Supercharge WebFOCUS Computes with Data Science: RStat and the New Rserve Adapter

Andy Kirby & Ira Kaplan
Information Builders

Information Builders
Summit 2019 User Conference
June 4 - 7, 2019

Presentation Abstract:

WebFOCUS RStat is a fully integrated business intelligence (BI) and data mining environment. It supports data modeling as well as data scoring. In this session, we'll review both data modeling and scoring, along with the recently introduced Rserve Adapter, which integrates WebFOCUS into the R environment for computing the values of virtual data fields. You'll be able to run your scripts and models to enhance your reports and visualizations, providing a deeper understanding and analysis of the information you have in your data.

Business Need:

RStat, an interface over the R language, integrates R with WebFOCUS, our BI workbench, to provide a robust, easy to use and accurate platform for data access, cleansing and preparation, model building, testing and for operationalizing results. RStat, enables business analysts to perform data mining and statistical modeling with ease. It is also useful to the data scientist as R scripts that build models may be uploaded and the models may be deployed for use in scoring applications as charts, reports, dashboards and visuals, in WebFOCUS. It brings the power of predictive analytics to any user in any department in any organization.
In this session, we will cover the following:

- Walk through of the model development process
- Access our data in App Studio
- Launching RStat
- Exploring our modeling data
- Build a Cluster model
  - Deploy the model
  - Create a Report using the model
- Build a Market Basket model
  - Deploy the Market Basket example
  - Create a Report using the example
- Build a predictive model
  - Compare and evaluate different models
  - Deploy the best model identified by the test
  - Create a scoring application using the model
- Build a WebFOCUS report based on a pre-written R script
  - Add fields to a WebFOCUS report based on the R Script
  - Run the report within an application
Section 1 – Unsupervised or exploratory models

Part I – Clustering

Clustering is organizing objects in groups based on their similarities. The goal of clustering is to discover the intrinsic grouping in a dataset. The groups discovered should be dissimilar to other groups. Clustering is used in many fields including machine learning, pattern recognition, image analysis and computer graphics.

1. Launch App Studio 8205.
   If not already launched, double-click on the WebFOCUS App Studio icon on the desktop.

2. In the Environments Tree, double click the following folders to open and expand them:
   Data Servers -> EDASERVE -> Applications -> _rstat_modeling
3. Double Click on the cluster_train.fex program to open it in the Report Canvas.

4. Select the Modeling tab above the Ribbon.

5. Click the RStat launch button.
WebFOCUS RStat 3.0 opens. The R Console remains open in the background.


Note: You may click the View button to view the columns and values for the data.
7. Clustering requires numeric fields. Ignore the alpha fields. Click the Education variable, the Shift key then click Occupation to multi-select.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Data Type</th>
<th>Input</th>
<th>Target</th>
<th>Risk</th>
<th>Ident</th>
<th>Ignore</th>
<th>Weight</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ID</td>
<td>Numeric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 1926</td>
</tr>
<tr>
<td>2</td>
<td>Age</td>
<td>Numeric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 62</td>
</tr>
<tr>
<td>3</td>
<td>Education</td>
<td>Categoric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 5</td>
</tr>
<tr>
<td>4</td>
<td>Marital</td>
<td>Categoric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 4</td>
</tr>
<tr>
<td>5</td>
<td>Gender</td>
<td>Categoric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 2</td>
</tr>
<tr>
<td>6</td>
<td>Occupation</td>
<td>Categoric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 12</td>
</tr>
<tr>
<td>7</td>
<td>Income</td>
<td>Numeric</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>Unique: 1926</td>
</tr>
</tbody>
</table>

8. Click the Ignore option.
9. Uncheck the Partition check box.

10. Click the Run button.
11. Click on the Cluster Tab.

12. Uncheck Re-Scale. Rescale is used to modify all the values to a decimal between 0 and 1. We will use the actual values in our Clusters.

13. Change the number of clusters to 2.

14. Click the Run button to view the summary of the clusters generated.
15. Click on the Data button to view the clusters for Age and Income in a graphical plot.
16. Close the data plot window.
17. Change the cluster size to 8 and click the Run button.

