Association Discovery with the BigML Dashboard

The BigML Team

Version 1.1

MACHINE LEARNING MADE BEAUTIFULLY SIMPLE

Last updated March 27, 2024

About this Document

This document provides a comprehensive description of how to solve association discovery tasks with the BigML Dashboard. Learn how to use the BigML Dashboard to configure, visualize, and interpret this unsupervised model.

This document assumes that you are familiar with:

- Sources with the BigML Dashboard. The BigML Team. June 2016. [5]
- Datasets with the BigML Dashboard. The BigML Team. June 2016. [4]

To learn how to use the BigML Dashboard to build supervised predictive models read:

- Classification and Regression with the BigML Dashboard. The BigML Team. June 2016. [2]
- Time Series with the BigML Dashboard. The BigML Team. July 2017. [6]

To learn how to use the BigML Dashboard to build other unsupervised models read:

- Cluster Analysis with the BigML Dashboard. The BigML Team. June 2016. [3]
- Anomaly Detection with the BigML Dashboard. The BigML Team. June 2016. [1]
- Topic Modeling with the BigML Dashboard. The BigML Team. November 2016. [7]

Contents

CHAPTER 1

Introduction

There are problems that require to find meaningful relationships among two or more values in large datasets across thousands of values, e.g., discovering which products are bought together by customers (i.e., market basket analysis), finding interesting web usage patterns, or detecting software intrusion. These problems can be solved using Association Discovery, a well-known unsupervised learning technique to find relevant associations among values in high-dimensional datasets.

The BigML associations algorithm was acquired from Professor Geoff Webb (Monash University), a globally acknowledged expert, who spent ten years developing the association discovery in Magnum Opus. Read more about BigML algorithm in Chapter 2.

Association Discovery (also called Association Mining) complements other Machine Learning techniques in two main ways as it:

- Avoids the problems associated with model selection. Most Machine Learning techniques produce a single global model of the data. A problem with such a strategy is that there will often be many such models, all of which describe the available data equally well. A typical model chooses between these models arbitrarily, without necessarily notifying the user that these alternatives exist. However, while the system may have no reason for preferring one model over another, the user may, e.g., two medical tests may be almost equally predictive in a given application. If so, the user is likely to prefer the model that uses the test that is cheaper or less invasive.
- A single model that is globally optimal may be locally suboptimal in specific regions of the problem space. By seeking local models, association mining can find models that are optimal in any given region. If there is no need for a global model, locally optimized models may be more effective.

This chapter provides a comprehensive description of the BigML associations, including how they can be created with 1-click (Chapter 3), all the configuration options (Chapter 4), and the twofold visualization provided by BigML, a network chart and a table (Chapter 5). BigML provides certain measures that rate each association; those are explained in Section 2.1. There is also a section devoted to how to structure your data (Section 2.2), which is very useful to get the best performance of your association's model. You can also export your associations into a CSV file Section 9.1), move your associations to another project (Chapter 12), or delete them permanently (Chapter 14).

In BigML, the sixth tab on the main menu of your Dashboard allows you to list all your available associations. The association list view shows (Figure 1.1), for each association, the **dataset** it was created from, the association's **Name**, the **K** (number of rules found), **Age** (time elapsed since it was created), and **Size**. The SEARCH menu option in the top right corner allows you to **search** your associations by name.

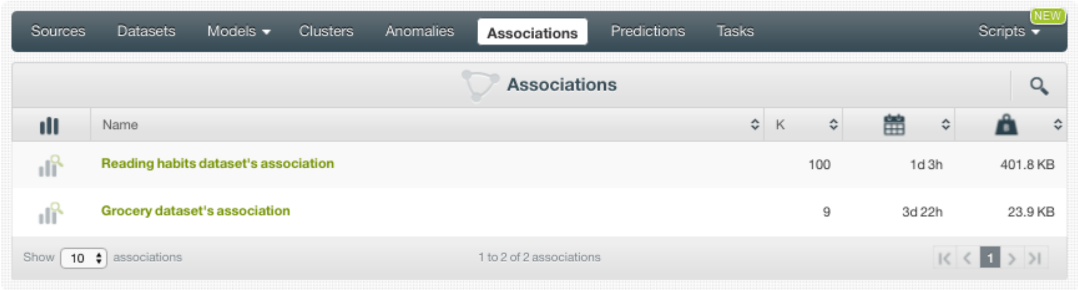


Figure 1.1: Associations list view

When you first create an account with BigML, or every time that you start a new project, your list view for associations will be empty. (See Figure 1.2.)

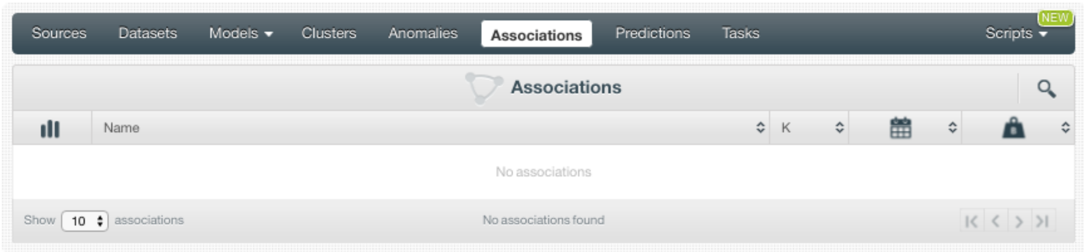


Figure 1.2: Empty Dashboard association view

Finally, in Figure 1.3 you can see the icon used to represent an association.

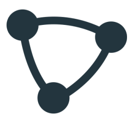


Figure 1.3: Associations icon

CHAPTER 2

Understanding Associations

This chapter describes internal details about the BigML associations, providing the foundations to understand the associations' configuration options. Association Discovery has been extensively researched over the last two decades. It is distinguished from existing statistical techniques for categorical association analysis in three respects:

- Association Discovery techniques scale to high-dimensional data. The standard statistical approach to categorical association analysis, log-linear analysis[1] has complexity that is exponential with respect to the number of variables. In contrast, Association Discovery techniques can typically handle many thousands of variables.
- Association Discovery concentrates on discovering relationships between values rather than variables. This is a non-trivial distinction. If someone is told that there is an association between gender and some medical condition, they are likely to immediately wish to know which gender is positively associated with the condition and which is not. Association Discovery goes directly to this question of interest. Furthermore, associations between values, rather than variables, can be more powerful (i.e., discover weaker relationships) when variables have more than two values. Statistical techniques may have difficulty detecting an association when there are many values for each variable and two values are strongly associated, but there are only weak interactions among the remaining values.
- Association Discovery focuses on finding associations that are useful for the user, whereas statistical techniques focus on controlling the risk of making false discoveries. In contexts where there are very large numbers of associations, it is critical to help users quickly identify which are the most important for their immediate applications.

Historically, the main body of Association Discovery research has concentrated on developing efficient techniques for finding frequent itemsets, and has paid little attention to the questions of what types of association are useful to find and how those types of associations might be found. The dominant association mining paradigm, frequent association mining, has significant limitations and often discovers so many spurious associations that it is next to impossible to identify the potentially useful ones.

The filtered-top-k[2] association technique that underlies the BigML associations implementation was developed by Professor Geoff Webb. It focuses on finding the most useful associations for the user specific application. This approach has been successfuly used in numerous scientific applications ranging from health data mining and cancer mortality studies to controlling robots and to improving e-learning.

[1] https://en.wikipedia.org/wiki/Log-linear_analysis
[2] http://i.giwebb.com/index.php/research-programs/filtered-top-k-association-discovery/

2.1 Association Measures

This section details the precise formulas that are utilized to compute the BigML association measures. Given the association rule $(A \rightarrow C)$ where A is the antecedent itemset of the rule and C is the consequent, and N is the total number of instances in the dataset, below are the mathematical definitions for the measures[3] utilized by the BigML associations:

- **Support**[4]: the proportion of instances in the dataset that contain an itemset.

$$Support(itemset) = \frac{|\, instance \in D\, itemset \subseteq instance\,|}{N}$$

$$Support(A \rightarrow C) = Support(A \cup C)$$

- **Coverage**[5]: the support of the antecedent of an association rule, i.e., the portion of instances in the dataset that contain the antecedent itemset. It measures how often a rule can be applied.

$$Coverage(A \rightarrow C) = Support(A)$$

- **Confidence**[6] (or Strength): the percentage of instances that contain the consequent and antecedent together over the number of instances that only contain the antecedent. Confidence is computed using the support of the association rule over the coverage of the antecedent.

$$Confidence(A \rightarrow C) = \frac{Support(A \rightarrow C)}{Support(A)}$$

- **Leverage**[7]: the difference between the probability of the rule and the expected probability if the items were statistically independent.

$$Leverage(A \rightarrow C) = Support(A \rightarrow C) - (Support(A) \times Support(C))$$

- **Lift**[8]: how many times more often antecedent and consequent occur together than expected if they were statistically independent.

$$Lift(A \rightarrow C) = \frac{Support(A \rightarrow C)}{Support(A) \times Support(C)}$$

2.2 How to Structure Your Data

Association Discovery models require the data to be structured in a specific way. In section Items of the Sources with the BigML Dashboard[9] document [5] there is an introduction to the **items field** (when a field contains an arbitrary number of items, i.e., categories or labels). This section shows some data structures that lend themselves particularly well for Association Discovery.

It is common in Association Discovery to have a great number of different values per instance, e.g., a commercial dataset containing the transactions with all the products bought by customers; or medical datasets containing all the medicines prescribed per patient.

[3] http://michael.hahsler.net/research/association_rules/measures.html
[4] http://michael.hahsler.net/research/association_rules/measures.html#support
[5] http://michael.hahsler.net/research/association_rules/measures.html#coverage
[6] http://michael.hahsler.net/research/association_rules/measures.html#confidence
[7] http://michael.hahsler.net/research/association_rules/measures.html#leverage
[8] http://michael.hahsler.net/research/association_rules/measures.html#lift
[9] https://static.bigml.com/pdf/BigML_Sources.pdf

See Figure 2.1 for an example of CSV file transactional data where each transaction-ID is associated to a set of purchased products.

```
trans-ID/12345, product_A, product_B, product_C, product_D
trans-ID/67890, product_A, product_E
trans-ID/67890, product_B, product_C, product_F
```

Figure 2.1: Example of transactional data

The transactional data from Figure 2.1 can be structured in several ways:

- Binary data representation:

Tran-ID	prod_A	prod_B	prod_C	prod_D	prod_E	prod_F
12345	1	1	1	1	0	0
67890	1	0	0	0	1	0
98540	0	1	1	0	0	1

Table 2.1: Example of binary representation for transactional data

- Vertical data layout:

Trans-ID	1st_prod	2nd_prod	3rd_prod	4th_prod
12345	prod_A	prod_B	prod_C	prod_D
67890	prod_A	prod_E		
67890	prod_B	prod_C	prod_F	

Table 2.2: Example of vertical layout for transactional data

- Horizontal data layout:

Trans-ID	Products
12345	product_A, product_B, product_C, product_D
67890	product_A, product_E
67890	product_B, product_C, product_F

Table 2.3: Example of horizontal layout for transactional data

The ideal way to structure your data for Association Discovery is the one shown in the **horizontal data layout** example. By using this data structure the field “Products” will be considered an **items** field, and each product will be a unique item.

Note: you need to separate your items by a unique separator (e.g., the above example items are separated by a comma).

CHAPTER 3

Creating Associations with 1-Click

To create an association in BigML you have two options: you can use the **1-click option** which uses the default values for all available configuration options, or you can tune the parameters in advanced by using the **configuration options** explained in Chapter 4. This chapter explains how to create an association with 1-click.

From the dataset view, select 1-CLICK ASSOCIATION in the **1-click action menu**. (See Figure 3.1.)

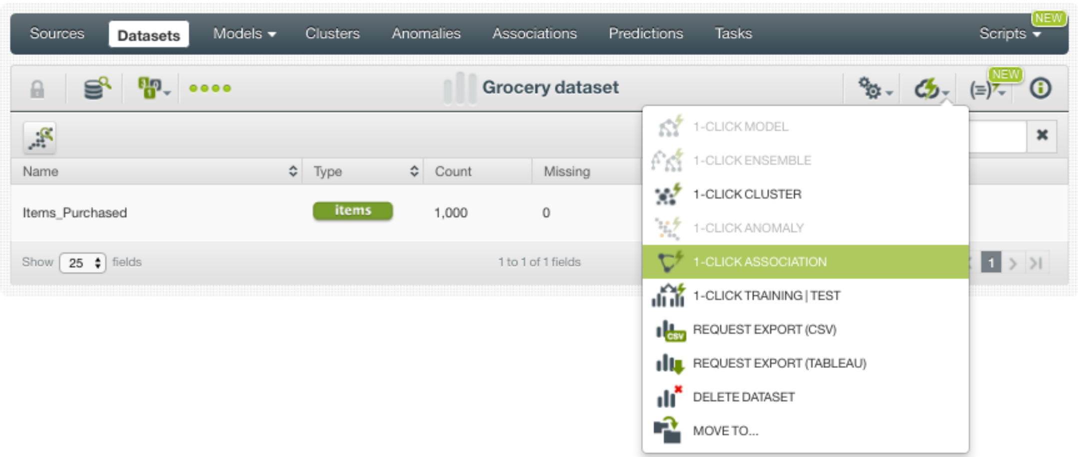


Figure 3.1: Creating an association from the 1-click action menu

Alternatively, you can select 1-CLICK ASSOCIATION from the **pop up menu** in the dataset list view. (See Figure 3.2.)

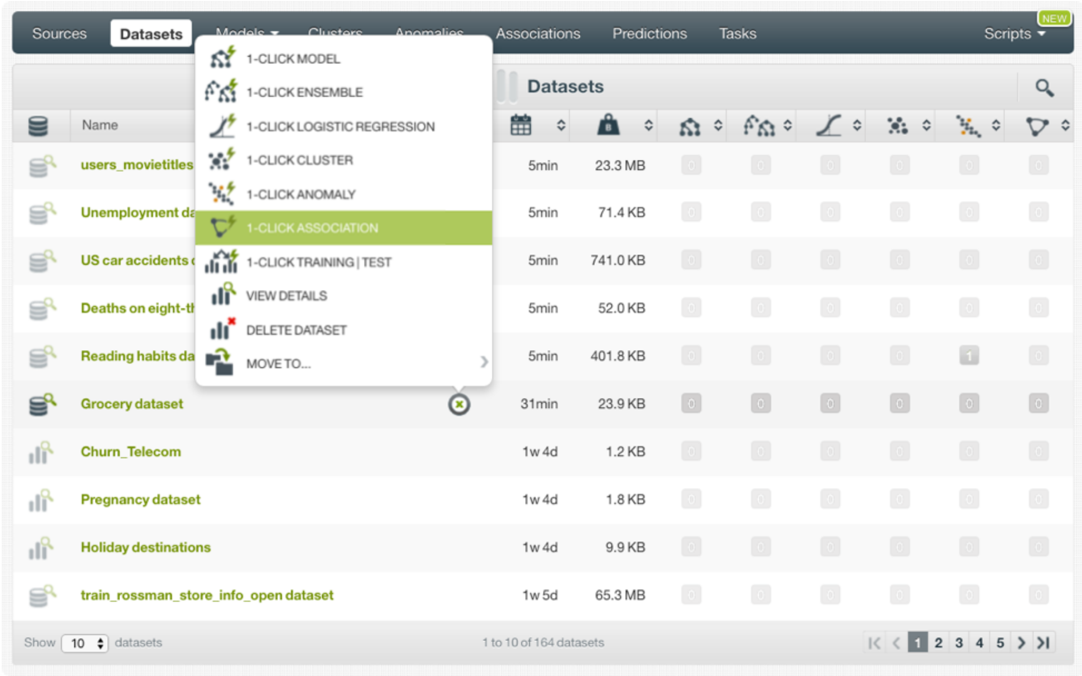


Figure 3.2: Creating an association from the pop up menu

Either option creates a new association by using **default** values for all available configuration options. (See Chapter 4.)

CHAPTER 4

Association Configuration Options

While 1-click creation (see Chapter 3) provides a convenient and easy way to create BigML associations from a dataset, there are cases when you want more control. This chapter explains a number of options you can use to configure your associations.

To display the configuration panel to see all options, from the dataset view, click the CONFIGURE ASSOCIATION menu option in the **configure option menu**. (See Figure 4.1.)

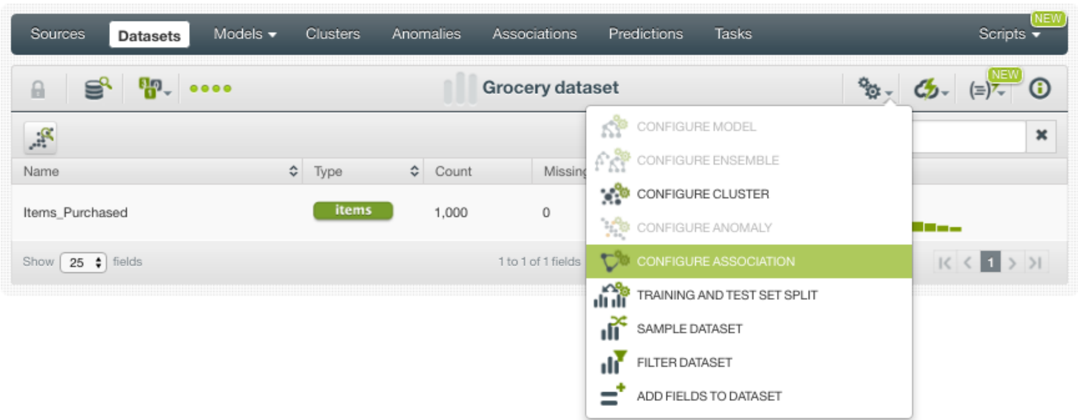


Figure 4.1: Access to configure your associations

The sections below provide a detailed explanation of the configuration options.

4.1 Maximum Number of Associations (K)

The **Max. number of associations (K)** option lets you specify the maximum number of associations to be discovered by BigML. You can set any value between 1 and 500 by moving the max. number of associations (K) slider or by typing the number you wish in the input box. For higher number of fields, values and instances in your dataset, the number of potential associations tends to increase exponentially. Thus, it makes sense to cap the number of associations. Keep in mind that higher K values will take longer to calculate. (See Figure 4.2.)

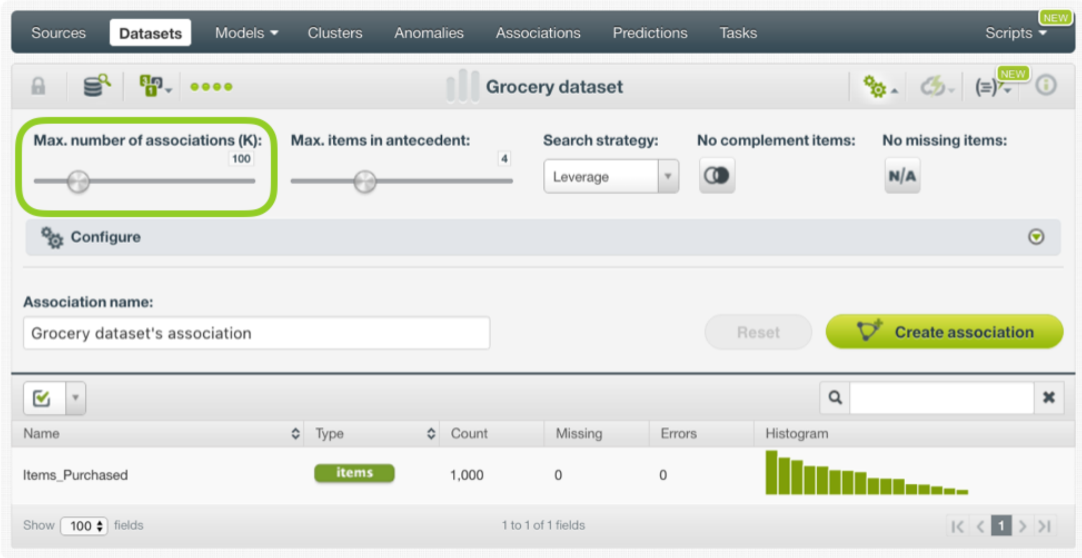


Figure 4.2: Maximum number of associations

4.2 Maximum Items in Antecedent

The **Max. items in antecedent** option lets you set the maximum number of items to be considered within the antecedent itemset. You can set values between 1 and 10 by moving the max. items in antecedent slider or by typing the number you wish in the input box. Larger numbers of items will naturally produce more complex association rules. However, the consequent itemset will always contain one item. (See Figure 4.3.)

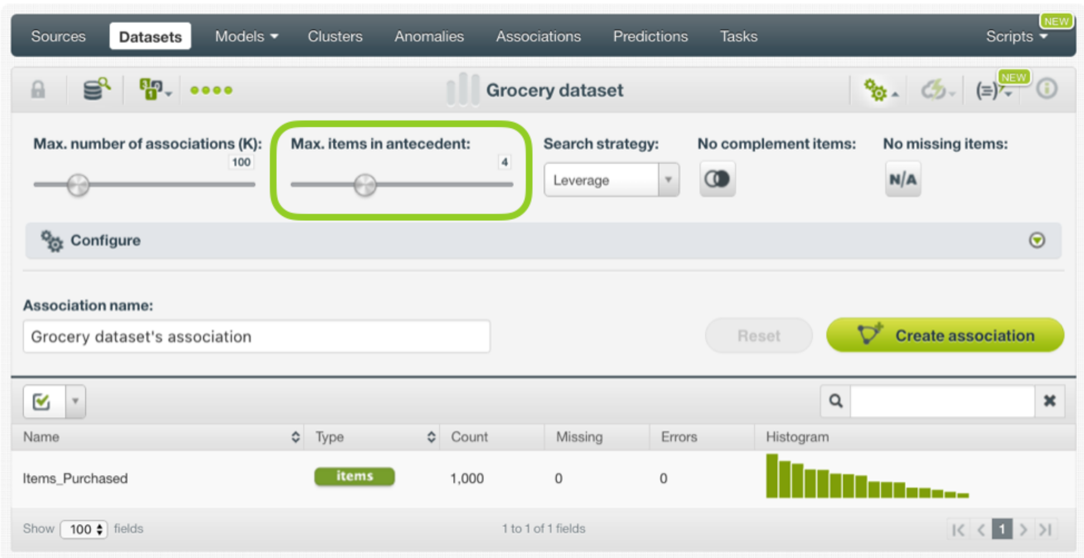


Figure 4.3: Maximum number of items in antecedent

4.3 Search Strategy

The **Search strategy** option lets you select the **measure** to prioritize the associations discovered. (See Figure 4.4.) You can use leverage, lift, coverage, support, and confidence (explained in Figure 4.4), so rules with higher values for the measure chosen will be prioritized.

By default the search strategy is **leverage** since it is one of the measures that gives relevant results in most cases. By choosing leverage, you will find associations of items that occur more frequently in your dataset. Another popular measure is lift. By choosing lift as your search strategy, you will find associations of less frequent items in your dataset, but strongly related with each other. The strategy you choose should be coherent with your application.

