## ACTUATE is now OPENTEXT

### Technical White Paper

BIRT Analytics 5.1 Technical Summary of New Features

The BIRT Analytics 5.1 release introduces the addition of the Naive Bayes algorithm to our powerful set of datamining algorithms. This algorithm belongs to the family of linear classifiers and is very useful in the manipulation of conditional probabilities. These classifiers are highly scalable, providing quick and easy linear probability classification that requires relatively few parameters to arrive at significantly accurate results.

The user experience has been further heightened for both administrators and end-users by the addition of new functions and instructions to all modules such as, among others, exporting a Crosstab analysis directly into the database as a table, no limit on the size of data files uploaded into our loader and full user control over logging out or locking their application.

# Notice

The information in this white paper is proprietary to Actuate Corporation ("Actuate") and may not be used in any form without the prior consent of Actuate.

 $\ensuremath{\mathbb{C}}$  2015 by Actuate Corporation. All rights reserved.

Version 1 – March 2015

# **Table of Contents**

Notice	2
Table of Contents	3
Introduction	4
Naive Bayes linear classification tool	4
Enhanced export possibilities for Crosstab analyses	5
Functional Enhancements in BAadmin, BAfrontend and QLoader	5
BAadmin	5
SMPT Protocol Setting	5
BAFrontend	6
Lock Screen	6
Rename Repository Objects	6
Sorting Charts in Explorer	6
DataTree	7
Database Deletion	7
Workflow - Tasks and Events	7
QLoader	9
Limitless file size	9
Ignoring items	9
ON ERROR - attribute and instructions	9
Upgrading to BIRT Analytics 5.110	0

## Introduction

BIRT Analytics provides remarkably fast, free-form visual data mining and predictive analytics. The uniqueness of BIRT Analytics results from a combination of the ease of use of its data discovery and data mining tools with the power and sophistication of its analytic tools. This uniqueness has been further enhanced in this release through the addition of the Naive Bayes classification algorithm to our already powerful set of linear and logistical datamining tools. A summary list of new BA 5.1 improvements appears below:

- Addition to our datamining tool set of a both simple and significantly accurate linear classification tool based on the Naive Bayes algorithm, used in the manipulation of conditional probabilities.
- Enhancement of export possibilities, allowing a user to send a Crosstab analysis directly into the FastDB database as a new database table.
- More new functions and features in our administrative, frontend and loader modules designed to greater enhance the user experience.

## **Naive Bayes linear classification tool**

The already powerful BIRT Analytics datamining tool set, has been extended to include a *linear classification tool* based on the Naive Bayes algorithm. This algorithm has an underlying probability model with *statistically independent characteristics* and the estimation of its parameters is based on *maximum likelihood*. It provides simple probability classification with highly scalable classifiers, that require only a minimum number of predictors for training that produces significant results.

The main advantage of the naive Bayesian classifier is that it requires *relatively little* training data. Also, likelihood training is *very fast* because it is done by evaluating a closed-form expression (taking *linear time*) rather than using iterative approximation which is used for most other types of classifiers. The hypothesis of independence of variables does not require knowing more than the *variance* of each variable for every class, *without* having to calculate a *covariance* matrix. The simple and intuitive use of the Naive Bayes tool is much the same as for the Linear and Logistics Regression tools. The following figure shows its "Parameters" view in the Advanced tab of the Analytics tool set.

Analysis	Advanced	♥ Selectio	ns						
S Preprocessing	- Clustering	க் Decision tree	😵 Naive Bayes	🚄 Forecasting	← Association Rules	🥂 Correlations	🗶 Linear regression	Logistic regression	Windows (1)
Naive Bayes									_ @ X
Train Save App	bly								
Parameters R	esults Test								
Filter	Drag a segment	t							
Deferred Model									
Domain colu	mns								
			Dra	g tables, filters or	selections onto this par	el to select the colu	umns to use.		
Classification									
Alias	Color Nan	ne							+
New classifie	er Remove selec	ted							Ŧ
	A								L
	Ŕ	<b>@</b>	C						
Start Explore	e Enrichment	Analytics (	Ballery Workflo	w					

# Enhanced export possibilities for Crosstab analyses

It is now possible to export a Crosstab analysis directly to the FastDB engine, creating a new table in the database. This is done by selecting the new option "Analytic DB" from the dropdown list of the "Export" tool found in the Crosstab toolbar.