<table>
<thead>
<tr>
<th>Cluster sizes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] 199 598 410 36 165 310 95 112</td>
</tr>
</tbody>
</table>

Data means:

<table>
<thead>
<tr>
<th>Age</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.30737</td>
<td>228904.38339</td>
</tr>
</tbody>
</table>

Cluster centers:

<table>
<thead>
<tr>
<th>Age</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>41.39196</td>
<td>285645.47</td>
</tr>
<tr>
<td>42.19900</td>
<td>183600.25</td>
</tr>
<tr>
<td>40.78780</td>
<td>209709.28</td>
</tr>
<tr>
<td>37.52778</td>
<td>473720.02</td>
</tr>
<tr>
<td>38.52727</td>
<td>335680.94</td>
</tr>
<tr>
<td>41.78287</td>
<td>244137.95</td>
</tr>
<tr>
<td>39.55208</td>
<td>393353.52</td>
</tr>
<tr>
<td>26.59621</td>
<td>21130.94</td>
</tr>
</tbody>
</table>

Within cluster sum of squares:

| [1] 36667866979 31651199305 30342223479 70517669230 35321737167 36976383335 |
| [7] 34164801001 10030414596 |

Time taken: 0.00 secs

Generated by RStat 2018-05-08 08:22:02
18. Click the Data button to view the plot.

19. Close the Data plot window.

**Accessing the Cluster in WebFOCUS.**

There are two stages to making the cluster algorithm available to WebFOCUS. First, we need to export the model from WebFOCUS RStat. Secondly, it must be deployed to an application directory on the WebFOCUS reporting server for use as a function.
20. Click the Export button on the Ribbon.

![Export button on the Ribbon](image)

**Note the Model Export dialog opens.**

21. Enter the name - cluster and click Save to export the model to the default folder, rstatdata. The model export can be accessed in C:\rstatdata.

![Model Export dialog](image)

22. Click Save button to complete the export.
Click the Quit button in WebFOCUS RStat to close the interface.

23. Click Yes to confirm.

25. Click the Model Deployment icon in the Modeling group on the ribbon.

In the Model Deployment window, click on the Add button, select cluster.c and click ‘Open’.
In the Deployment Destination Window, expand Localhost -> EDASERVE and select _rstat_modeling’ as the destination.

26. Click Deploy.
Click OK on the notification that the deployment was successful and close the Model Deployment dialog.

27. On the Report canvas, click to the right of the last field, Income, to place the cursor at this position.
At this point, we will apply the Kmeans cluster algorithm as a function to a Compute field in the report.

28. Select the Data tab and click the Summary (Compute) button.

29. In the Computed Field Creator dialog, enter the name, cluster and change the format to A5.

30. Click the Retrieve Scoring Routines button.
Supercharge WebFOCUS Computes with Data Science: RStat and the new Rserv Adapter

Copyright © 2019 Information Builders
Page 19 of 68
31. Select the cluster function in the Select a function list box.

![Cluster function selection](image)

32. Select the Age field and drag it to the Value box for the corresponding Input Parameter.

33. Select the Income field and drag it to the Value box for the corresponding Input Parameter.
34. For Target, enter the name assigned to the Compute field, cluster.

35. Click Ok on both dialogs to close them.
36. Click the Run button to execute the report.

The report shows the cluster into which each individual in the new data set is assigned.

