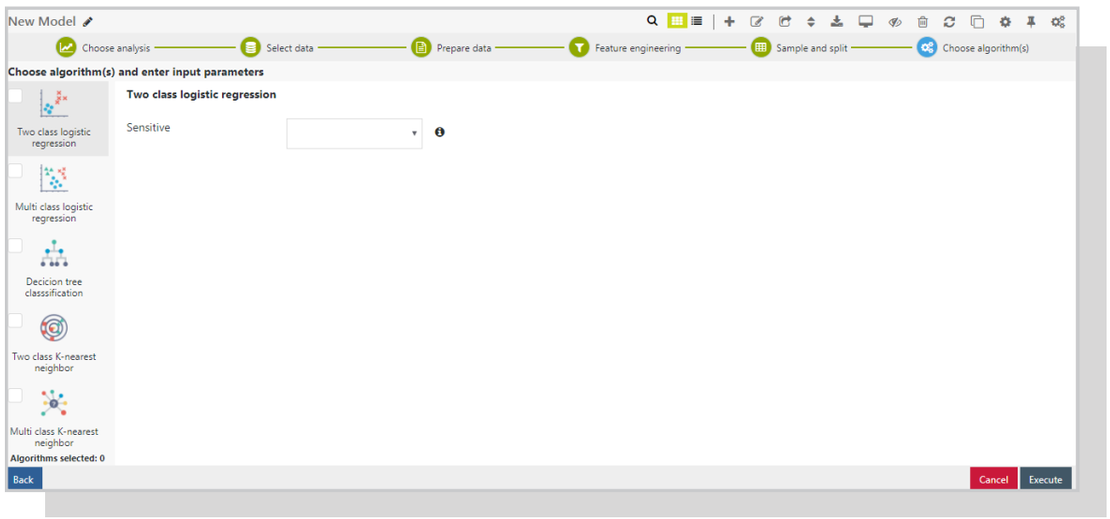
Tree based :
Specify Number of Trees and Depth of Trees.
Principle component: Select Matrix generation based on: Correlation / Covariance.

Proportion of variance to :
Use the slider to increase or reduce the value as required.

Normalization :
Select Z-score to standardize the values.

5





6

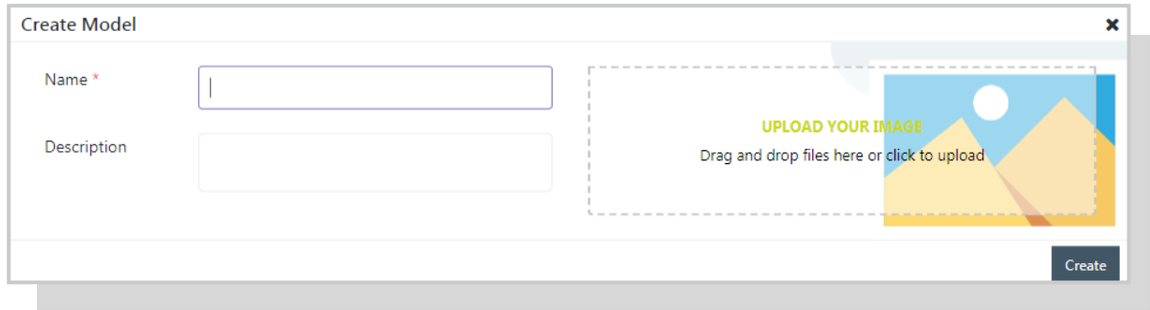
Choose algorithms :
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

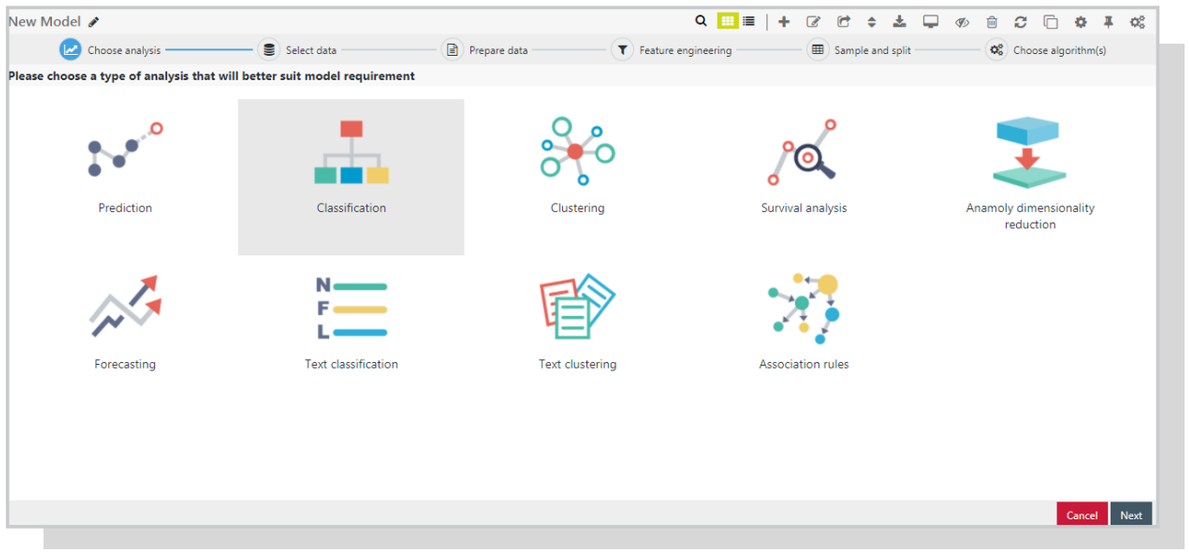
ALGORITHMS	INPUT PATRAMETERS
Multiple linear regression	No input parameter
Decision Tree regression	Minimum split, Minimum bucket, Maximum Depth
Bayesian linear regression	No input parameter
Neural network regression	Sensitive value
Multi class k-nearest neighbor	Hidden layer size, Maximum iterations

PROCEDURE TO ADD CLASSIFICATION MODEL

Create Model :
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1



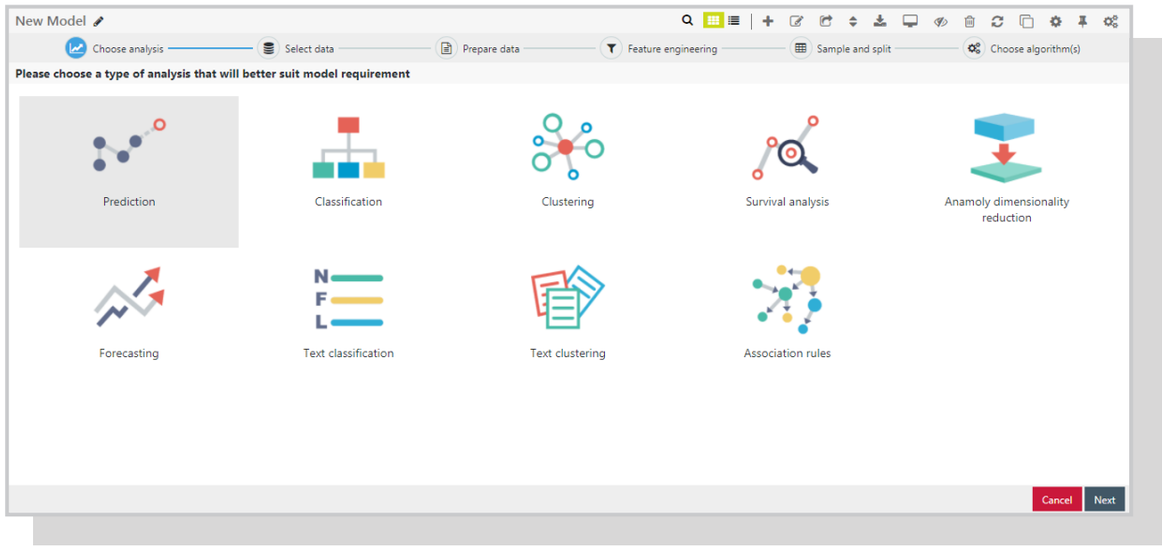


2

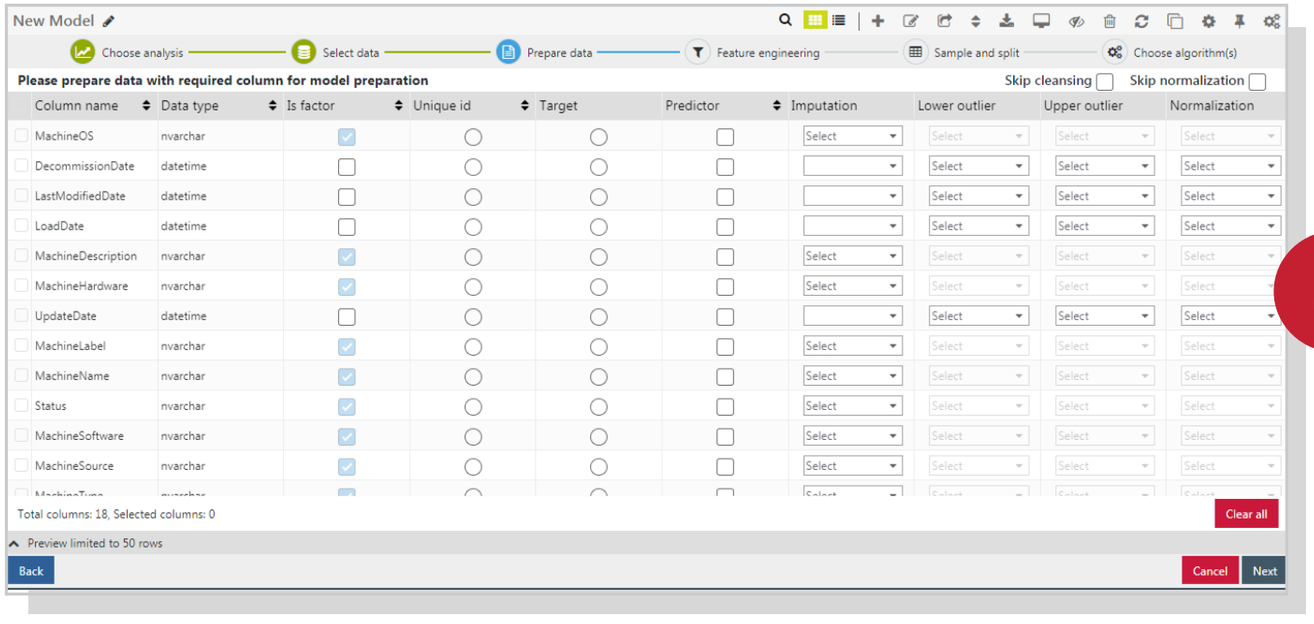
Choose Analysis :
Select Classification analysis from the list of options.

Choose existing datasource :
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

3



Quick Start Reference Card - MODEL



4

Prepare Data : Prepare the data with the required columns.

Is factor: Select if the variable is to be treated as categorical.

Unique id: Select the unique variable if the datasource has an unique value.

Target: This variable value is mandatory. Select a variable that you will like to predict that is dependent for the analysis.

Predictor: This column value is mandatory and independent. Select columns to influence the target variable.

Imputation: Select substitute value in case there are missing values for each variable. Options: Most frequent, Least frequent, User Input.

Lower outlier: Select the lower value that is less than the variable value, the options are: Percentile 2nd, Percentile 5th, User input.

Upper outlier: Select the upper value that is greater than the column value, the options are: Percentile 95th , Percentile 98th , User input.

Normalization: Select from the options Z-score, Min Max, Logistic, Log normal, Tanh to normalize data.

Skip Cleansing and Skip Normalization if available data is cleansed and normalized or user wants to view raw data to be analyzed.

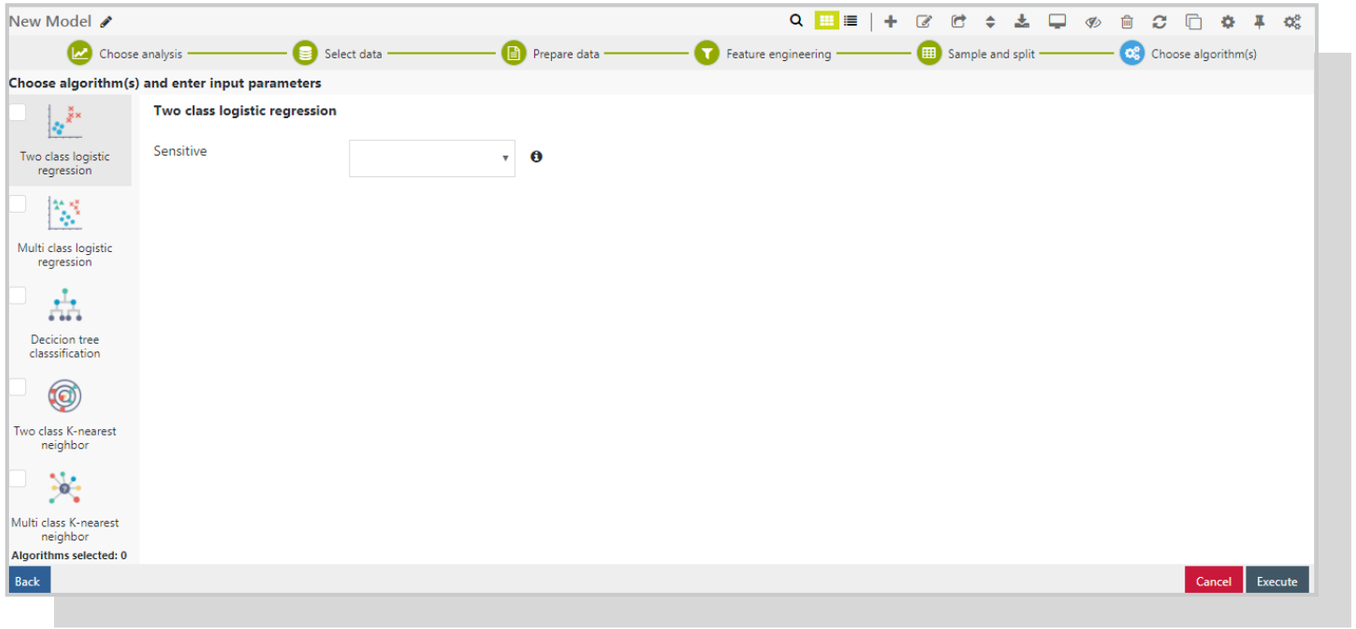
Feature engineering: Select the reduction method. No reduction, PCA based dimensionally reduction.

PCA based dimensionally reduction:
Select Matrix generation based on: Correlation / Covariance.

