# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
<th>Model Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>04</td>
<td>Add a Prediction Model</td>
</tr>
<tr>
<td>09</td>
<td>Add a Classification Model</td>
</tr>
<tr>
<td>14</td>
<td>Add a Clustering Model</td>
</tr>
<tr>
<td>18</td>
<td>Add a Survival Analysis</td>
</tr>
<tr>
<td>22</td>
<td>Add an Anomaly / Dimensionality Reduction Model</td>
</tr>
<tr>
<td>26</td>
<td>Add a Forecasting Model</td>
</tr>
<tr>
<td>30</td>
<td>Add a Text Classification Model</td>
</tr>
<tr>
<td>34</td>
<td>Add a Text Clustering Model</td>
</tr>
<tr>
<td>38</td>
<td>Add an Association Rules Model</td>
</tr>
</tbody>
</table>
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.

---

**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.

**PREREQUISITES**

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

**RELATIONAL DATABASES**

- Oracle
- MySQL
- PostgreSQL
- Teradata
- SQL Server
- Access
- DB2
- SQLite
- MariaDB

**BIG DATA**

- Hortonworks
- Impala
- MapR
- Apache Drill
- Solr
- Elasticsearch
- Hadoop

**API**

- SharePoint
- REST API
- OData

**FILE**

- Excel
- JSON
- XML

**SOCIAL MEDIA**

- facebook
- twitter

**CUBES**

- SQL Server

**Model QUICK START GUIDE**

2. A business challenge must exist.
Add a Prediction Model

01 SELECT A BUSINESS CHALLENGE:
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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.
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**
  - **Treat as categorical**: Select to include the variable in the data preparation process.
  - **Unique id**: Select if the variable is to be treated as categorical.
  - **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.
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.

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.
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.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple linear regression</td>
<td>No input parameter</td>
</tr>
<tr>
<td>Decision Tree regression</td>
<td>Minimum split, Minimum bucket, Maximum Depth</td>
</tr>
<tr>
<td>Bayesian linear regression</td>
<td>No input parameter</td>
</tr>
<tr>
<td>Neural network regression</td>
<td>Sensitive value</td>
</tr>
<tr>
<td>Random forest regression</td>
<td>Maximum depth, Number of trees</td>
</tr>
</tbody>
</table>
Add a Classification Model

01 SELECT A BUSINESS CHALLENGE:

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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.
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 categorical
- 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:
   - **Tree Based**: Type the **Number of Trees** and **Depth of Trees**.
   - **Principle Component Analysis**: Type a value or use the slider to increase or reduce the **Proportion of variance to** value as required.
   - **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.
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 and Input Parameters

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two class logistic regression</td>
<td>Sensitive value</td>
</tr>
<tr>
<td>Multi class logistic regression</td>
<td>No input parameter</td>
</tr>
<tr>
<td>Two class decision tree classification</td>
<td>Sensitive value</td>
</tr>
<tr>
<td>Two class K-nearest neighbor</td>
<td>Sensitive value</td>
</tr>
<tr>
<td>Multi class K-nearest neighbor</td>
<td>No input parameter</td>
</tr>
<tr>
<td>Two class random forest</td>
<td>Maximum depth, number of trees, sensitive, minimum split, minimum bucket, maximum leaf node</td>
</tr>
<tr>
<td>Two class neural network</td>
<td>Sensitive value</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two class support vector machine</td>
<td>Sensitive, Cost function C, Gamma value, Kernel</td>
</tr>
<tr>
<td>Two class gradient boosting machine</td>
<td>Minimum split, Maximum depth, Sensitive, Bagging size, Estimators, Feature strategy, Learning rate, Loss function</td>
</tr>
<tr>
<td>Ada boosting</td>
<td>Minimum split, Maximum depth, Sensitive, Criterion, Estimators, Feature strategy, Learning rate, Type</td>
</tr>
<tr>
<td>Multi class random forest</td>
<td>Minimum split, Minimum bucket, Maximum depth, Number of trees, Maximum leaf node</td>
</tr>
<tr>
<td>Multi class Ordinal regression</td>
<td>No input parameter</td>
</tr>
<tr>
<td>Two class naïve bayes classification</td>
<td>Sensitive class</td>
</tr>
<tr>
<td>Multi class naïve bayes classification</td>
<td>No input parameter</td>
</tr>
</tbody>
</table>
Add a Clustering Model

01 SELECT A BUSINESS CHALLENGE:
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

02 Click Add (+) > Model.
Type the **Name** and **Description** of the model and then click **Create**.

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 categorical**: 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.
1. Choose the type of algorithm(s) for calculating the values for the model and specify the input parameters.
2. Click \textit{Execute}.

