FasterAnalytics – A Diagnostic Case Study

DecisionQ has developed FasterAnalytics, a unique analytics package that enables researchers, analysts, and managers to use sophisticated predictive analytics from the desktop. FasterAnalytics is fast and creates high quality, predictive models from data that enable efficient review of clinical data, real-time hypothesis testing, and rapid decisions.

FasterAnalytics uses a modeling approach called Bayesian Networks to provide a mapping of the complex relationships in data, which can then be used to make high quality predictions. Users can:

- Get an instant global view of their data.
- Understand the driving factors in the data.
- Test hypotheses in real time in our model Explorer.
- Produce reports that can be exported to other applications.
- Make determinations that can help prioritize the use of scarce research resources.

Market Overview

Clinicians consistently desire better tools to enhance their diagnosis and prognosis efforts. The more accurate they can be in identifying diseases and expected outcomes, the better they can be at designing appropriate treatment protocols. To assist in this effort, DecisionQ has developed FasterAnalytics, a tool that can help users discover actionable insight from healthcare data. DecisionQ's FasterAnalytics software can combine data and clinician experience to create powerful predictive models, models that can be used to improve patient treatments and outcomes.

Value to the Customer

FasterAnalytics enables both experts and non-experts in statistics to discover and leverage knowledge from large quantities of data quickly. Examples include:

- Automatically mapping data where targets are unknown to reveal correlations.
- Discovering new relationships between variables and identifying new opportunities to improve care or reduce cost.
- Identifying potential morbidities early.
- Discovering populations that may have substantially different responses from the population at large.
- Predicting the behavior of any factor or combination of factors in the model.

FasterAnalytics is designed for real-time environments. Bayesian models are highly effective at identifying emerging trends that can be used to either to identify potential adverse advents or improve quality of outcomes.

Product and Technology

DecisionQ Corporation has produced a range of modules that include data analysis, modeling, visualization, reporting, and decision optimization. FasterAnalytics modules include:

- *Discretizer.* Automatically configures the data for modeling.

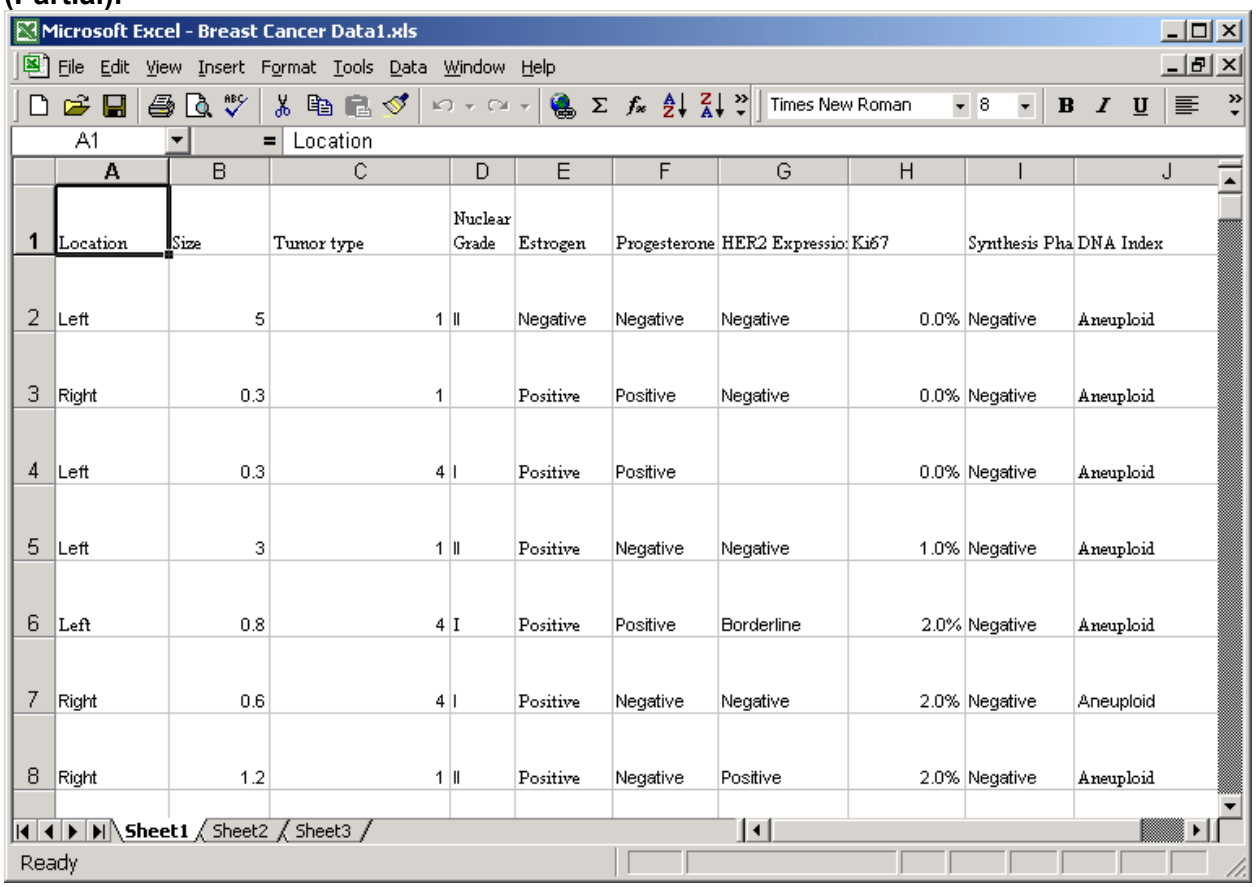
- *Modeler.* Quickly creates a visual model of the data.
- *Explorer.* Allows real-time generation and testing of hypotheses.
- *Reporter.* Extracts insights and key points for inclusion in reports and presentations.