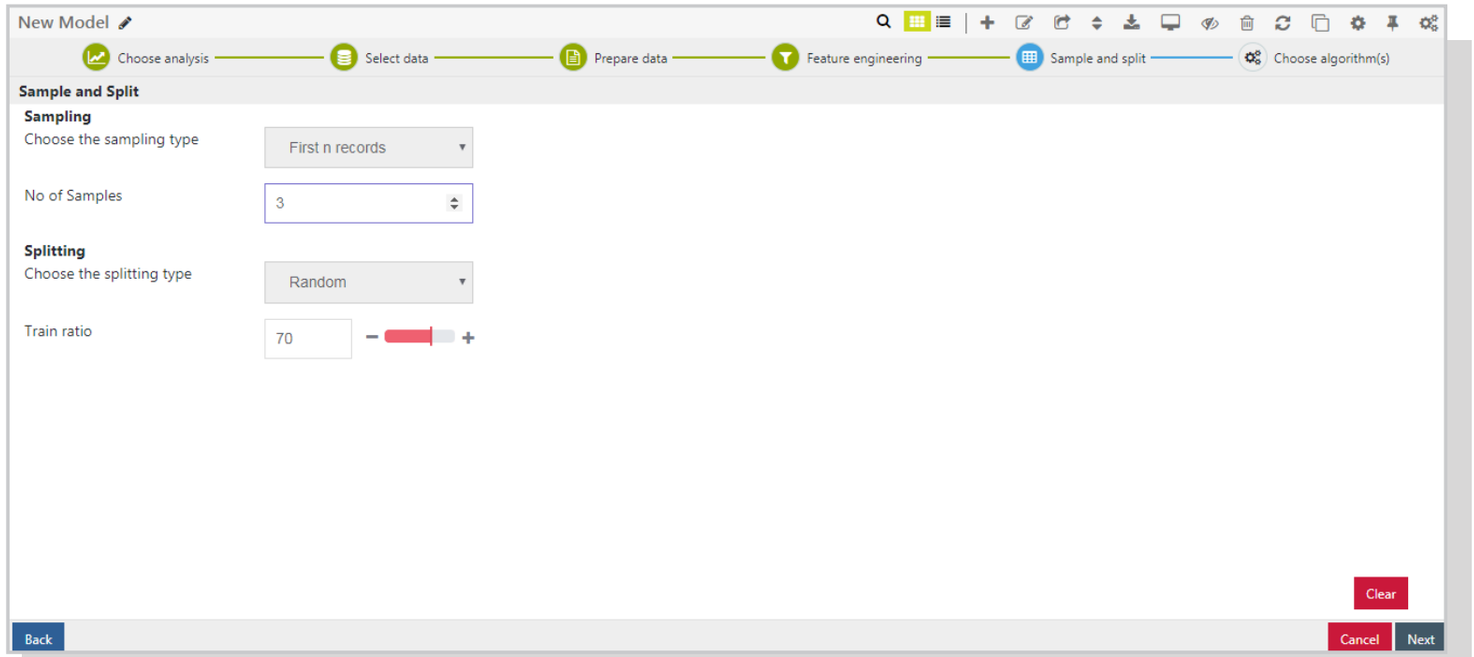
Proportion of variance to:
Use the slider to increase or reduce the value as required.

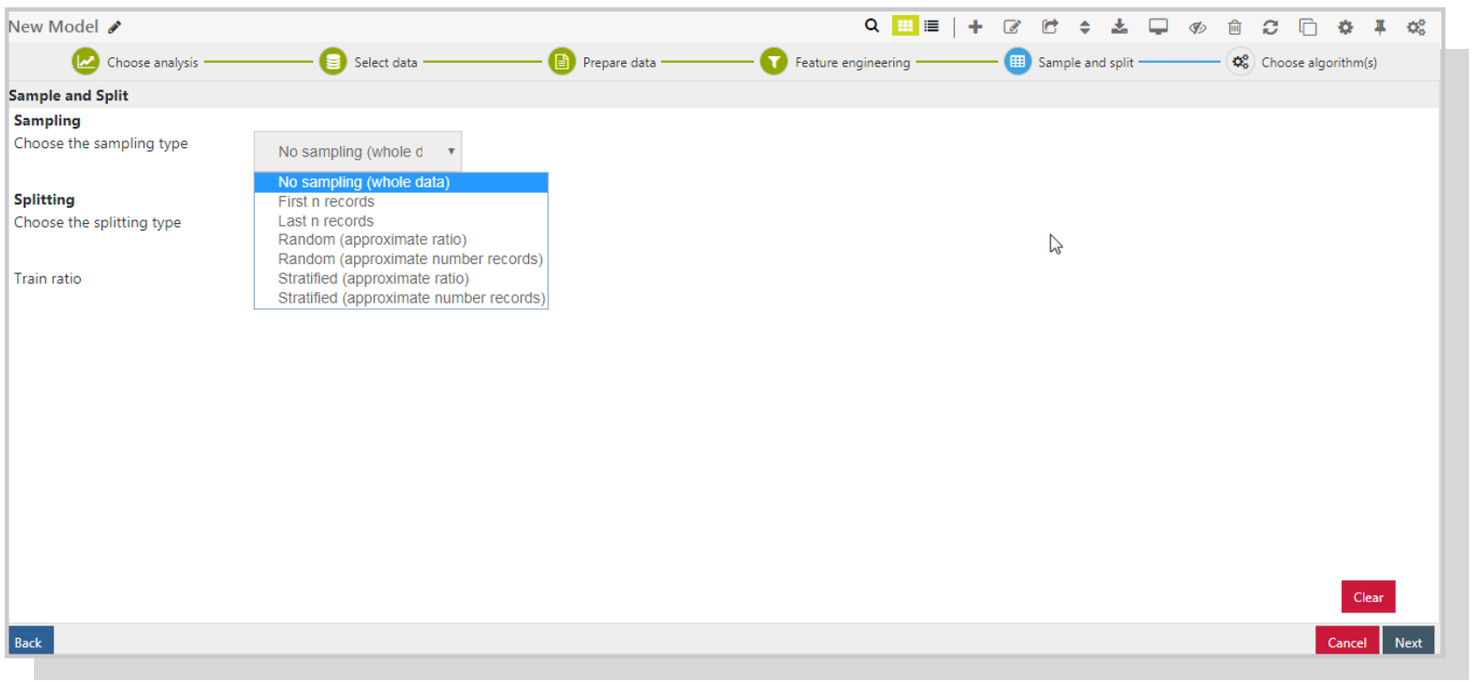
Normalization:
Select Z-score to standardize the values.

5



Quick Start Reference Card - MODEL





6

Sampling and Split : Choose the sampling and splitting the data method from the options.

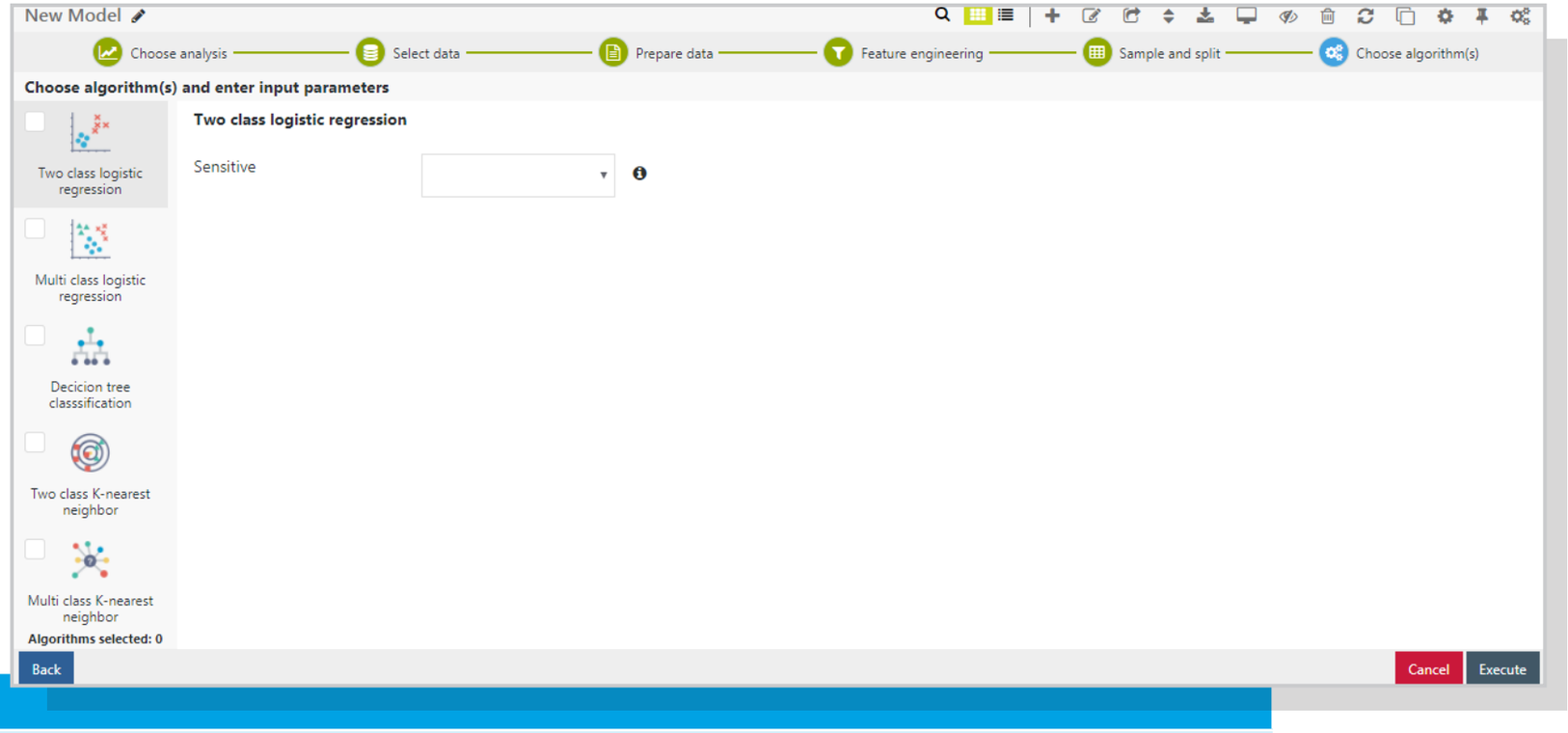
Sampling: Options for selecting the Sampling method:
- First n records
- Last n records
- Random (approximate ratios)
- Random (approximate no.of.records)
- Stratified (approximate ratio): Specify - No.of. samples, Column Information: select from the drop down list.
- Stratified (approximate number of records): Specify - No.of.samples, Column Information: select from the drop down list.

Splitting: Choose the Splitting type. Random
- Train ratio: The defaulted value is specified, the ratio can be increased or decreased using the slider.

Quick Start Reference Card - MODEL

Choose algorithms :
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

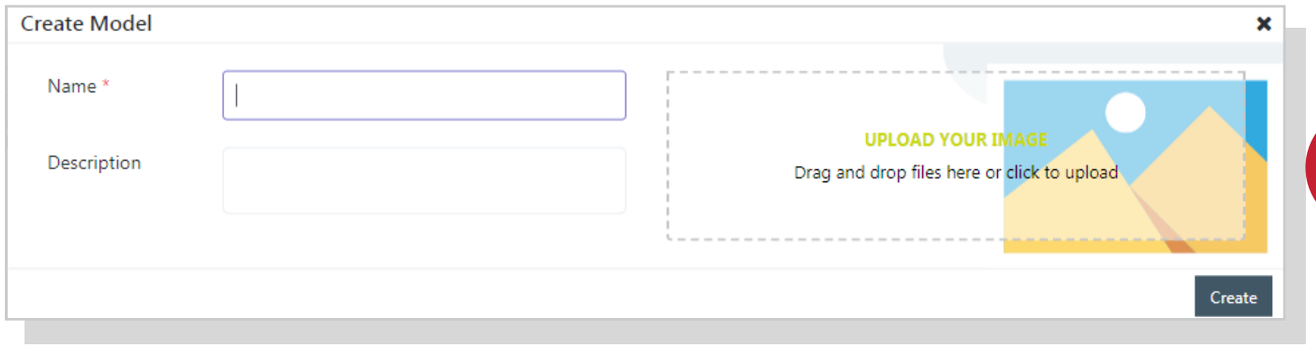
7



ALGORITHMS	INPUT PATRAMETERS
Two Class logistic regression	Sensitive value
Multi class logistic regression	No input parameter
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
Two class support vector machine	Sensitive, Cost function C, Gamma value, Kernel
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
Ordinal regression	No input parameter

Quick Start Reference Card - MODEL

PROCEDURE TO ADD CLUSTERING MODEL

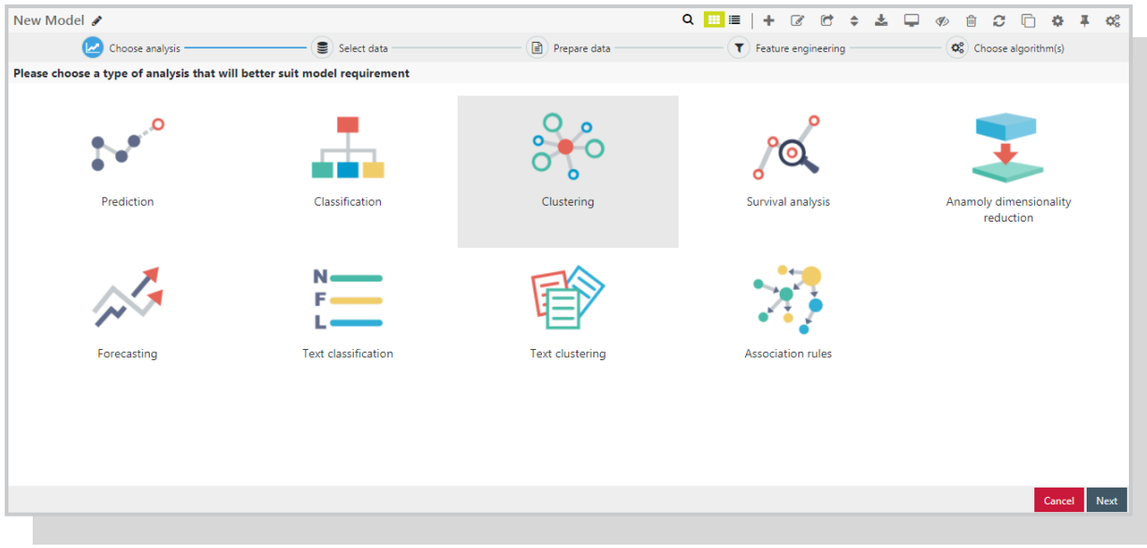


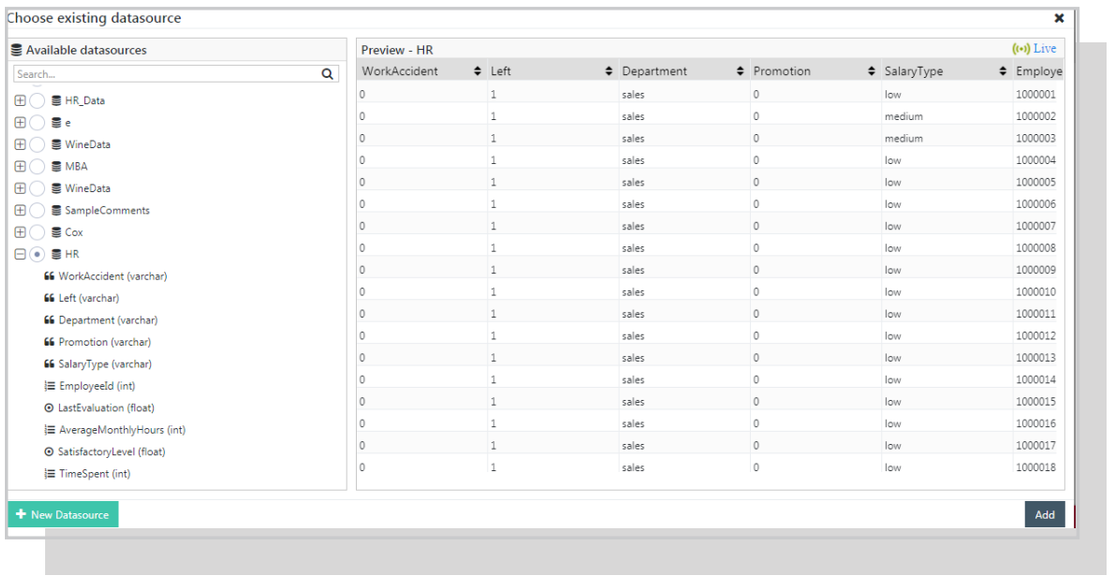
1

Create Model :
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

2

Choose Analysis :
Select Clustering analysis from the list of options.





