Implementing a Customer Lifetime Value Predictive Model: Use Case

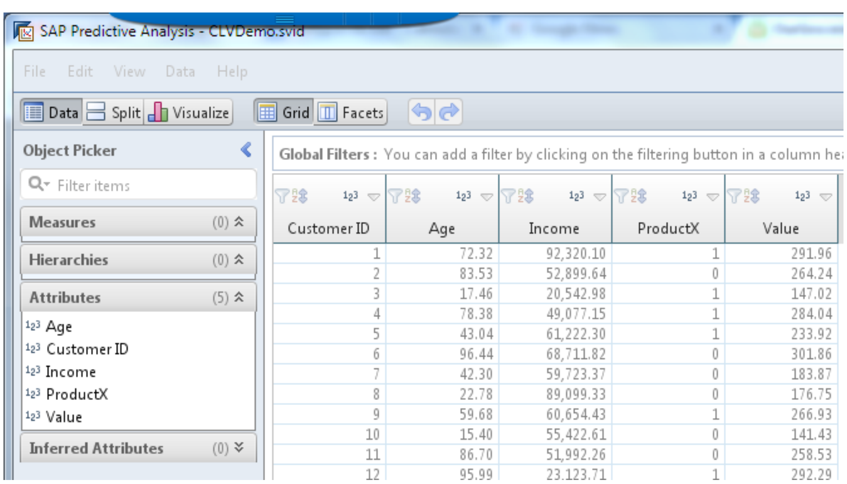
Since predictive algorithms are really just mathematical formulas that can be applied to many different problems, many organizations have a difficult time understanding how they can be applied and implemented within their existing Business Intelligence environment. Fortunately, SAP provides several tools that offer an end-to-end solution for generating and visualizing analysis sets, fitting predictive models, and implementing those models into the BI platform. This example shows a case study of an online retailer to demonstrate fitting and implementing a predictive model for customer lifetime value.

Background

AdventureWorks Cycle Company is an online retailer of bicycle supplies and parts. Through their e-commerce channel, they sell to individual customers around the world. AWCC sells a wide range of products, including both inexpensive parts and complete bicycles worth thousands of dollars. They have already acknowledged a large variance in the value of an individual customer, from low value one-time parts customers, to very high value repeat purchase customers. Therefore, AdventureWorks would like to devise a lifetime value model that takes into account a few simple variables to return an estimate of the customer's value potential.

Building a Customer Lifetime Value Model

To start this project, we've estimated and imported into SAP Predictive Analysis the lifetime value of a set of customers with extensive purchase history and a few factors we have available on newer customers: age, income, and an indicator of whether they purchased a specific product (in this case, we'll just name it Product X). This information could be from a spreadsheet or a Web Intelligence report in of the existing BusinessObjects reporting environment.



Customer ID	Age	Income	ProductX	Value
1	72.32	92,320.10	1	291.96
2	83.53	52,899.64	0	264.24
3	17.46	20,542.98	1	147.02
4	78.38	49,077.15	1	284.04
5	43.04	61,222.30	1	233.92
6	96.44	68,711.82	0	301.86
7	42.30	59,723.37	0	183.87
8	22.78	89,099.33	0	176.75
9	59.68	60,654.43	1	266.93
10	15.40	55,422.61	0	141.43
11	86.70	51,992.26	0	258.53
12	95.99	23,123.71	1	292.29

Figure 1: CLV Input Data

We then bring the R-Multiple Linear Regression algorithm into the predictive workflow on the Predict pane, since the value we want to predict (Value) is continuous.

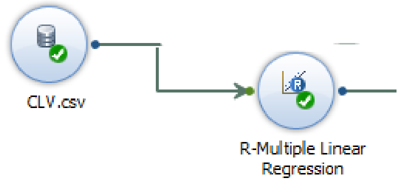


Figure 2: Predictive Workflow

Then, configure the R-Multiple Linear Regression component with independent columns Age, Income, and ProductX to predict the dependent column Value. We'll also save our model as CLVLRModel.