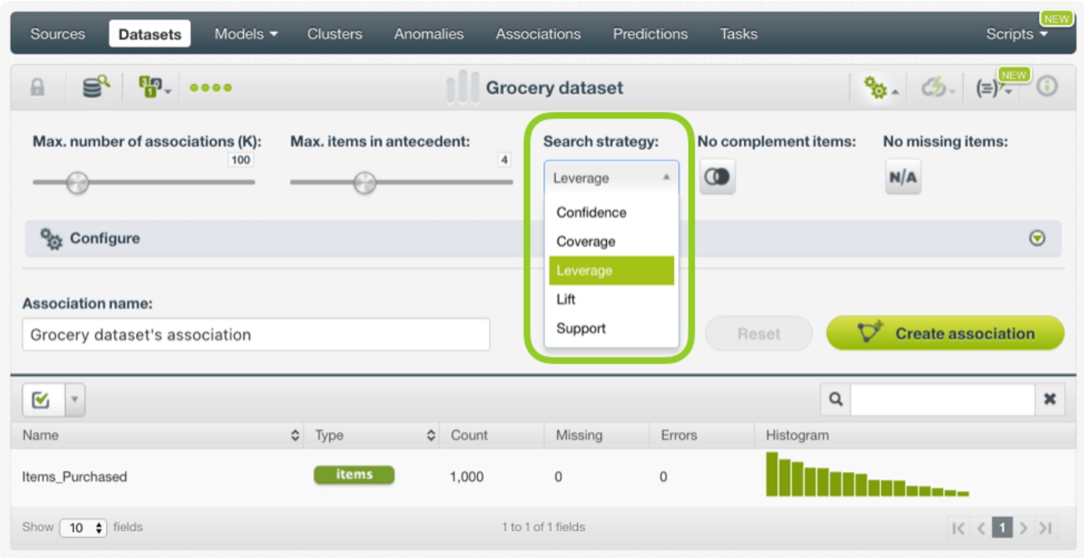


Figure 4.4: Search strategy

4.4 Complementary Items

If you enable the **complement items** option, complementary items are also taken into account. e.g., for the item $(coffee)$, the complement would be $(NOTcoffee)$. In this case, apart from the association $(milk, coffee) \rightarrow (sugar)$, complementary rules such as $(milk, NOTcoffee) \rightarrow (chocolate)$ may also be detected. BigML represents complementary items with an exclamation point $(coffee \rightarrow !coffee)$. (See Figure 4.5.)

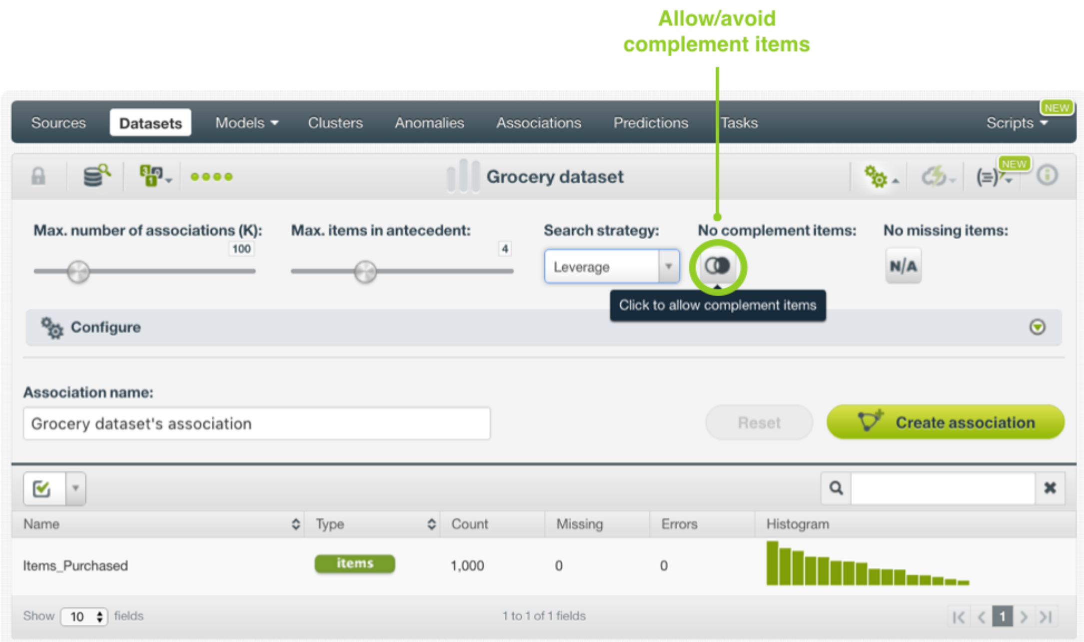


Figure 4.5: Allow or avoid complementary items

4.5 Missing Items

If you enable the **Missing items** option, missing values will be considered as another valid item when computing associations. For instance, a rule such as $(income < 39500, job\_titleisMISSING) \rightarrow (loan\_default = YES)$ can be discovered. (See Figure 4.6.)

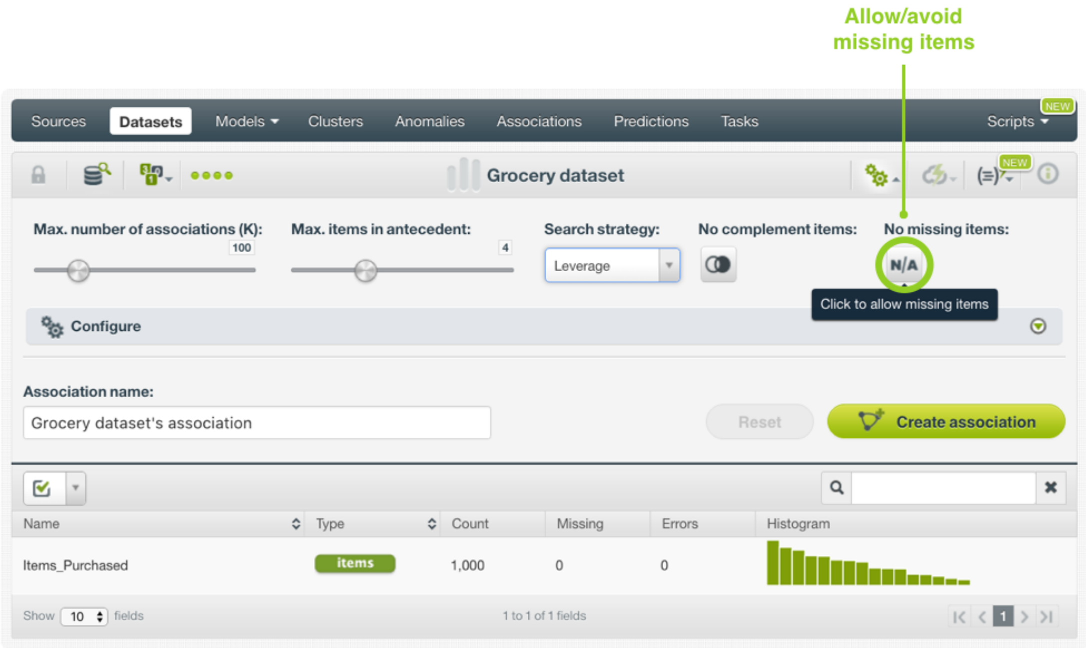


Figure 4.6: Allow or avoid missing items

4.6 Minimum Levels for the Association Measures

You can set minimum levels for a number of association measures (Figure 4.7) that let you focus on more interesting association rules, while filtering out potentially spurious ones. As for interestingness of an association rule, there is no single measure that is always more important than others. Similarly, there are no general thresholds to consider as essential rules. Analyze your results according to your main goals, which may be different depending on the problem you are trying to solve.

For example, you may be interested in very frequent associations, so you will have to pay more attention to the support rule. Perhaps you want to find some more infrequent associations, but with a stronger relationship between the items (i.e., rules with higher lift). Usually it is not one single measure, but the combination and coherence of all measures that makes one rule more relevant and useful than others.

The following subsections explain the meaning of each association measure.

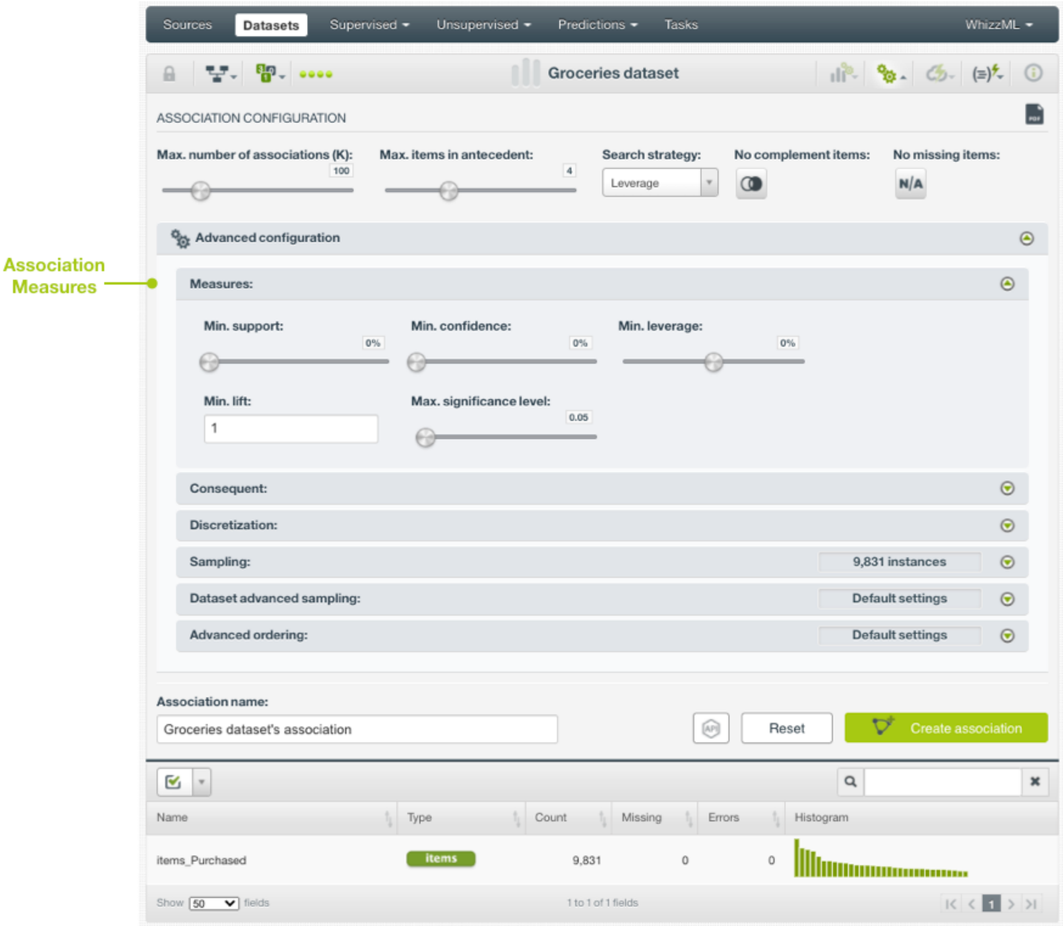


Figure 4.7: Association measures

4.6.1 Minimum Support

In Figure 4.7, **support** is the portion of instances in the dataset which contain the rule's antecedent and rule's consequent together, divided by the total number of instances (N) in the dataset. It gives a measure of the prevalance of the rule in your dataset.

You can set a support threshold between 0% and 100% by moving the min. support slider or by typing the percentage in the input box. BigML will automatically discard associations below this support level. As the minimum support percentage increases, your association rules will be based on higher occurance in your dataset.

4.6.2 Minimum Confidence

In Figure 4.7, **confidence** is the percentage of instances which contain the consequent and antecedent together over the number of instances which only contain the antecedent. Think of it as an estimate of the probability that the consequent will occur in case the antecedent occurs. Some publications also refer to confidence as strength.

You can set a confidence threshold between 0% and 100% by moving the min. confidence slider or by typing the percentage in the input box. Associations below this confidence will be automatically discarded.

4.6.3 Minimum Leverage

In Figure 4.7, **leverage** measures the difference between the probability of the rule and the expected probability if the items were statistically independent. Leverage ranges between [-1, 1]. A leverage of 0 suggests there is no association between the items. Higher positive leverage values suggest a stronger positive association between the antecedent and consequent. Negative values for leverage suggest a negative relationship.

You can set a leverage threshold between -100% and 100% by moving the min. leverage slider or by typing the percentage in the input box. Associations below this leverage will be discarded.

4.6.4 Significance Level

In Figure 4.7, **significance level** is the maximum level of risk you are willing to take to discover a spurious association. BigML applies statistical tests to control the risk of finding spurious associations. The lower the significance level, the less likely this rule is spurious, either because the antecedent and consequent are unrelated to one another, or because one or more of the values in the antecedent do not contribute to the association with the consequent. It is set to 5% (or 0.05) by default, but you can change this value by moving the max. significance level slider or by typing the number you wish in the input box.

4.6.5 Minimum Lift

Finally, in Figure 4.7, **lift** represents how much more often antecedent and consequent occur together, than expected, if they were statistically independent, e.g., a lift of 5 for the following rule $(onions \rightarrow potatoes)$ means that buying onions makes it 5 times more likely the shopper will buy potatoes. Lift is always a real positive number. A lift of 1 suggests there is no association between the items. A lift between 0 and 1 indicates a negative correlation. Higher values suggest stronger relationships between the items.

You can set any positive real number by typing the number in the input box. Associations below this lift will be discarded.

4.7 Discretization

Associations do not support numeric fields. Your numeric fields will be automatically converted into categorical fields to create your association. This process is called discretization. For instance, a numeric field like “Age”, with values between 0 and 50, can be discretized in 5 different segments or classes: 1-10, 11-20, 21-30, 31-40, and 41-50. These five segments will be the classes for your new categorical field.

BigML allows you to configure the following discretization options. If you do not configure them, BigML will apply the default values. (See Figure 4.8.)

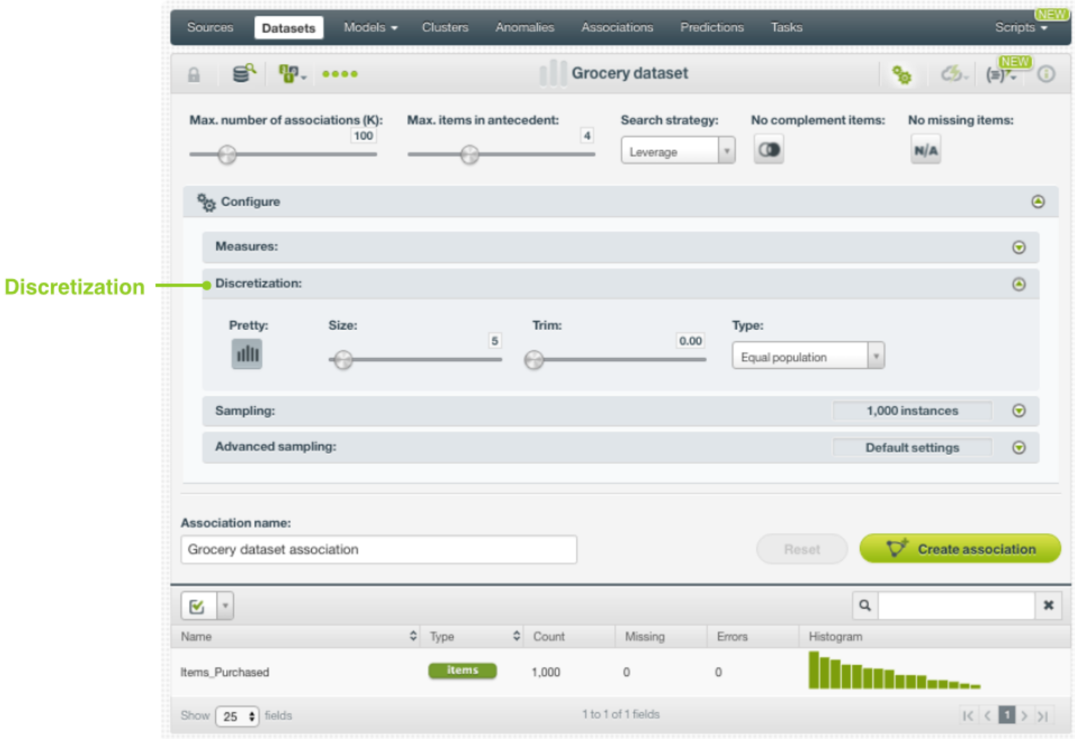


Figure 4.8: Discretization options

4.7.1 Pretty

It is highly likely that during discretization, numeric fields may have boundaries that are decimal numbers. By enabling the **Pretty** discretization option (Figure 4.8), you can force segment boundaries and widths for numeric fields to be set in a way that are easy to read, e.g., instead of $segment > 20.678$ you will get $segment > 20$. If Pretty is enabled, the specified **Size** may act as a maximum. (See Subsection 4.7.4 and Subsection 4.7.2.)

4.7.2 Size

The **Size** discretization option (Figure 4.8) lets you specify the number of groups (or classes) for your numeric fields, e.g., if you set **Size** = 2 and **Type** = width, for a field ranging from 1 to 10 containing integer values, you will get two equal width segments, from 1 to 5, and from 6 to 10. The default value is **Size** = 5. You can set up to 50 segments by moving the size slider or by typing the number of segments you wish in the input box.

If the **Pretty** option is enabled, then this value acts as a maximum size.

4.7.3 Trim

The **Trim** discretization option (Figure 4.8), is the portion of the overall population that may be removed from either tail of the distribution. You can set a number between 0% and 10% by moving the trim slider or by typing the percentage in the input box.

For example, 0.01 indicates that 1% of the data may be removed from either tail. A trim of 1% usually gives good results, because it tends to eliminate most of the outliers.

4.7.4 Type

Finally, the **Type** discretization option (Figure 4.8), lets you select whether you want to discretize the field by using an equal width or equal population strategy for each segment. The right choice depends on the distribution of your numeric field.

4.8 Sampling Options

If you do not want to use all your dataset to create associations, BigML lets you create associations for a sample of your dataset. You may configure the sampling options explained in the following subsections. (See Figure 4.9).

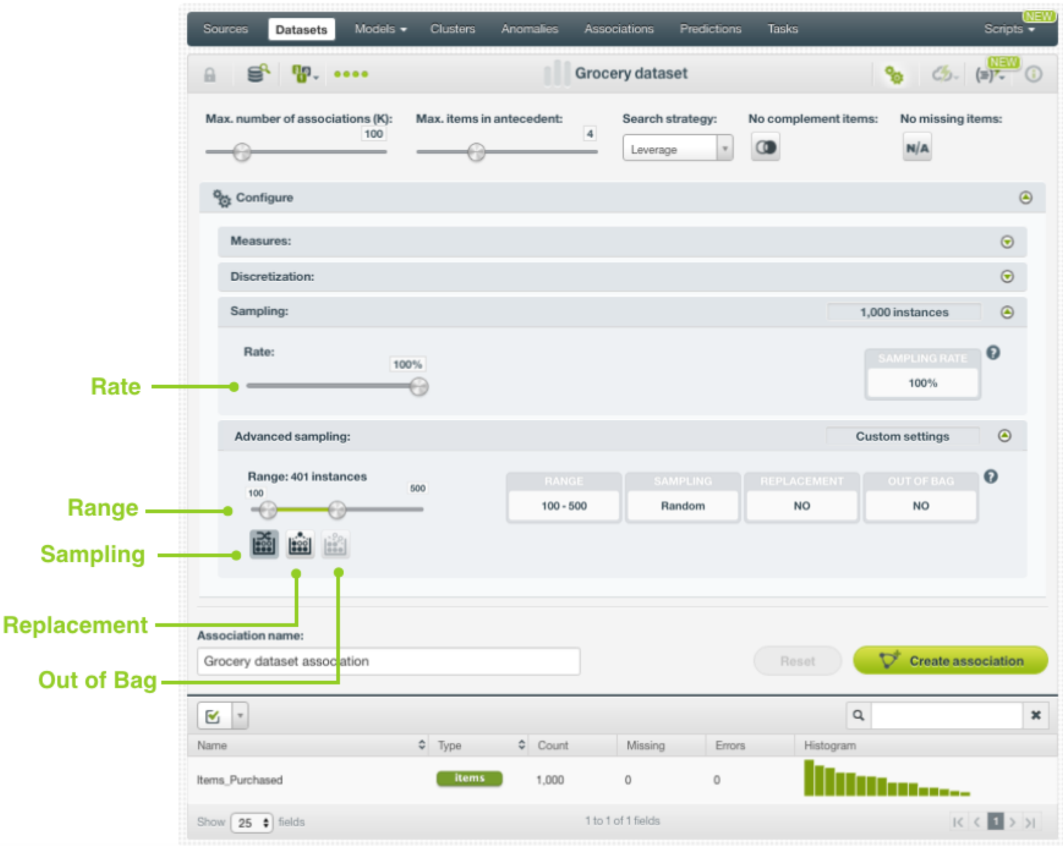


Figure 4.9: Configuration panels to sample your dataset

4.8.1 Rate

The **Rate** option allows you to set the proportion of instances to include in your sample. It is a value between 0% and 100% and it defaults to 100%. You can change this value by moving the rate slider shown in Figure 4.9 or by typing the percentage in the input box.

4.8.2 Range

The **Range** option lets you specify a linear subset of the instances that you want to consider for your sample, e.g., from instance 100 to instance 500. Select the desired range by moving the range slider shown in Figure 4.9 or by typing the percentage in the input box. The **rate** value that you set will only be computed over the **range** you specify.

4.8.3 Sampling

By default, BigML selects your instances for the sample by using a random number generator, which means two samples from the same dataset will likely be different even when using the same rates and row ranges. Choose between a **random sampling** or **deterministic sampling**. If you choose **deterministic sampling**, the random-number generator will always use the same seed, thus producing repeatable results. This lets you work with identical samples from the same dataset.

4.8.4 Replacement

Sampling with replacement allows a single instance to be selected multiple times. **Sampling without replacement** ensures that each instance cannot be selected more than once. By default, BigML generates samples without replacement.

4.8.5 Out of Bag

This option creates a sample containing only out-of-bag instances for the currently defined rate. If an instance is not selected as part of a sampling, it is considered an out-of-bag instance. Thus, the final total percentage of instances for your sample will be 100% minus the rate configured for your sample (when replacement is false). This can be useful for splitting a dataset into training and testing subsets. It is only electable when a sample rate is less than 100%.

4.9 Creating Associations with Configured Options

After finishing the configuration of your options, you can change the default association name in the editable text box. Then you can click on the Create association button to create the new association, or reset the configuration by clicking on the Reset button.