3

Choose existing datasource :
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

Quick Start Reference Card - MODEL

Prepare Data: Prepare the data with the required columns.

Is factor: Select if the variable is to be treated as categorical.

Unique id: Select the unique variable if the datasource has a unique value.

Cluster Variable: This variable value is mandatory. Select a variable that you will like to cluster for the analysis.

Imputation: Select substitute value in case there are missing values for each variable. Options: Most frequent, Least frequent, User Input.

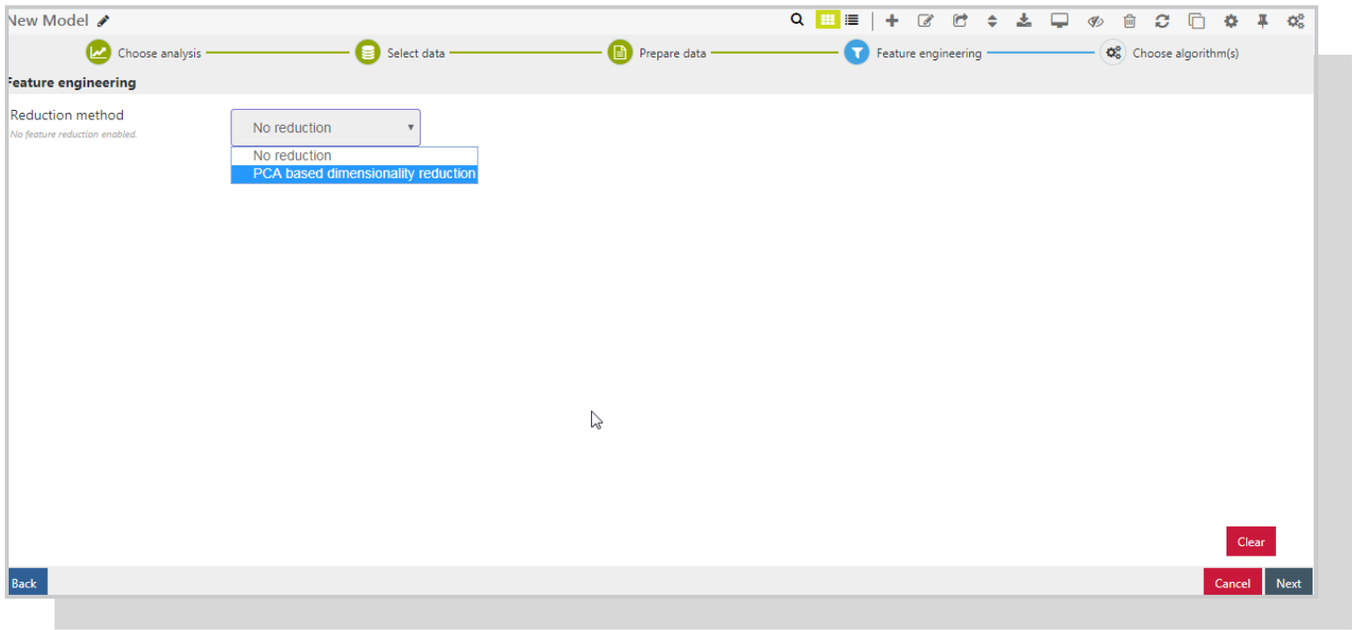
Lower outlier: Select the lower value that is less than the variable value, the options are: Percentile 2nd, Percentile 5th, User input.

Rescaling: Select from options to apply rescaling

Apply rescaling for the entire matrix: Enable to apply rescaling and select from the options listed.

Skip cleansing: Enable to skip cleansing. On selecting Imputation, Lower outlier, Upper outlier columns will be deactivated.

4



5

Feature engineering: Select the reduction method. Tree based, Principle component analysis, Lasso Regression.

Tree based: Specify No.ofTrees and Depth of Trees.

Principle component: Select Matrix generation based on: Coorelation / Covariance.

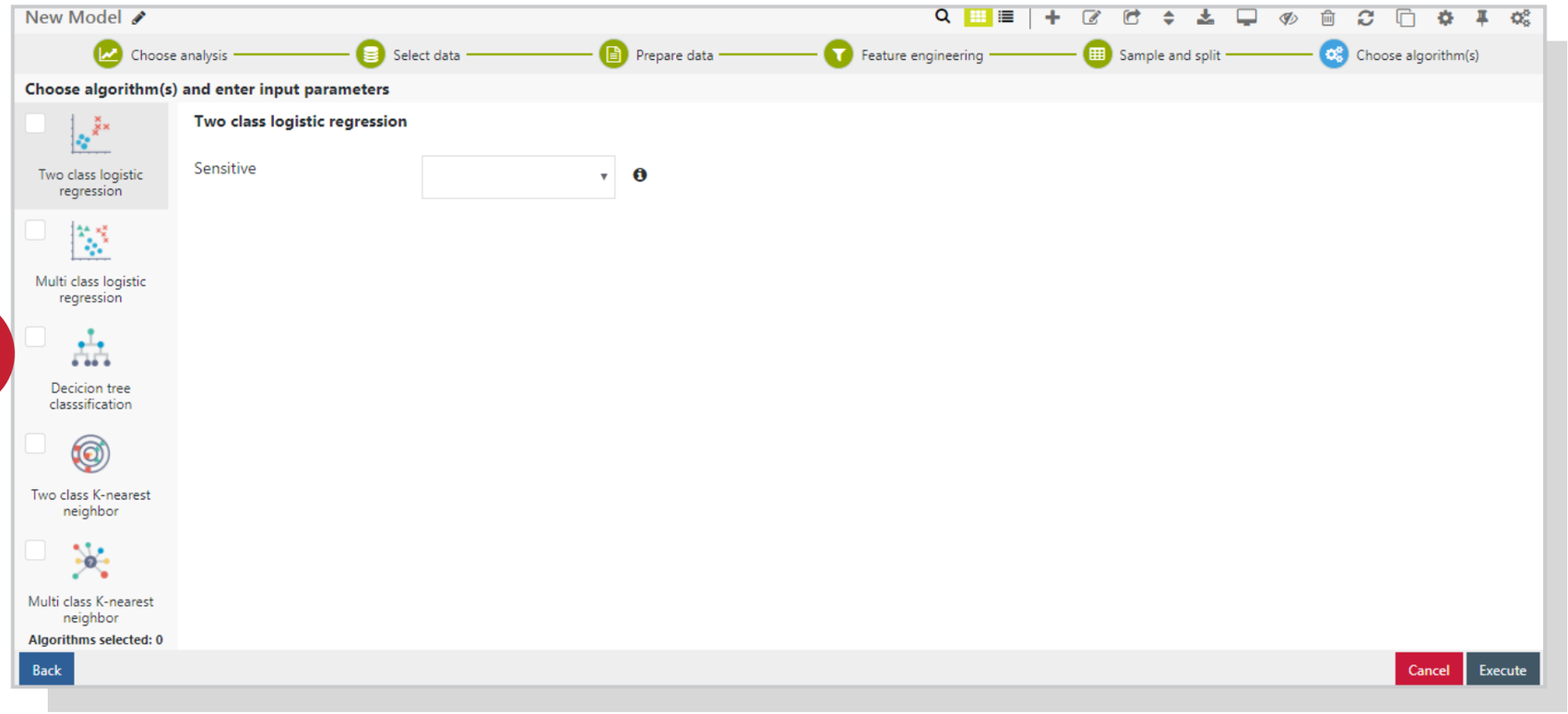
Proportion of variance to: Use the slider to increase or reduce the value as required.

Normalization: Select Z-score to standardize the values.

Quick Start Reference Card - MODEL

Choose algorithms :
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

6

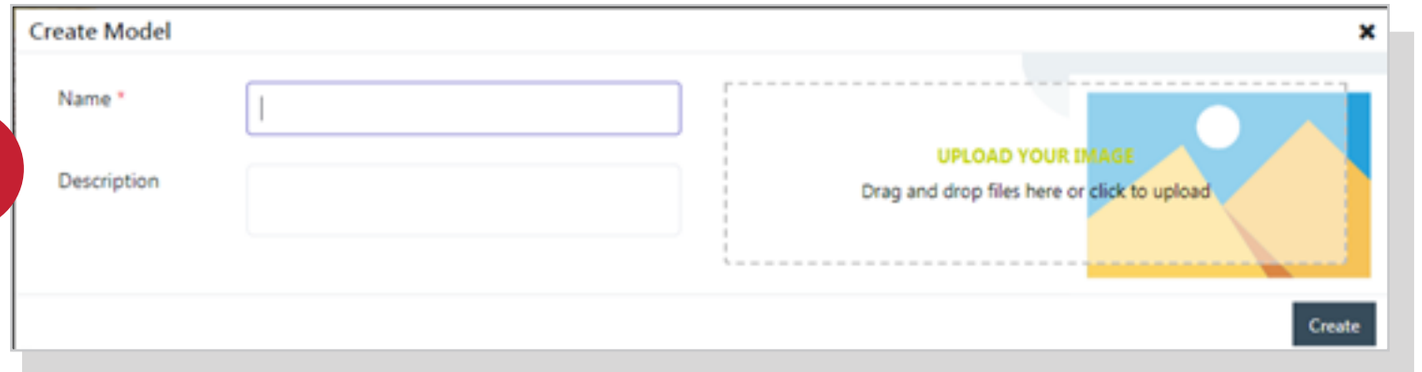


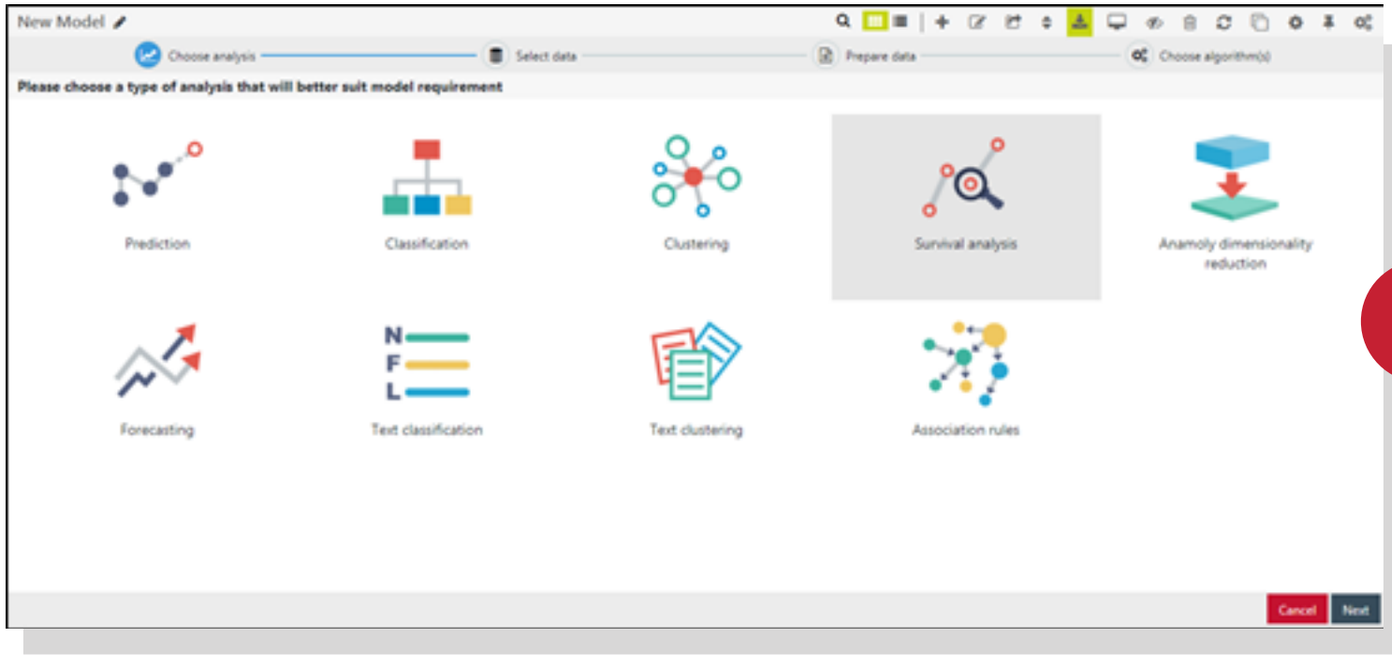
ALGORITHMS	INPUT PATRAMETERS
K-means clustering	Exact K
Mini batch K – means clustering	Exact K, Batch size, Maximum iterations, Number of initialization, Initialization fraction, Initialization
Gaussian mixture model	Minimum iterations, Number of components
Partition around mediods	Exact K, Distance 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

PROCEDURE TO ADD SURVIVAL ANALYSIS MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1





