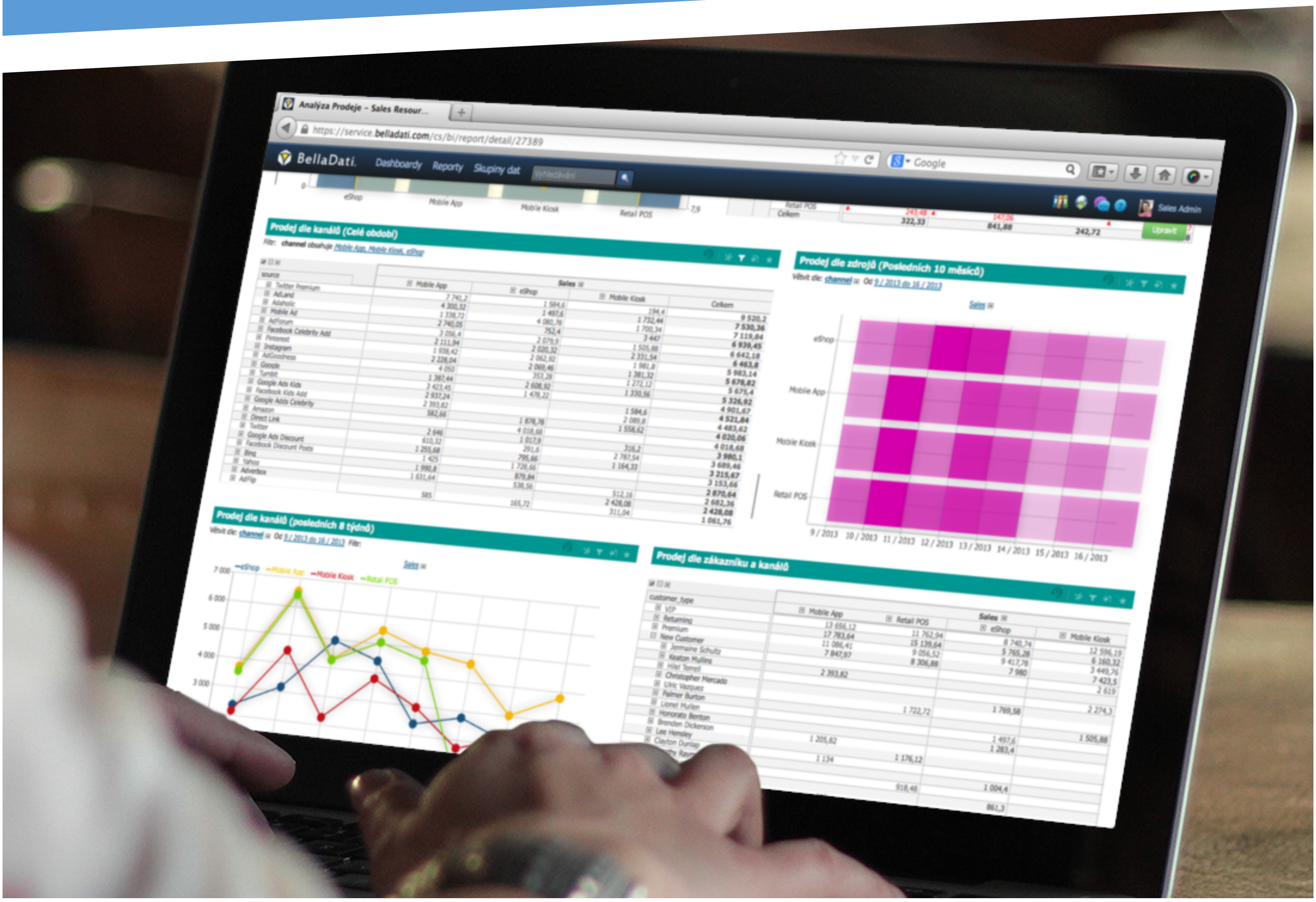
BellaDati

BellaDati Advanced Analytics Machine Learning for Retailers Market Basket Analysis and Associations