Using the System: A Diagnostic/Prognostic Example

The following is an example application of our software to a set of breast cancer data. We have used a set comprised of 457 breast cancer patients with 10 attributes or markers. FasterAnalytics built the model in this example, from start to finish, in less than 5 minutes.

To build predictive models, our learning engine requires the data to be in a flat tabular format. The data can be numerical, or variable character strings. Our software also handles missing values automatically and will either impute a value or treat missing values as a special category, at the user's discretion.

Figure 1: This example uses a data set held in an Excel spreadsheet as shown below (Partial).



Microsoft Excel - Breast Cancer Data1.xls

File Edit View Insert Format Tools Data Window Help

Times New Roman 8

A1 = Location

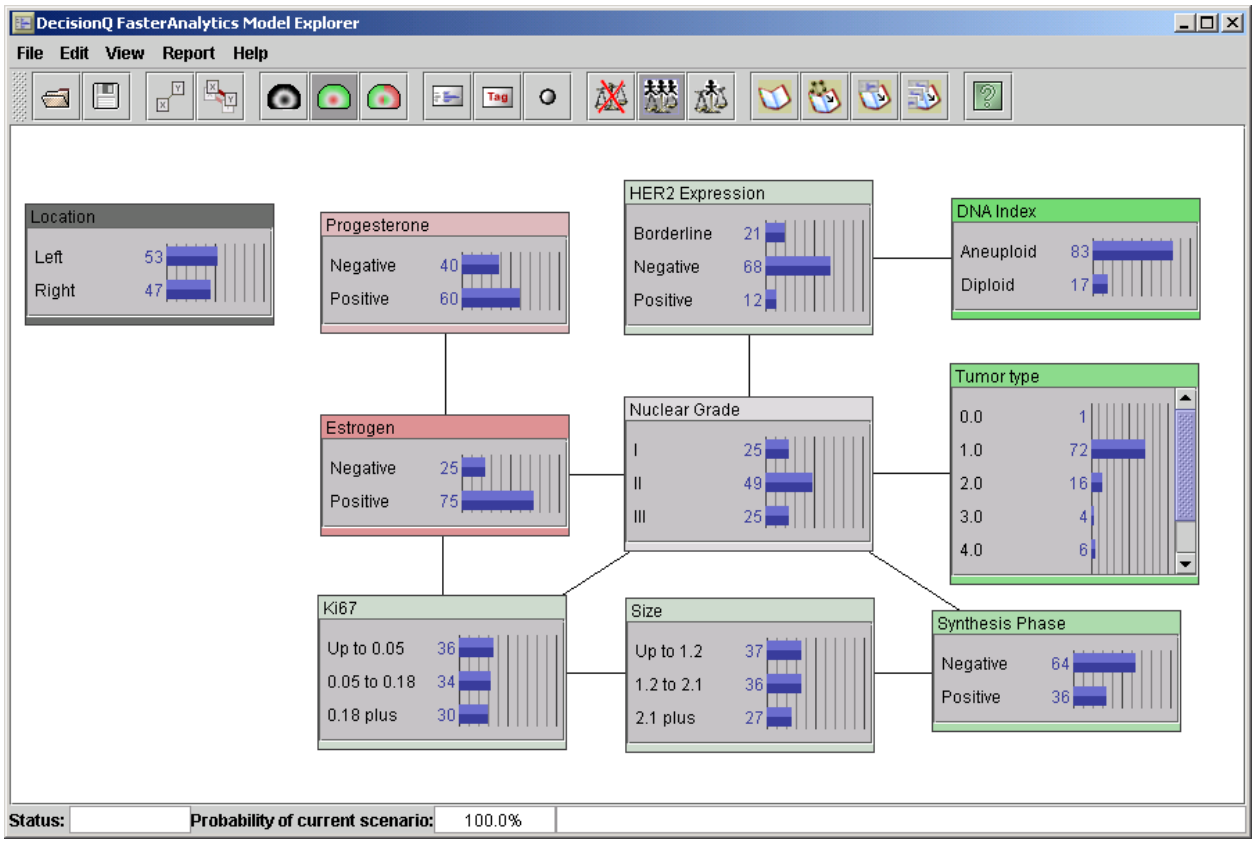
	A	B	C	D	E	F	G	H	I	J
1	Location	Size	Tumor type	Nuclear Grade	Estrogen	Progesterone	HER2 Expressio	Ki67	Synthesis Pha	DNA Index
2	Left	5	1	II	Negative	Negative	Negative	0.0%	Negative	Aneuploid
3	Right	0.3	1		Positive	Positive	Negative	0.0%	Negative	Aneuploid
4	Left	0.3	4	I	Positive	Positive		0.0%	Negative	Aneuploid
5	Left	3	1	II	Positive	Negative	Negative	1.0%	Negative	Aneuploid
6	Left	0.8	4	I	Positive	Positive	Borderline	2.0%	Negative	Aneuploid
7	Right	0.6	4	I	Positive	Negative	Negative	2.0%	Negative	Aneuploid
8	Right	1.2	1	II	Positive	Negative	Positive	2.0%	Negative	Aneuploid

Sheet1 Sheet2 Sheet3

Ready

Having selected the data, a fully automated process will continue until a full model is presented, or the user can stop each part of the process to manually change parameters. The software begins by categorizing the data and 'binning' in accordance with the default settings. The data is then passed seamlessly to the Modeler for automated model development. Once the software has mapped the complex correlations in the data a model is presented in the Explorer.

Figure 2: Base case model of the data presented in Explorer



The display illustrates conditional dependence between variables and the pathways existing in the prognostics model. Notice that the network has multiple branches, and that the data is interrelated in a "web", one of the strengths of multivariate Bayesian networks.

In the example below, we examine likely prognostic and predictive indicators that correlate with breast cancer patient outcome, and correlate with the subjective finding of nuclear grade of tumor cells. We begin by selecting our target variable, Nuclear Grade. We can see that the coloring and distribution of the surrounding nodes has changed to indicate the effect on other prognostic markers associated with the current case. Note that Estrogen Expression has decreased dramatically, HER2 Expression has increased, and cells in the Synthesis phase have increased.