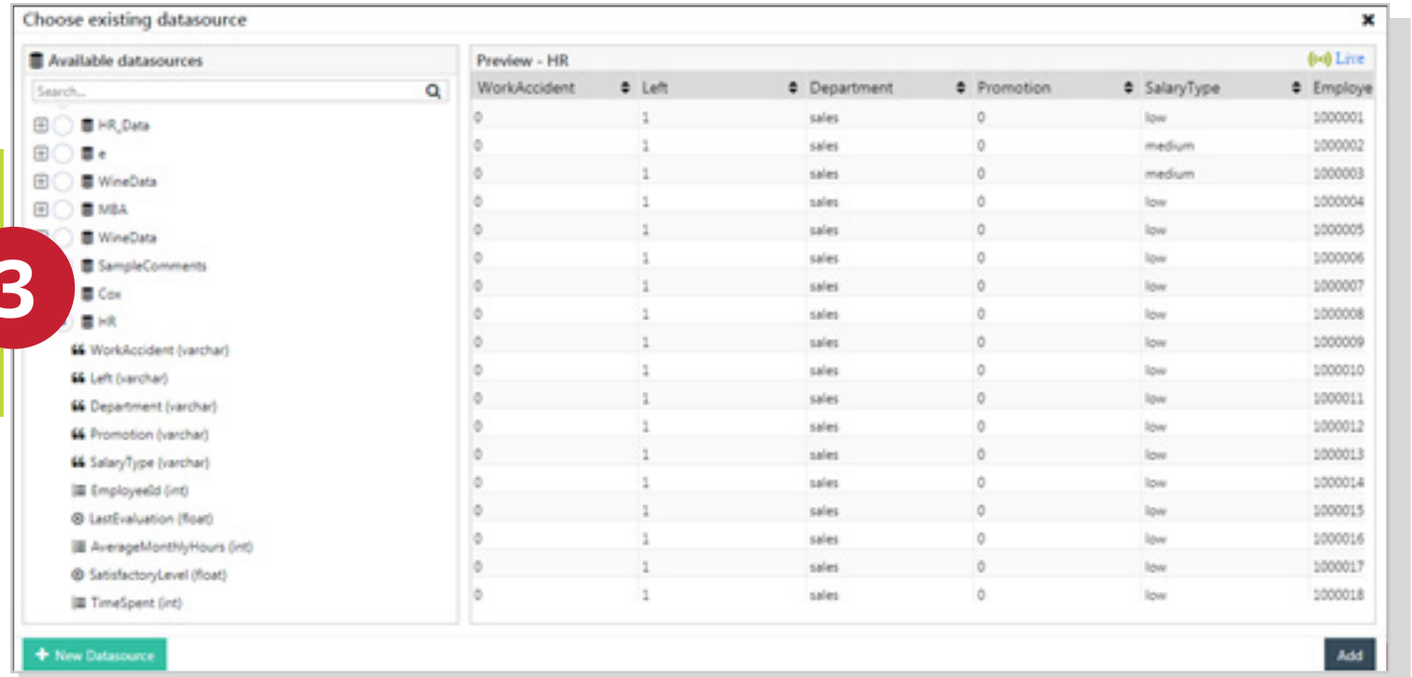
2

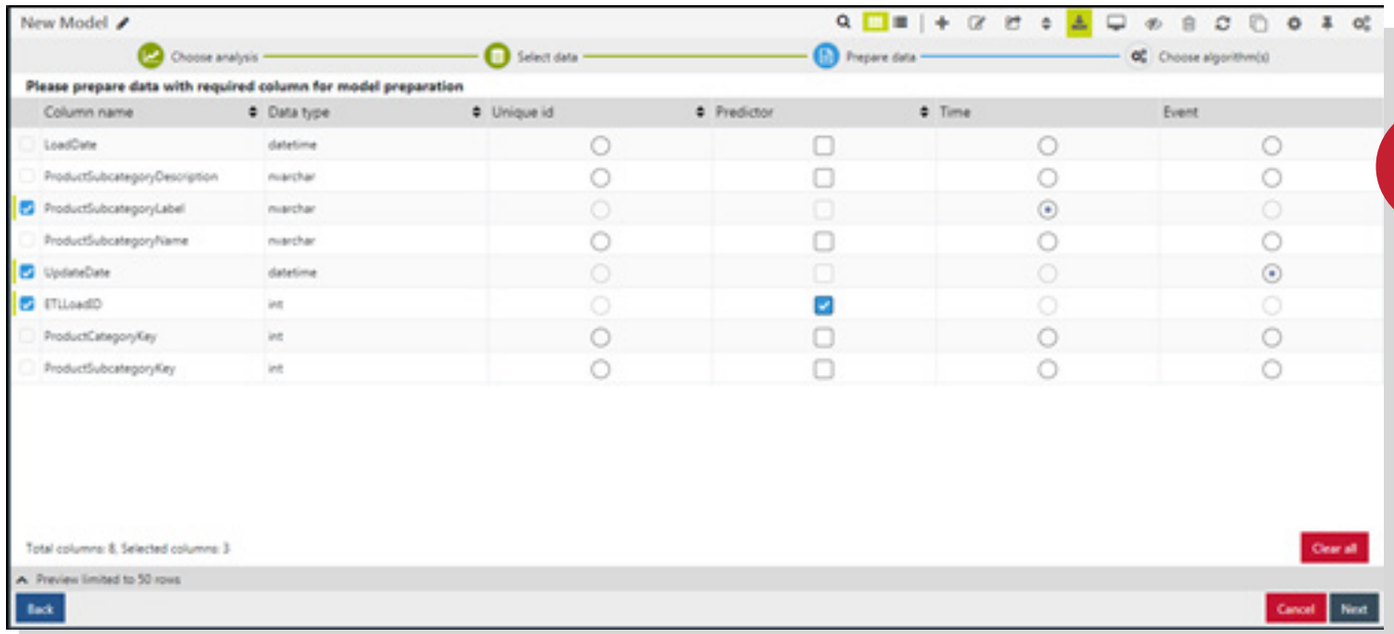
Choose Analysis:
Select Survival analysis from the list of options.

Quick Start Reference Card - MODEL

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

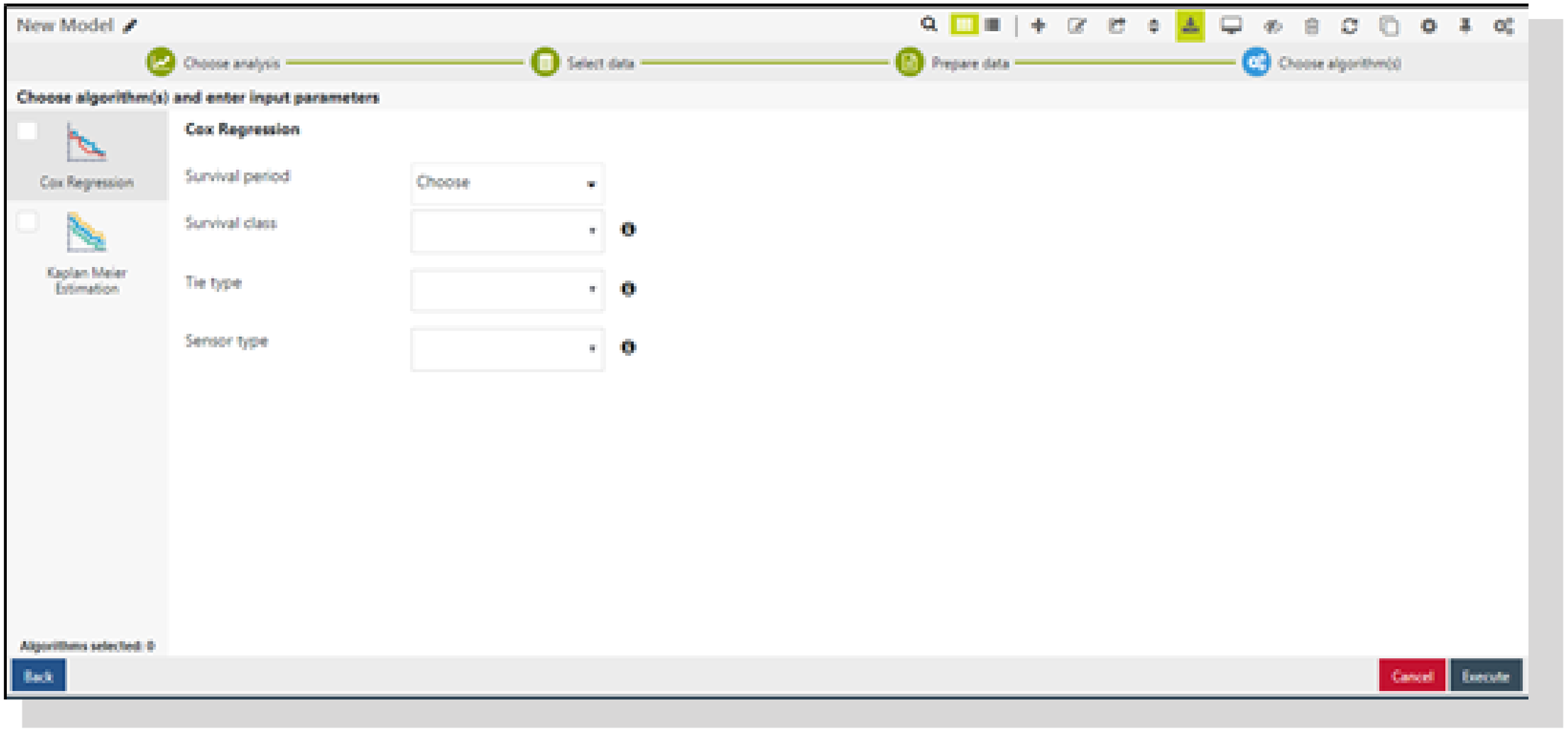
3





4

Prepare Data: Prepare the data with the required columns.
Data type: List of data type from the selected table.
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 parameter are disabled for the particular column.
Time: This column value is mandatory. On selecting Time for a particular variable all the other parameter are disabled for the particular column.
Event: This column is mandatory. On selecting an event for a particular variable all other factors for the variable such as Predictor, Time are disabled.



5

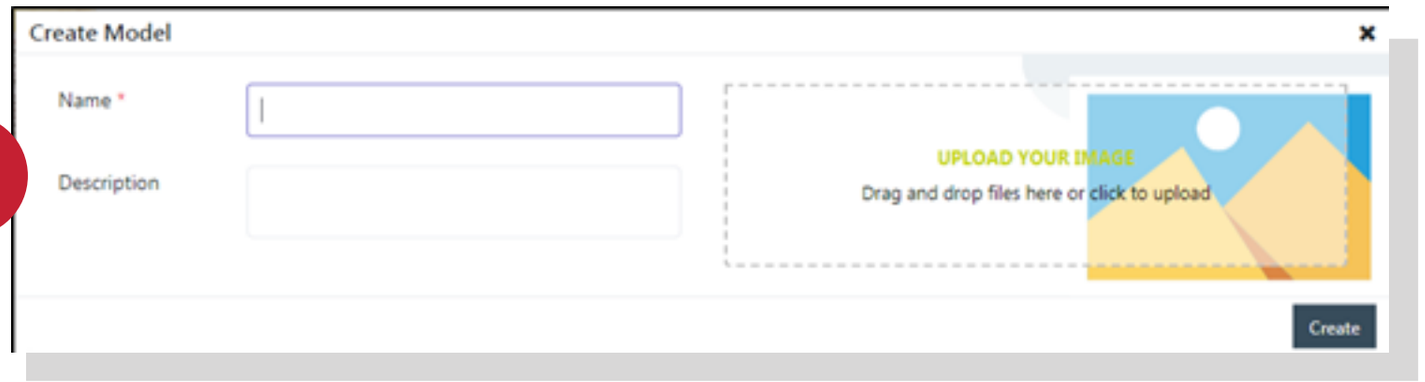
Choose Algorithm(s):
Select the algorithm(s) to analyze data:

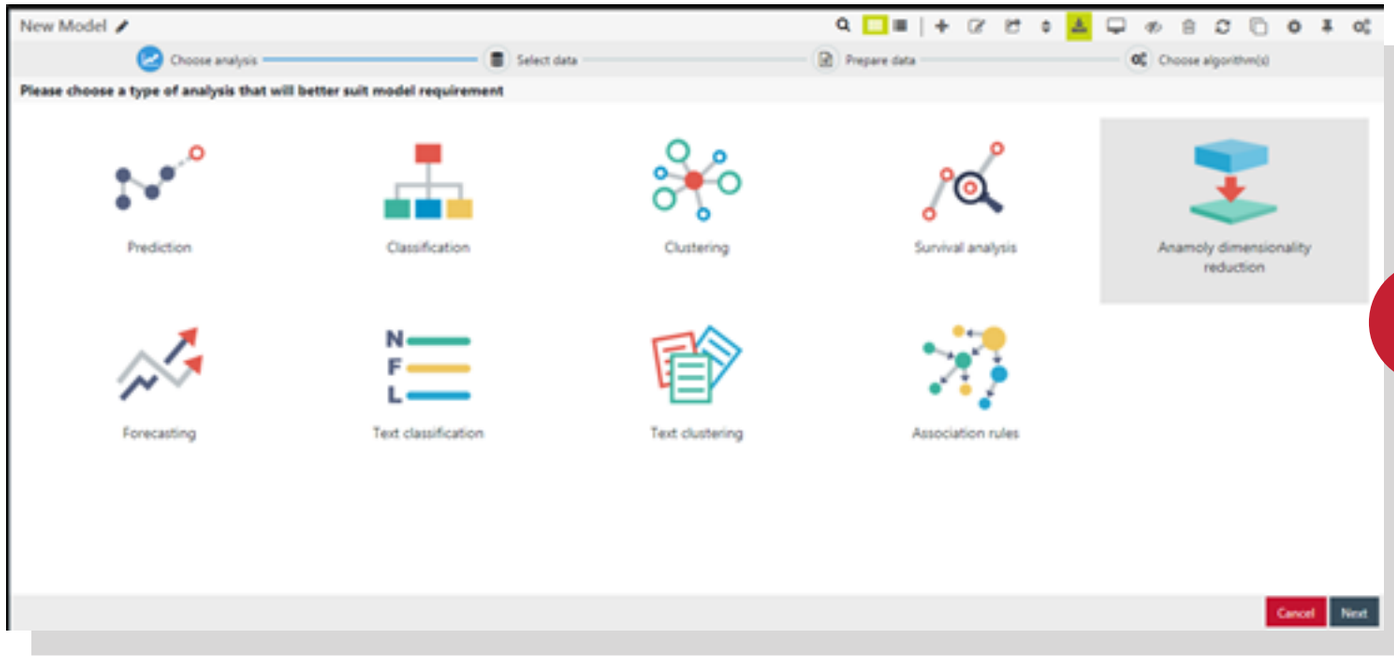
ALGORITHMS	INPUT PATRAMETERS
Cox Regression	Survival period, Survival class, Tie Type, Sensor Type
Kaplar Meier Estimation	Censor Type, Variable Type, Confidence Interval.

PROCEDURE TO ADD ANOMALY / DIMENSIONALITY REDUCTION MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1



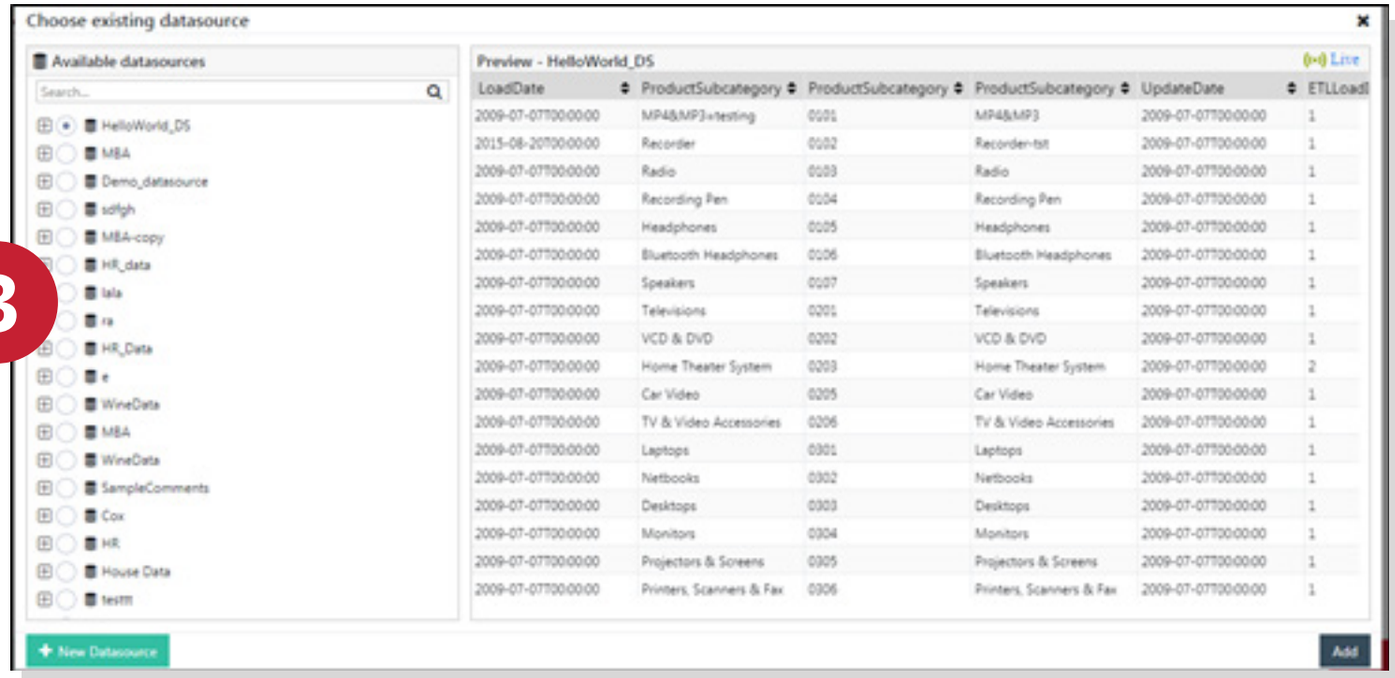


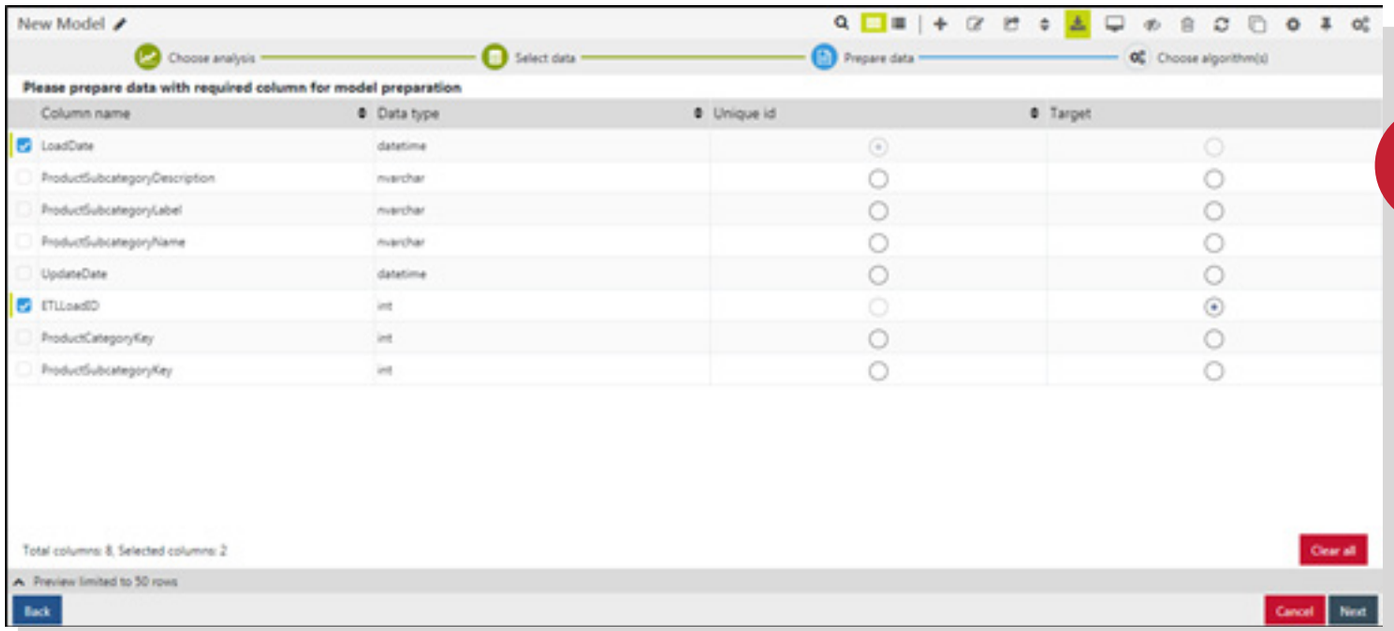
2

Choose Analysis:
Select Anomaly / dimensionality reduction from the list of options.

