Modelling – R Models as customised HTML reporting

Release Date

2017

Summary

Modelling licence holders have the ability to create models in R and import them into Market Insight as well as creating their own RMarkdown files to display html outputs alongside R results.

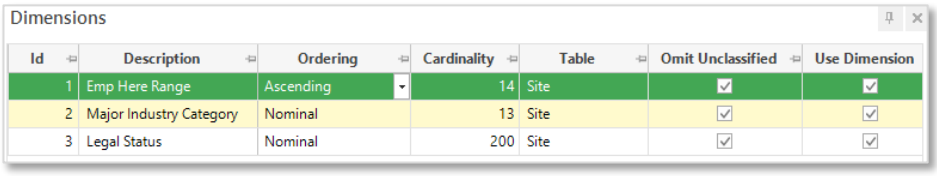
This feature is available to Modelling licence holders through the Modelling Environment and is primarily aimed at experienced R users.

For more information and help with R visit their official website - https://www.r-project.org/

The following example will illustrate the functionality of this feature using the Market Insight Training System. It will use Logistic Regression in R, to build a model that can then generate an expression in Market Insight, which will return the probability of a Customer having a Policy Product D or not.

- Drag the **Modelling Environment** tool onto the workspace
- Create a **Customer** level selection of Customers with any **Policy Product Type** of **Product D** and drag it on to the selections tab as your **Analysis** group
- Create a **Customer** level selection and set a **Random limit** of **100,000.** Drag this on the selections tab and set it as the **Base** group.
- Click onto the **Dimensions** tab and then from the **System Explorer** drag and drop the variables **Economic Region**, **Major Industry Sector** and **Legal Status Code**

If you wish to compare the R Model with the equivalent Profile and Decision Tree build those now and create the respective Virtual Variables.

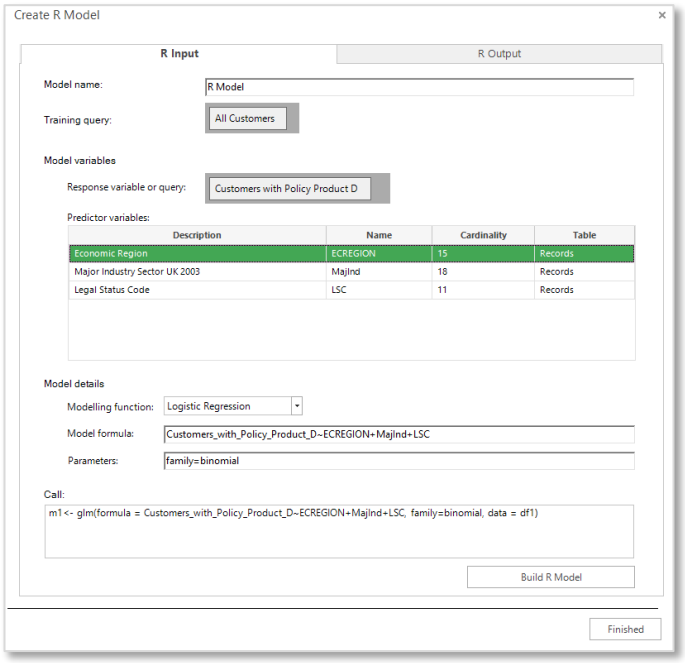

Dimensions

Id	Description	Ordering	Cardinality	Table	Omit Unclassified	Use Dimension
1	Emp Here Range	Ascending	14	Site	☑	☑
2	Major Industry Category	Nominal	13	Site	☑	☑
3	Legal Status	Nominal	200	Site	☑	☑

- To prepare your environment to be transferred to R click on the **R** icon

- The **Create R Model** window is pre-populated using the information from your Modelling environment

- Change the Model Name to **R Model**



Training Query – this identifies the records to be used to build the model

Response variable or Query – this would be a numeric value which could also be a 0 or 1 to denote a yes/no query response

Predictor variables – variables that the user has determined to be used in the building of the model

The above information will form columns in a Data Grid which will be uploaded to R.