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

### Algoritms

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means clustering</td>
<td>Exact K</td>
</tr>
<tr>
<td>Mini batch K – means clustering</td>
<td>Exact K, Batch size, Maximum iterations, Number of initialization, Initialization fraction, Initialization</td>
</tr>
<tr>
<td>Gaussian mixture model</td>
<td>Minimum iterations, Number of components</td>
</tr>
<tr>
<td>Partition around mediods</td>
<td>Exact K, Distance metric</td>
</tr>
<tr>
<td>Ward hierarchical clustering</td>
<td>Exact K, Distance metric, Method</td>
</tr>
<tr>
<td>DB Scan</td>
<td></td>
</tr>
<tr>
<td>Adaptive K-means clustering</td>
<td>Maximum iterations, Distance metric, Minimum K, Maximum K, Threshold value, N Start</td>
</tr>
</tbody>
</table>

### CHOOSE THE FEATURE ENGINEERING OPTIONS:

1. In the \textit{Reduction method} list, choose a reduction method.
2. If you choose \textit{PCA based dimensionality reduction}:
   2a. Type the Number of Trees and Depth of Trees.
   2b. For Matrix generation based on, select \textit{Correlation} or \textit{Covariance}.
   For \textit{Proportion of variance to keep}, enter a number between 0 and 100.
3. Click \textit{Next}. 

© Copyright 2018 | Ducen IT | All Rights Reserved
Add a Survival Analysis Model

**QUICK START GUIDE**

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.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox Regression</td>
<td>Survival period, Survival class, Tie Type, Sensor Type</td>
</tr>
<tr>
<td>Kaplan Meier Estimation</td>
<td>Censor Type, Variable Type, Confidence Interval.</td>
</tr>
<tr>
<td></td>
<td>The UniqueID and Predictor column information is not applicable for this algorithm.</td>
</tr>
</tbody>
</table>
Add an Anomaly / Dimensionality Reduction Model

01 SELECT A BUSINESS CHALLENGE:
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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 and Input Parameters

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle Component Analysis</td>
<td>Matrix generated based on: Correlation / Covariance</td>
</tr>
<tr>
<td>Linear Discriminant Analysis</td>
<td>Proportion of variance to keep: Increase or decrease the slider</td>
</tr>
<tr>
<td>No input parameter</td>
<td></td>
</tr>
</tbody>
</table>
Add a Forecasting Model

01 SELECT A BUSINESS CHALLENGE:

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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 a 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, the 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.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA (Auto Regressive Integrated Moving Average)</td>
<td>Cycle length, Horizon, Confidence 1, Confidence 2, Type</td>
</tr>
<tr>
<td>Exponential smoothening</td>
<td>Cycle length, Horizon, Confidence 1, Confidence 2</td>
</tr>
<tr>
<td>Seasonal decomposition</td>
<td>Cycle length, Horizon, Confidence 1, Confidence 2</td>
</tr>
<tr>
<td>Holt winters</td>
<td>Cycle length, Horizon, Confidence 1, Confidence 2, Type</td>
</tr>
<tr>
<td>Multi seasonality (TBATS)</td>
<td>Cycle length, Horizon, Confidence 1, Confidence 2</td>
</tr>
</tbody>
</table>
Add a Text Classification Model

01 SELECT A BUSINESS CHALLENGE:
1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two class naïve bayes</td>
<td>Sensitive class</td>
</tr>
<tr>
<td>Two class K-nearest neighbor</td>
<td>Sensitive class</td>
</tr>
</tbody>
</table>
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.
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.
Add an Association Rules Model

1. Login to Analance using your Analance URL.
2. Click Advanced Analytics.

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.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Input Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market_Basket_Analysis</td>
<td>Criterion, Value</td>
</tr>
</tbody>
</table>