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

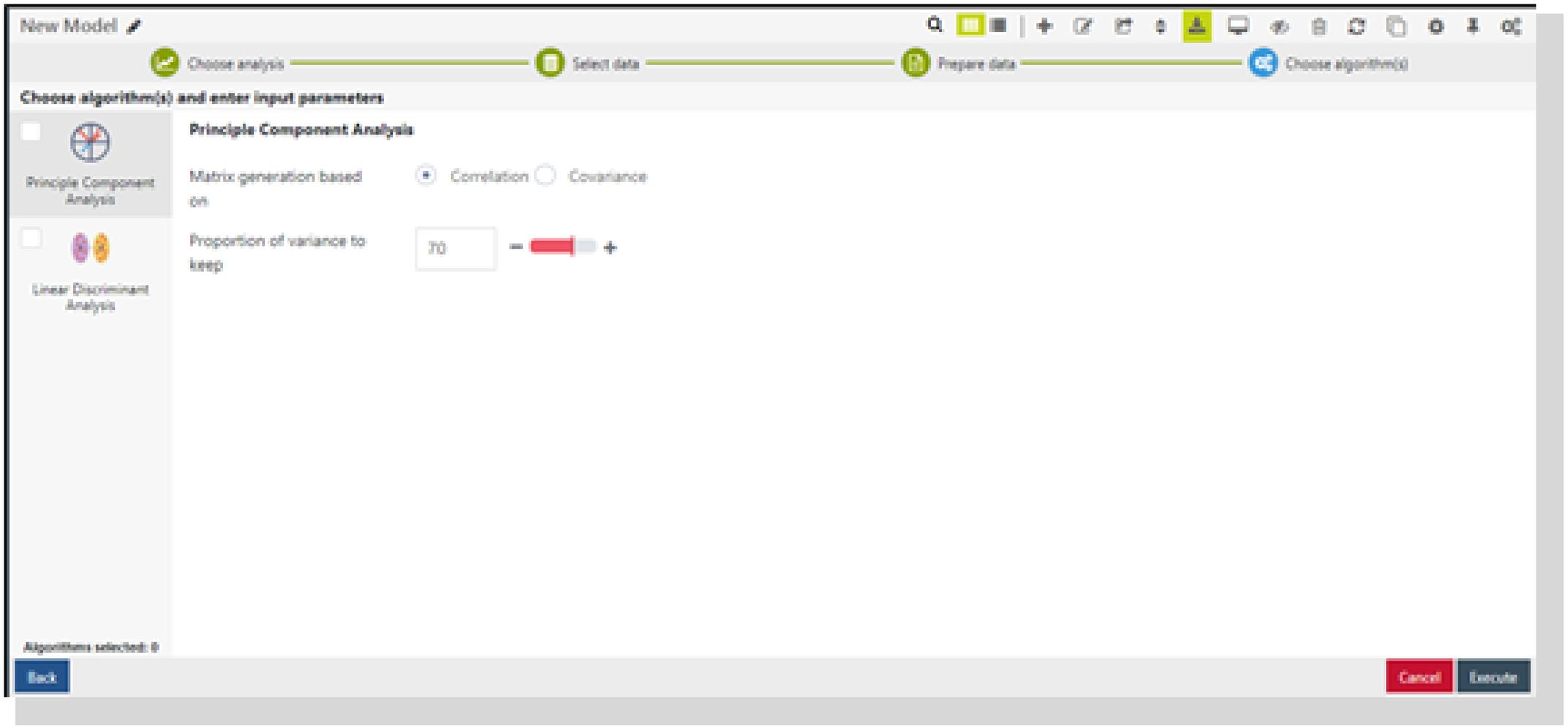
3





4

Prepare Data: Prepare the data with the required columns.
Data type: List of data type from the selected table.
Unique id: Select the unique variable if the datasource has a unique value.
Target: This variable value is mandatory. Select a variable that you will like to predict that is dependent for the analysis.
On selecting either of Unique id or Target the variable will be disabled.



5

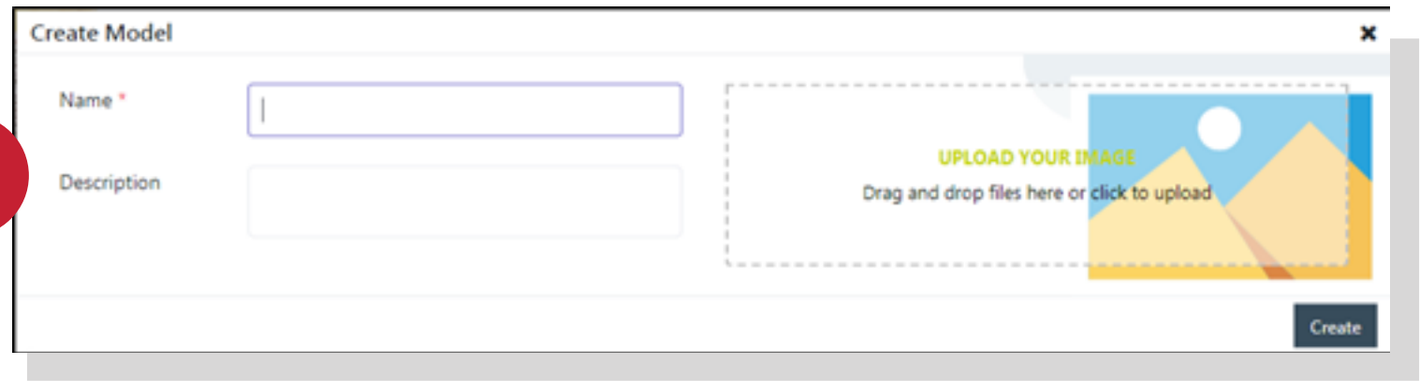
Choose Algorithm(s):
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

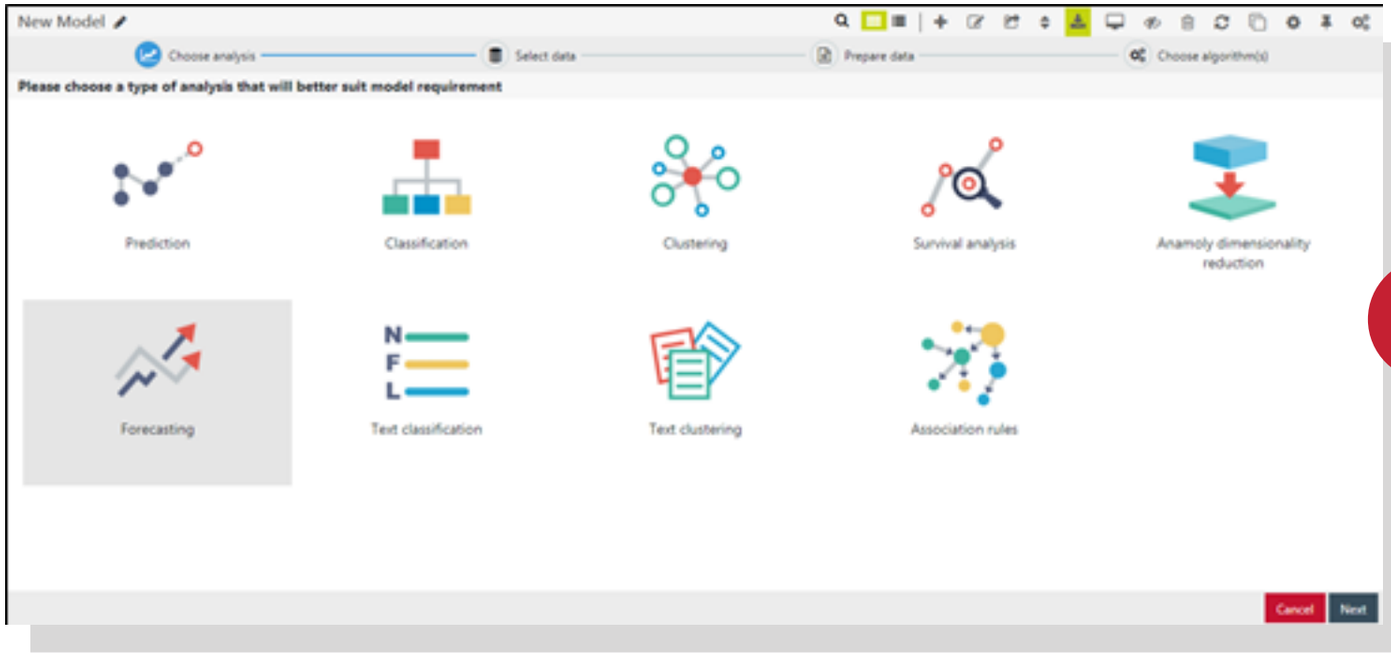
ALGORITHMS	INPUT PATRAMETERS
Principle Component Analysis	Matrix generated based on: Correlation / Covariance Proportion of variance to keep: Increase or decrease the slider
Linear Discriminant Analysis	No input parameter

PROCEDURE TO ADD FORECASTING MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1





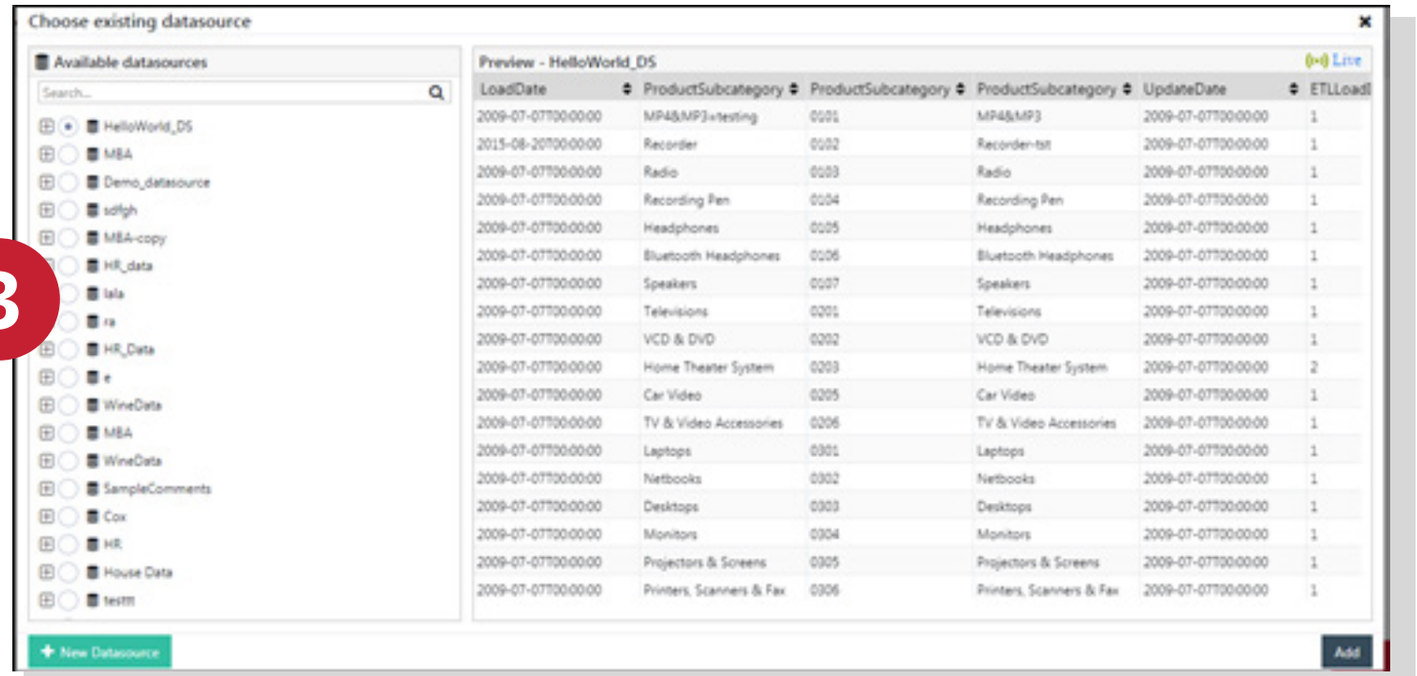
2

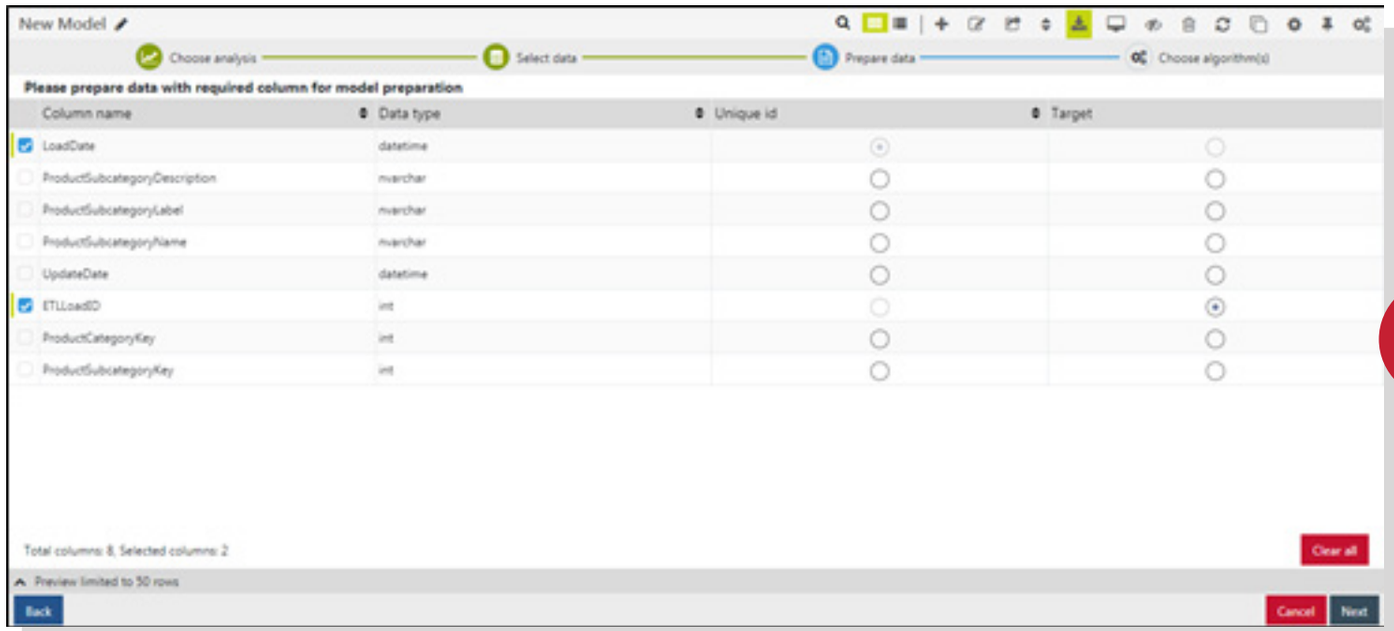
Choose Analysis:
Select Forecasting from the list of options.