BellaDati Advanced Analytics module for Retailers Market Basket Analysis and Associations

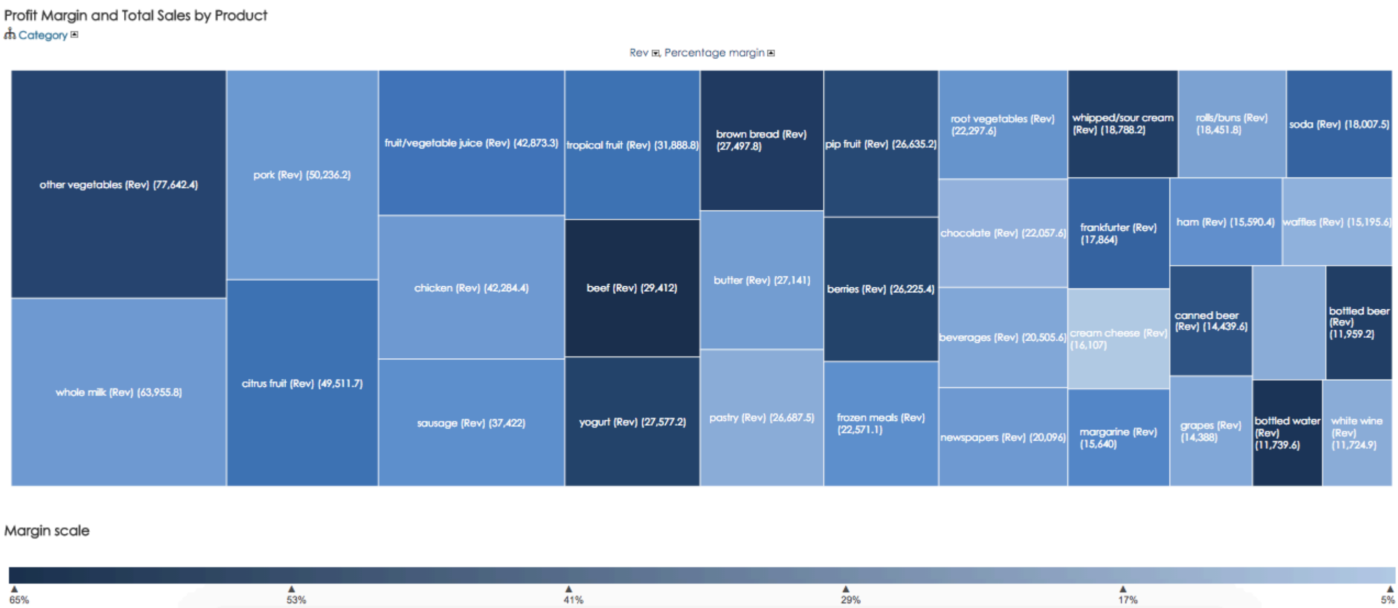
Business case:

Market basket analysis is used behind the scenes for the recommendation systems used in many brick-and-mortar and online retailers. The learned association rules indicate the combinations of items that are often purchased together. Knowledge of these patterns provides insight into new ways a retail chain might decrease costs and increase revenue, for example:

- optimize the inventory,
- advertise promotions
- organize the physical layout of the store (for instance, if shoppers frequently purchase coffee or orange juice with a breakfast pastry, it may be possible to increase profit by relocating pastries closer to coffee and juice).

The techniques could be applied to many different types of problems.

BellaDati Advanced Analytics module for Retailers – Market Basket Analysis use Association rule analysis to search for interesting connections among a very large number of elements. Human beings are capable of such insight quite intuitively, but it often takes expert-level knowledge or a great deal of experience to do what a rule learning algorithm can do in minutes or even seconds. Additionally, some datasets are simply too large and complex for a human being to find the needle in the haystack.



Barcode scanners, computerized inventory systems, and online shopping trends have built a wealth transactional data, machine learning has been increasingly applied to learn purchasing patterns. Generally, the demo involves: Using simple performance measures to find associations in large databases and identify the useful and actionable patterns. The results of a market basket analysis are actionable patterns.