Crosstab - 02 Complex Cros	stab _ @	×
Calculate Export - Clear	Convert - Save Save as	
Table 💾 File		
🔺 Produ 🎯 BIRT	+ 🛛 Count(Order Detail) 📚	
Rows S Analytic DB	H 44 1 /1	
Occupation Decode	Occupation HI(Count(Or Dire Destination	>
	Hou Man	T
	Man Offic	
Filter Drag a segment	Cancel	ОК

# **Functional Enhancements in BAadmin, BAfrontend and QLoader**

The user experience has been further heightened by the addition of the following new functions and instructions to our modules.

### **BAadmin**

#### **SMPT Protocol Setting**

Now when Users are configuring the SMTP server they can select the type of protocol to use from the following list: Plain, SSL and TLS.

Security User mana	gement Configuration	Monitoring use
	Configure	SMTP server
Map management	Actions	
Styles admin	Use this option to set up th	e internal SMTP configuration used to send e-mails.
Configure SMTP server	Server	
	Port	25
	From	
	User	
	Password	
	Requires SSL	Plain 👻
		Plain
		SSL
		TLS

## **BAFrontend**

#### Lock Screen

The application no longer automatically logs out the User but only locks his/her screen when there has been no use for a prolonged period. Simply re-entering the password unlocks it again.

The User not only has total control over logging-out but also has the possibility of locking the screen on demand via the User icon at the top right-hand side of the screen.

#### **BIRT ANALYTICS**

Session locked by: Administrator	Administrator 👤 🔻
Password	Preferences
Password	BIRT Analytics Admin
	QLoader ) •
Log out	Lock
	Log out

#### **Rename Repository Objects**

In this version, Users can rename any repository object (*database, table or column*) by simply right-clicking on it in the Data Tree and choosing "Rename" in the dropdown list that opens.



#### Sorting Charts in Explorer

By default, and for performance reasons, the Chart view displays your records sorted by Value (ascending). This default sort is determined by how the values are sorted in the Discrete Values view. If you wish to display your Chart sorted by Count (descending) instead of by Value (ascending), you need to go to the Discrete Values view and click on the "Count" column label. Your Chart will then be displayed by Count (descending).

Record View	Summary	Discrete Values	Chart Statistics
H 44	1 / 1 Click	on "Count" to s	ort by
		nding count	
Value	Count <	Percentage	Percentage
47	10,476	4.03	
46	10,372	3.99	
44	10,309	3.97	
48	10,215	3.93	
45	10,197	3.92	
49	9,997	3.85	
50 🕂	9,774	3.76	
43	9,766	3.76	
42 Default so ascending	9,303	3.68	
41 ascending	g values 9,211	3.54	
51	9,152	3.52	

### DataTree

Data Tree use has been enhanced by significantly decreasing the speed of the discrete values searching process. It now takes into account multiple keystrokes instead of individual strokes, giving instant results to your search.

9 🖪			
H 44	1 / 895	► H	
(Demo].[Cust	omer].[Surna	me]	
Value	Count		
A Ahmed		1	•
A Ali		4	
A Alsaad		5	
A Barlow		3	
A Crossley		3	
A Davies		3	
A Khan		3	
A Mukheini		3	J.
A Nakashi		2	•
Search	Ĥ		
Enter name to s almost instantly			ar

#### **Database Deletion**

Database deletion can now be done directly from the Data Tree by right-clicking on the table and choosing "Delete".

Ø ACTI	JATE BI	RT Analytics
8 🖩	-	
<ul> <li>♥ Demo</li> <li>▶ Ⅲ (</li> <li>▶ Ⅲ +</li> <li>▶ Ⅲ (</li> <li>▶ Ⅲ (</li> </ul>	Summary Rename Delete	

#### **Workflow - Tasks and Events**

Several enhancements have been made to the "Tasks and events" tab of the "Workflow" module:

1. **New action: delete database** through the "Actions" field of the "Scheduled task detail" screen in the "Tasks and events" tab of the "Workflow" module.