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Education</th>
<th>Marital</th>
<th>Gender</th>
<th>Occupation</th>
<th>Income</th>
<th>cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1004641</td>
<td>38</td>
<td>Bachelor</td>
<td>Single</td>
<td>Female</td>
<td>Consultant</td>
<td>231,838.00</td>
<td>6</td>
</tr>
<tr>
<td>1010229</td>
<td>35</td>
<td>Engineerate</td>
<td>Divorced</td>
<td>Male</td>
<td>Doctor</td>
<td>222,099.00</td>
<td>3</td>
</tr>
<tr>
<td>1024587</td>
<td>32</td>
<td>Master</td>
<td>Divorced</td>
<td>Male</td>
<td>Sales</td>
<td>304,676.74</td>
<td>1</td>
</tr>
<tr>
<td>1038288</td>
<td>21</td>
<td>HSgrad</td>
<td>Widowed</td>
<td>Male</td>
<td>Professional Services</td>
<td>177,743.82</td>
<td>2</td>
</tr>
<tr>
<td>1047095</td>
<td>74</td>
<td>College</td>
<td>Widowed</td>
<td>Male</td>
<td>Doctor</td>
<td>183,144.40</td>
<td>2</td>
</tr>
<tr>
<td>1047698</td>
<td>43</td>
<td>Engineerate</td>
<td>Widowed</td>
<td>Male</td>
<td>Education</td>
<td>193,391.17</td>
<td>2</td>
</tr>
<tr>
<td>1053888</td>
<td>35</td>
<td>College</td>
<td>Widowed</td>
<td>Male</td>
<td>Sales</td>
<td>209,906.65</td>
<td>3</td>
</tr>
<tr>
<td>1061323</td>
<td>56</td>
<td>Bachelor</td>
<td>Divorced</td>
<td>Female</td>
<td>Consultant</td>
<td>276,888.91</td>
<td>1</td>
</tr>
<tr>
<td>1062363</td>
<td>22</td>
<td>HSgrad</td>
<td>Single</td>
<td>Female</td>
<td>Student</td>
<td>52,466.49</td>
<td>8</td>
</tr>
<tr>
<td>1068642</td>
<td>48</td>
<td>Engineerate</td>
<td>Divorced</td>
<td>Female</td>
<td>Attorney</td>
<td>441,416.11</td>
<td>4</td>
</tr>
<tr>
<td>1071815</td>
<td>60</td>
<td>College</td>
<td>Married</td>
<td>Male</td>
<td>Professional Services</td>
<td>174,155.31</td>
<td>2</td>
</tr>
</tbody>
</table>

37. Close the Output window.

38. Close the report, cluster_train.fex, and click Yes to save the changes.
Part II - Market Basket Analysis

Market Basket Analysis is another exploratory technique that analyzes the co-occurrences of transactions with the appropriate algorithm.

1. In the Environments Tree, right click on mkt_bskt_train.fex and select Open.
2. Click the Run button to view the data.

![Image of WebFOCUS interface](image)

Items or products that have been purchased are placed in baskets with the items they have been purchased with. Here we have 24 baskets. This data will be used to train our market basket model prior to deployment.

![Image of market basket data](image)

3. Close the browser.
4. Select the Modeling tab and click the Launch button to open RStat.

WebFOCUS RStat opens.

5. Uncheck the Partition checkbox.

6. Set Basket as an Identifier.

7. Set Product as a Target.

8. Click Run.
9. Select the Associate Tab, check the Basket checkbox and click on Run.

10. Click on the Show Rules button and scroll down to view the rules.
Market Basket or Affinity Analysis focuses on purchase coincidence. The output is a set of rules that can be explored to make business decisions.

The rules determine the co-occurrences of items in transactions. Example, if an item or product on the left hand side (LHS) is purchased then it is likely that a client or customer will be interested in a product or item on the right hand side (RHS).

Market Basket Analysis uses the appropriate algorithm. The model can be exported from WebFOCUS RStat and deployed as a function for use in a report or chart.
11. Click the Export button on the ribbon.

12. Enter `association_analysis` as the name of the C file that will be generated when the model is exported.
13. Ensure that Item is selected as the Export Type.
14. Click Save.

Now we are going to deploy a second version of the model that will give us the confidence level that someone would be interested in the suggested Item.

15. In WebFOCUS RStat, click on Export again.

16. In the Model Export dialog, enter the name `association_confidence`.

17. Select the Confidence Radio Button for Export Type.
18. Click Save.

![Model Export dialog box]

19. Click on Quit to close RStat and to return to the App Studio canvas.

![RStat window with Quit button highlighted]

20. Close the mkt_bskt_train.fex

We will use our exported models to analyze new transactions.
21. In the Environments Tree, double click to open the mkt_bskt_score.fex

![Environments Tree](image)

22. Run the mkt_bskt_score report by right clicking on the report and selecting Run.

![WebFOCUS Report](image)

<table>
<thead>
<tr>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket 1</td>
<td>Beverage</td>
<td>Candy</td>
<td>Chips</td>
</tr>
<tr>
<td>Basket 2</td>
<td>Chips</td>
<td>Lottery</td>
<td>Medicine</td>
</tr>
<tr>
<td>Basket 3</td>
<td>Candy</td>
<td>Chips</td>
<td>Lottery</td>
</tr>
<tr>
<td>Basket 4</td>
<td>Beverage</td>
<td>Candy</td>
<td>Lottery</td>
</tr>
</tbody>
</table>

23. Close the output window.
24. Click on the Home tab.

![Home tab in WebFOCUS](image)

25. Click on the Model Deployment button in the Modeling group on the Ribbon.

![Model Deployment button](image)

26. In the Model Deployment dialog, click Add.

![Model Deployment dialog](image)
27. Multi-select the models exported, association_analysis.c and association_confidence.c

28. Click the Open button on the dialog.

29. In the Deployment Destination section of the Model Deployment dialog, expand EDASERVE and select _rstat_modeling.

30. Click Deploy.
31. Click OK on the deployment status dialog.

32. On the report canvas, click an empty space on the right of the last field to reposition the cursor.
33. Right click the Computed Fields label and select New Compute Virtual Field.