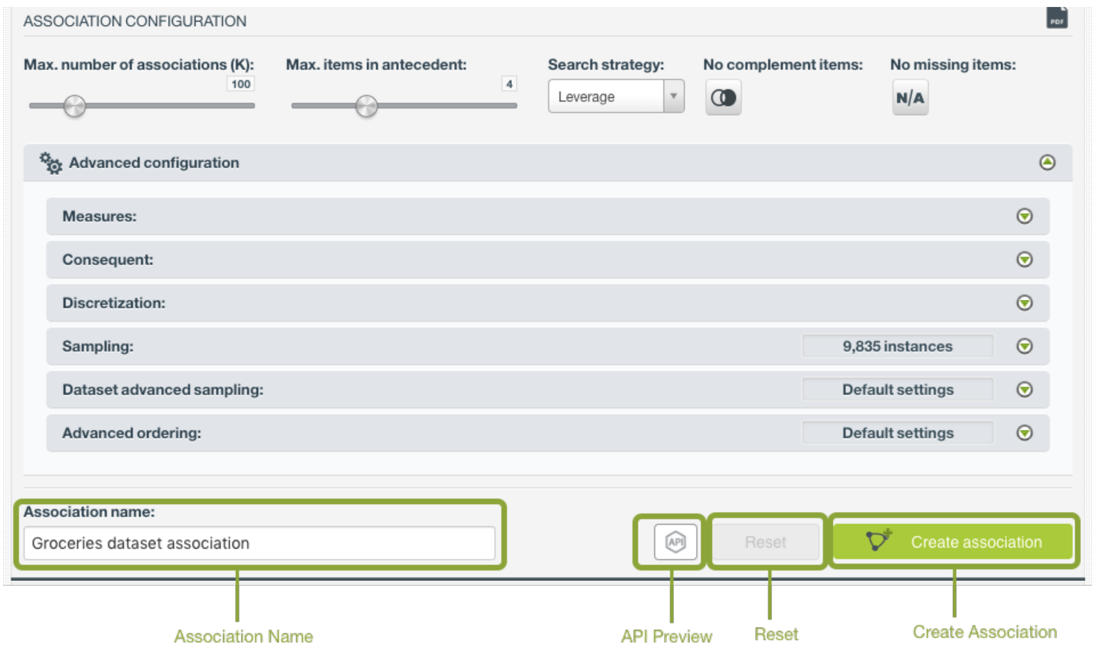


Figure 4.10: Create association after configuration

4.10 API Request Preview

The API Request Preview button is in the middle on the bottom of the configuration panel, next to the Reset button (See (Figure 4.10)). This is to show how to create the association programmatically: the endpoint of the REST API call and the JSON that specifies the arguments configured in the panel. Please see (Figure 4.11) below:

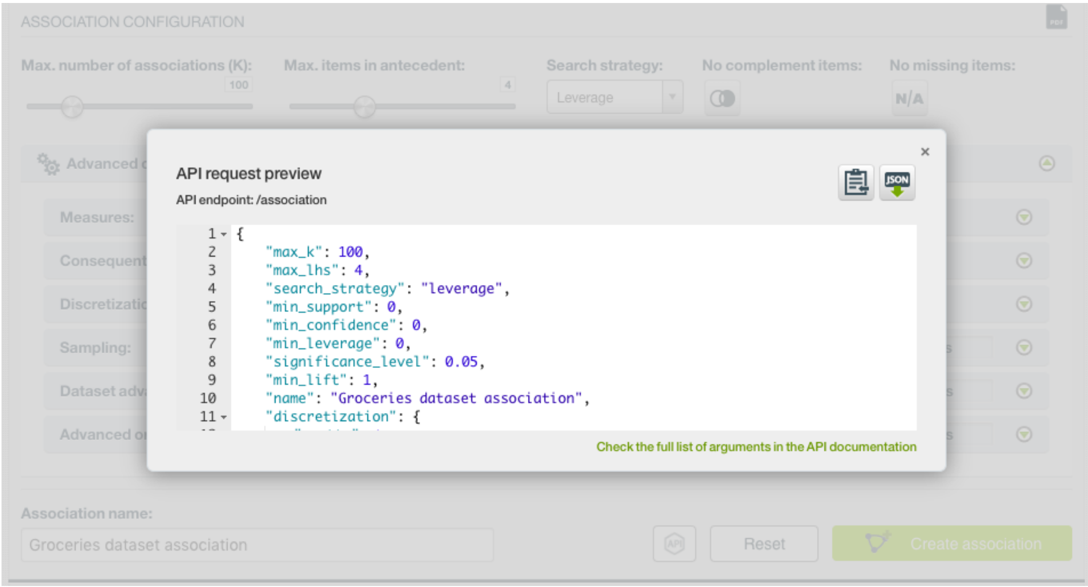


Figure 4.11: Association API request preview

There are options on the upper right to either export the JSON or copy it to clipboard. On the bottom there is a link to the API documentation for associations, in case you need to check any of the possible values or want to extend your knowledge in the use of the API to automate your workflows.

Please note: when a default value for an argument is used in the chosen configuration, the argument won't appear in the generated JSON. Because during API calls, default values are used when arguments are missing, there is no need to send them in the creation request.

CHAPTER 5

Visualizing Associations

BigML provides two different visualizations of the association rules discovered: a **table** and a **network chart**, explained in (Section 5.1) and (Section 5.2) respectively.

To better understand the conventions that BigML **association rules** follow, see below a simple association rule example (Figure 5.1):

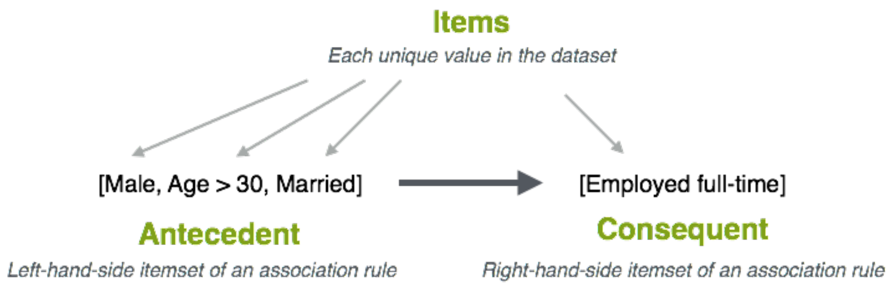


Figure 5.1: Association rule example

This rule indicates that if the person is male, more than 30 years old, and married (antecedent), it is likely that he is also a full-time employee (consequent).

Note: association rules look for co-occurrences between items and don't imply causality. In this example, being a full-time employee is not a direct consequence of being a 30-year-old married male, it's just a co-occurrence that appears more often than expected.

5.1 Associations Table View

After associations are created, you will get a table (Figure 5.2) that summarizes all the rules discovered.

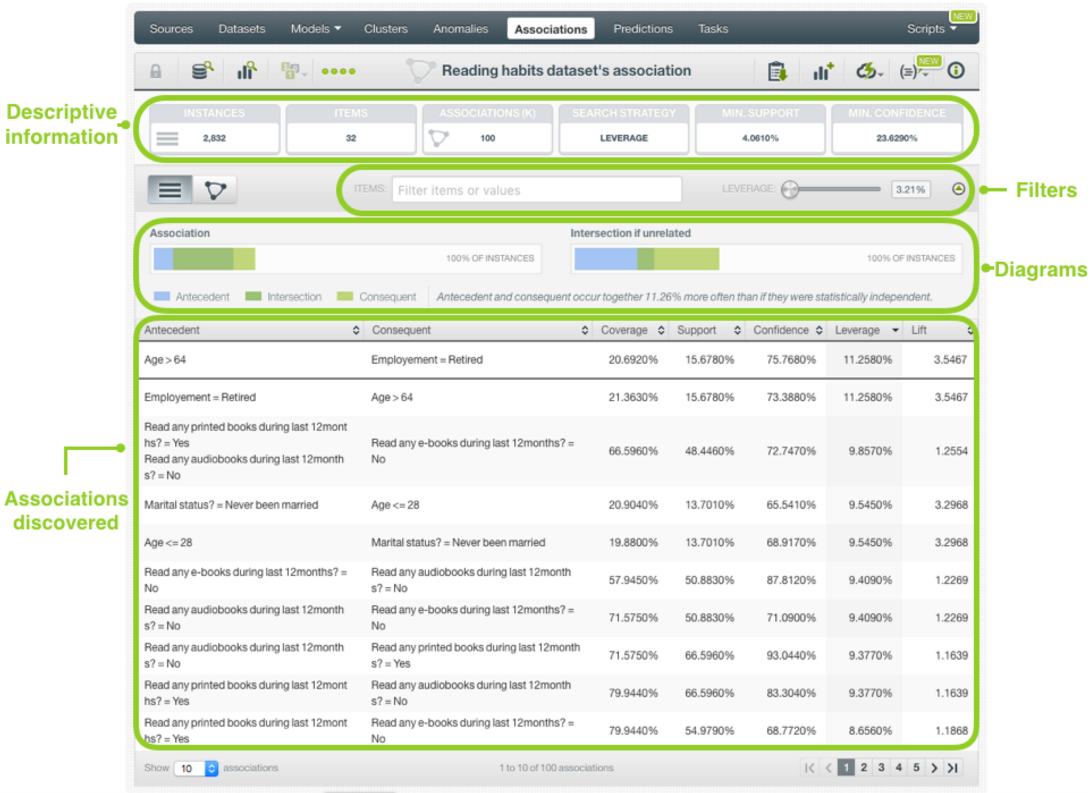


Figure 5.2: Associations table overview

At the top part, from left to right, you can see some basic **descriptive information**, such as the number of instances contained in this dataset (2,832), the number of item fields (32), number of associations discovered (100), the search strategy chosen (leverage), the percentage set for the minimum support (4.0610%), and the percentage set for the minimum confidence (23.6290%).

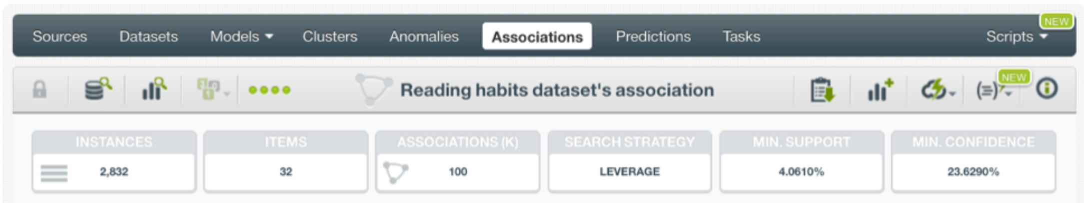


Figure 5.3: Descriptive information

Below this descriptive information, you can **filter your associations** by typing the items or values in the input box, or by moving the slider to filter rules by the measure used as the search strategy, leverage in this example. (See Figure 5.4.)

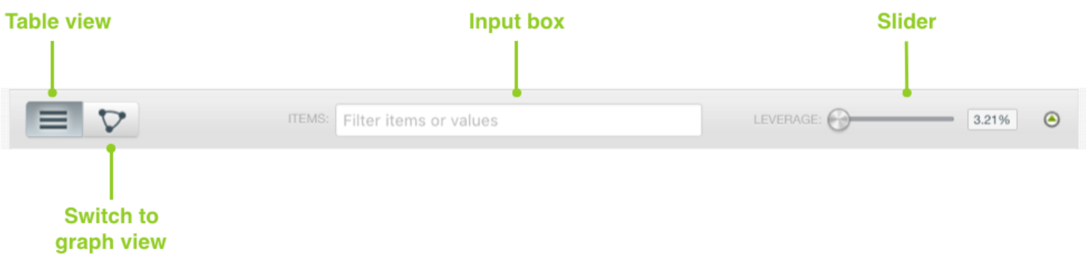


Figure 5.4: Filter your associations

If you open the **diagram panel** (Figure 5.5) and select an association in the table, you will get two **graphic representations** of this rule: the **association diagram** on the left indicating the actual intersection between the antecedent and consequent itemsets of this rule, and the **intersection if unrelated diagram** on the right, indicating the intersection if both itemsets were independent. These diagrams provide a visual overview of the importance of the selected association rule. The **blue bar** represents the portion of instances in the dataset that contain the antecedent items (coverage), and the **green bar** is the portion of instances that contain the consequent itemset. The **intersection** between them is the portion of instances that contain both itemsets (i.e., the support of the rule). You can also get a visual insight of the lift and leverage rules, which are represented by the differences between both diagrams intersections, i.e., the differences of the actual association intersection versus the intersection if both itemsets were independent.

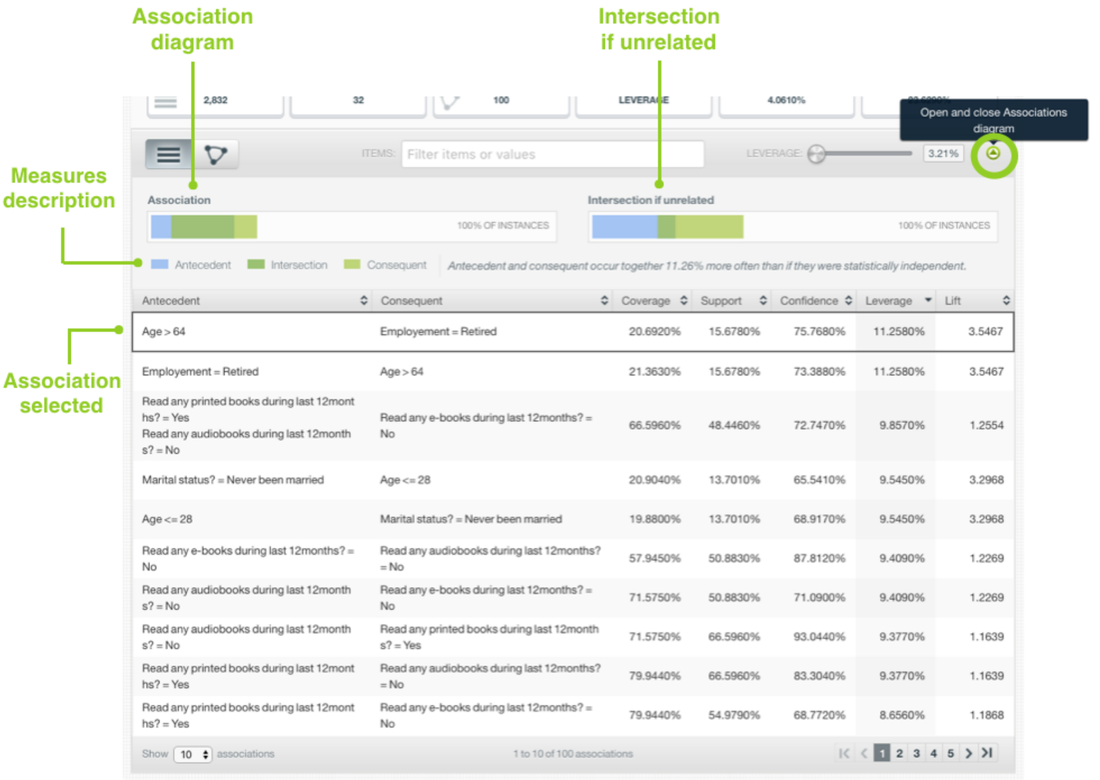
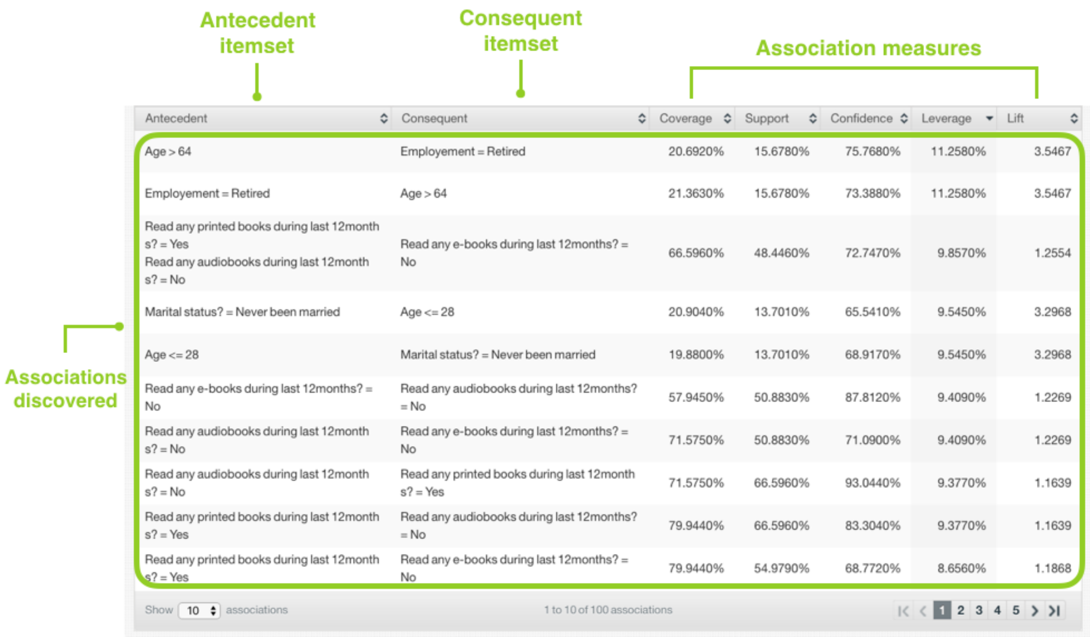


Figure 5.5: Association rule diagrams

Regarding the **table** (Figure 5.6), the main part of this view, each row contains a rule which is composed of two parts: the **Antecedent** itemset, with one or more items, and the **Consequent** itemset, which will always contain one item. For each rule you will find five different measures (**Coverage, Support, Confidence, Leverage, and Lift**) that describe the relationship between both parts of the rule. (The technicalities behind these rules are explained in Section 2.1.)



Antecedent	Consequent	Coverage	Support	Confidence	Leverage	Lift
Age > 64	Employement = Retired	20.6920%	15.6780%	75.7680%	11.2580%	3.5467
Employement = Retired	Age > 64	21.3630%	15.6780%	73.3880%	11.2580%	3.5467
Read any printed books during last 12months? = Yes Read any audiobooks during last 12months? = No	Read any e-books during last 12months? = No	66.5960%	48.4460%	72.7470%	9.8570%	1.2554
Marital status? = Never been married	Age <= 28	20.9040%	13.7010%	65.5410%	9.5450%	3.2968
Age <= 28	Marital status? = Never been married	19.8800%	13.7010%	68.9170%	9.5450%	3.2968
Read any e-books during last 12months? = No	Read any audiobooks during last 12months? = No	57.9450%	50.8830%	87.8120%	9.4090%	1.2269
Read any audiobooks during last 12months? = No	Read any e-books during last 12months? = No	71.5750%	50.8830%	71.0900%	9.4090%	1.2269
Read any audiobooks during last 12months? = No	Read any printed books during last 12months? = Yes	71.5750%	66.5960%	93.0440%	9.3770%	1.1639
Read any printed books during last 12months? = Yes	Read any audiobooks during last 12months? = No	79.9440%	66.5960%	83.3040%	9.3770%	1.1639
Read any printed books during last 12months? = Yes	Read any e-books during last 12months? = No	79.9440%	54.9790%	68.7720%	8.6560%	1.1868

Figure 5.6: Associations discovered

5.2 Associations Chart View

If you prefer a graph view, BigML lets you switch the view to visualize the rules in a **network chart** (Figure 5.7). This chart will give you a nice visual overview of which items are connected to which other items. You can apply a filter based on the measure you used as the search strategy, color the chart points by field, and show and hide item labels as you wish.

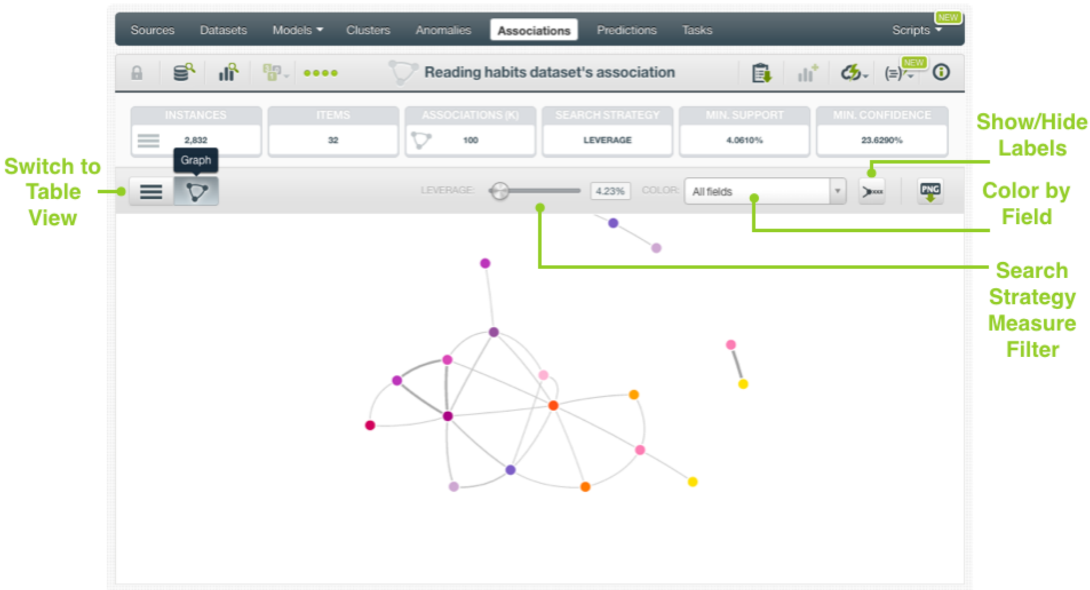


Figure 5.7: Associations network chart

CHAPTER 6

Association Summary Report

BigML provides a summary report to get an overview of the most important associations. From the association view, you can access the **association summary report** by clicking the ASSOCIATION REPORT menu option highlighted in Figure 6.1.

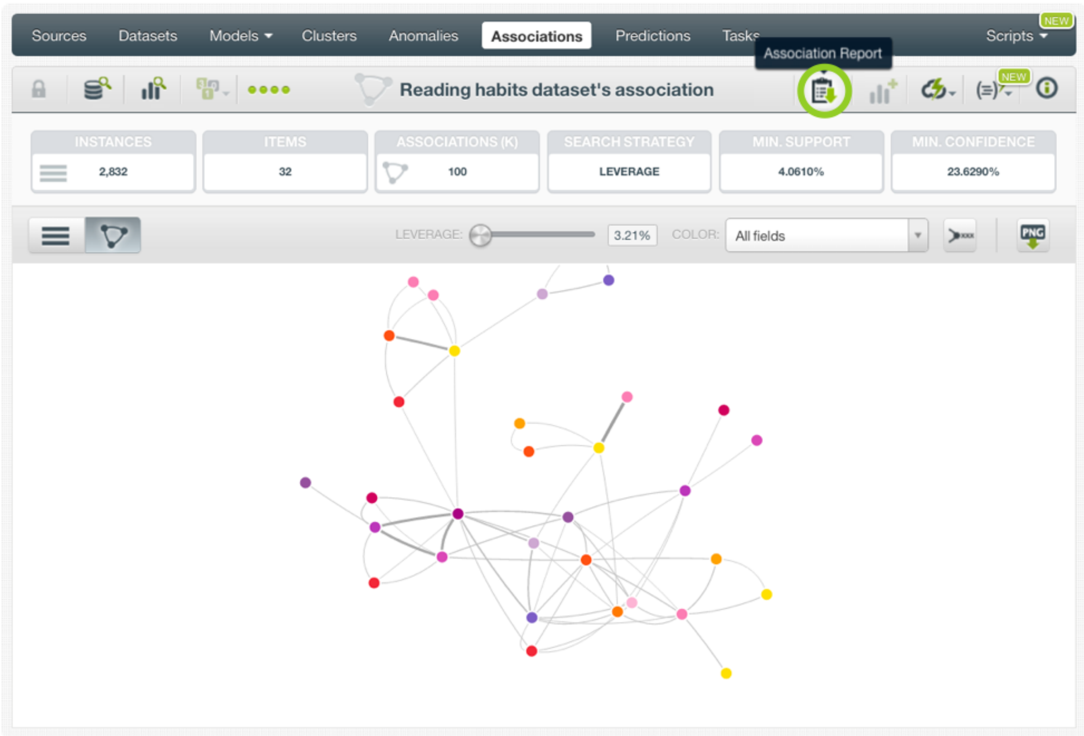


Figure 6.1: Associations report menu option

The association summary report (Figure 6.2) includes:

- Total number of rules: this states the number of association rules discovered.
- Top 10 by Coverage: top 10 association rules according to coverage.
- Top 10 by Support: top 10 association rules according to support.
- Top 10 by Confidence: top 10 association rules according to confidence.
- Top 10 by Leverage: top 10 association rules according to leverage.
- Top 10 by Lift: top 10 association rules according to lift.

- Top 10 by p-value: top 10 association rules according to p-value.