Quick Start Reference Card - MODEL

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

3





4

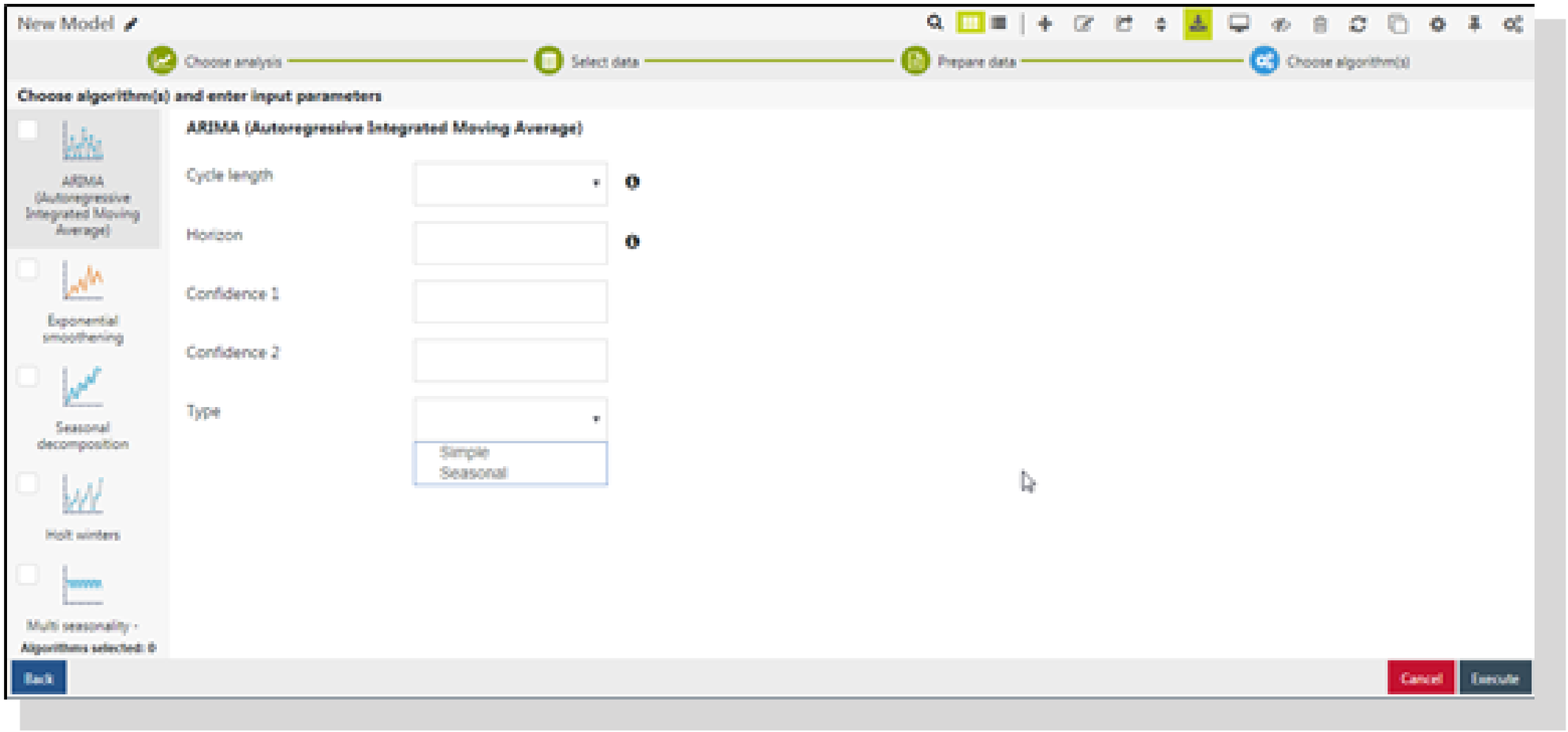
Prepare Data: Prepare the data with the required columns.

Data type: List of data type from the selected table.

Date 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.

Quick Start Reference Card - MODEL



5

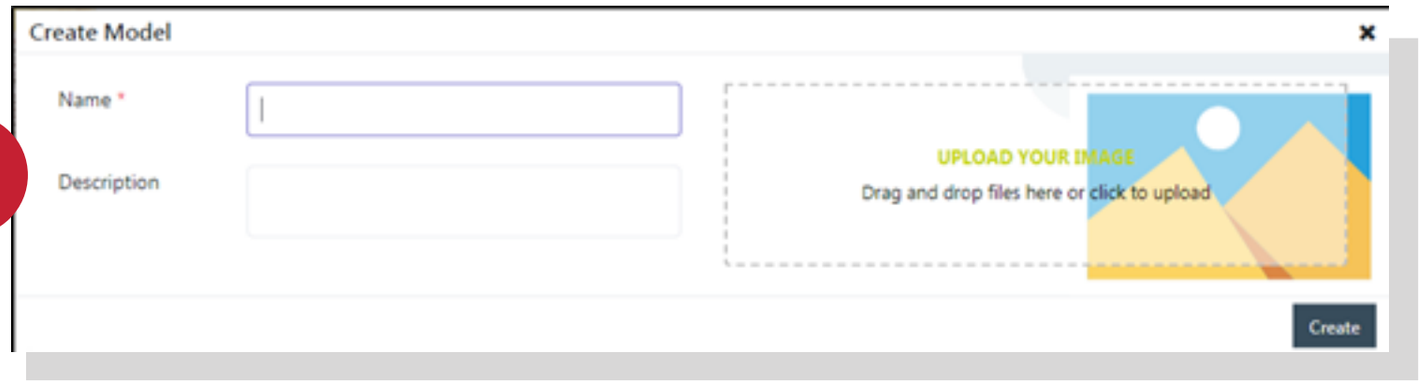
Choose Algorithm(s):
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

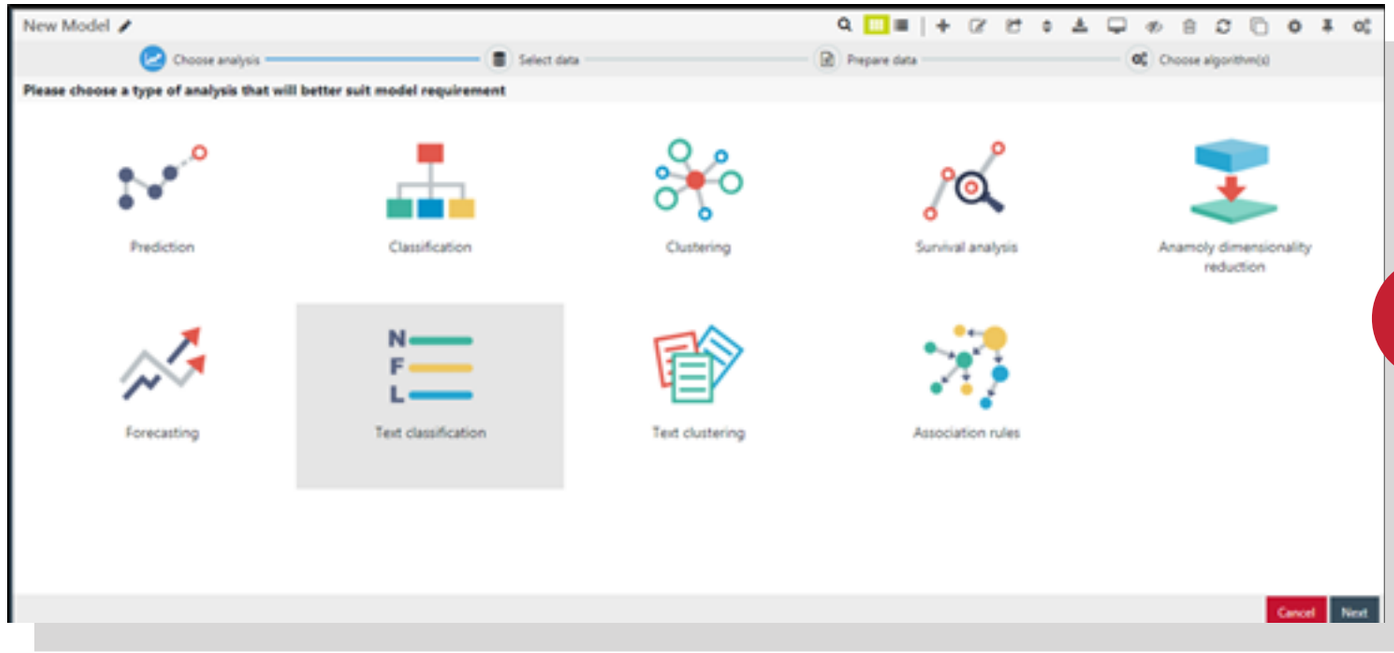
ALGORITHMS	INPUT PATRAMETERS
ARIMA (Auto regressive integrated moving average)	Cycle length, Horizon, Confidence 1, Confidence 2, Type
Exponential smoothening	Cycle length, Horizon, Confidence 1, Confidence 2
Seasonal decomposition	Cycle length, Horizon, Confidence 1, Confidence 2
Holt winters	Cycle length, Horizon, Confidence 1, Confidence 2, Type
Multi seasonality (TBATS)	Cycle length, Horizon, Confidence 1, Confidence 2

PROCEDURE TO ADD TEXT CLASSIFICATION MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1





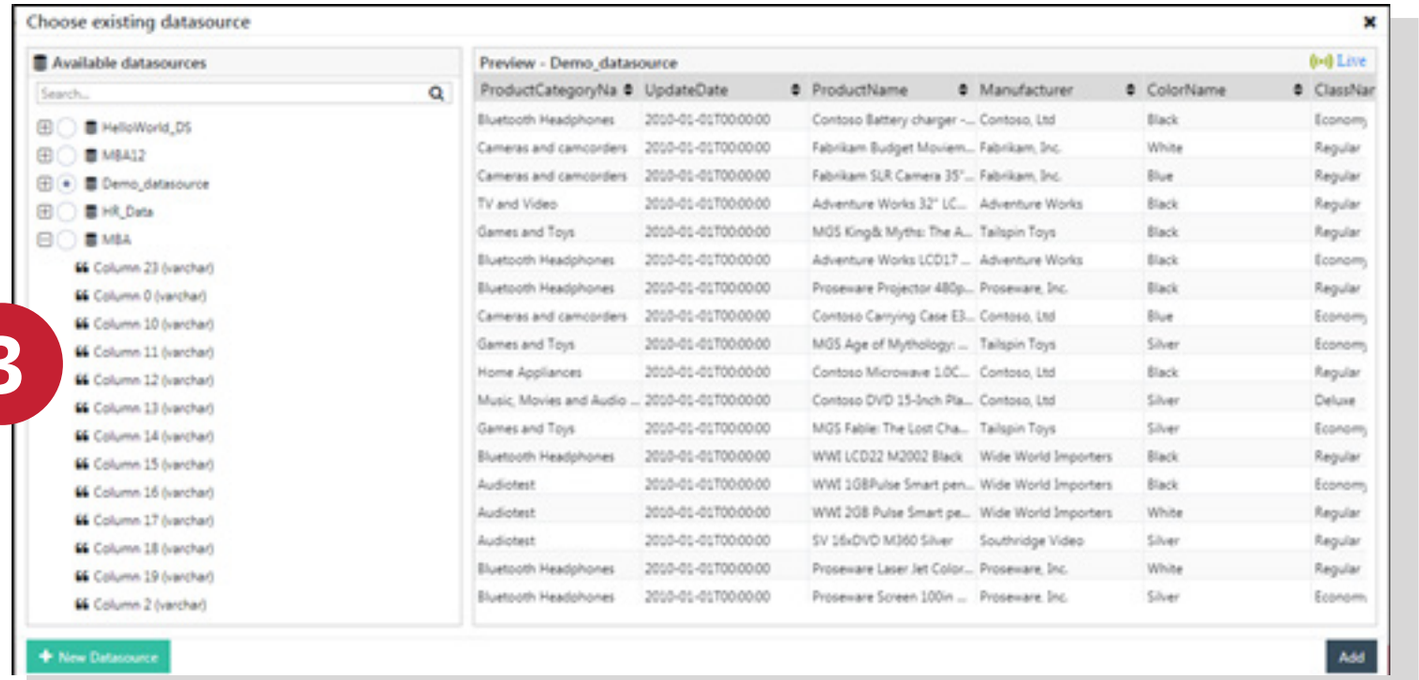
2

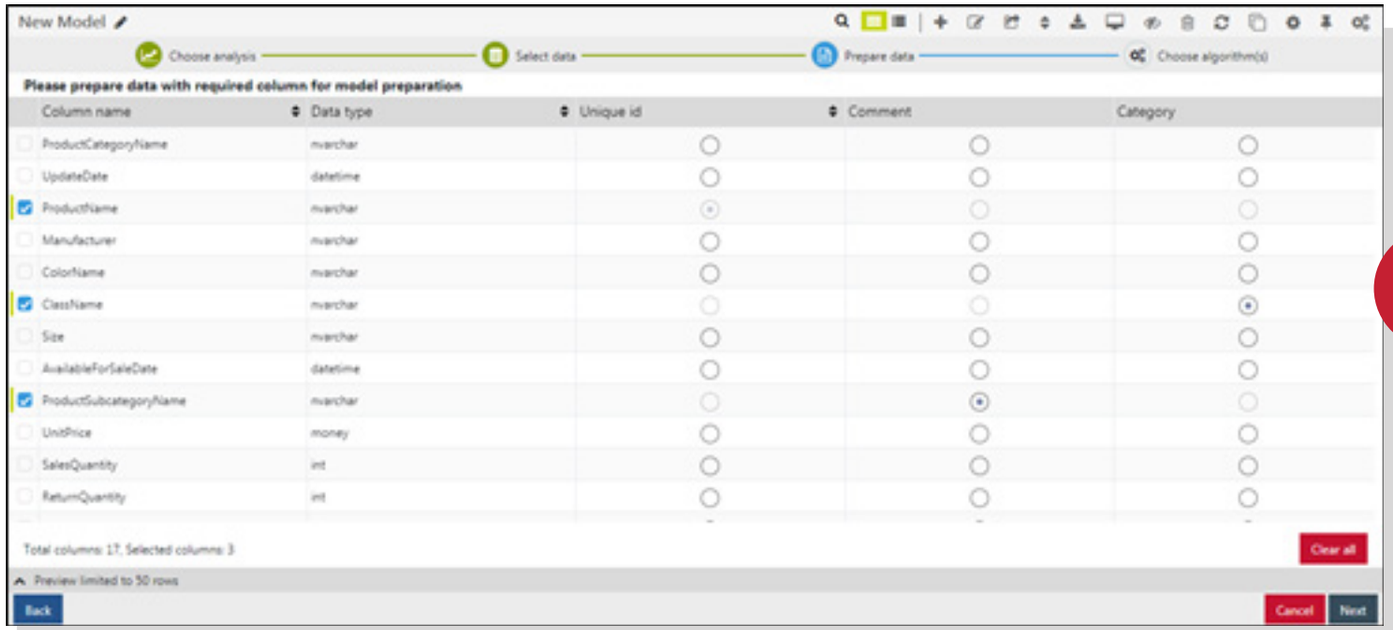
Choose Analysis:
Select Text Classification analysis from the list of options.