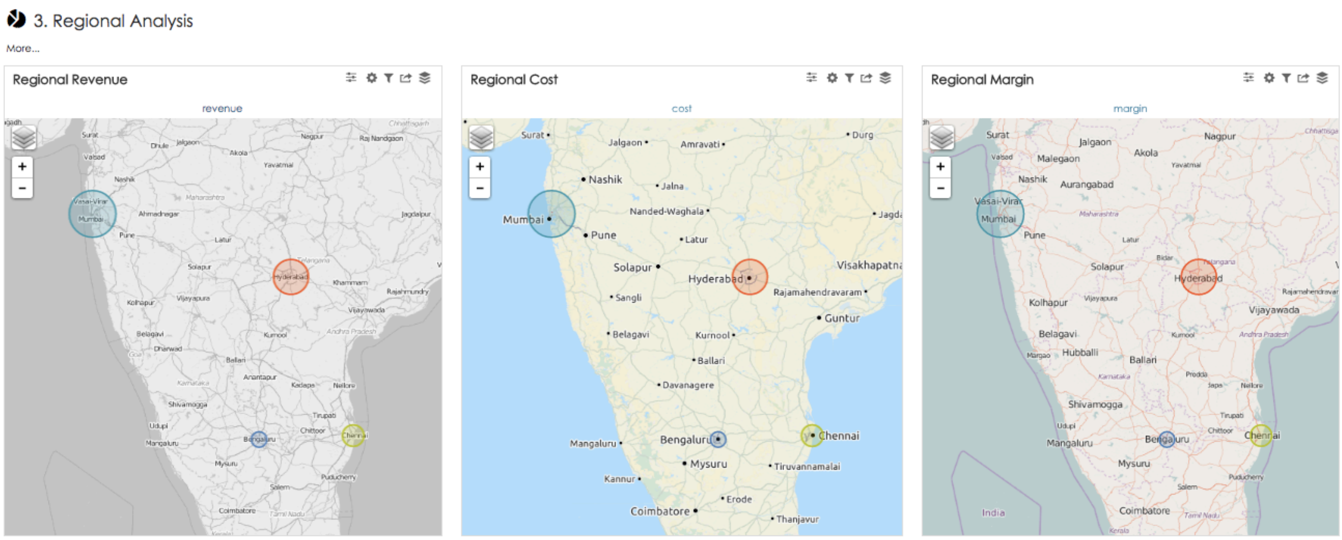
Just as it is challenging for humans, transactional data makes association rule mining a challenging task for machines as well. Transactional datasets are typically extremely large, both in terms of the number of transactions as well as the number of items or features that are monitored.

Rather than evaluating each of these itemsets one by one, BellaDati Machine Learning for Retailers- Market Basket Analysis and Associations use the algorithm for association rule learning.

Smarter rule learning algorithm takes advantage of the fact that, in reality, many of the potential combinations of items are rarely, if ever, found in practice. For instance, even if a store sells both automotive items and women's cosmetics, a set of {motor oil, lipstick} is likely to be extraordinarily uncommon. By ignoring these rare (and, perhaps, less important) combinations, it is possible to limit the scope of the search for rules to a more manageable size.

BellaDati Machine Learning for Retailers- Market Basket Analysis and Associations Module identifies actionable rules that provide a clear and useful insight. Whether or not an association rule is deemed interesting is determined by correlation statistical measures. These are used to sort association rules and to identify cross-sell opportunities.

BellaDati Machine Learning for Retailers- Market Basket Analysis evaluates for each interesting cross-selling opportunity (= interesting correlation rule) the value of cross-sell opportunity and also the profitability of designed marketing campaign and display results on Geomaps.



BellaDati incorporates full-range Java/Groovy capabilities for statistical computing, data mining and machine learning areas.