The information contained in the lower part of the window entitled **Model details** is as follows:

Modelling functions – a list of available functions are listed here including the ability to call out to user defined functions in R

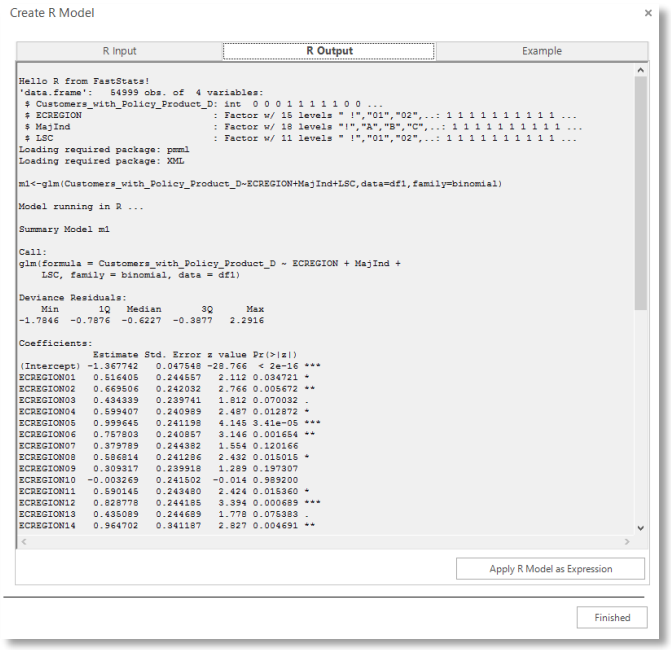
Model formula – Discoverer will automatically generate the model formula, however experienced users have the ability to edit the formula in this window

Parameters – extra parameters are required for some of the model building methods, experienced users may need to edit this especially if they have defined their own function

Call – the information that is sent to R based upon the settings made above

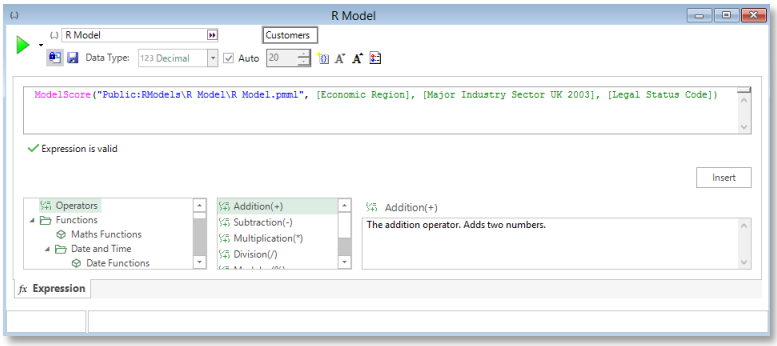
- Click the button **Build R Model**

This will export the data frame out and then bring the data back in, as shown in the tabbed window titled R Output.



A summary of the model activity is displayed and a pmml file is written into the Public directory and then accessed by:

- Click the button **Apply R Model as Expression**



Based upon the pmml file a probability score is calculated on whether we think the Customer Data has Policy Product D. The score will be between 0 and 1.

To make a comparison of the R model with say, a Profile model or Decision Tree model in the Market Insight Model Report tool, the expression will need to be converted into a virtual variable.