2. **New task: "Repository item renamed".** Now whenever a User renames a repository object it will trigger the application to raise a new type of event that can then be used in the "Tasks and events" tab. (See figure below).

-	Tasks and events	Campaign workflow		
Start	Scheduled tasks     Onew scheduled tasks	neduled task		
$\odot$	Scheduled task detail			
-	Save			
Explore	Scheduled task Actions			
R	Name	Starting date	ø	
Enrichment	Description	🗖 Ending date	Ð	
۲	<ul> <li>Active</li> <li>Trigger for task</li> </ul>			
Analytics	► P Scheduled task on tir	ne		Drag and drop an item from the left tree to use as a trigger for the task.
<str>&lt;</str>	<ul> <li>Scheduled task on ev</li> </ul>	ent		
Gallery	<ul> <li>BIRT Analytics</li> <li>Engineering</li> </ul>			
G	A Definition mod	ified		
	A Repository iter	n renamed		
Workflow	A Repository iter	n deleted		
	🕨 🖿 Data Mining			
	<ul> <li>Campaign workflo</li> </ul>	nv .		

3. New action: Rename (to be used for renaming Repository items) can now be performed in the "Tasks and events" tab.

-	Tasks and events Campaign workflow	
Start	Scheduled tasks 🔇 New scheduled task	
() Explore	Scheduled task detail Save Scheduled task Actions	Campaign workflow     Tasks and events     Campaign workflow     Start     Scheduled tasks     CNew scheduled task     Scheduled task
	Actions    Sending   Conditional	Explore Action detail
Analytics	Execute     Export     Data model     Delete column	Enrichment Full Name New name New name
Gallery	Delete table     Delete database     Rename	Gallery
Workflow	+ 🖿 Apply	<b>G</b> workflow

4. **More Algorithm Model Types** have been added for use in Task datamining. They are Linear Regression, Logistical Regression and the Naïve Bayes algorithm.

					vents 🛈 Campaign workflow	
			Start	0	New scheduled task	
	Tasks and events (1) Can	npaign workflow	$\odot$	Scheduled task de	tail	
Start	Scheduled tasks     New scheduled		Explore	Scheduled task	Actions	
0	Scheduled task detail		Ř	Name		Starting
	Save		Enrichment	Description		🗆 Ending d
Explore	Scheduled task Actions		8	Active Model applied		
R	Name	Starting date	Analytics	Condition deta	a	
Enrichment	Description	Ending date	<ul> <li><ul> <li><ul> <li><ul></ul></li></ul></li></ul></li></ul>	Attribute	Model type	
8	Active		Gallery	Operation	= •	
Analytics	Trigger for task		G		- •	
	Scheduled task on time		Workflow	Value	Clustering	•
4@b	<ul> <li>Scheduled task on event</li> </ul>		WORKNOW		Clustering	
	<ul> <li>BIRT Analytics</li> </ul>				Decision tree	
Gallery	👻 🖿 Data Mining				Linear regression	
G	🗯 Model applied				Logistic regression	
Workflow	Campaign workflow				Naive Bayes	

## QLoader

### Limitless file size

No limit to file size when uploading into QLoader. This is due to the fact that the *Default value* in *configuration.xml* is **-1**.

### Ignoring items

Users can now choose to have items (tables, instructions) ignored during the loading process in both the Data and Transform tabs. Clicking on the "check" or "minus" symbols in the column beside the "On error" column in the Data tab (or in the column beside the "Command" column in the Transform tab) allows Users to have chosen elements ignored by the loading process (toggling them on and off). The contents of ignored item rows appear in italics in both cases.