Quick Start Reference Card - MODEL

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

3





4

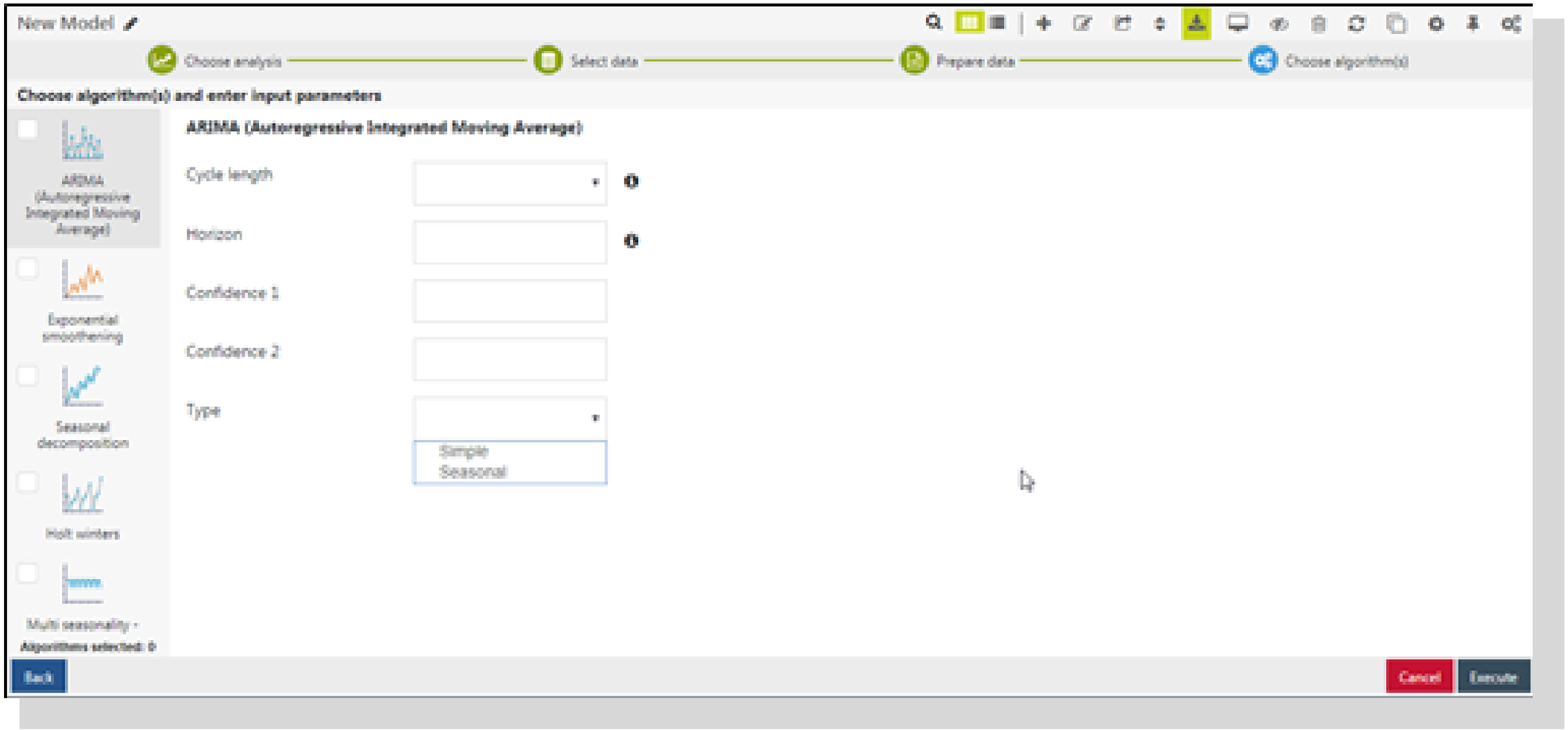
Prepare Data: Prepare the data with the required columns.

Unique id: This column is mandatory. Select the unique label to identify it as unique comment.

Comment: This column is mandatory. Select the column that has text comments. If 'Comment' is selected for a column then 'Unique id' and 'Category' for that column are disabled. User will not be able to define more than one criteria for a column.

Category: This column is mandatory. Select a column to define it as Category.

Quick Start Reference Card - MODEL



5

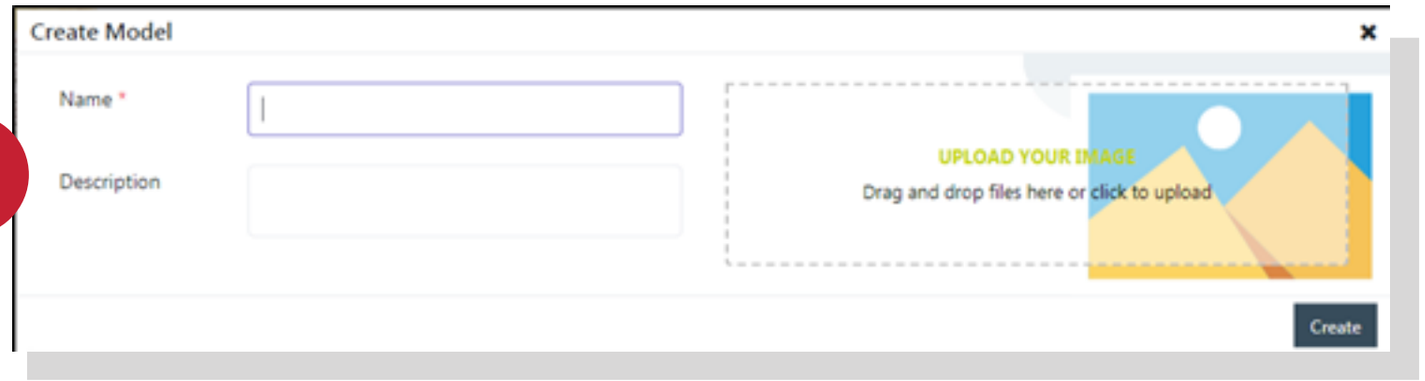
Choose Algorithm(s):
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

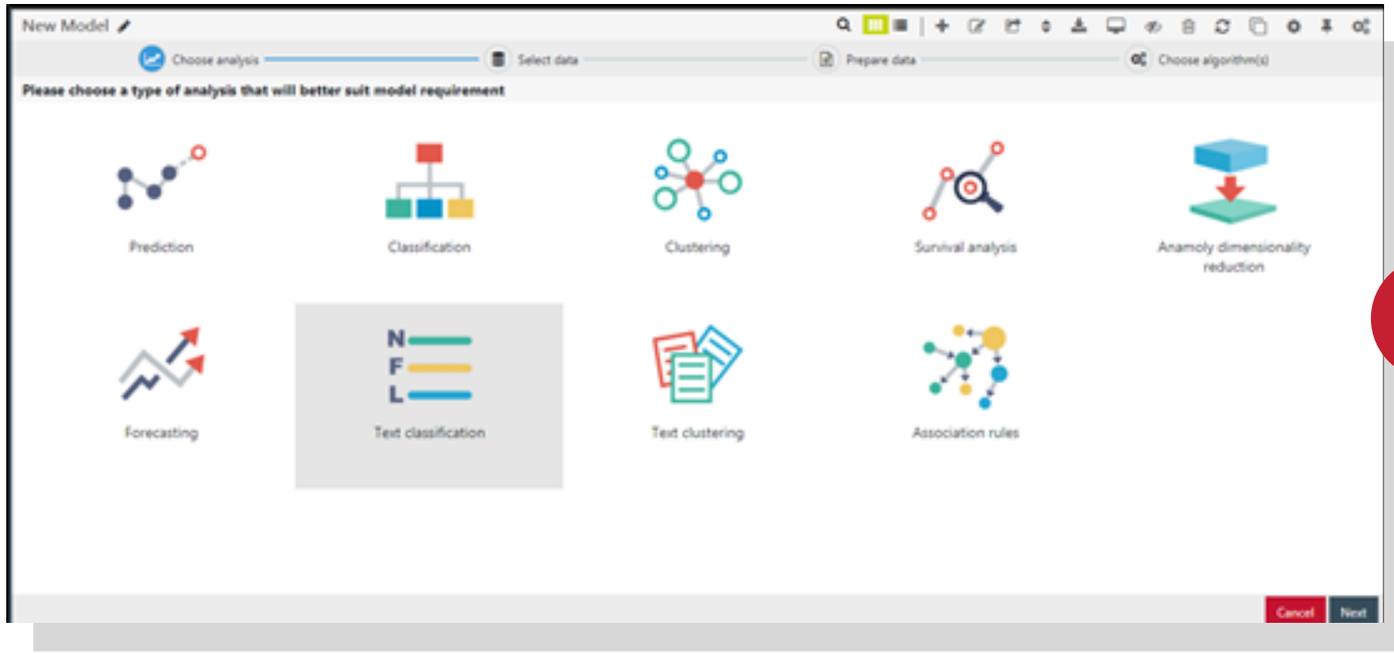
ALGORITHMS	INPUT PATRAMETERS
Two class naïve bayes	Sensitive class
Multi class naïve bayes	Alpha (Laplace Correlation)
Two class K-nearest neighbor	Sensitive class
Multi class support vector machine	Cost function 'C', Gamma value, Kernel

PROCEDURE TO ADD TEXT CLUSTERING MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1





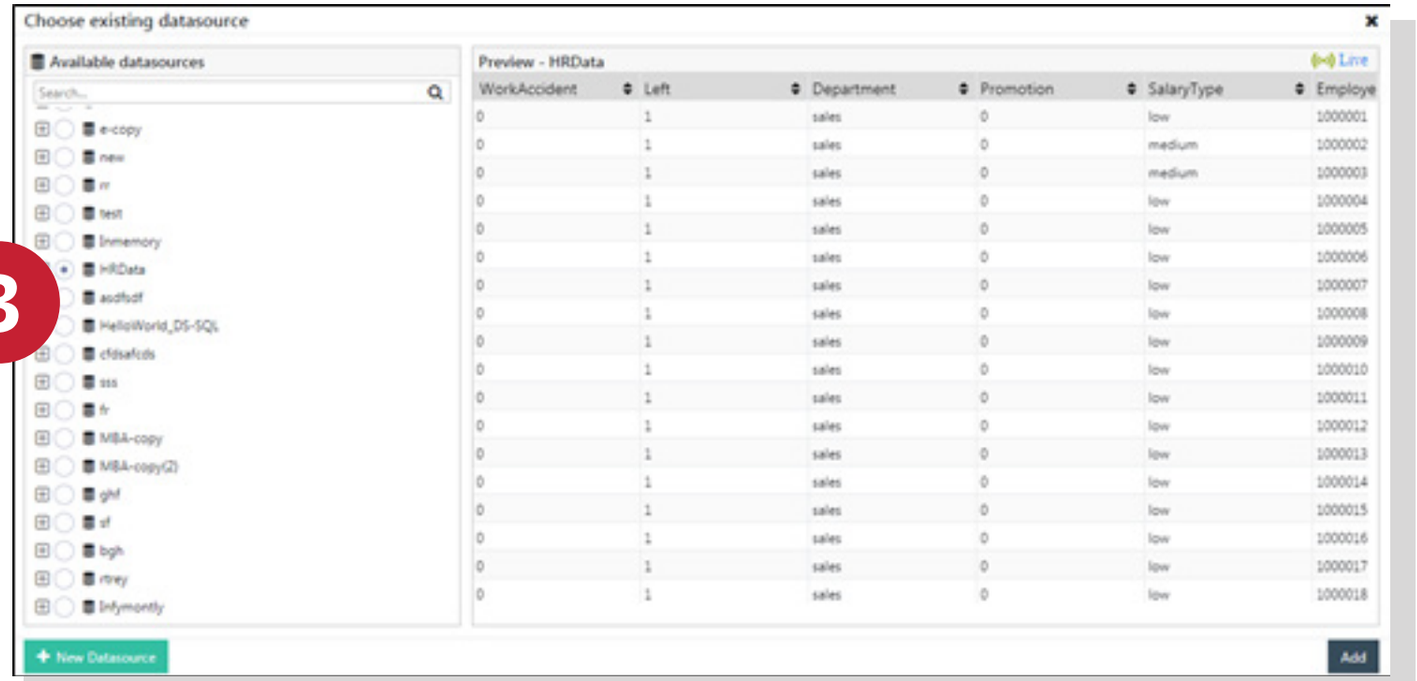
2

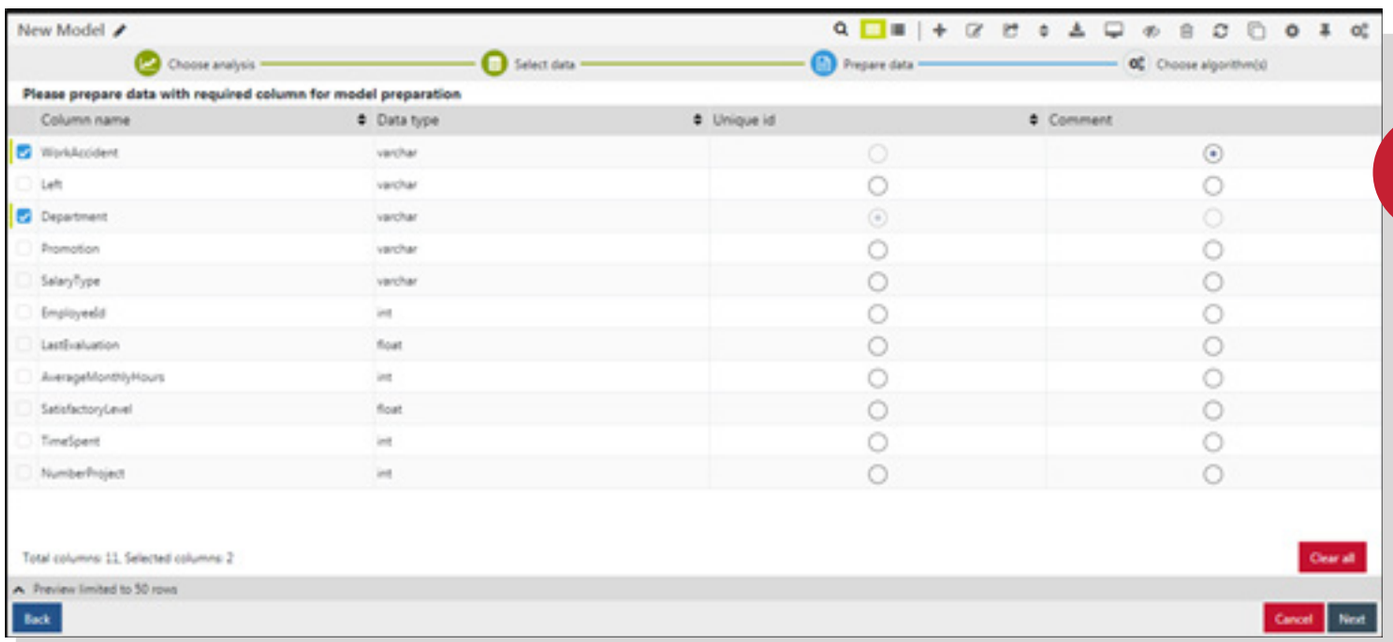
Choose Analysis:
Select Text Clustering analysis from the list of options.

