SAP Predictive Analysis 1.0.9—What’s New?

SAP Predictive Analysis is the latest addition to the SAP BusinessObjects suite and introduces entirely new functionality to the existing Business Objects toolbox. Predictive Analysis integrates SAP’s Visual Intelligence data visualization tool with new predictive functionality powered by both open source R and SAP-written algorithms. Predictive Analysis includes algorithms for time series forecasting (for predicting sales, demand, price, and other time-dependent metrics), clustering (for identifying distinct groups of individuals based on numeric descriptive data), decision trees (for creating a tree-like set of decision support rules to categorize observations), and linear regression (for fitting linear relationships between a dependent variable and one or more predictors). These predictive algorithms can be used to extract insights and predictions, improving the value and actionable of the existing Business Intelligence infrastructure. Predictive Analysis combines these powerful predictive algorithms with a familiar and easy-to-use tool that integrates with existing BusinessObjects tools to make data preparation, model building, and implementation faster and easier than ever before.

SAP Predictive Analysis is installed locally on the user’s machine and accesses data for processing locally (from a CSV, Excel, or ODBC connection to a database) or on SAP HANA. Predictive Analysis can be used alone to analyze data on the client machine or can be paired with HANA, allowing PA to leverage the powerful in-memory processing power of SAP HANA.

On March 29, 2013, SAP released the 1st Quarter 2013 release of Predictive Analysis, version 1.0.9. This upgrade includes bug fixes and increased stability, as well as some new visualizations, specifically for apriori and decision tree algorithms.

New Model Result Visualizations

Decision Trees

Of particular note is the new confusion matrix view included in the decision tree output. This is a new visualization, automatically-generated after a decision tree is run for both the HANA and local R decision tree algorithms. The confusion matrix is one of the key evaluation tools for a decision tree model—accuracy, false positive, and false negative rates are generally the first things at which an analyst looks after the visualization of the tree itself. Prior to this feature being added, the user had to create the confusion matrix manually in the Visualize portion of the Results output. However, it was unattractive and only allowed a vertical orientation, whereas the traditional cross-tab confusion matrix is much easier to read. I wish that SAP had included a few more metrics in the confusion matrix display, such as row/column percentages and an overall accuracy measure.
Apriori

The apriori association algorithm has a new visualization as well. Previously, the only output visual for the apriori algorithm was the R summary print out.

The new word tag visualization shows the Apriori rules as a “word cloud” (font size and color coded according to the lift and confidence metrics for each rule). The list below the word cloud allows users to filter for rules based on acceptable ranges for Support, Confidence, and Lift. This allows users to filter more easily through the thousands of rules that can be created with the apriori algorithm to better digest only those with a high likelihood.
R Output Fix
In addition to algorithm-specific fixes, the new version of Predictive Analysis prints the full R output for models, rather than a truncated version. This is extremely important for implementation, as the full model equation was not previously available in the output.

Figure 3--New Version 1.0.9 Apriori Tag Cloud Output

Figure 4--MLR Output
New Algorithms
The only new predictive algorithm added with version 1.0.9 is HANA PAL Multiple Linear Regression (multiple predictors), which replaces the HANA PAL Linear Regression (single predictor) algorithm. Simple linear regression is not used very commonly in practice, so the HANA MLR algorithm is much more likely to be of use. However, the HANA R-MLR algorithm was already available, so this essentially duplicates functionality for those with HANA Rserve set up. This does provide useful new functionality for those without HANA-R capabilities.

![Figure 5--New HANA MLR Algorithm](image)

Visual Intelligence Additions
In addition to the Predictive Analysis updates, there are updates to the data manipulation and Visual Intelligence features (see [http://scn.sap.com/docs/DOC-26507](http://scn.sap.com/docs/DOC-26507)), including being able to append worksheets to a dataset, create calculated measures from existing measures, and create a trellis chart from multiple measures.

![Figure 6--New Calculated Measure Dialog](image)
Other Updates

In working with the new release, I have noticed fewer errors when saving while in the **Predict** pane. Previously, users could not save while in the prediction mode, but I do not see that error any longer. SAP also reported that version 1.0.9 has the ability to export multiple predictive model objects within an SVID document. The export interface is now much easier to understand, clearly giving the option to export either the PMML or SPAR version, rather than only the SPAR version and having to right click on the analysis component in the workflow to export the PMML.

![New Export Dialog and Multiple Model Selection for SVID Export](image)

Hillary Bliss, Business Intelligence Consultant  
Decision First Technologies  
[Hillary.bliss@decisionfirst.com](mailto:Hillary.bliss@decisionfirst.com)  
twitter [@HillaryBlissDFT](https://twitter.com/HillaryBlissDFT)

Hillary Bliss is a Business Intelligence Consultant specializing in data warehouse design, ETL development, statistical analysis, and predictive modeling. Hillary works with clients and vendors to integrate business analysis and predictive modeling solutions into data warehouses based on their data and business needs. With Decision First Technologies, Hillary uses Data Services, Web Intelligence, Predictive Analysis, and HANA. Hillary has a Master’s in Statistics and an MBA from Georgia Tech.