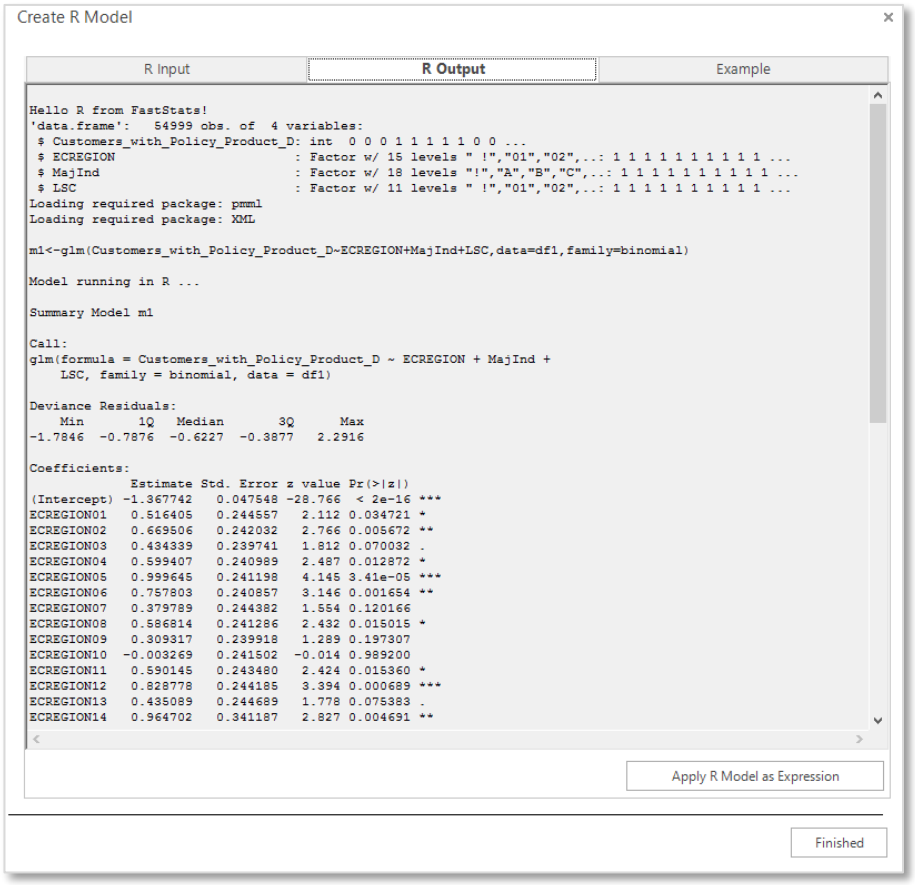
- Use the **Calculate Expression** wizard to create a numeric virtual variable of the model

The Model Report tool can only use Selector variables so the numeric virtual variable will need to be converted.

- Use the **Numeric Banding** wizard to convert the numeric virtual variable to a selector virtual variable

R Markdown Files

This feature is also available to Modelling licence holders through the Modelling Environment and is primarily aimed at experienced R users.



R users can create their own RMarkdown files (.Rmd) which will generate an HTML page with embedded R code. This code can refer to the dataframe used to train the model and the model itself. These files are placed in the Market Insight Public/RModels directory.

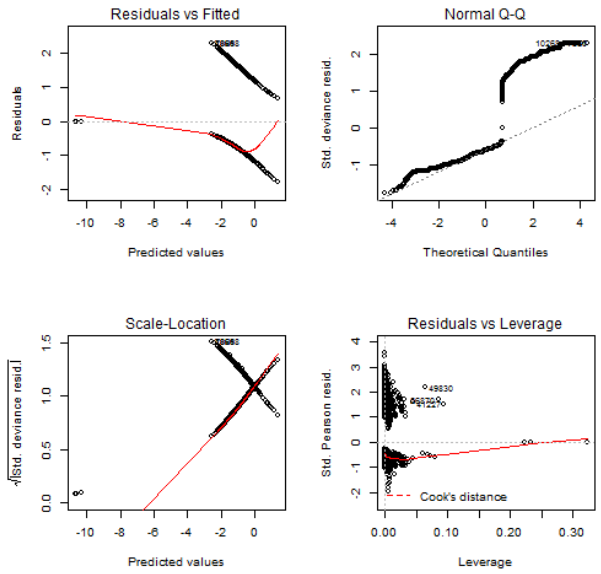
The file Example.Rmd has been created to display 4 plot graphs based upon the R model generated:

Create R Model

R Input | R Output | **Example**

```
## 4          1       !       !       !
## 5          1       !       !       !
## 6          1       !       !       !
```

Use m1 to refer to the model that has been created.



Finished