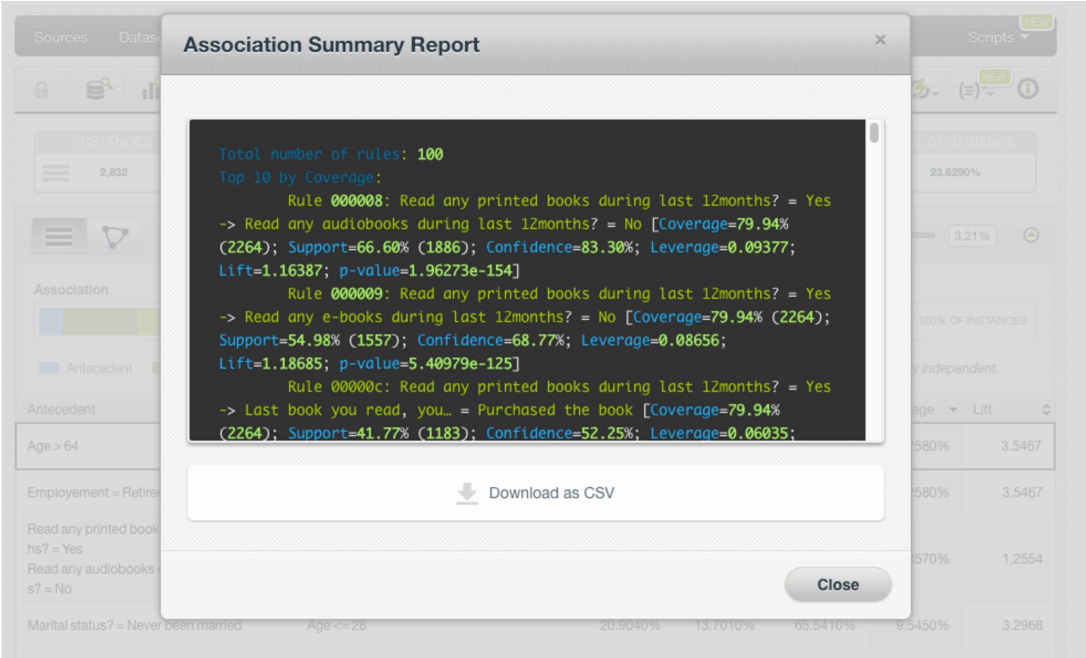


Figure 6.2: Associations summary report

CHAPTER 7

Create a Dataset From an Association

BigML lets you create a new dataset including or excluding the instances containing the associations discovered.

Access this option from the association (table) view, by clicking the CREATE DATASET FROM ASSOCIATION menu option (Figure 7.1)

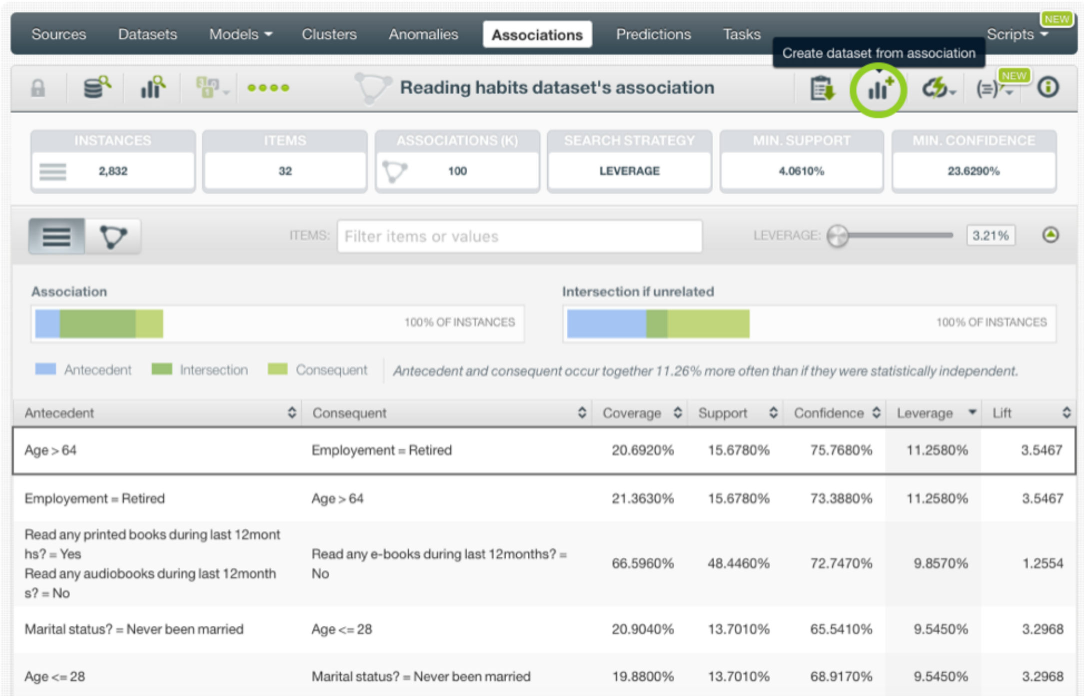


Figure 7.1: Create dataset from associations

Then, choose the rules you want to include in your dataset by checking the corresponding check boxes for them, and click the Create dataset button. Alternatively, you can click the highlighted button to create a new dataset removing the selected rules. (See Figure 7.2.)

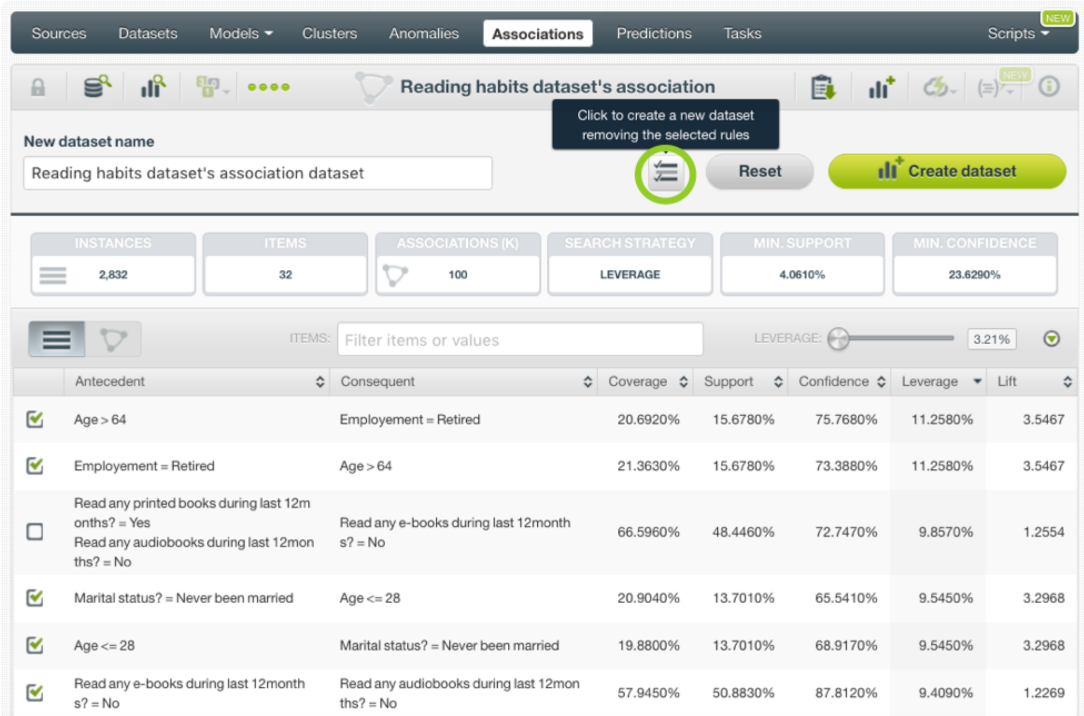


Figure 7.2: Create a new dataset including or excluding associations

CHAPTER 8

Association Predictions: Association Sets

8.1 Introduction

You can use your associations to find the items which are strongly correlated with a given set of inputs. Predictions for associations are referred to as **association sets** in BigML. Association sets are only available to predict items for **single instances** for the moment. The main goal of creating an association set is to obtain a set of predicted items associated to some input data. For example, given a set of products bought by a person, which others are very likely to be bought? Each predicted item comes with a score to indicate the strength of the association between the input data and the predicted items.

The predictions tab in the main menu of the BigML Dashboard is where all of your saved predictions are listed (Figure 8.1). In the association set list view, you can see the icon for the **Association** used for each prediction, the **Name** of the prediction, the **K** number of rules matching the input data, the **Scored by** measure and the **Age** (time since the association set was created). You can also search your association sets by name through clicking in the search menu option on the top right menu.

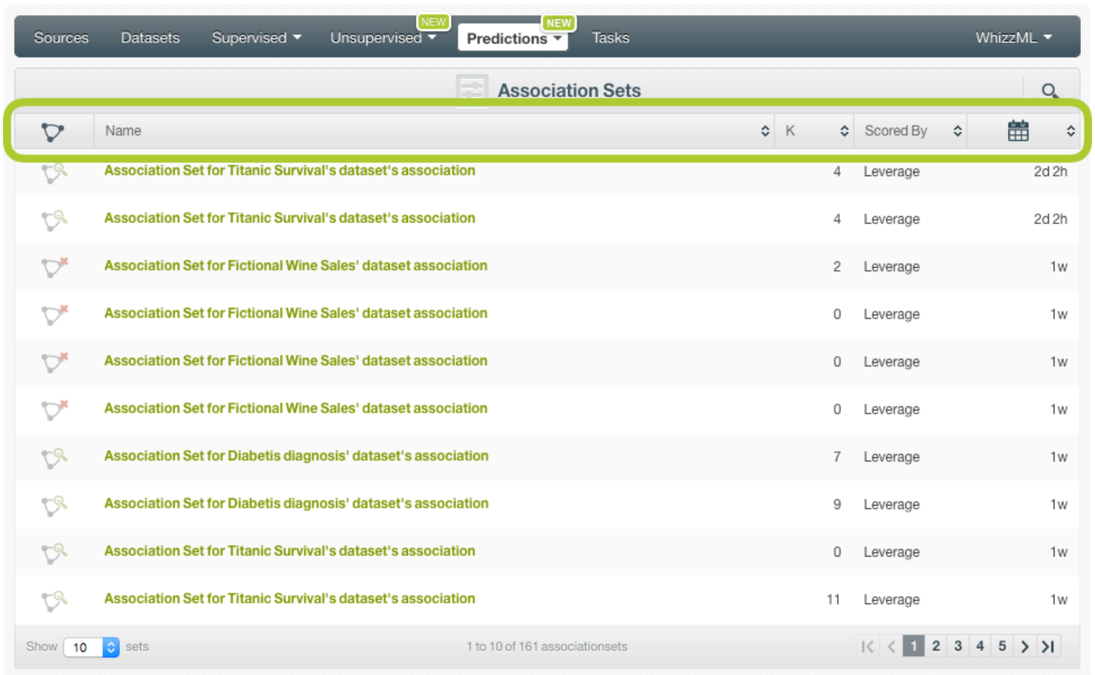


Figure 8.1: Predictions list view

By default, when you first create an account at BigML, or every time that you start a new project, your list view for predictions will be empty. (See Figure 8.2.)

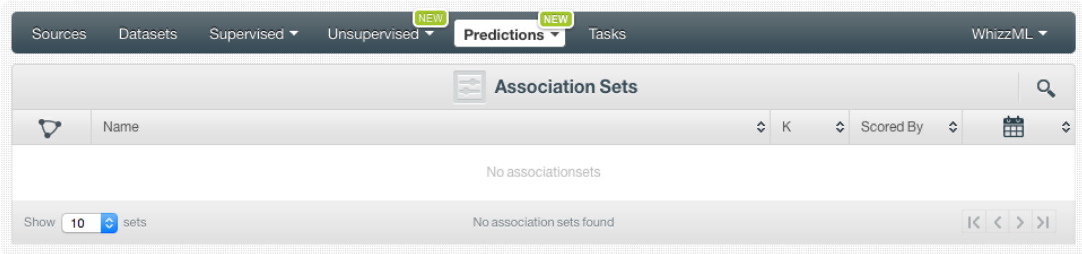


Figure 8.2: Empty Dashboard predictions view

See below the corresponding icon for association sets. (See Figure 8.3.)

Figure 8.3: Association set icon

8.2 Creating Association Sets

To get the items associated with some input data, you need to follow these steps:

1. Click the PREDICT ASSOCIATION SET option from the association **1-click action menu**. (See Figure 8.4.)

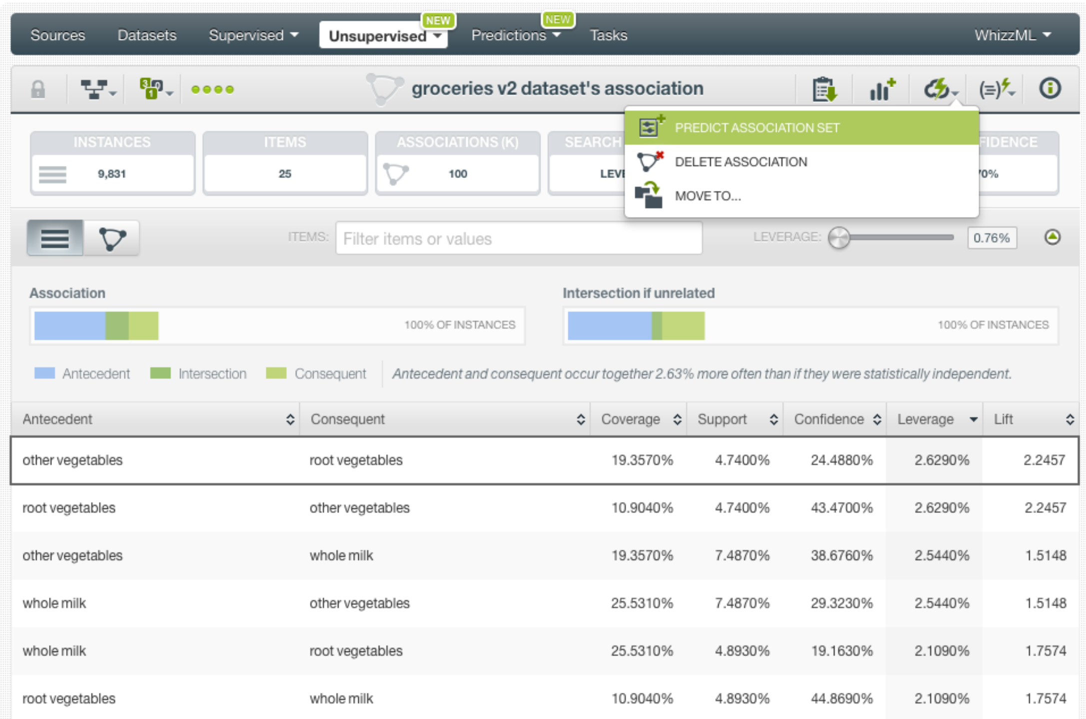


Figure 8.4: Predict option from association 1-click menu

Alternatively, click PREDICT ASSOCIATION SET in the **pop up menu** from the association list view as shown in Figure 8.5.

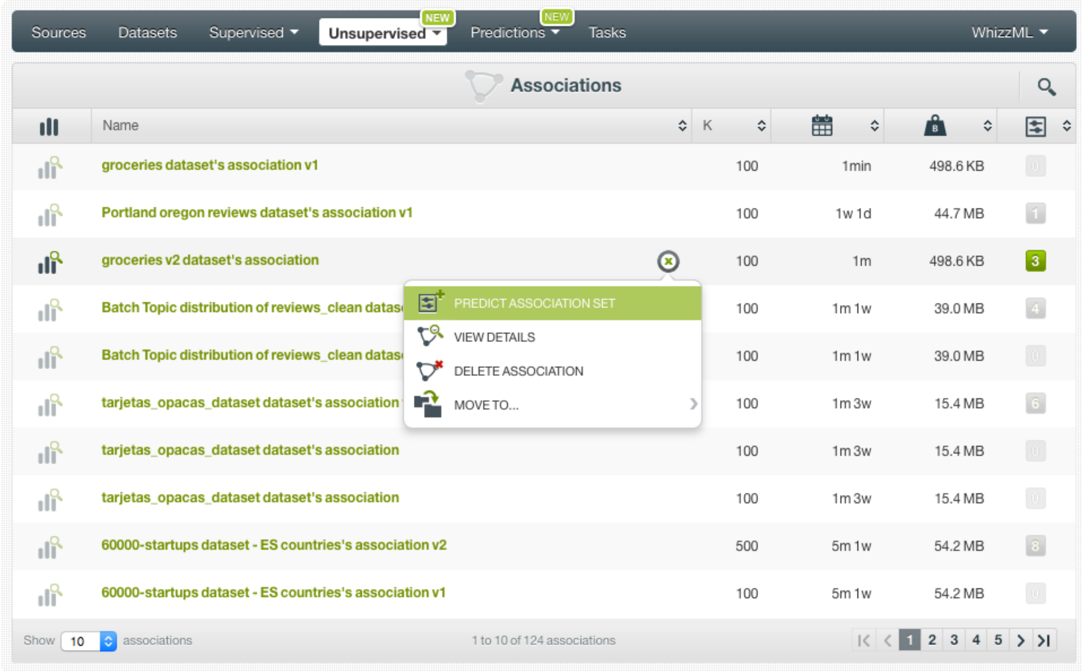


Figure 8.5: Predict option from association pop up menu

2. You will be redirected to the **prediction form**, where you will find all the input fields used by the association. (See Figure 8.6.)

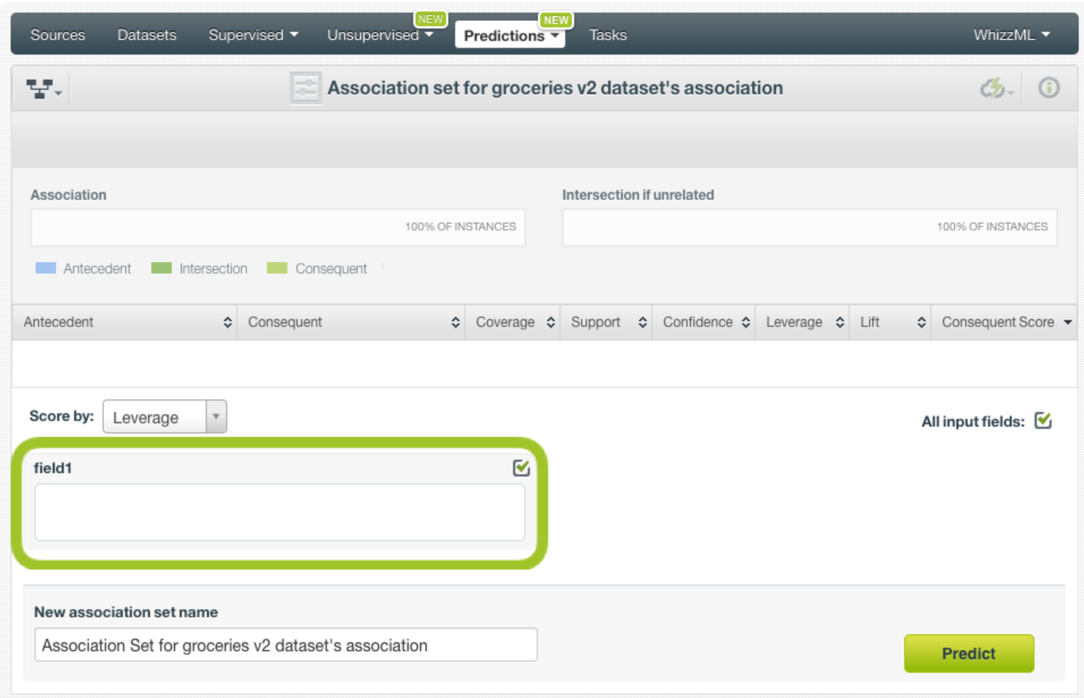


Figure 8.6: Association set form

3. **Select** the input fields that you want and **set their values**. For text and items fields, the values set as inputs will not be returned as predicted values.

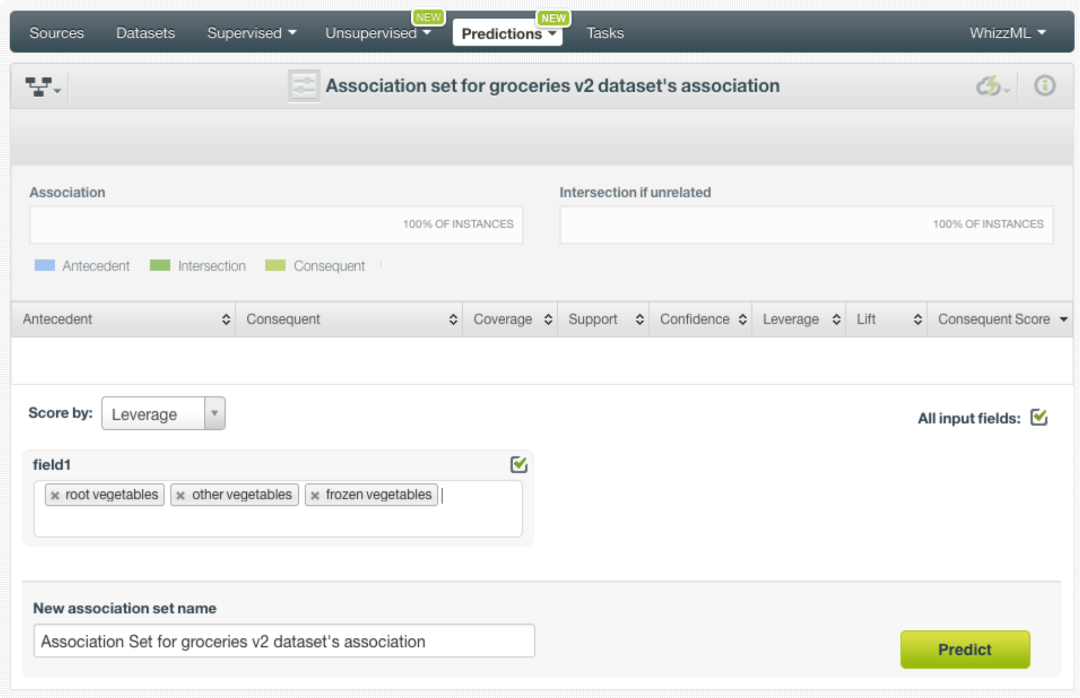


Figure 8.7: Association set inputs for text and items fields

Any categorical or numeric fields used as an input will be excluded from the predicted association set. So if you only have numeric and categorical input fields and you set them all, the predicted association set will be empty. (See Section 8.4.)

4. **Select** your preferred score measure to rank the predicted items. (See Figure 8.8.) By default, BigML uses the same measure used to create the association (see Section 4.3). See Section 8.3 for a detailed explanation of the score measure.

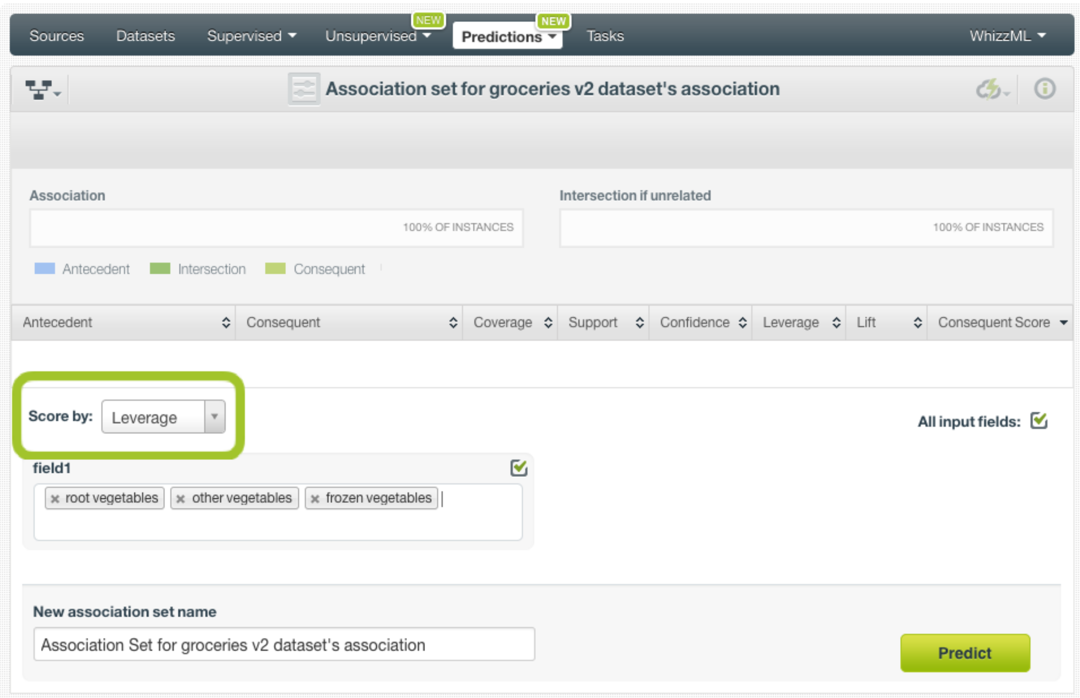


Figure 8.8: Association set scoring measure

5. Click Predict and you will get the **predicted items** along with their rules on top of the form. (See Figure 8.9.) See Section 8.4 for a detailed explanation to understand the association set results.