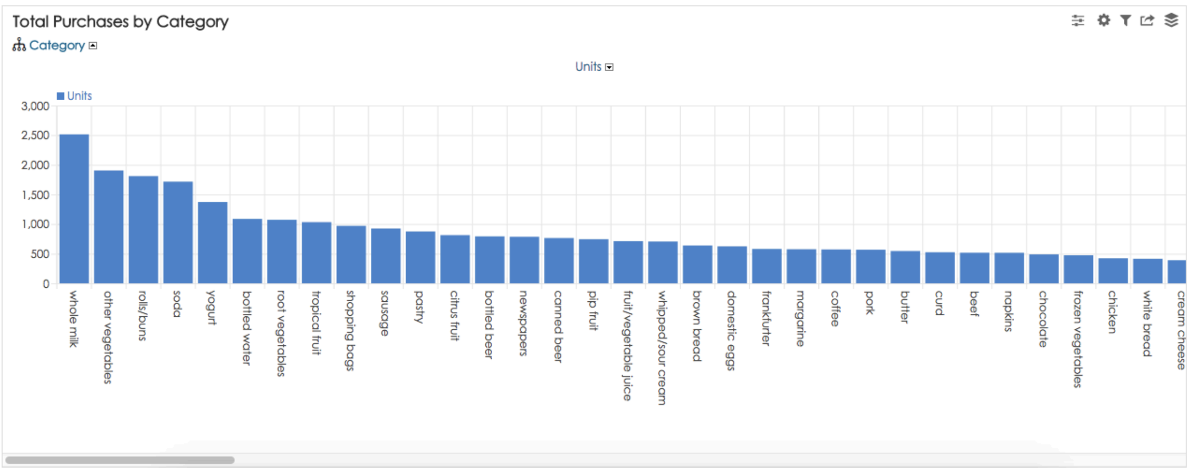
Using well designed BellaDati Machine Learning studio users can combine various methods to achieve optimum results. They can:

- easily use BellaDati ready made packages and functions,
- use packages from various Java based libraries as Apache Commons Math 3, Java ML or others
- write own advanced scripts and
- visualize intermediate results before further modeling
- results of modeling are then immediately available in BellaDati visualizations for publishing into reports or dashboards.

BellaDati Machine Learning Studio allows transparently interact with data stored inside data sets or gather the data from remote database, cloud services. 100+ ready made data connectors are already available in BellaDati. Implicit parallelism and just-in-time compilation in Java offers top performance (much faster then in R) and enables to achieve very effectively actionable insights.

Rich and agile BellaDati BI layer is used for further visualization, analysis of stock optimization, marketing campaign design and management information.

Architecture of Belladati Machine Learning for Retailers- Market Basket Analysis allows easy and fast (=agile) tailoring to specific customer needs by partner or by a customer.



Additional explanatory notes:

A/ Classification of association rules

A common approach is to take the association rules and divide them into the following three categories:

- Actionable
- Trivial
- Inexplicable

The goal of a market basket analysis is to find actionable rules that provide a clear and useful insight. Some rules are clear, others are useful; it is less common to find a combination of both of these factors. So-called trivial rules include any rules that are so obvious that they are not worth mentioning—they are clear, but not useful. Suppose you were a marketing consultant being paid large sums of money to identify new opportunities for cross-promoting items. If you report the finding that {diapers} → {formula}, you probably won't be invited back for another consulting job.

Rules are inexplicable if the connection between the items is so unclear that figuring out how to use the information is impossible or nearly impossible. The rule may simply be a random pattern in the data, for instance, a rule stating that {pickles} → {chocolate ice cream} may be due to a single customer, whose pregnant wife had regular cravings for strange combinations of foods. The best rules are hidden gems—those undiscovered insights into patterns that seem obvious once discovered. Given enough time, one could evaluate each and every rule to find the gems. However, we (the one performing the

B/ Measuring rule interest – support confidence and lift

Whether or not an association rule is deemed interesting is determined by two statistical measures: support and confidence measures. By providing minimum thresholds for each of these metrics and applying the Apriori principle, it is easy to drastically limit the number of rules reported, perhaps even to the point where only the obvious or common sense rules are identified. For this reason, it is important to carefully understand the types of rules that are excluded under these criteria.