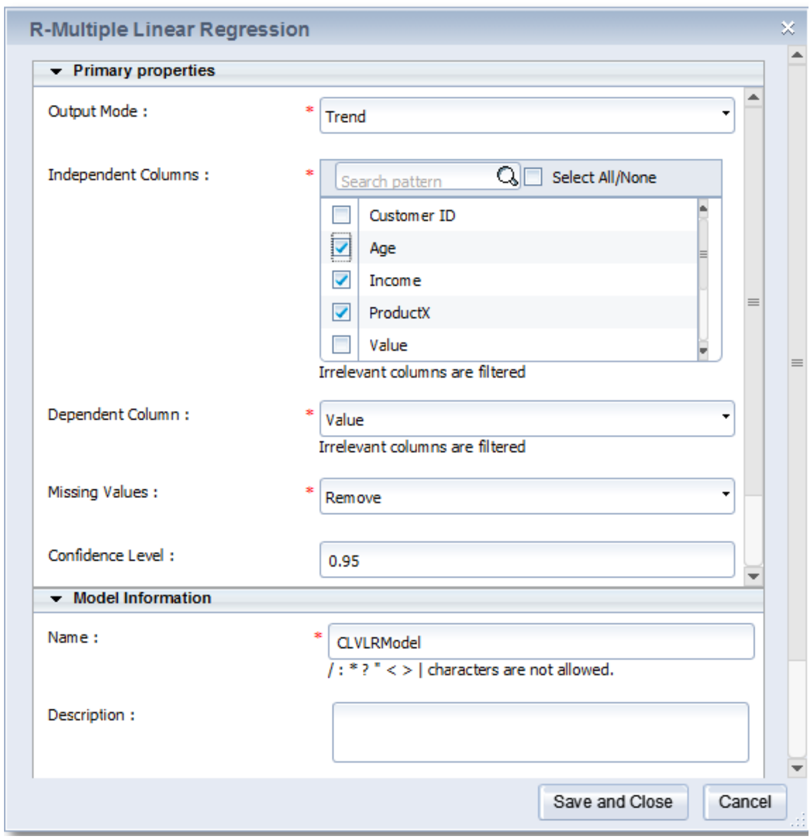


Figure 3: R-Multiple Linear Regression Algorithm Properties

Once the model has been run, the fit statistics appear on the Results pane. In this case, I'm just using all 3 variables and not too concerned about the actual fit, since this is just simulated customer value data. The 2 most important results are the saved Predictive Analysis model object on the Saved Models tab and the R model output (specifically the Coefficients section) on the Results pane.

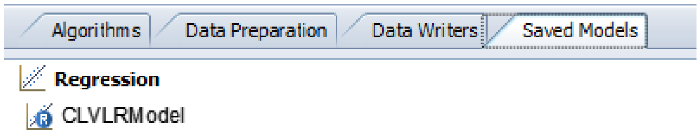
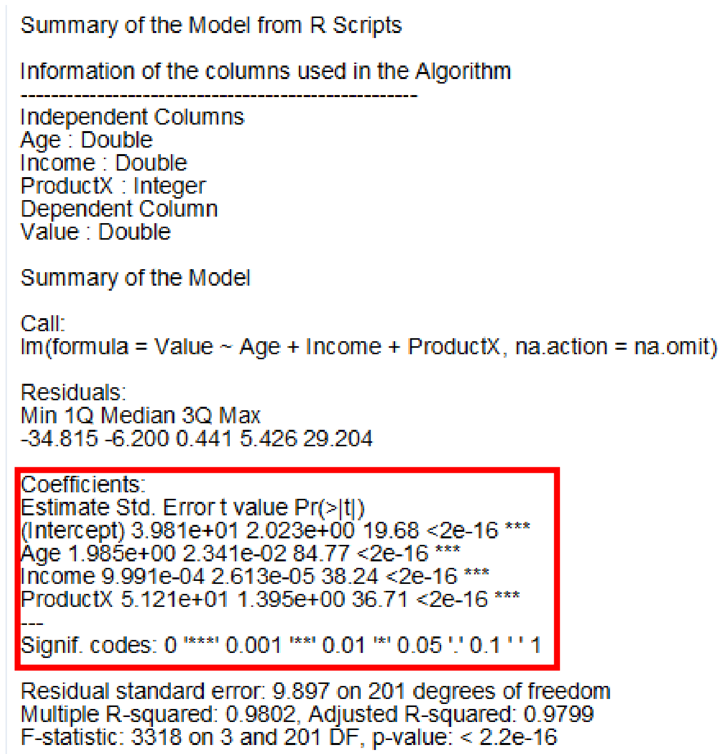


Figure 4: Saved SAP Predictive Analysis Model Object



```
Summary of the Model from R Scripts

Information of the columns used in the Algorithm
----------------------------------------------------
Independent Columns
Age : Double
Income : Double
ProductX : Integer
Dependent Column
Value : Double

Summary of the Model

Call:
lm(formula = Value ~ Age + Income + ProductX, na.action = na.omit)

Residuals:
Min 1Q Median 3Q Max
-34.815 -6.200 0.441 5.426 29.204

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.981e+01 2.023e+00 19.68 <2e-16 ***
Age 1.985e+00 2.341e-02 84.77 <2e-16 ***
Income 9.991e-04 2.613e-05 38.24 <2e-16 ***
ProductX 5.121e+01 1.395e+00 36.71 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.897 on 201 degrees of freedom
Multiple R-squared: 0.9802, Adjusted R-squared: 0.9799
F-statistic: 3318 on 3 and 201 DF, p-value: < 2.2e-16
```

Figure 5: R-Multiple Linear Regression R Results Output

With these two outputs, we have all the information we need to either score all customers in our database or generate customer values for new customers coming in.