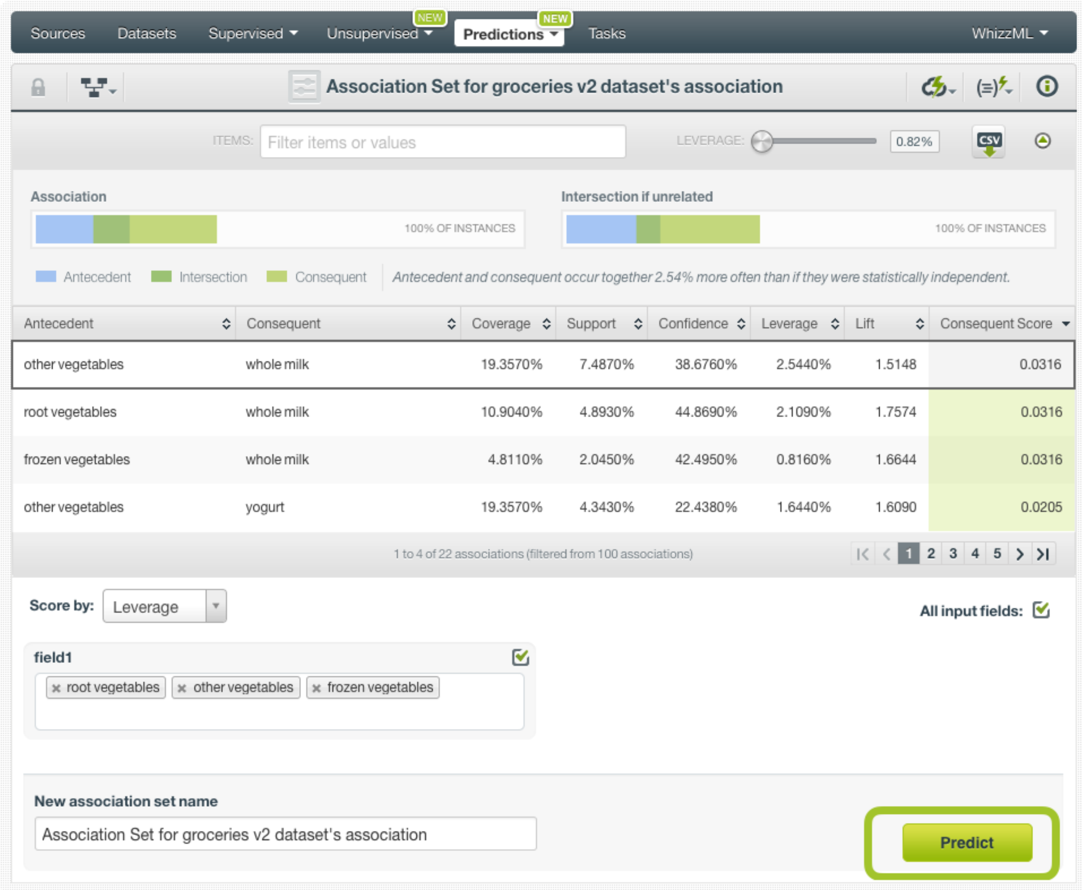


Figure 8.9: Click Predict to get the predicted items

6. The association set is saved automatically so you can find it afterwards in the prediction list view. (See Figure 8.1.)

8.3 Association Set Score

Each predicted item has an **score** associated. This score is used to rank the predicted items returned. (See Figure 8.10.) The score measures the **similarity** between the left-hand-side of the discovered rules, a.k.a **antecedent**, and the **input data**.

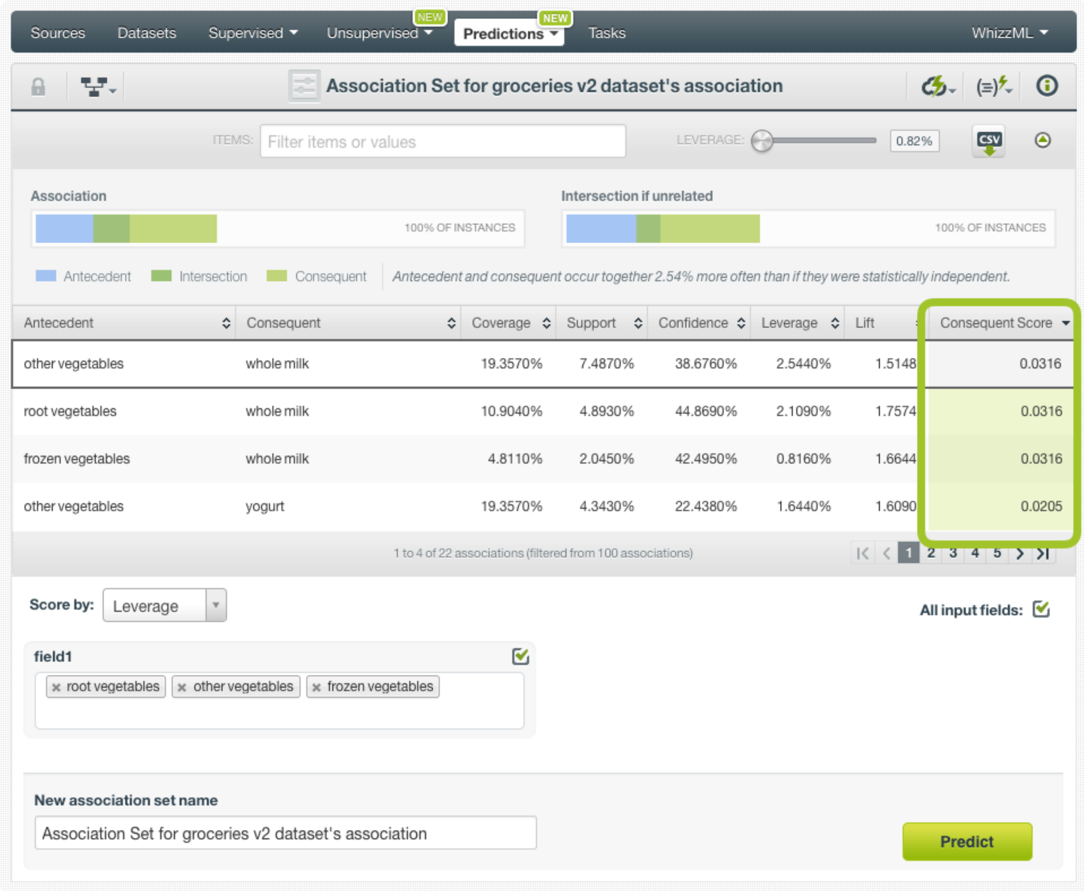


Figure 8.10: Score for predicted items

The score uses the **cosine similarity** to measure the level of coincidence between the **input data** of the association set and the **antecedent** of the association rules.

$$sim(inputs, antecedent) = \frac{|inputs \cup antecedent|}{\sqrt{|inputs|}\sqrt{|antecedent|}}$$

If the rule's antecedent does not contain any of the input items, the score will be zero. If the rule's antecedent contains at least one item from the ones given in the input data, the score will be greater than zero. If the antecedent matches the input items exactly, then it will yield the maximum similarity score, which is one. For example, if we have the following rules:

- Rule 0: $[pears] \rightarrow [kiwis]$
- Rule 1: $[bananas, pears] \rightarrow [kiwis]$
- Rule 2: $[oranges, bananas] \rightarrow [peaches]$
- Rule 3: $[oranges] \rightarrow [apples]$

Given the input itemset $oranges$ and $bananas$, the "Rule 0" will have a score equal to zero, while the "Rule 2" will yield a score equal to one since its antecedent perfectly matches the input data. "Rule 1" and "Rule 3" will have a score between zero and one because they have partial matches.

This similarity score is then multiplied by a given rule **measure** to produce a **similarity-weighted score**. You can select any of the measures explained in Section 2.1 to weight the score: coverage, support, confidence, leverage or lift. (See Figure 8.11.) By default, BigML uses the same measure used to create the association (see Section 4.3).

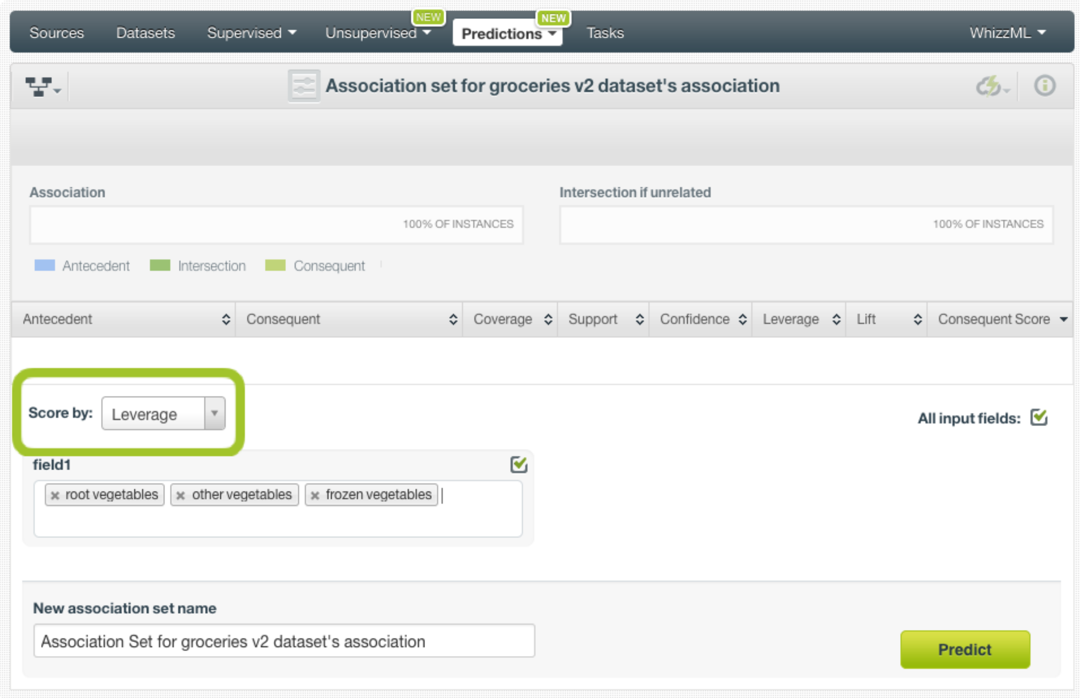


Figure 8.11: Association set scoring measure

For each rule with a non-zero score, its consequent is added to the prediction, as long as it is not already contained in the input set. If a consequent is predicted by multiple rules, its score will be the sum of the individual rule's scores.

For a further reading about the association set score refer to this paper[1].

8.4 Visualizing Association Sets

Association sets return a set of **predicted items** given some **input data** for a single instance. As we explained in Section 8.3, association sets computes the similarity score between the input data and the rule's antecedent from the original association model. Then, for rules with a similarity score greater than zero, the consequent part of the rule is returned as a predicted item (as long as the items in the consequent are not part of the input data).

There are two cases in which you will not obtain any predicted items for your input data:

- If the **input data** is not found among the **rules** discovered in the original association model. Then the following warning message will be displayed:

[1] `http://lethalletham.com/Letham_SimConf.pdf`

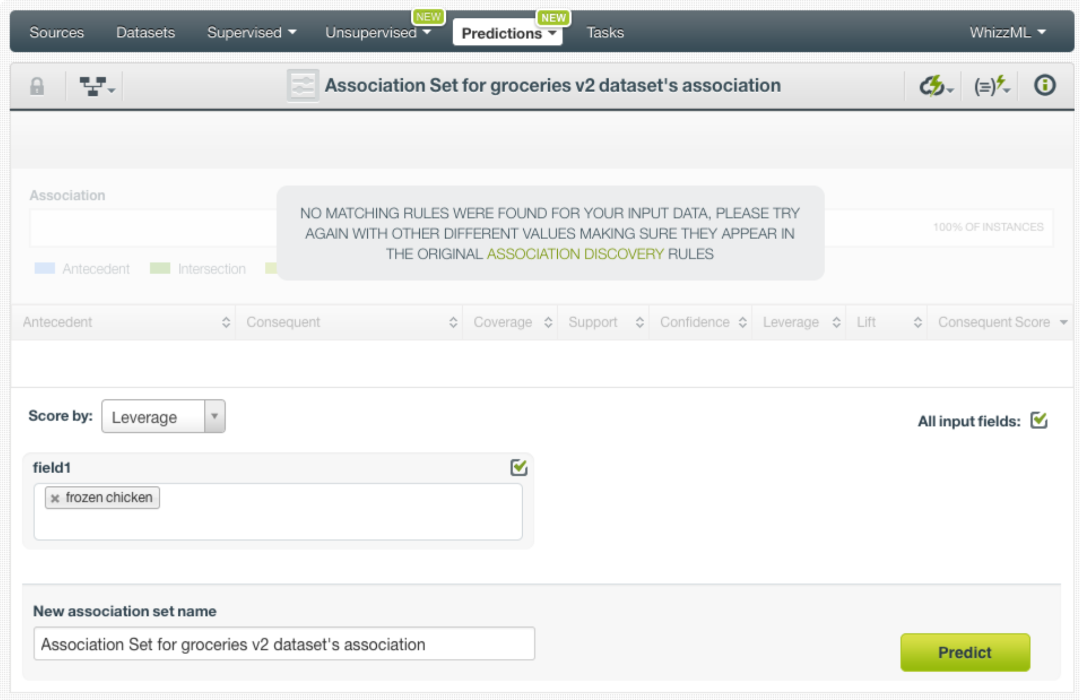


Figure 8.12: Unable to find matching rules for the given input data

- If the original association model contains only **categorical** and **numeric** fields and you set values for **all** of them. This is due to the fact that categorical and numeric fields only have one single value per instance. So if they are given as input fields, they cannot be returned as predicted items at the same time. For example, if you already set age=20 as input, BigML will not return the age as output since a person cannot have two different ages at the same time. In those cases the following warning message will be displayed:

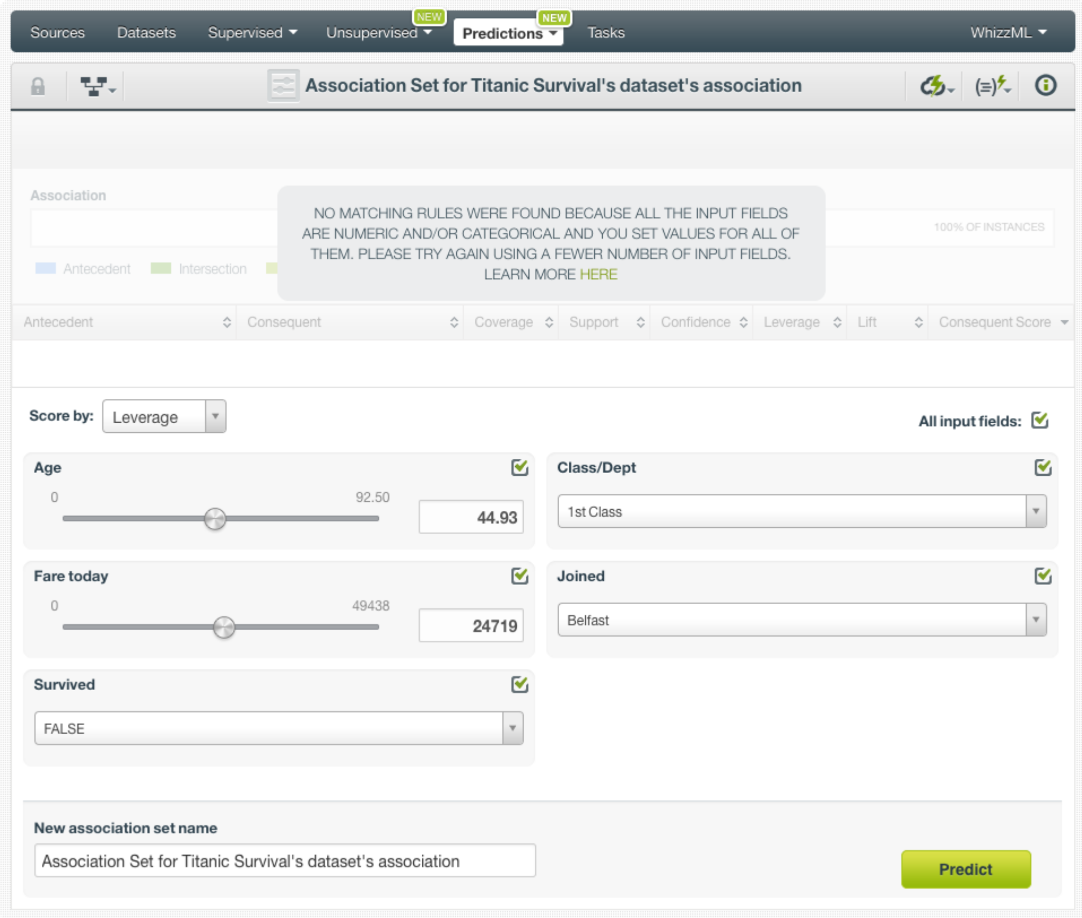


Figure 8.13: Unable to find matching rules because all the fields are set as inputs

BigML provides two different views for your predicted items, the **table** and the **diagrams** explained below.

8.4.1 Association Set Table

The table contains the **rules** from the original association model which **antecedent** matches the input data. (See Figure 8.14.)

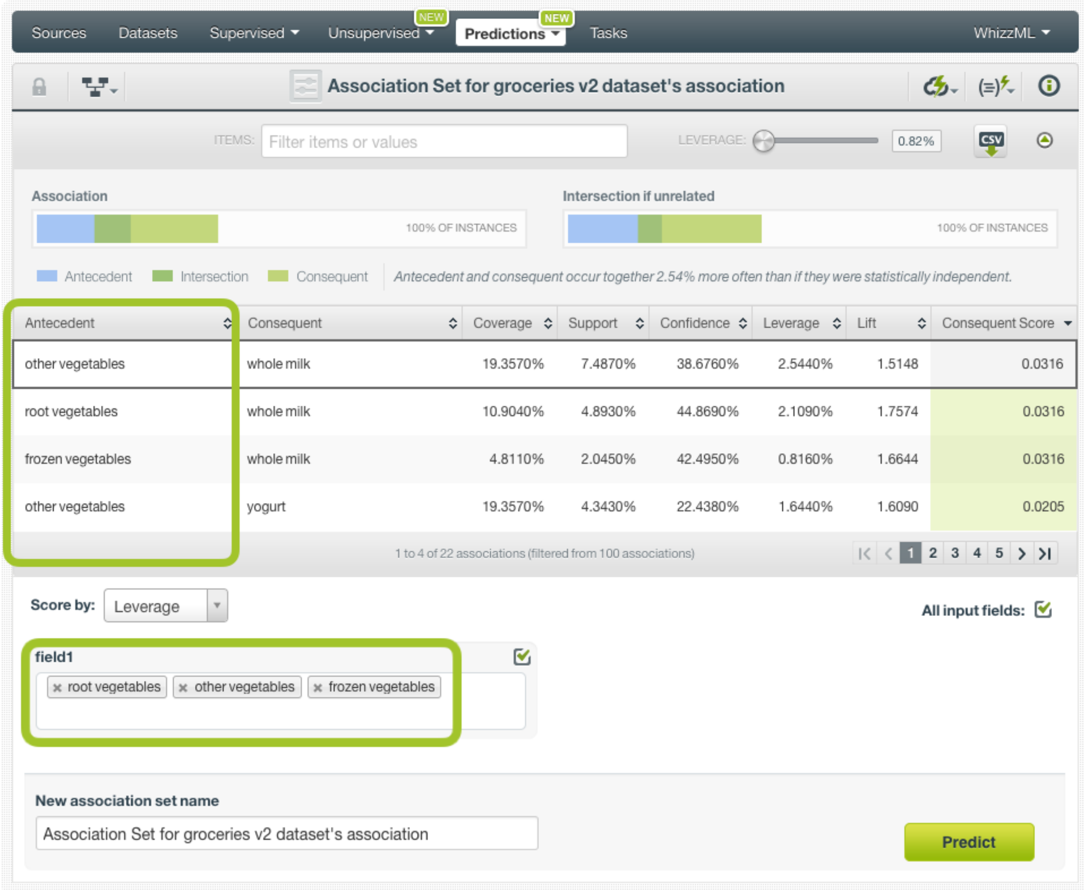


Figure 8.14: Match between the predicted rule's antecedent and the input data

The **consequent** part of the rules contains the predicted items associated to their **score**. (See Figure 8.15.) See Section 8.3 for a full explanation of the consequent score.

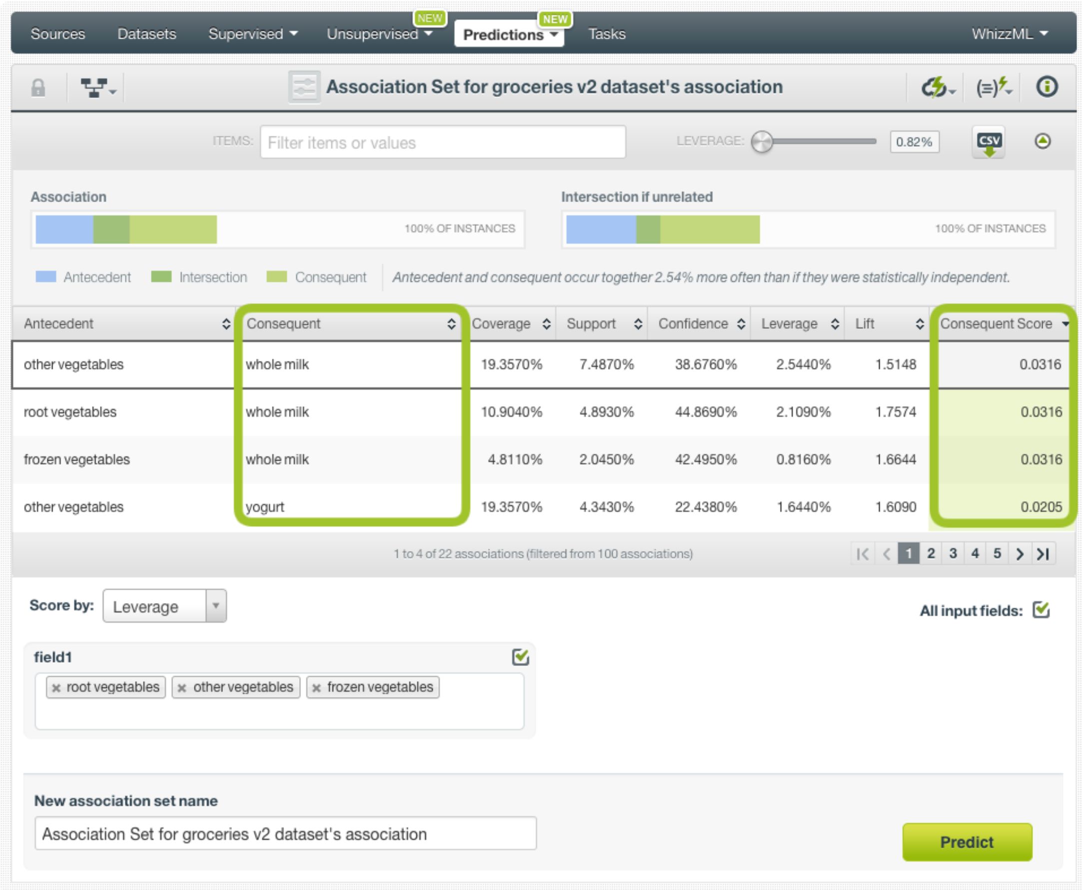


Figure 8.15: Predicted items and their similarity-weighted score

The table also contains the **measures** for each of the matching rules: coverage, support, confidence, leverage or lift. (See Figure 8.16.) See Section 2.1 for a detailed explanation of each measure.