34. In the Computed Field Creator dialog, enter new_item as the field name and click the Format button.

35. In the Format dialog, select Alphanumeric as the Format Type. We will use the default length of 20 characters. Click OK.
36. Click Functions.

37. Click Refresh Scoring Routines.

38. Select association_analysis in the Select a Function list box.

39. Drag the columns from the Fields tree to the corresponding Input Parameters Value box.

   Example: Drag the field, Basket1 to the Value box corresponding to LHS1.
40. Repeat this so that it is done for all four fields.

41. Enter the name of the Compute field as the Target.

42. Click Ok

43. Click Ok

Note:
A new field is added to the report.

44. Select the Data tab and click the Compute (Summary) button. The cursor is positioned after the new field.

45. Enter the name, confidence, for the new Compute field and leave the default format, D12.2

46. Click the Functions button.
47. Click Refresh Scoring Functions.

48. Select association_confidence in the Select a Function window.

49. Drag the corresponding columns from the Field tree to the Input Parameters Value box.
50. Type the fieldname, confidence, in the Target Recommendation input box.

51. Click Ok.
   
   Note the function that has been generated:
   
   ```
   association_confidence(MARKETBASKETFORSCORE.MARKETBASKETFORSCORE.Basket1, MARKETBASKETFORSCORE.MARKETBASKETFORSCORE.Basket2, MARKETBASKETFORSCORE.MARKETBASKETFORSCORE.Basket3, MARKETBASKETFORSCORE.MARKETBASKETFORSCORE.Basket4, confidence)
   ```
   
   52. Click Ok.
53. Click Run to execute the new report.
   Note that confidence is the conditional probability that when a specific product is purchased, a certain other product will also be purchased.

<table>
<thead>
<tr>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>new_item</th>
<th>confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basket 1</td>
<td>Beverage</td>
<td>Candy</td>
<td>Chips</td>
<td>Lottery</td>
<td>Medicine</td>
</tr>
<tr>
<td>Basket 2</td>
<td>Chips</td>
<td>Lottery</td>
<td>Medicine</td>
<td>Beverage</td>
<td>Candy</td>
</tr>
<tr>
<td>Basket 3</td>
<td>Candy</td>
<td>Chips</td>
<td>Lottery</td>
<td>Medicine</td>
<td>Beverage</td>
</tr>
<tr>
<td>Basket 4</td>
<td>Beverage</td>
<td>Candy</td>
<td>Lottery</td>
<td>Medicine</td>
<td>Chips</td>
</tr>
</tbody>
</table>

54. Close the output window.

55. Close the report, mkt_bskt_score.fex.

56. Click Yes to save.
Section 2 - Supervised or Predictive Models

Part III – Decision Tree and Regression models

WebFOCUS RStat offers the ability to build and compare some of the most commonly used predictive models. These can be categorized as:
- Rule Based – Decision Tree, Random Forest
- Statistical – Regression
- Machine Learning – Neural Network, Support Vector Machine

We will look at a data set that shows information about customers who have departed from First Bank. We will then train and test a model and deploy it for use against a dataset of new customers in reports, charts or dashboards for internal analysis and business decision making.

1. In App Studio, double click on churn_train.fex in the Environments Tree to open it.

2. Click on the Run button to run the report.

3. Note the historic data and the Churn field, indicating which customers no longer do business with the company, First Bank.
4. Close the output window.

5. Select the Modeling tab and click the Launch button to open WebFOCUS RStat.

6. Maximize the RStat window.

7. The data tab is selected by default. Select the following as Input variables:
   a. Age
   b. Education
   c. Marital
   d. Gender
   e. Occupation
   f. Income
   g. Credit Score
   h. Years at Current Job
   i. Customer Life Time Value

Target Variable:
   - Churn
Note: The Partition is enabled and this is required to divide the data set into a training data set and testing or evaluation data set. The training data will be used to create the model and the testing data set will be used to evaluate how well the model predicts.