### **ON ERROR** - attribute and instructions

Users can decide *not to stop* the loading process when errors occur in either the DATA or TRANSFORM tabs. By default, the loading process stops whenever an error is encountered. However, thanks to both a new *ONERROR attribute* in data table definition and a new *ONERROR instruction* for the transformation script, it is now possible to choose whether you want the loading process to stop in the case of error or to continue loading. This allows the loading process to ignore any tables or instructions that present errors during the process, leaving them to be fixed and loaded later if desired. This can be particularly useful in the case of loading high volume tables needing overnight processing.

**In the DATA tab**, clicking on the "minus" symbol (in the "On error" column) beside a table row, changes the "minus" symbol to an "arrow" symbol and de-activates the default "Fail on error" setting in the list of tables to be loaded. Double clicking on the table opens its "Table declaration" window where you can see that the "Fail on error" checkbox is now unchecked. Clicking on the "arrow" symbol changes it back to "minus" and resets the default "Fail on error" setting, checking the box once again. (See figure below on page 10).

ptions													
🗹 Load 🛛	C Transfe	orm 🗹 Post I	oading	Notify	0		Sam	ple 0					
New from fi	le Ne	w from databa	se Edit	Remo	ve Dup	licate							
ctive	On error	Source	Databas	e		N	ame			Descri	ption		N Colu
0	•	delimited	Demo			C	ustom	ner		Custo	mers		
0	•	delimited (	Table decla	ration									
0	•	delimited	Table decia	auon									
0	•	delimited	Database				Actio						
0	•	delimited	butubust	Demo	)	•	100101		Create	*	0 12	verwrite	
0	•	delimited	Name	Custo	mer		Descr	iption C	ustomers		0 Ig	gnore	Fail on erro
0	•	delimited	Туре	_									_
0	0 0	delimited	туре	Delin	nited							•	
S	•	delimited	Encoding	ASCI		-	Skip	1			п н	eader	
												cooci	_
			End Line	CrLf		*	Offse	t					
			File		ustomer.cs			•	Is list View	6.	t colum		
					ustomentes	v			is list view	UC	colum	1115	
			Delimiter			•	Quali	fier		•	💌 W	/ell forme	d
			New E	dit F	Remove	Duplicat	e						+ 1
			Name		Туре	Width		Decimals	Format		Index	Skip	
			Cust ID		integer		18	(			$\checkmark$	×	
			DOB		date		8	(	yyyymmdd		$\checkmark$	×	
			Gender		string		1	(			$\checkmark$	×	
			Household	ID	integer		18	0			$\checkmark$	×	
			Income		real		10	2			~	×	
			Initials		string		4	(			$\checkmark$	X	
			Occupation		string		2	(			1	×	

In the TRANSFORM tab, you can create new instructions ON ERROR continue and

*ON ERROR break*. This is done by clicking on "New", selecting "ON ERROR" from the list of available instructions that appears and then choosing either "continue" or "break" as needed.

Use the "up" and "down" arrows at the top of the tab to position your instructions wherever you want them to be in your rows of instructions.

Data	Transform Post loading	Data Management	Admin	
New	Duplicate Remove 🛛 🗢 🕇 🖡 🛛	· ·	On error	
Analy	sis_Results.xml		On error Break	
	Command	Description	break	•
۲	LINK [Decode],[Customer Results],[Cust ID],[D		Description	
0	DROPCOLUMN [Demo].[Customer].[RFM Clust		Description	
0	EXPRESSION [Demo].[Customer],RFM Clusteri			
0	ON ERROR continue			
0	MAKEPERMANENT [Demo].[Customer].[RFM CL			
Ø	ON ERROR break			
0	DROPCOLUMN [Demo].[Customer].[Is target f			
0	EXPRESSION [Demo].[Customer],Is target for			
0	MAKEPERMANENT [Demo].[Customer].[Is targ			
0	DROPCOLUMN [Demo].[Customer].[Scoring to		Cancel OK	
0	EXPRESSION [Demo].[Customer],Scoring to Of			

# **Upgrading to BIRT Analytics 5.1**

New installation of BIRT Analytics 5.1 requires a license key specific to BA 5.1.

More detailed information on upgrading from BA5.0 to BA 5.1 will be available upon request.