Quick Start Reference Card - MODEL

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

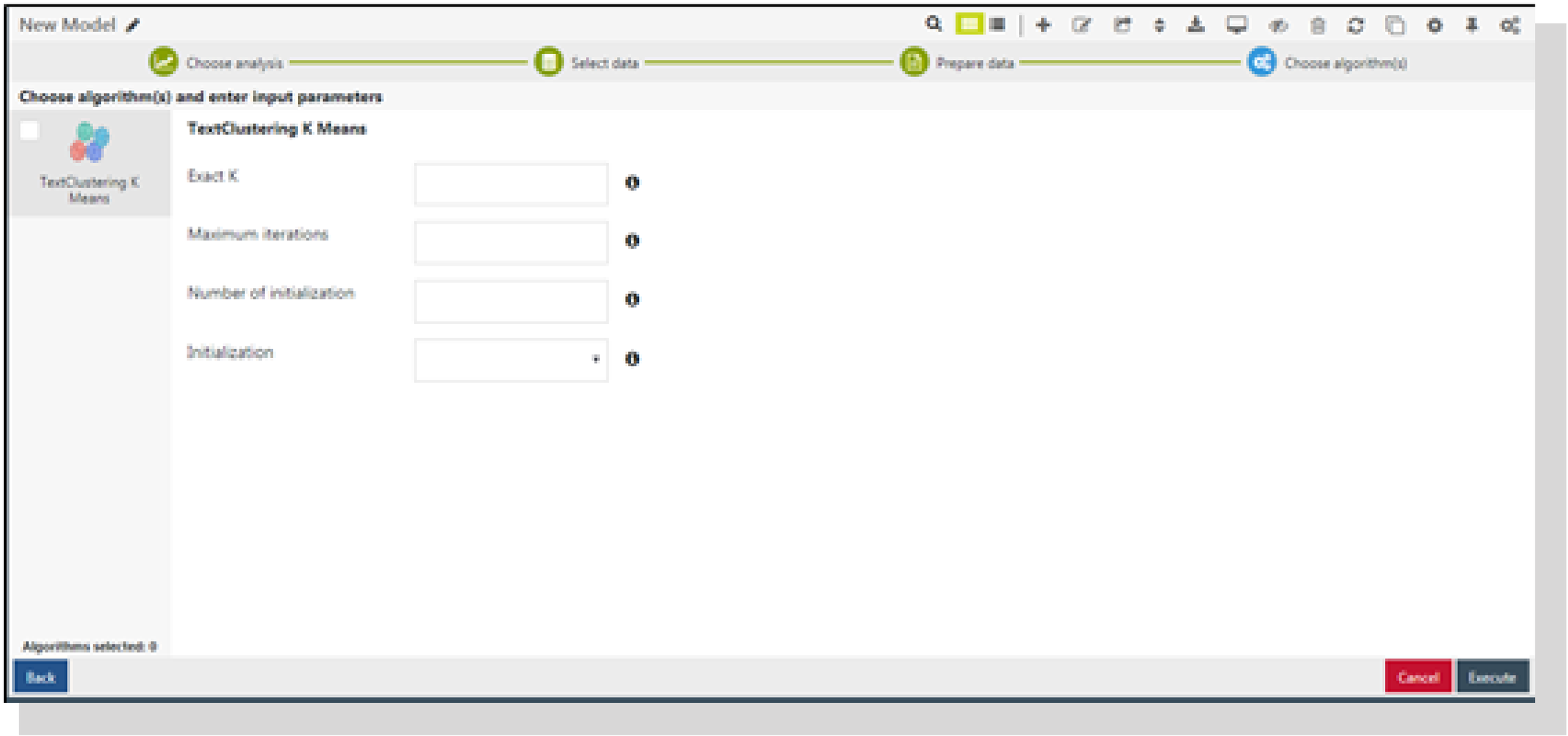
3





4

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



5

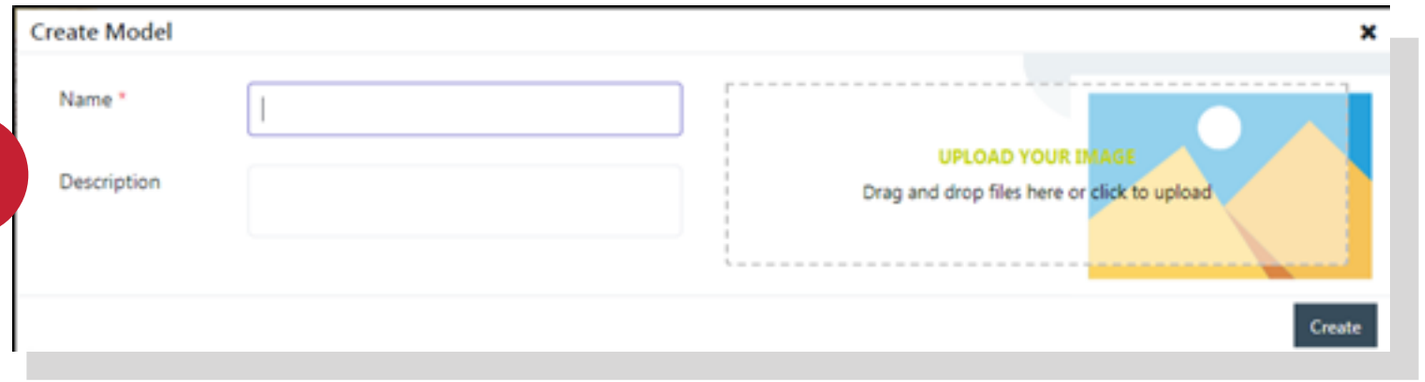
Choose Algorithm(s):
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

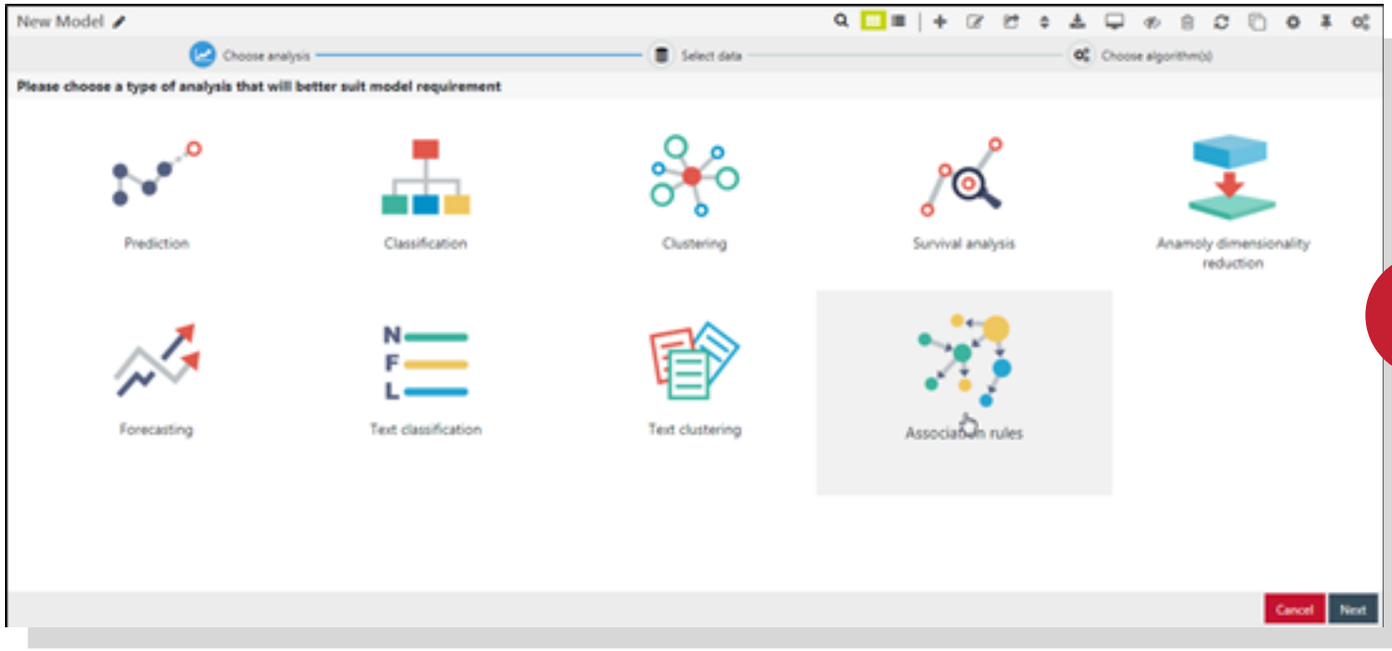
ALGORITHMS	INPUT PATRAMETERS
TextClustering K Means	Exact K, Maximum number of iterations, Number of initialization, Initialization

PROCEDURE TO ADD ASSOCIATION RULES MODEL

Create Model:
Associate a model with a name, description and if required identify it with an image to view as a thumb nail icon.

1





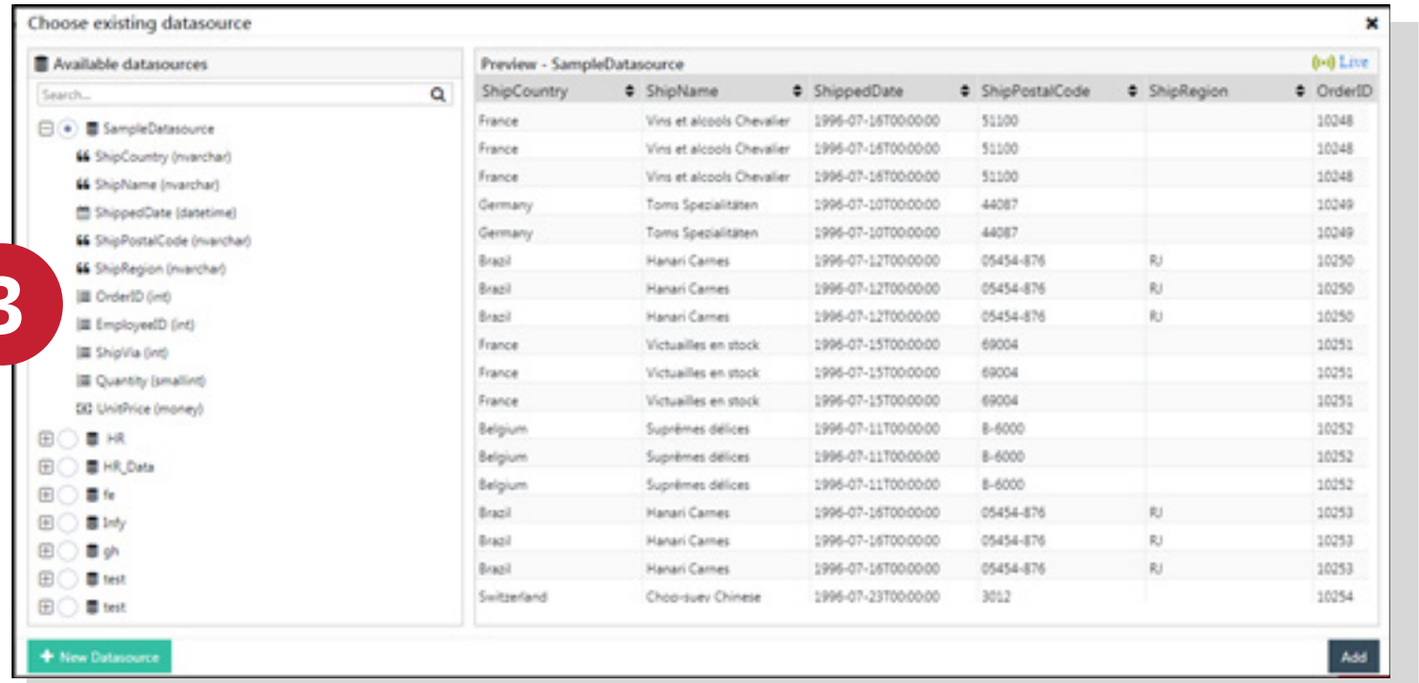
2

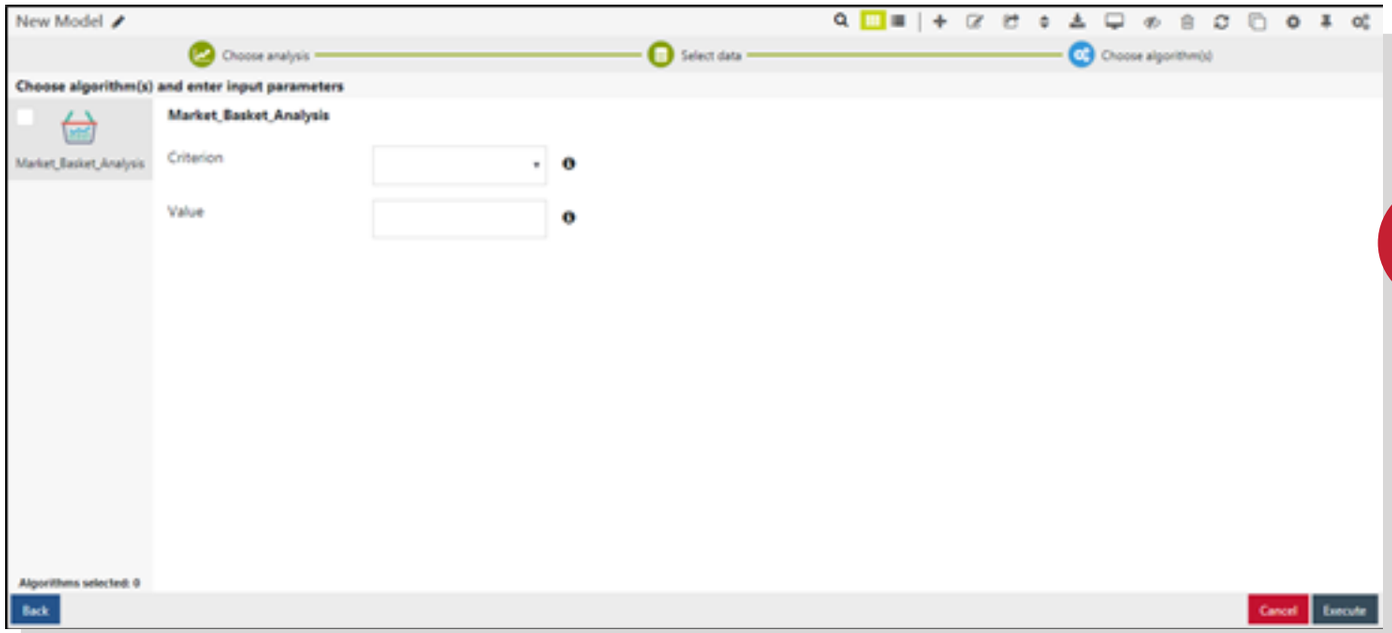
Choose Analysis:
Select Association rules analysis from the list of options.

Quick Start Reference Card - MODEL

Choose existing datasource:
Select the database to add data and the preview will be shown. Option to add a new datasource is also available.

3





4

Choose algorithms:
Choose the type of algorithm(s) from the available algorithms for calculating the values for the model and specify the input parameters.

ALGORITHMS	INPUT PATRAMETERS
Market_Basket_Analysis	Criterion, Value