The support of an itemset or rule measures how frequently it occurs in the data. For instance the itemset {get well card, flowers}, has support of 3 / 5 = 0.6 in the hospital gift shop data. Similarly, the support for {get well card} → {flowers} is also 0.6. The support can be calculated for any itemset or even a single item; for instance, the support for {candy bar} is 2 / 5 = 0.4, since candy bars appear in 40 percent of purchases. A function defining support for the itemset X can be defined as follows:

$$support\ (X) = count\ (X) / N$$

Here, N is the number of transactions in the database and count(X) is the number of transactions containing itemset X.

A rule's confidence is a measurement of its predictive power or accuracy. It is defined as the support of the itemset containing both X and Y divided by the support of the itemset containing only X:

confidence (X-Y) = support (X,Y)/support(X)

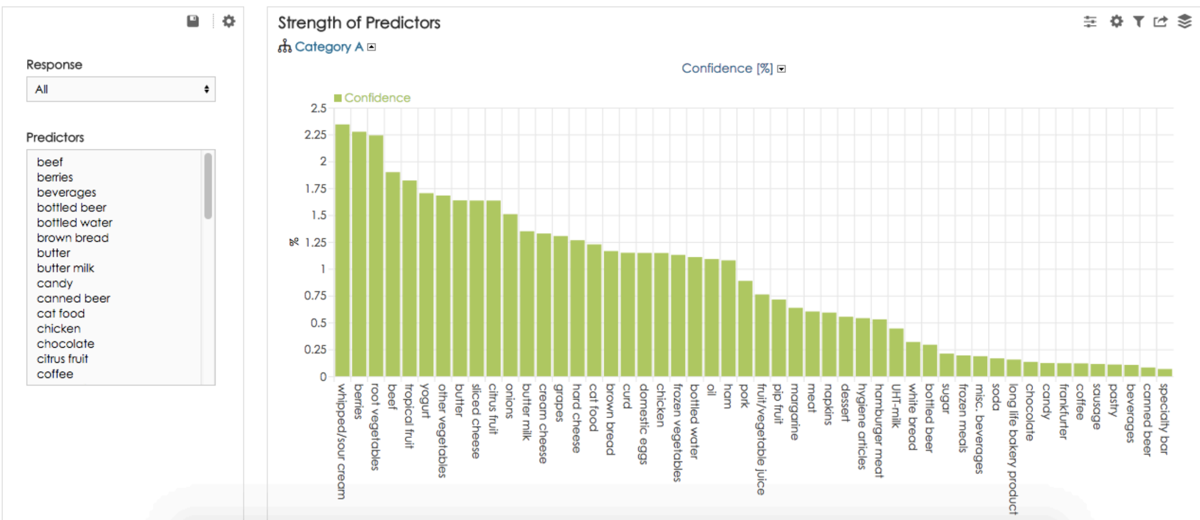
Essentially, the confidence tells us the proportion of transactions where the presence of item or itemset X results in the presence of item or itemset Y. Keep in mind that the confidence that X leads to Y is not the same as the confidence that Y leads to X. For example, the confidence of {flowers} → {get well card} is 0.6 / 0.8 = 0.75. In comparison, the confidence of {get well card} → {flowers} is 0.6 / 0.6 = 1.0. This means that a purchase involving flowers is accompanied by a purchase of a get well card.

The lift of a rule measures how much more likely one item or itemset is purchased relative to its typical rate of purchase, given that you know another item or itemset has been purchased. This is defined by the following equation:

Lift (X-Y) = confidence (X-Y)/support (Y)

C/ Other area where Association rules can be applied

The same technique could be applied to many different types of problems, from movie recommendations, to dating sites, to finding dangerous interactions among medications, prediction of customer churn or insurance policy misuse.



Available resources:

- Market basket analysis demo – YouTube
- Market basket analysis demo – Youku
- Demo reports (public)
- Demo reports – BellaApp download (Partner portal)
- Demo script (Partner portal)