Implementing Predictive Results

AdventureWorks's marketing and customer services teams have identified 2 core applications of the Customer Lifetime Value model in their organization. Firstly, the Customer Service team would like to know the customer lifetime value of all existing customers so they can prioritize service calls and resolution activities. Secondly, the Marketing team has identified the need to evaluate potential future customers on a case-by-case basis to determine the potential value of specific prospects. Therefore, we will implement the customer lifetime value model in 2 ways:

1. We will score and record every customer's lifetime value and store this score in the reporting mart. This can be updated regularly to reflect changes in customer profiles and score new additions to the database.

2. We will program the customer lifetime value scoring function into an SAP Dashboards application that the marketing team can run "what-if" scenarios on to evaluate specific customers.

This first task is easy—we can locate the new data (this must have Age, Income, and ProductX columns for each customer) and run our saved model and write the results back to a flat file or directly to the customer database. At that point, we can integrate the score table in the existing Universe to expose to the Customer Service organization.

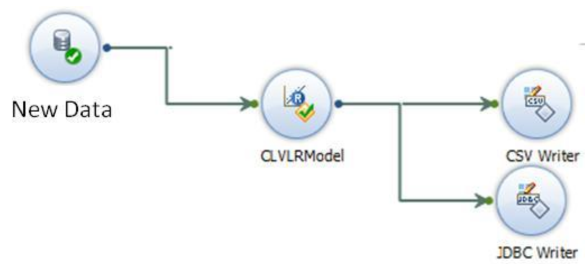
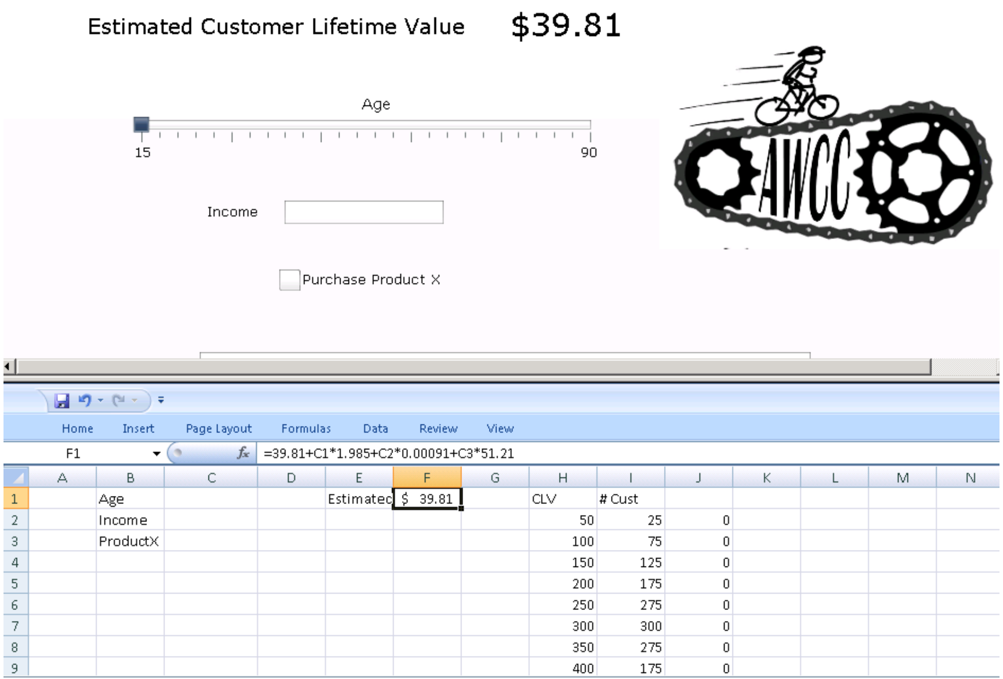


Figure 6: Scoring Full Customer Database

The second task requires defining the lifetime value scoring function, which can be extracted from the R output in Figure 5. The Customer Lifetime Value scoring equation is:

$$Value = 39.81 + Age * 1.985 + Income * 0.0009991 + ProductX * 51.21$$

With this information, the Business Intelligence developer can program this model into a BusinessObjects dashboard that allows users to interact with this function and understand the impact of changing each of these variables.



In this example, we've added a chart to show the value of a particular customer in relation to the entire customer base of AdventureWorks Cycle Company. This dashboard can be customized with the company logo and accessed through the internal BusinessObjects Launchpad or a static version can be distributed as a Flash file that any user can open and interact with.

You can experiment with this dashboard by clicking on this link:

CLV Dashboard

Hillary Bliss, Analytics Practice Lead
Decision First Technologies
Hillary.bliss@decisionfirst.com
twitter @HillaryBlissDFT

Hillary Bliss is the Analytics Practice Lead at Decision First Technologies, and specializes in data warehouse design, ETL development, statistical analysis, and predictive modeling. She works with clients and vendors to integrate business analysis and predictive modeling solutions into the organizational data warehouse and business intelligence environments based on their specific operational and strategic business needs. She has a master's degree in statistics and an MBA from Georgia Tech.