<table>
<thead>
<tr>
<th>No.</th>
<th>Variable</th>
<th>Data Type</th>
<th>Input</th>
<th>Target</th>
<th>Risk</th>
<th>Ident</th>
<th>Ignore</th>
<th>Weight</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CUSTOMER_ID</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 1704</td>
</tr>
<tr>
<td>2</td>
<td>AGE</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 56</td>
</tr>
<tr>
<td>3</td>
<td>EDUCATION</td>
<td>Categoric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 6</td>
</tr>
<tr>
<td>4</td>
<td>MARITAL</td>
<td>Categoric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 4</td>
</tr>
<tr>
<td>5</td>
<td>GENDER</td>
<td>Categoric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 2</td>
</tr>
<tr>
<td>6</td>
<td>OCCUPATION</td>
<td>Categoric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 11</td>
</tr>
<tr>
<td>7</td>
<td>INCOME</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 1689</td>
</tr>
<tr>
<td>8</td>
<td>CREDIT_SCORE</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 343</td>
</tr>
<tr>
<td>9</td>
<td>YEARS_AT_CURR_JOB</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 10</td>
</tr>
<tr>
<td>10</td>
<td>CUSTOMER_LTV</td>
<td>Numeric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 50</td>
</tr>
<tr>
<td>11</td>
<td>CHURN</td>
<td>Categoric</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Unique 2</td>
</tr>
</tbody>
</table>
8. With the data tab selected, click Run.