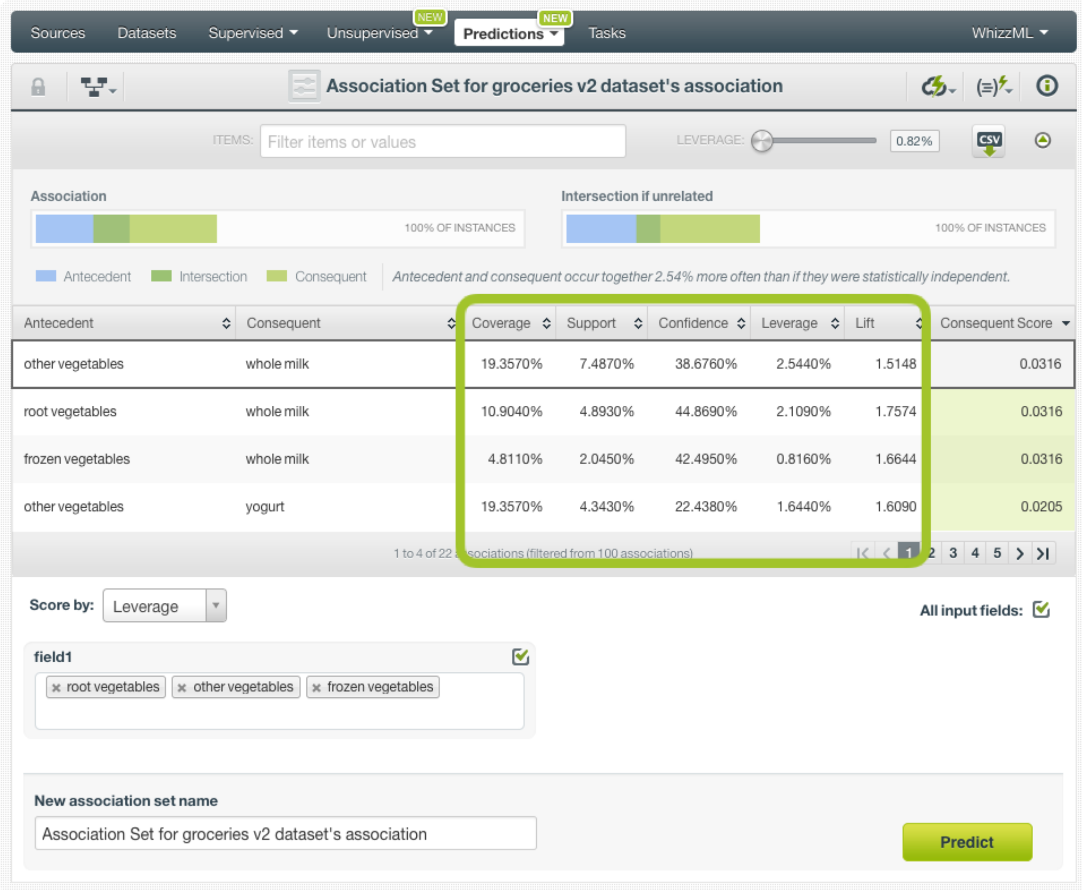


Figure 8.16: Predicted rules measures

BigML displays up to four rules in the same view. To view more rules, use the **pagination** options at the bottom of the table. (See Figure 8.17.)

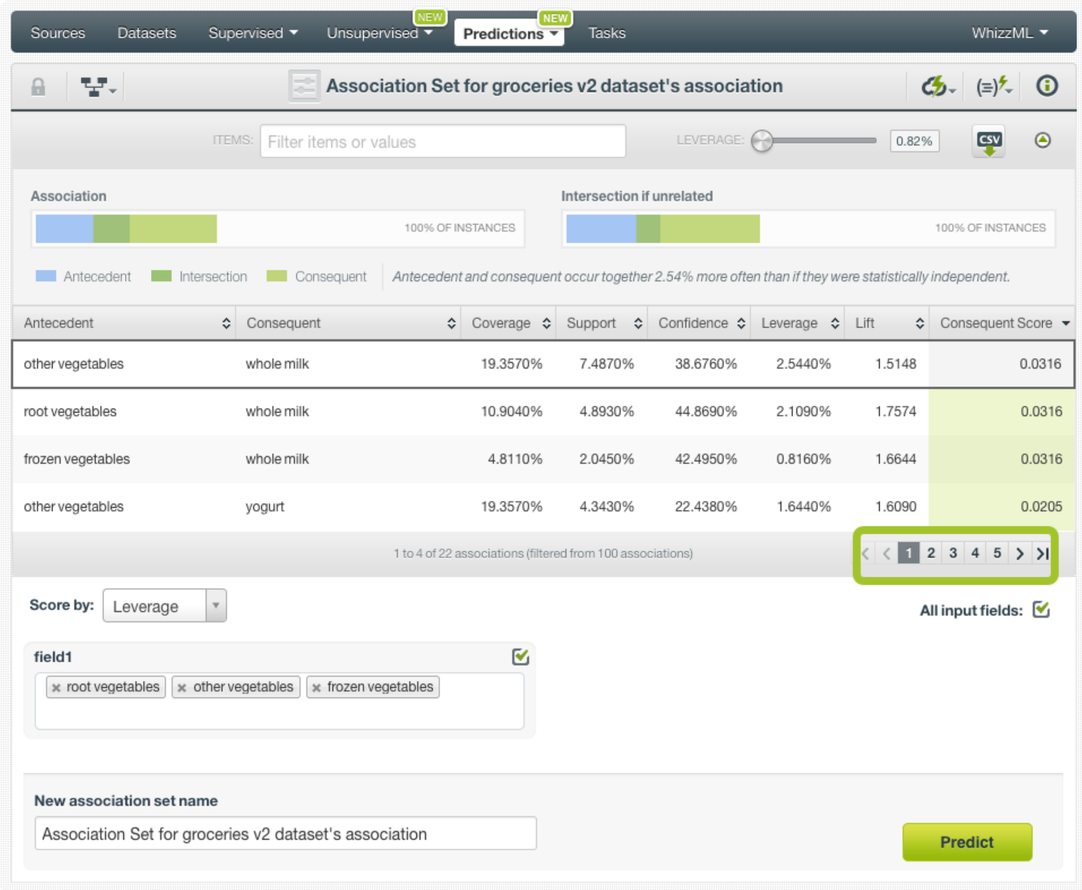


Figure 8.17: Predicted rules pagination

You can also **filter** the rules by typing any item or field name within the search box or using the measure slider. (See Figure 8.18.)

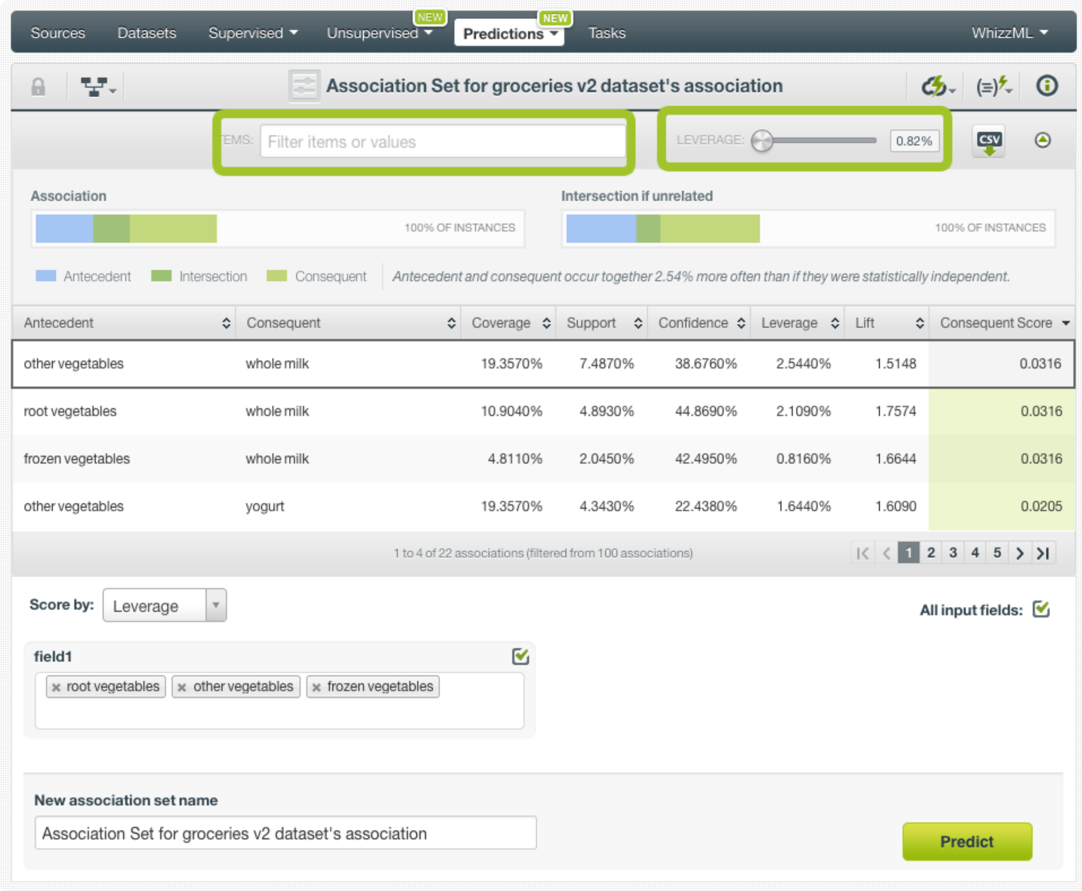


Figure 8.18: Filter the predicted rules table

Finally, **export** the table in **CSV format** by clicking on the option show in Figure 8.19

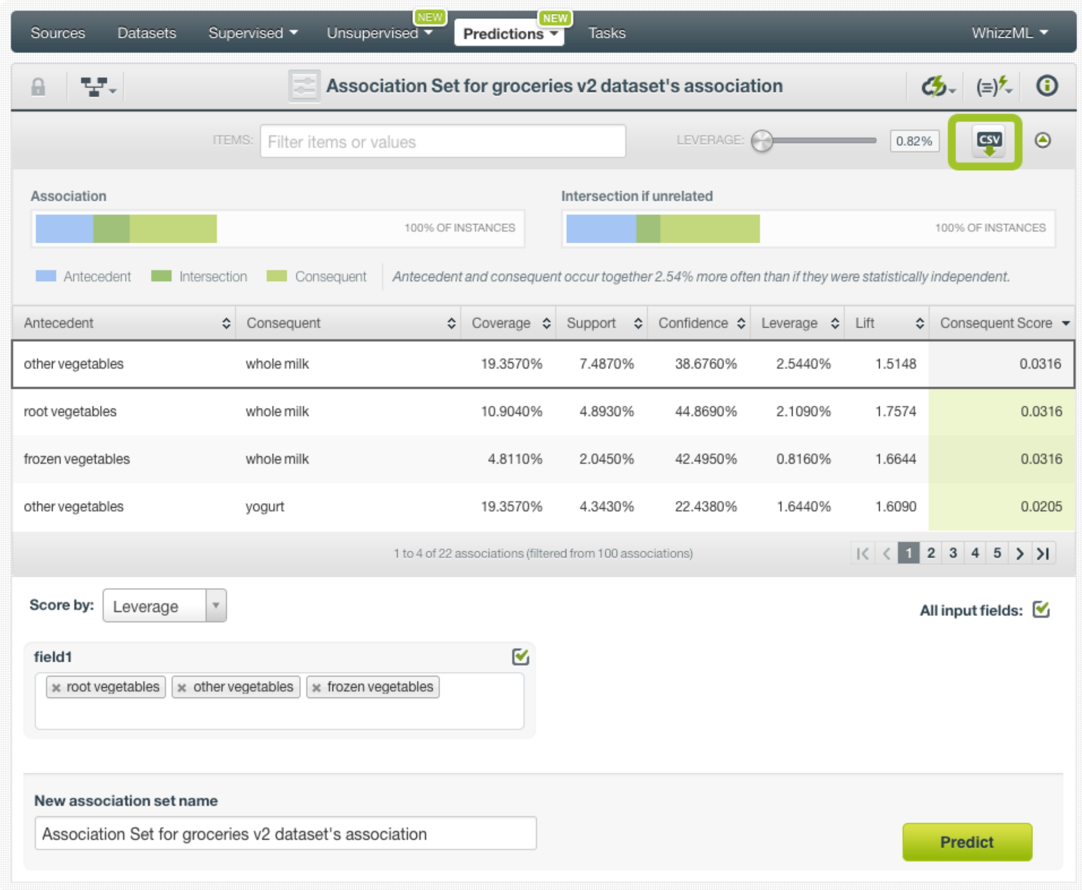


Figure 8.19: Export the predicted rules in CSV file

8.4.2 Association Set Diagrams

When you select a rule from the table, you will see two Venn diagrams displayed above the table. The **association diagram** on the left indicates the actual intersection between the antecedent and consequent itemsets of this rule, and the **intersection if unrelated diagram** on the right indicates the intersection if both itemsets were independent. These diagrams provide a visual overview of the importance of the selected association rule. See Section 5.1 for a full explanation.

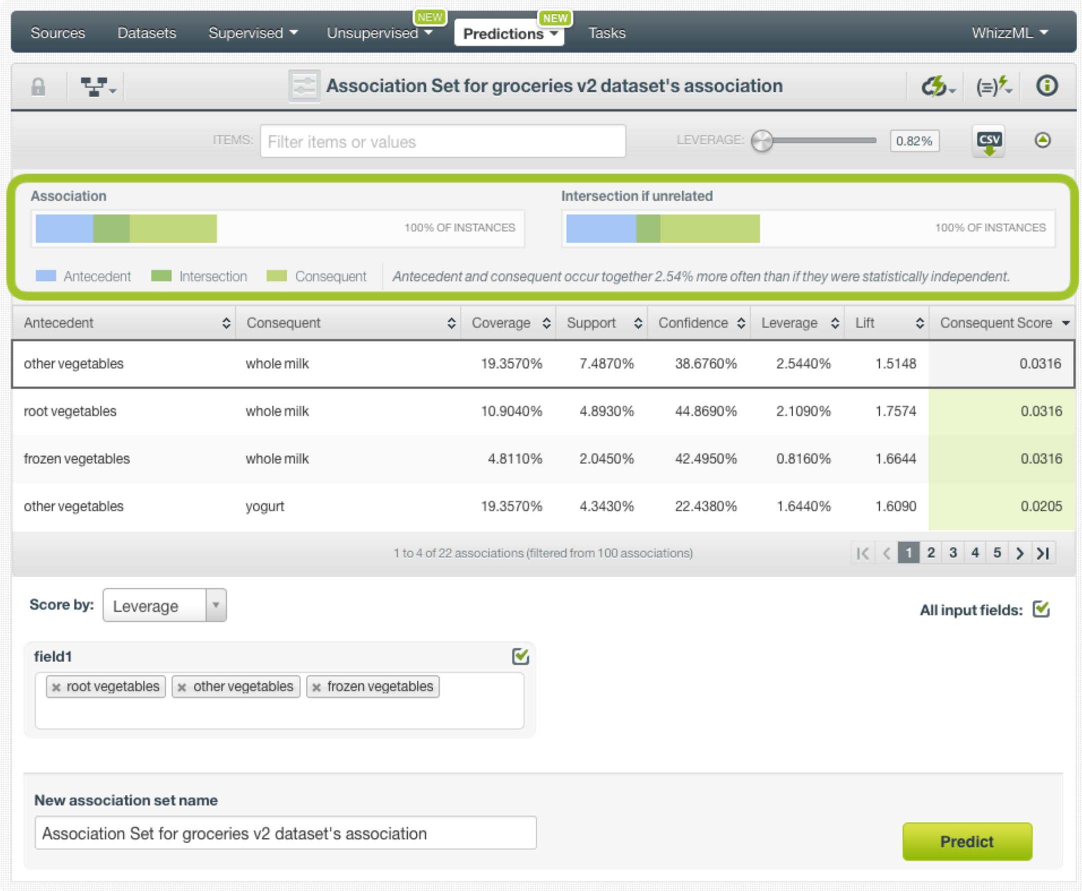


Figure 8.20: Predicted rules diagrams

You can hide or show this view by clicking in the corresponding option. (See Figure 8.21.)

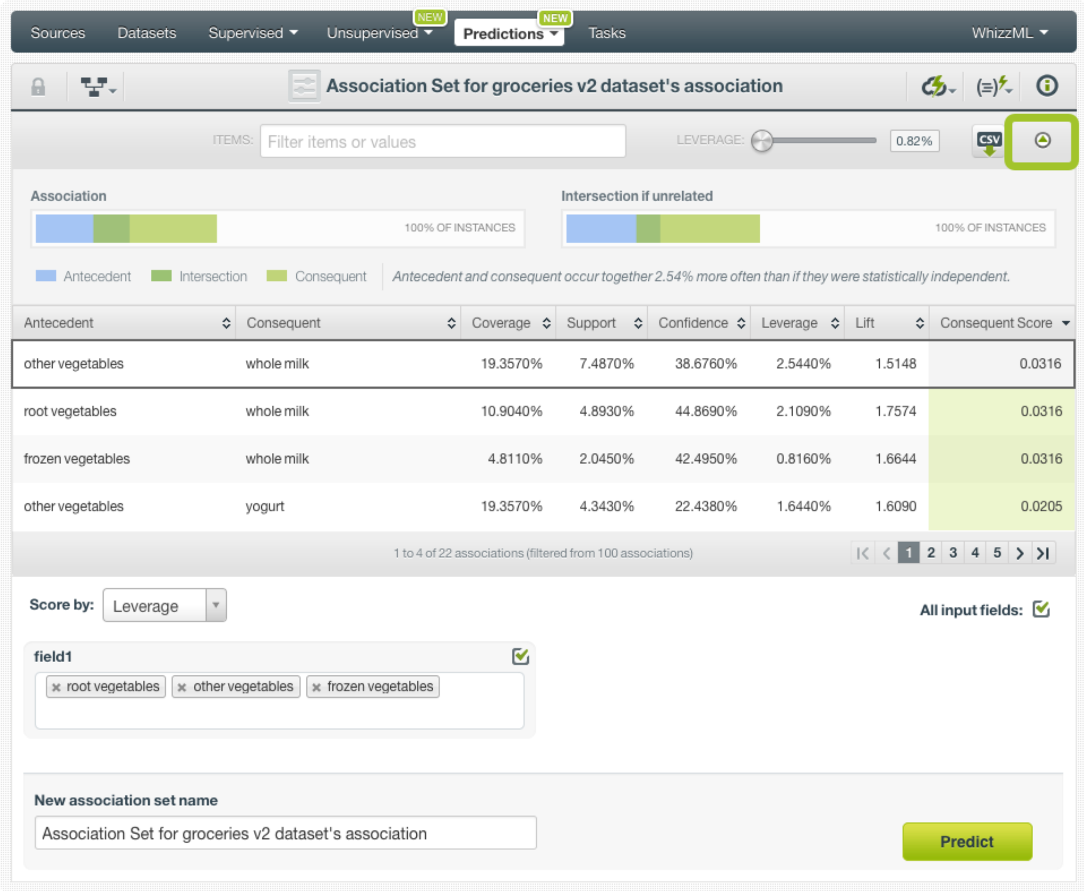


Figure 8.21: Show and hide diagrams

8.5 Consuming Association Sets

You can perform all the association set actions explained in this document such as creating, retrieving, listing, updating, and deleting association sets via the BigML API.

The example below shows how to create an association set with the definition of the input data after the BIGML_AUTH environment variable, which was previously set with your authentication credentials:

```
curl "https://bigml.io/associationset?$BIGML_AUTH" \
     -X POST \
     -H 'content-type: application/json' \
     -d '{"association": "association/5423625af0a5ea3eea000028",
          "input_data": ["oranges", "apples"}'
```

For more information on using association sets through the BigML API, please refer to association sets REST API documentation[2].

8.6 Descriptive Information

Each association set has an associated **name**, **description**, **category** and **tags**. Those options are editable through the MORE INFO menu on the top right of the association set view. (See Figure 8.22.)

[2] https://bigml.com/api/associationsets

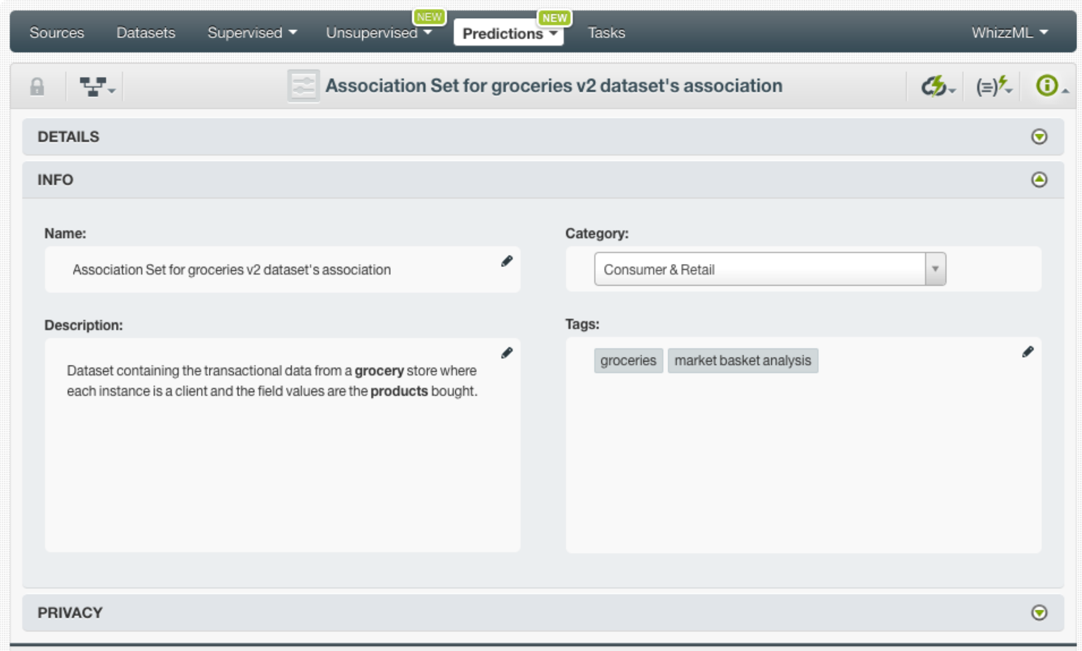


Figure 8.22: Edit association sets metadata from More info panel

8.6.1 Association Set Name

If you do not specify a **name** for your association sets, BigML assigns a default name. The name always follow the structure "Association Set for <association name>".

Association set names are displayed on the list view and also on the top bar of the association set view. Association set names are indexed to be used in searches. You can rename your association sets at any time from the MORE INFO menu option.