9. Click on the Explore tab and ensure the Summary box is checked.

10. Click Run to view the summary of the dataset.

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUCATION</td>
<td>Associates, BS, M, PhD, MEMP</td>
</tr>
<tr>
<td>MARITAL</td>
<td>Divorced, Married, Single, Widowed</td>
</tr>
<tr>
<td>GENDER</td>
<td>Female, Male</td>
</tr>
</tbody>
</table>
```

This data can also be viewed in the format of a Chart.
11. Select the Distributions radio button.

12. Check the Histogram checkbox for Age.

13. Click Run.

The graphic shows the distribution of Age and how our churners compare to the non-churners.

14. Close the Chart.

15.

16. Select the Model Tab.

This is where all the included predictive algorithms are selected and applied to the data. These techniques uncover the relationships of input variables to the target variable and produce models that can be deployed to score new records for the target behavior.
17. Expand the Model Type drop list and select Decision Tree.

18. Click Run.

Notice the input fields that are significant to calculating the target, Churn, are Age, Marital and Years at Current Job.
19. Click the Rules button on the upper right.

Looking at Rule 29 of “If a person is divorced, married or single and their years at current job between 1.5 and 5.5 and they are older than 30.5” then churn = yes with probability 0.8
20. Click the Draw button. Note the visual of the decision tree.

21. Close the decision tree window.
22. Click on the Clear All button, under the Select Model(s) label.
Build a Logistic Regression Model

23. Expand the Model Type drop list and check the Logistic Regression model.

24. Click Run.

The significant variables identified by the logistic regression model are different than those identified by the Decision Tree. It is normal to find differing patterns across algorithms. Models built are compared in the Evaluate tab.

25. Check the both the Decision Tree and the Logistic Regression models in the Model Type Dropdown.

26. Click the Evaluate Tab

From the Evaluate tab, we may test the model with the remaining 30% of the dataset that was previously partitioned. We will compare the Decision Tree and the Logistic Regression model.
27. Check the Error Matrix radio button and click Run.

RStat generates an error matrix of how the selected models perform. The results show the Decision Tree model with a slightly lower error rate of 19%, compared to the Regression model error rate of 21%. So the Decision Tree model will be used as the predictive model.
Re-creating the Model with useful variables

We have selected the Decision Tree as the more effective model to deploy to WebFOCUS, we will now re-create the model using only significant variables identified.

28. Return to the data tab and since the Decision Tree model we have built only uses three variables, Age, Marital and Years at Current Job in predicting churn, select Ignore for all inputs EXCEPT the following:
   a. Age
   b. Marital
   c. Years at Current Job

29. Click Run.

30. Select the Model Tab, on the Model Type drop list, select the Decision Tree model.
31. Click run to execute the Decision Tree model only with the new variables as inputs.

32. Click the Export button.
33. Enter a name for the model export, churn_class.

![Model Export](image)

Name: churn_class
Save in folder: rstatdata

34. Ensure that the Export Target Type is Class.

35. Click Save.

We will export versions of the model that produces predicted values of yes or no for Class and values between 0 and 1 for probability for the Target variable, Churn.

36. Click the Export button again.

37. Enter a name for the model exported as a C file, churn_probability.

38. Check the Probabilities option for the Export Target Type.
39. Click Save and click Yes to confirm.

This exports the model that produces predicted values in a range of 0 to 1.

40. In WebFOCUS RStat, click Quit.
Deploying the Model to WebFOCUS

41. Select the Home Tab and click the Model Deployment button.

42. In the Model Deployment dialog, click Add in the Deployment Source section.

43. Browse to the folder where the decision tree models were exported (c:\rstatdata). Multi-select the two files:
   a. churn_class.c
   b. churn_probability.c
44. Select the Deployment Destination on the WebFOCUS Reporting Server: EDASERVE
   -> _rstat_modeling

45. Click Deploy.
46. Click OK to close the notification that deployment was successful.

47. Click the Close button to close the Model Deployment Window.

48. Close churn_train.fex

49. In the Environments Tree, double click to open churn_score.fex, a new dataset.
50. Click an empty space after the last field on the canvas.

51. Select the Data tab and click the Summary (Compute)

52. In the Compute Field Creator dialog, enter the name, churn_classification.

53. Select the Format button.

54. Select Alphanumeric for the Format Type and change the length to 5.

55. Click OK.
56. Click the Functions button.

57. In the Function Arguments dialog, click the Refresh Scoring Routines button.

58. Select churn_class from the Select a Function list box.

59. Drag and drop the fields from the tree to the Value box for Input parameters.
60. Type `churn_classification` in the Target input box.

61. Click OK.

62. Click Ok.
Create Compute field – Probability of Churn

63. On the ribbon, click the Summary (Compute) button on the Data tab.

64. Enter field name, churn_prob.

65. Set format to D12.2.

66. Click the Functions button.

67. Click on Refresh Scoring Routines.
68. Select churn_probability routine from the Select a Function list box.

69. Drag and drop the fields from the tree to the Value boxes for the Input Parameters.

70. Enter churn_probability in the Target input box.
71. Click OK on both dialogs to close them.

Note both Compute fields have been included in the report.

72. Click the Save button above the ribbon to save the report.

73. Click the Run button.

Notice the two new columns, predicted churn with Yes and No values and Churn Probability with values between 0 and 1 depicting likelihood of churn.
Section 3 – Using the RStat Adapter

Part I – Wine Price Prediction (Logarithmic)

For time limitations most of the work has been done for you and we will simply be leveraging the new Rserve adapter in this portion of the Lab.

The Rserve adapter configuration in the WebFOCUS Reporting Server looks like this:

Essentially it leverages 3 or 4 components:

- `wine_run_model.r` – This is the r script that has been created for us and predicts the wine prices
- `wine_sample_data.csv` – is the ‘training’ data used by the script to develop and test the model
- `wine_model.mas` and `wine_model.acx` – which are the metadata tables generated and used by the adapter
1. In App Studio double click on wine_consume_model.fex which is running a report against the wine data we need to add pricing information too.
2. Highlight the last position in the report as the spot we want to add the new price field in, by clicking to the right of Wine Age (Yrs) and under the Data tab select Summary Compute to open the dialog box.

3. Since the Rserv adapter is ‘hot off of the press’ we have not yet been able to update the compute dialog to accommodate the new syntax, (scheduled for the next release in late 2019) so we are going to have to type in the compute manually.
4. In the Computed Field Name use Price as the fieldname and make sure the format is D12.2, the default. Then type the following exactly as written into the main dialog box.

RSERVE(wine_model,AGST,HarvestRain,WinterRain,Age,Price)

Where wine_model is the metadata or master file for the r model are using and AGST, HarvestRain, WinterRain, Age and Price. Price being the output field. Then click OK.
5. Click Run

The price, as a logarithm is displayed.
Socialize to Win. Daily Prizes Awarded!

Share your favorite Summit moments on social media! Use #IBSummit in your posts – you just might make it on the Leaderboard.

Check our social channels for contests, updates, photos, and announcements.

@infobldrs

@informationbuildde

https://linkedin.com/company/information-builders

https://facebook.com/informationbuild