The name cannot be longer than **256** characters. More than one association set can have the same name even within the same project, but they will always have different identifiers.

8.6.2 Description

Each association set also has a **description** that it is very useful for documenting your Machine Learning projects. Association sets take the description from the association used to create them.

Descriptions can be written using plain text and also markdown[3]. BigML provides a simple markdown editor that accepts a subset of markdown syntax. (See Figure 8.23.)

[3] https://en.wikipedia.org/wiki/Markdown

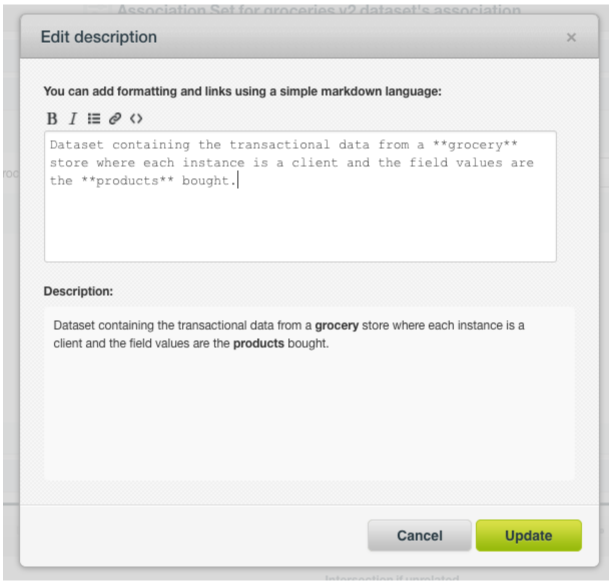


Figure 8.23: Markdown editor for association set descriptions

Descriptions cannot be longer than **8192** characters and can use almost any character.

8.6.3 Category

Each association set has associated a **category**. Categories are useful to classify association sets according to the domain which your data comes from. This is useful when you use BigML to solve problems across industries or multiple customers. By default, association sets take the category from the association used to create them.

An association set category must be one of the categories listed on Table 8.1.

Table 8.1: Categories used to classify association sets by BigML

Category
Aerospace and Defense
Automotive, Engineering and Manufacturing
Banking and Finance
Chemical and Pharmaceutical
Consumer and Retail
Demographics and Surveys
Energy, Oil and Gas
Fraud and Crime
Healthcare
Higher Education and Scientific Research
Human Resources and Psychology
Insurance
Law and Order
Media, Marketing and Advertising
Miscellaneous
Physical, Earth and Life Sciences
Professional Services
Public Sector and Nonprofit
Sports and Games
Technology and Communications
Transportation and Logistics
Travel and Leisure
Uncategorized
Utilities

8.6.4 Tags

An association set can also have a number of **tags** associated with it that can help to retrieve it via the BigML API or to provide some extra information. An association set inherits the tags from the association used to create it.

Each tag is limited to a maximum of 128 characters. Each association set can have up to 32 different tags.

8.7 Association Set Privacy

The link displayed in the **privacy panel** is the private URL of your association set, so only a user logged into your account is able to see it. Association sets cannot be shared from the BigML Dashboard by sharing a link as you can with other resources. (See Figure 8.24.)

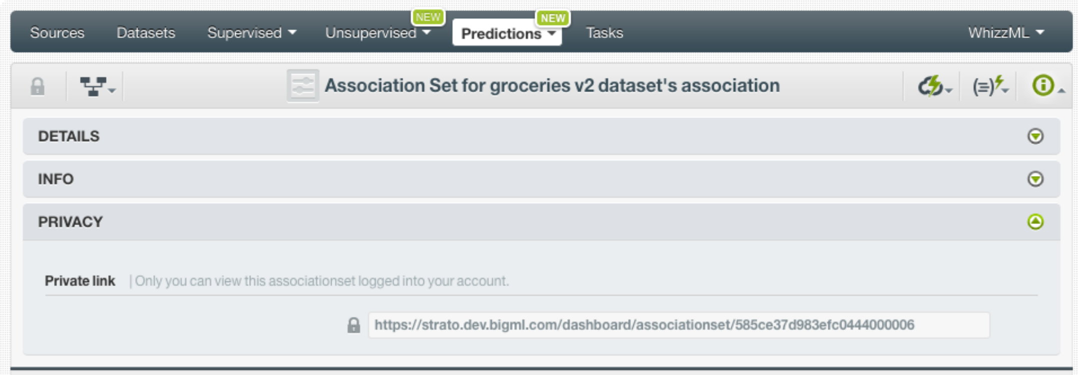


Figure 8.24: Private link of an association set

8.8 Moving Association Sets

When you create an association set it will be assigned to the same project where the original association is located. You cannot move association sets between projects as you do with other resources.

8.9 Deleting Association Sets

You can delete your association sets by clicking on the DELETE ASSOCIATION SET option in the **1-click action menu** (see Figure 8.25).

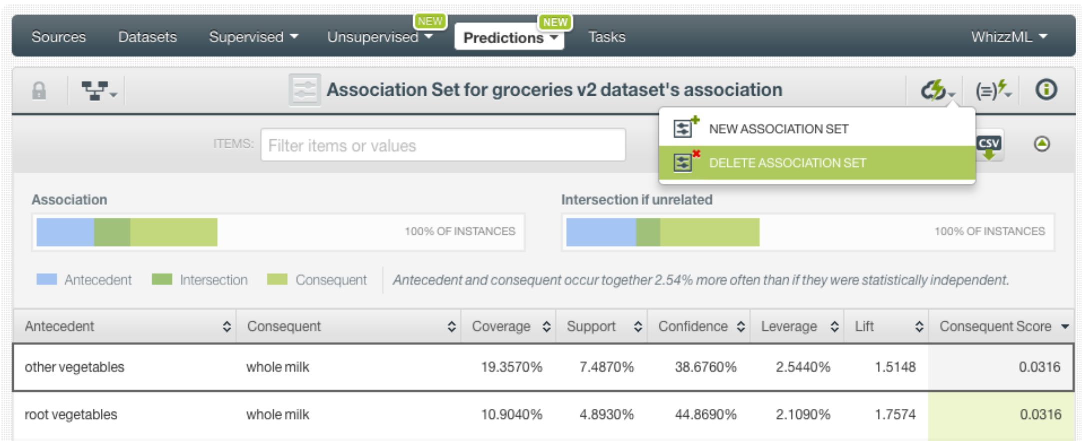


Figure 8.25: Delete batch association set from the 1-click menu

Alternatively, you can click the DELETE ASSOCIATION SET in the **pop up menu** from the list view (see Figure 8.26.)

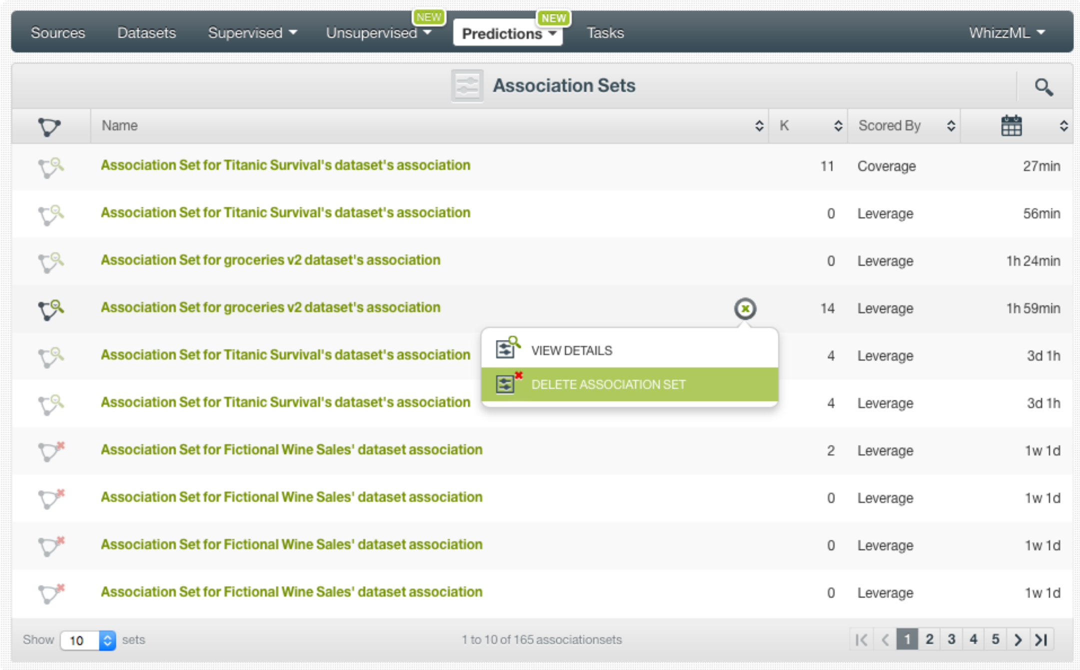


Figure 8.26: Delete association set from pop up menu

A modal window will be displayed asking you for confirmation. Once an association set is deleted, it is permanently deleted and there is no way you (or even the IT folks at BigML) can retrieve it. (Figure 8.27.)

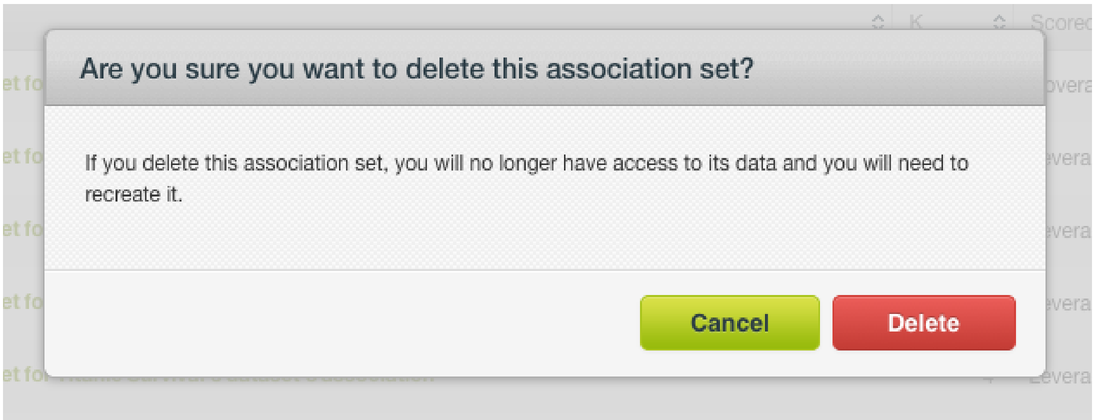


Figure 8.27: Delete association set confirmation

CHAPTER 9

Consuming Associations

Similarly to other resources in BigML, you can **download** associations to your local environment. You can also create and consume your associations programmatically via the **BigML API** and the **BigML bindings**. The following subsections explain these three options.

9.1 Exporting and Downloading Associations

Export your associations table as a CSV file from the ASSOCIATION REPORT menu option. (See Figure 9.1.)

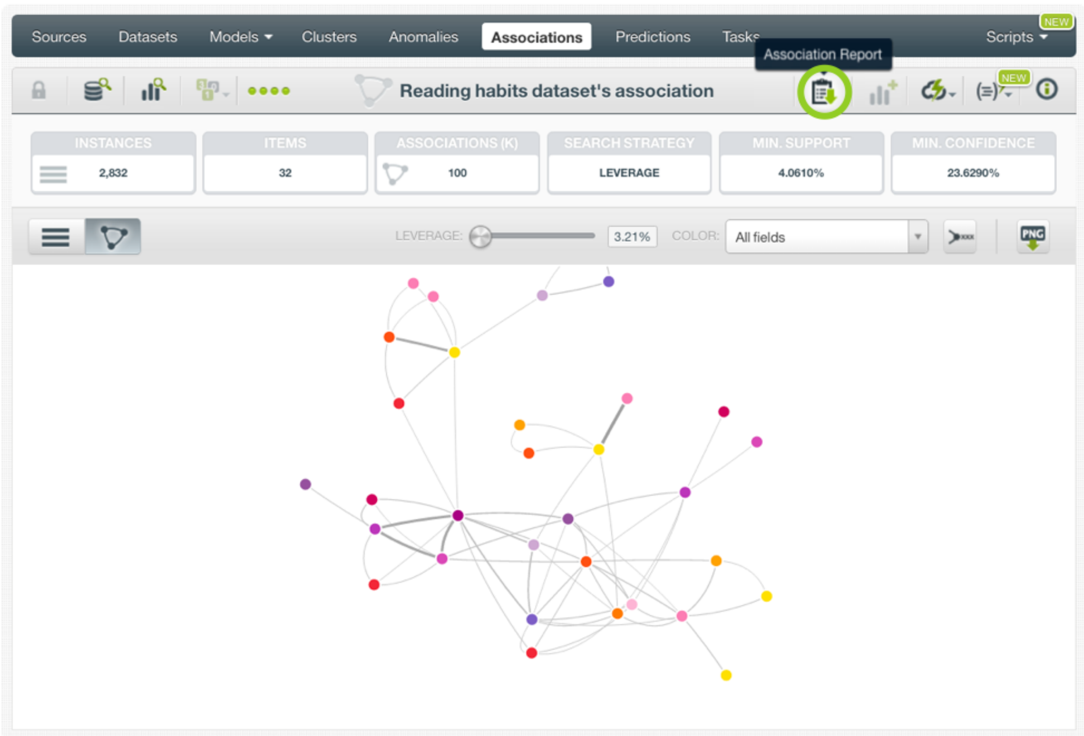


Figure 9.1: Associations report menu option

A modal window will display with the **Download as CSV** link. (See Figure 9.2)

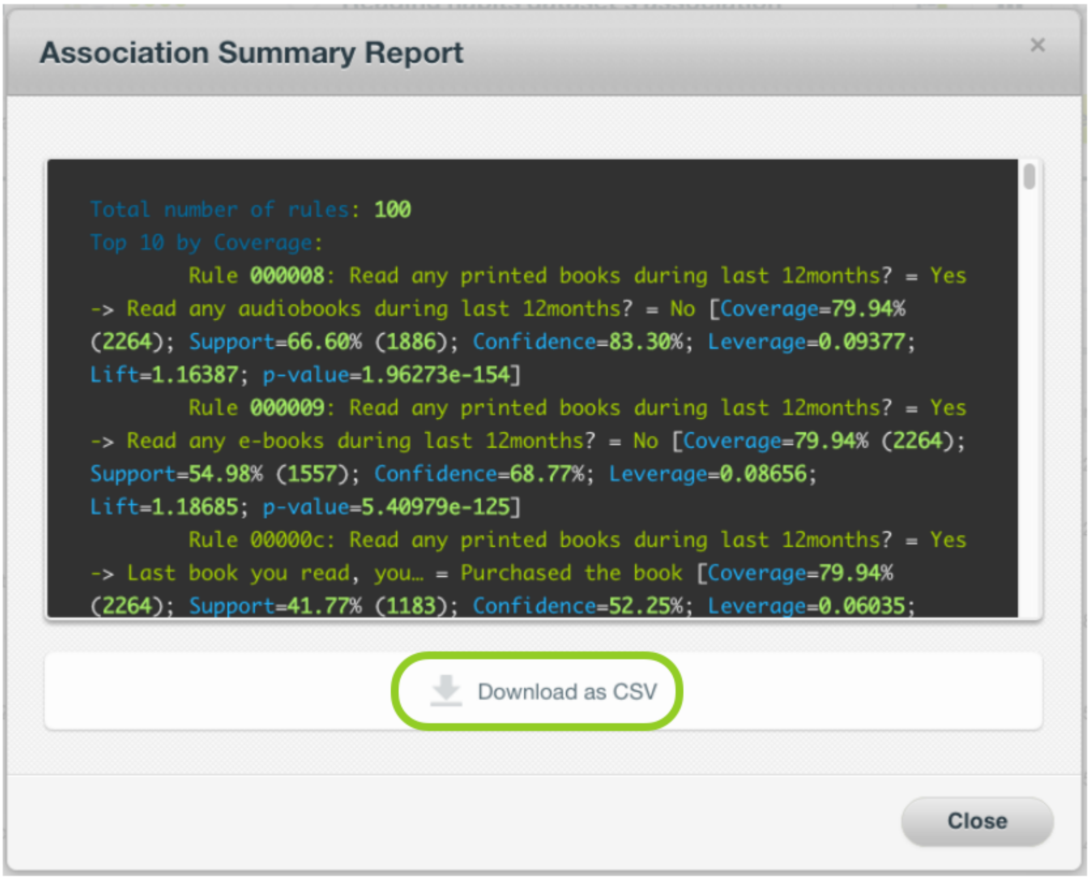


Figure 9.2: Download association rules in a CSV file

You can also download your association visualization as a PNG image. From the association view, click the highlighted PNG in Figure 9.3.

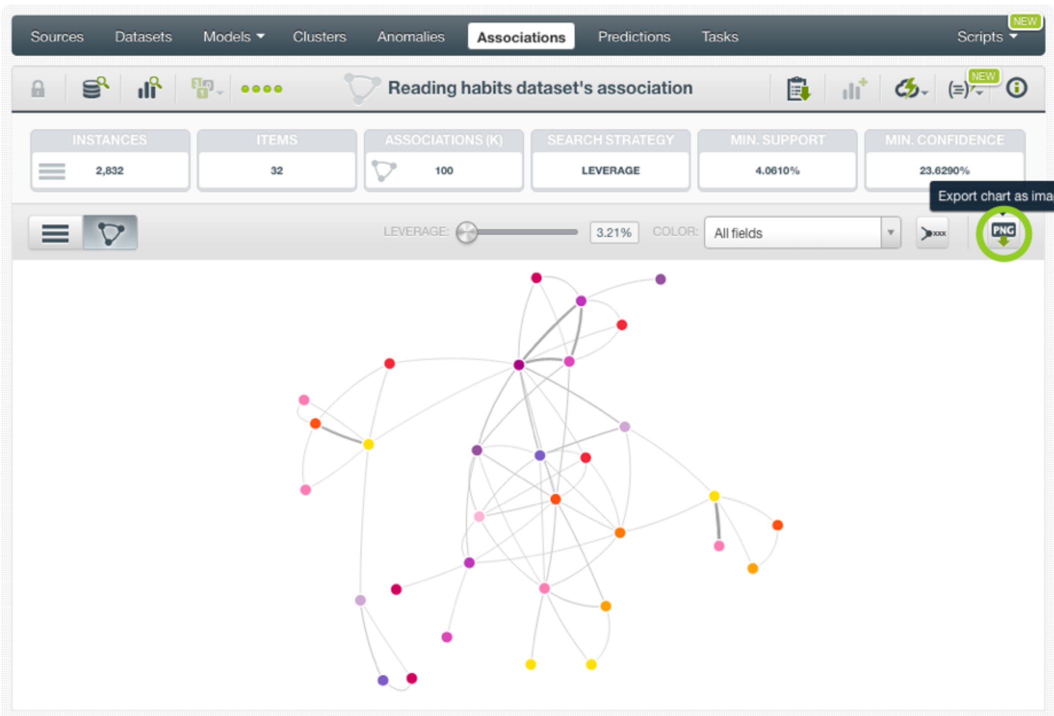


Figure 9.3: Export network chart as image

9.2 Using Associations Via the BigML API

Associations have full citizenship in the **BigML API**, which allows you to programmatically create, update, list, and delete them.

See in the example below how you can create an association after you properly set the BIGML_AUTH environment variable, which was previously set with your authentication credentials:

```
curl "https://bigml.io/association?$BIGML_AUTH" \
     -X POST \
     -H 'content-type: application/json' \
     -d '{"dataset": "dataset/4f66a80803ce8940c5000006"}'
```

Each association has a unique identifier in the form of “association/ID”, where ID is a string of 24 alpha-numeric characters that you can use to retrieve and further manipulate the association programatically. For more information on using associations through the BigML API, please refer to associations REST API documentation[1].

9.3 Using Associations Via the BigML Bindings

You can also create and use associations via the **BigML bindings**, which are libraries aimed to make it easier to use the BigML REST API from your language of choice. BigML offers bindings for a number of languages, including: Python, Node.js, Java, Swift or Objective-C. See below an example to create a dataset with the Python bindings:

```
from bigml.api import BigML
api = BigML()
association = api.create_association('dataset/4f66a80803ce8940c5000006')
```

For more information on using associations through the BigML bindings, please refer to the BigML bindings page[2].

[1] https://bigml.com/api/associations
[2] https://bigml.com/tools/bindings

CHAPTER 10

Associations Limits

BigML imposes a few limits to the characteristics of associations that it can handle:

- **Fields**: there is no enforced limit to the number of fields that can be present in an association.
- **Instances**: there is no enforced limit to the number of instances that can be handled.
- **Total Associations**: a maximum of 500 associations are allowed on the BigML Dashboard. There is no enforced limit when using the BigML API.
- **Items in Antecedent**: a maximum of 10 antecedent items are permitted on the BigML Dashboard. There is no enforced limit when using the BigML API.
- **Total items**: a maximum of 10,000 total different items in your dataset is permitted.

CHAPTER 11

Descriptive Information

Each association model has an associated **name**, **description**, **category**, and **tags**. A brief description follows for each concept. From the association view, the MORE INFO menu option lets you edit this metadata. (See Figure 11.1.)

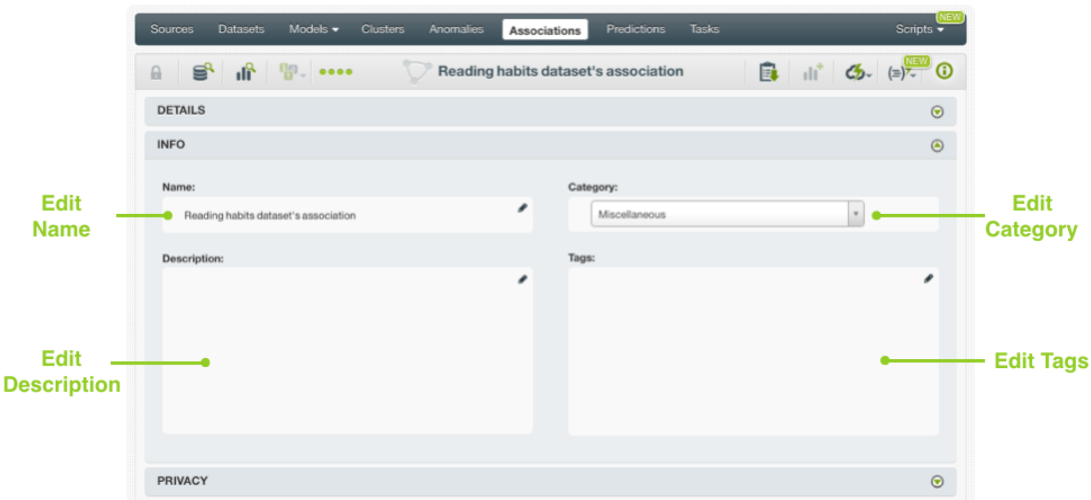


Figure 11.1: Panel to edit an association's name, category, description and tags

11.1 Association Name

Each association has a **name** that is displayed in the associations list view and also on the top bar of the association view. Association names are indexed to be used in searches. When you create an association, it gets a default name. Change it using the MORE INFO menu option (see Figure 11.1). The name of an association cannot be longer than **256** characters. There is no restriction on the characters that can be used in an association name. More than one association can have the same name, even within the same project. They will always have different identifiers.

11.2 Description

Each association also has a description that is useful for documenting your Machine Learning projects. Descriptions can be written using plain text and also markdown[1]. BigML provides a simple markdown editor that accepts a subset of markdown syntax. (See Figure 11.2.)

[1] https://en.wikipedia.org/wiki/Markdown

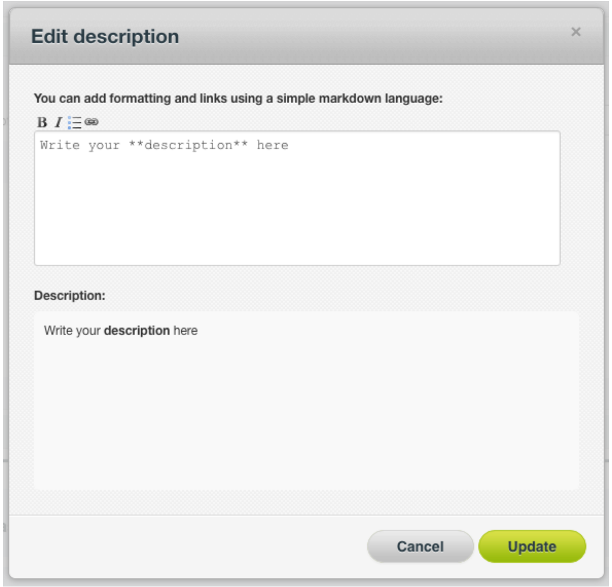


Figure 11.2: Markdown editor for the association description

Descriptions cannot be longer than **8192** characters and can use almost any character.

11.3 Category

Each association has a category. Categories are useful to classify associations according to the domain which your data comes from. This is useful when you use BigML to solve problems across industries or multiple customers.

An association category must be one of the categories listed in table Table 11.1.

Category
Aerospace and Defense
Automotive, Engineering and Manufacturing
Banking and Finance
Chemical and Pharmaceutical
Consumer and Retail
Demographics and Surveys
Energy, Oil and Gas
Fraud and Crime
Healthcare
Higher Education and Scientific Research
Human Resources and Psychology
Insurance
Law and Order
Media, Marketing and Advertising
Miscellaneous
Physical, Earth and Life Sciences
Professional Services
Public Sector and Nonprofit
Sports and Games
Technology and Communications
Transportation and Logistics
Travel and Leisure
Uncategorized
Utilities

Table 11.1: Categories used to classify associations by BigML

11.4 Tags

An association can also have a number of **tags** associated with it that can help in retrieving it via the BigML API or in providing associations with some extra information. Each tag is limited to a maximum of **128** characters. Each association can have up to **32** different tags.

11.5 Association Privacy

Privacy options for an association can be defined in the MORE INFO menu option, displayed in Figure 11.3. There are two **levels of privacy** for the BigML associations:

- **Private**: only accessible by authorized users.
- **Shared**: accessible by any user with whom the owner shares a secret link. You can choose to share your associations by enabling the **secret link** from the information panel. (See Figure 11.3.) The first one is a **sharing link** that you can copy and send to others so they can visualize and interact with your association model. The second one is **a link to embed** your association model directly on your web page.

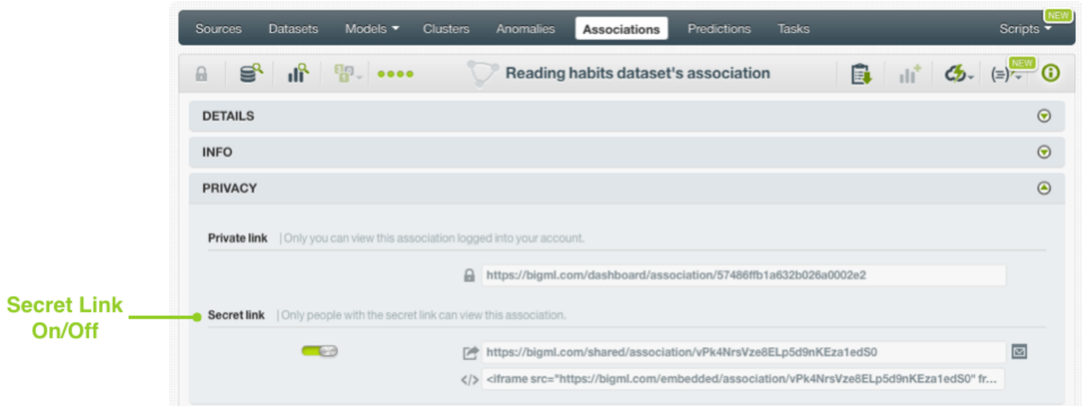


Figure 11.3: Share your association

CHAPTER 12

Moving Associations to Another Project

By default, when you create an association, it will be assigned to the same project as the dataset used to create the association. If you did not assign any project to the dataset used to create your association, the new association will not be assigned to any project, and it will be shown when the project selector bar shows "**All**", as seen in Figure 12.1.

Associations can only be assigned to a single project. However, you can move associations between projects. The menu option to do this can be found in two places:

1. In the **association view**, within the **1-click action menu**. (See Figure 12.1.)

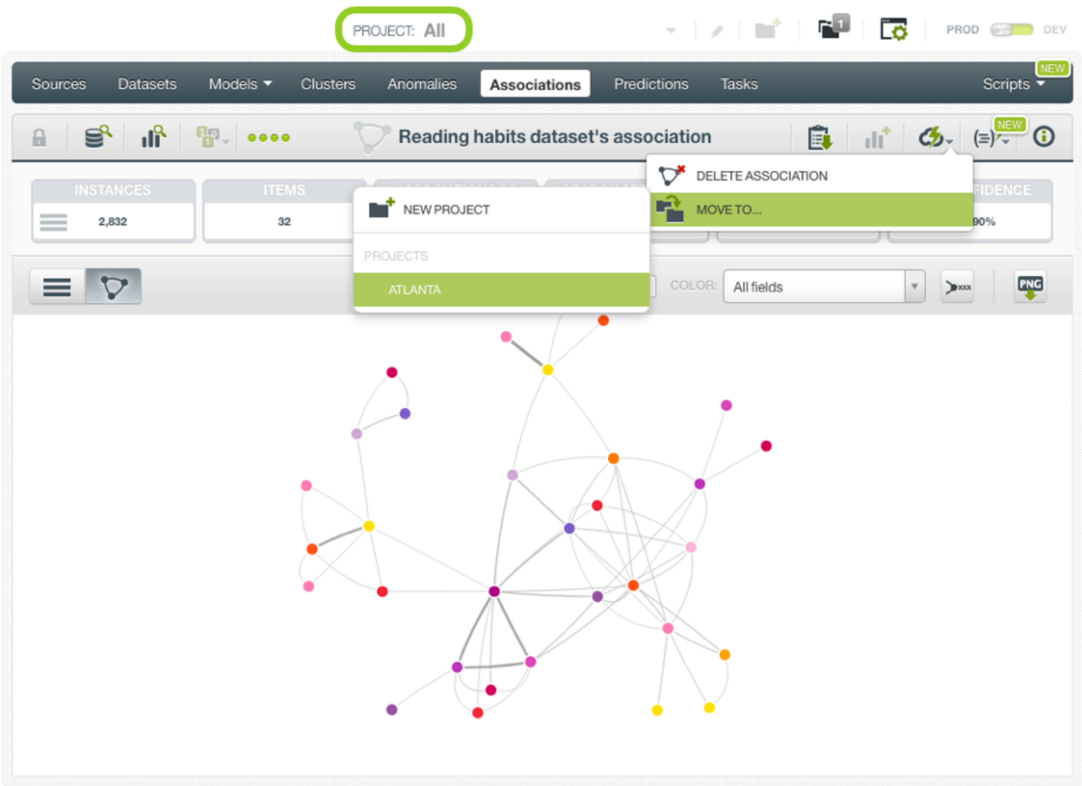


Figure 12.1: Menu option to move associations fromt he 1-click action menu

2. In the **association list view**, within the **pop up menu**. (See Figure 12.2.)

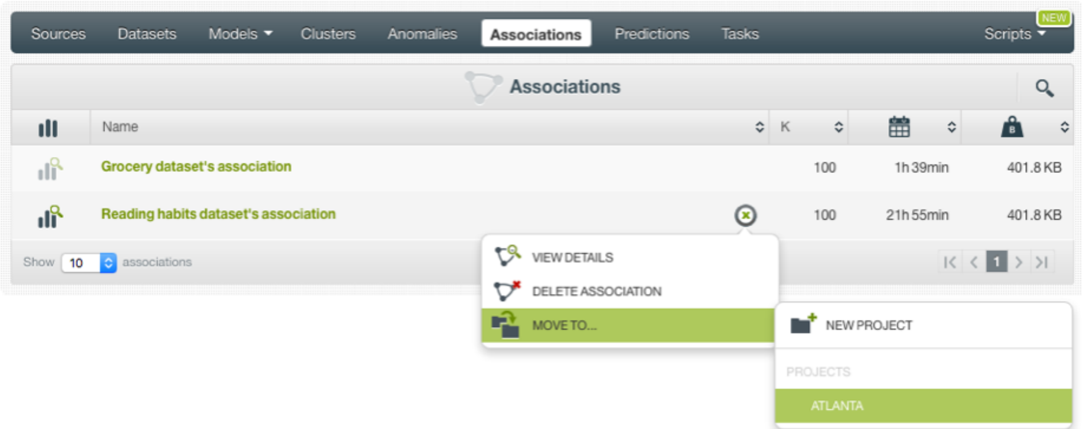


Figure 12.2: Menu option to move associations from the pop up menu

CHAPTER 13

Stopping Association Creation

BigML lets you stop an association creation before the task is finished. You can do this in two ways:

1. Select DELETE ASSOCIATION from the **1-click action menu** on the **association view** while BigML is processing your request. (See Figure 13.1.)

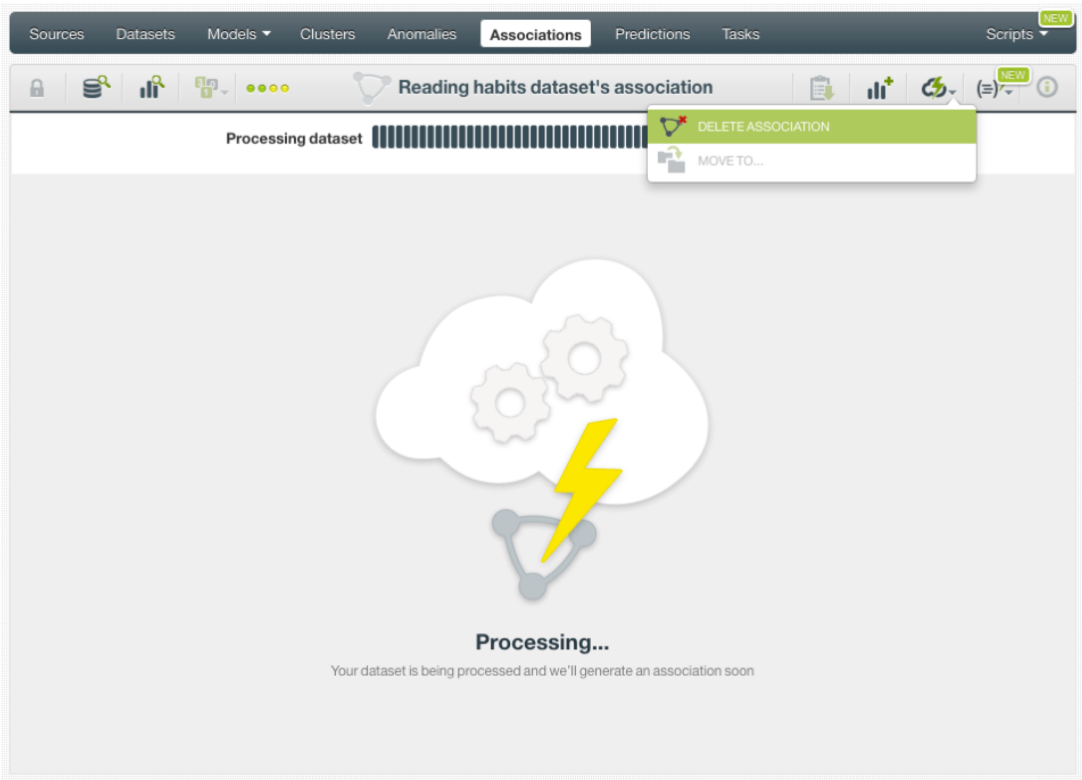


Figure 13.1: Stop the association creation from the 1-click action menu options

2. Or select DELETE ASSOCIATION from the **pop up menu** on the **association list view**. (See Figure 13.2.)

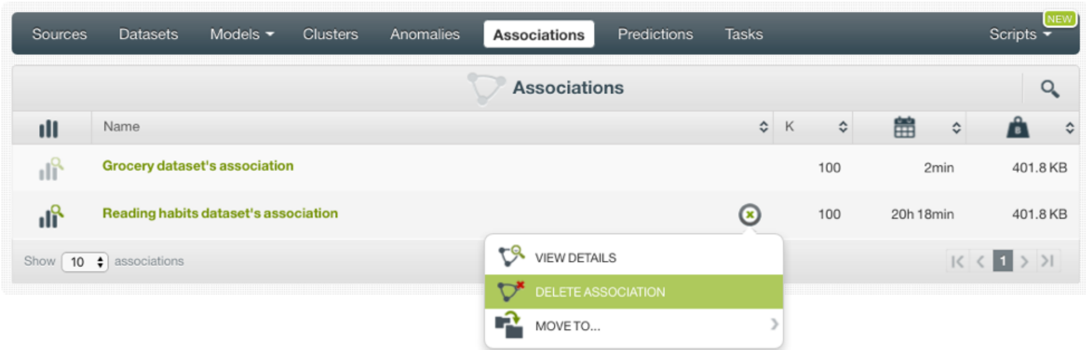


Figure 13.2: Stop the association creation from the pop up menu

In both cases, a modal window (see Figure 13.3) will be displayed asking you for confirmation.

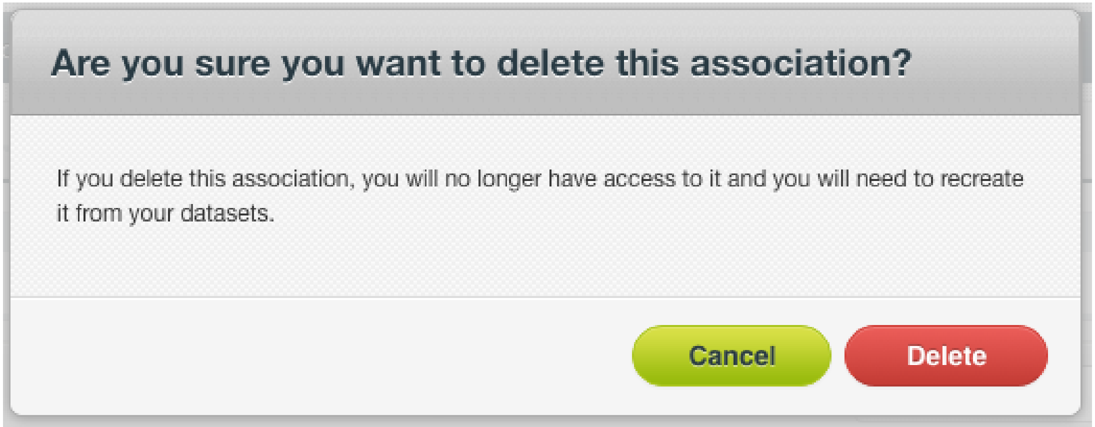


Figure 13.3: Confirmation window to stop the association creation

The next section describes how to delete datasets once they have been created.

CHAPTER 14

Deleting Associations

If you no longer need your associations, BigML lets you delete them permanently. You can delete your associations in two ways:

1. From the association view, using the **1-click action menu**, and selecting DELETE ASSOCIATION. (See Figure 14.1.)

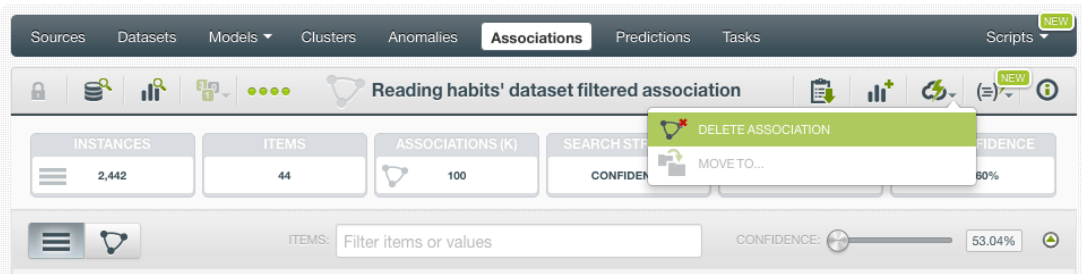


Figure 14.1: Delete an association from the 1-click menu

2. From the association list view, using the **pop up menu**, and selecting DELETE ASSOCIATION. (See Figure 14.2.)

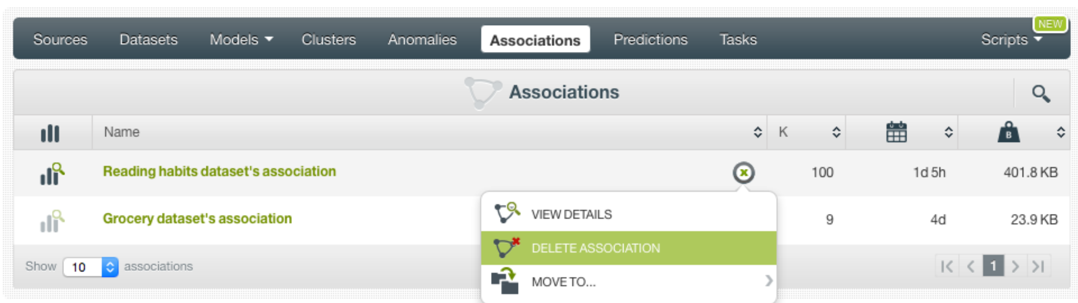


Figure 14.2: Delete an association from the pop up menu

In both cases, a modal window (Figure 14.3) will be displayed asking you for confirmation. Once you delete an association, it is deleted permanently, and there is no way you (or even the IT folks at BigML) can retrieve it.

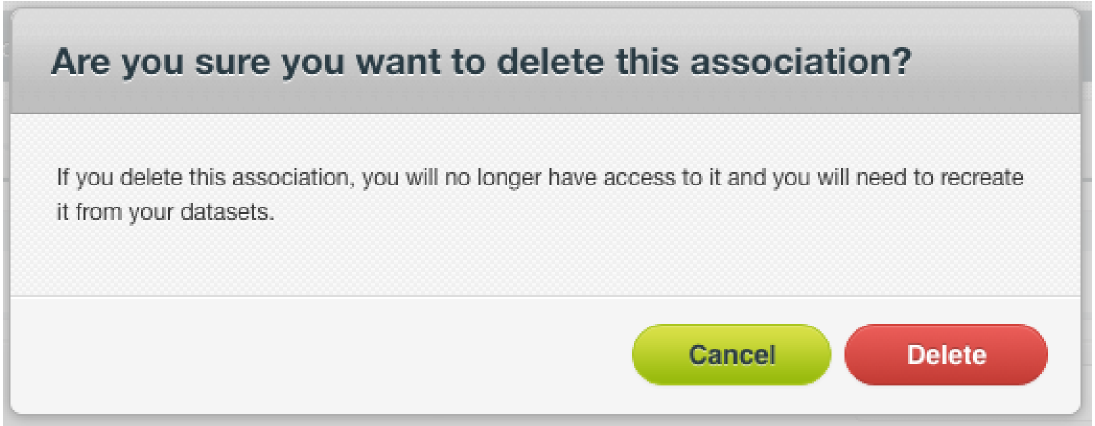


Figure 14.3: Association deletion modal window

CHAPTER 15

Takeaways

This document explains associations in detail. We finish it with a list of key points:

- Association Discovery (or associations) finds **meaningful relationships** among fields and their values in high-dimensional datasets, whereas statistical techniques focus on controlling the risk of making false discoveries.
- Associations output is easily expressed as rules that can be understood by non-experts.
- You can create associations from datasets that have been created in BigML, and then create a new dataset from the association rules that you discover. (See Figure 15.1.)
- Associations require the data to be structured in a specific way, using the **items** field type.
- You can create an association with just **1-click or configure** it as you wish.
- There is no single measure (support, coverage, confidence, leverage, or lift) that is always more important than others. This will depend on your main goals.
- You can set minimum levels for a number of association measures that let you focus on more interesting association rules, while filtering out potentially spurious ones.
- You can control multiple interestingness measures, yet easy to tune without having to configure difficult to comprehend parameters.
- You can easily discretize your numeric fields to transform them into categorical fields.
- BigML lets you create associations for a sample of your dataset.
- After associations are created, you will get a **table** that summarizes all the rules discovered, and you can visualize these rules in a **network chart**.
- You can download your association rules in a CSV file, and export the network chart as an image.
- You can programmatically create, list, delete, and use your associations through the BigML API and the BigML bindings.
- You can furnish your associations with **descriptive information** (name, description, tags, and category).
- You can stop an association creation before the task is finished.
- You can permanently delete an association.

Figure 15.1: Association Workflow

Please use the noidx option in the documentclass invocation.

List of Figures

List of Tables

Glossary

Antecedent the left-hand-side itemset of an association rule. 9, 18, 30, 32

Association Discovery an unsupervised Machine Learning task to find out relationships between values in high-dimensional datasets. It is commonly used for market basket analysis. ii, 1, 62

Confidence an indicator of the prediction's certainty for classification models and ensembles. It takes into account the class distribution and the number of instances at a certain node. It is a value between 0% and 100%. 31, 36

Confidence (Associations) the percentage of instances which contain the consequent and antecedent together over the number of instances which only contain the antecedent. 9, 62

Consequent the right-hand-side itemset of an association rule. 9, 18, 32

Coverage the support of the antecedent of an association rule, i.e., the portion of instances in the dataset which contain the antecedent itemset. 9, 20, 31, 36, 62

Dashboard The BigML web-based interface that helps you privately navigate, visualize, and interact with your modeling resources. ii

Discretization the process of transforming a numeric field into a categorical field. 13

Leverage the difference between the probability of the rule and the expected probability if the items were statistically independent. 9, 31, 36, 62

Lift how many times more often antecedent and consequent occur together than expected if they were statistically independent. 9, 31, 36, 62

Predictive Model a machine-learned model that has been created using statistical learning. It can help describe or infer some statistical properties of an entity using the instances provided by a dataset. ii

Project an abstract resource that helps you group related BigML resources together. 26

Support the proportion of instances in the dataset which contain an itemset. The support of an association is the portion of instances in the dataset which contain the rule's antecedent and rule's consequent together over the total number of instances (N) in the dataset. 9, 31, 36, 62

Task the process of creating a BigML resource, such as creating a dataset, or training a model. A given task can also create subtasks, as, in the case of a WhizzML script that contains calls to create other resources. ii, 58, 62

Unsupervised learning a type of Machine Learning problem in which the objective is not to learn a predictor, and thus does not require each instance to be labeled. Typically, unsupervised learning algorithms infer some summarizing structure over the dataset, such as a clustering or a set of association rules. ii, 1

References

[1] The BigML Team. *Anomaly Detection with the BigML Dashboard*. Tech. rep. BigML, Inc., Jan. 2016.
[2] The BigML Team. *Classification and Regression with the BigML Dashboard*. Tech. rep. BigML, Inc., May 2016.
[3] The BigML Team. *Cluster Analysis with the BigML Dashboard*. Tech. rep. BigML, Inc., May 2016.
[4] The BigML Team. *Datasets with the BigML Dashboard*. Tech. rep. BigML, Inc., Jan. 2016.
[5] The BigML Team. *Sources with the BigML Dashboard*. Tech. rep. BigML, Inc., Jan. 2016.
[6] The BigML Team. *Time Series with the BigML Dashboard*. Tech. rep. BigML, Inc., July 2017.
[7] The BigML Team. *Topic Models with the BigML Dashboard*. Tech. rep. BigML, Inc., Nov. 2016.