Figure 3: Nuc Grade set to III (high)

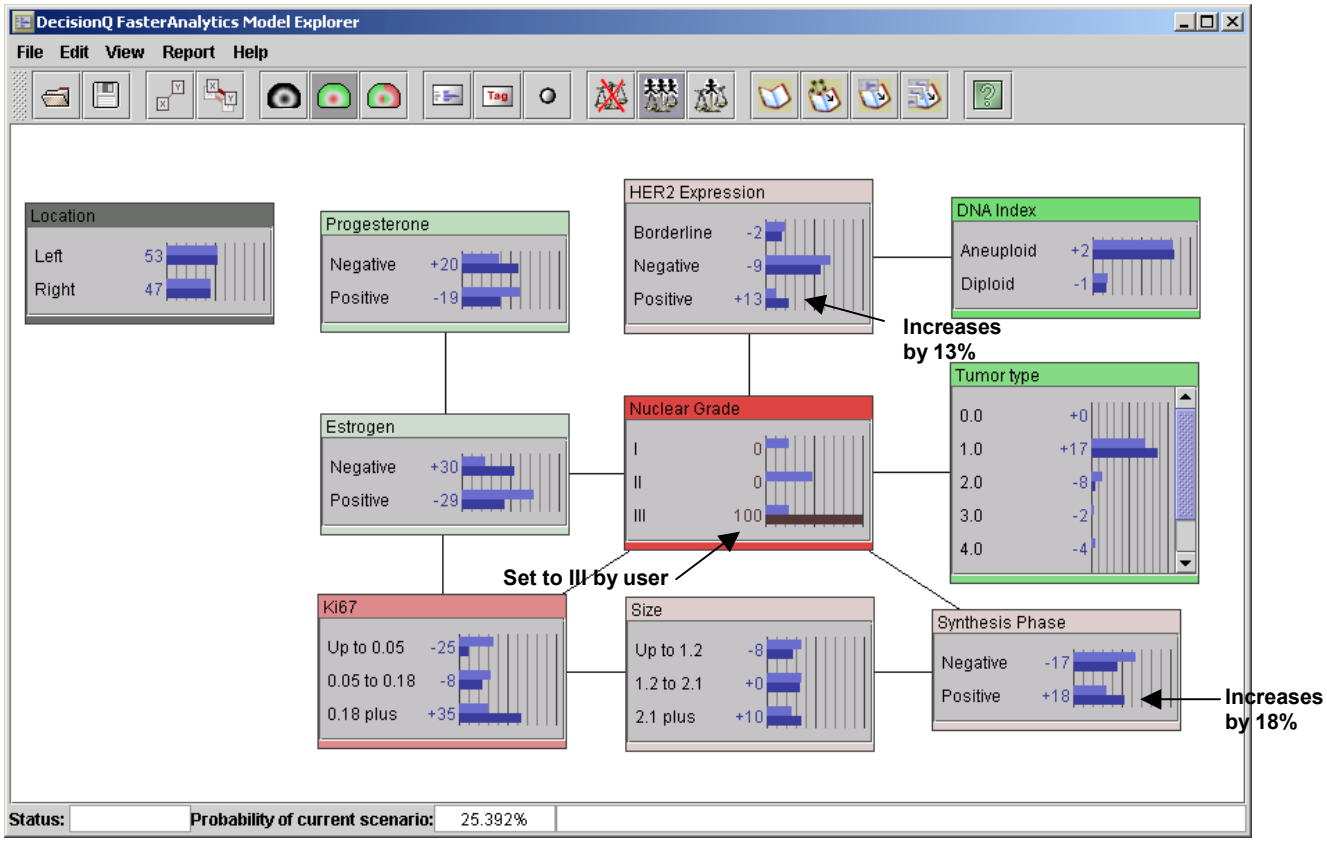
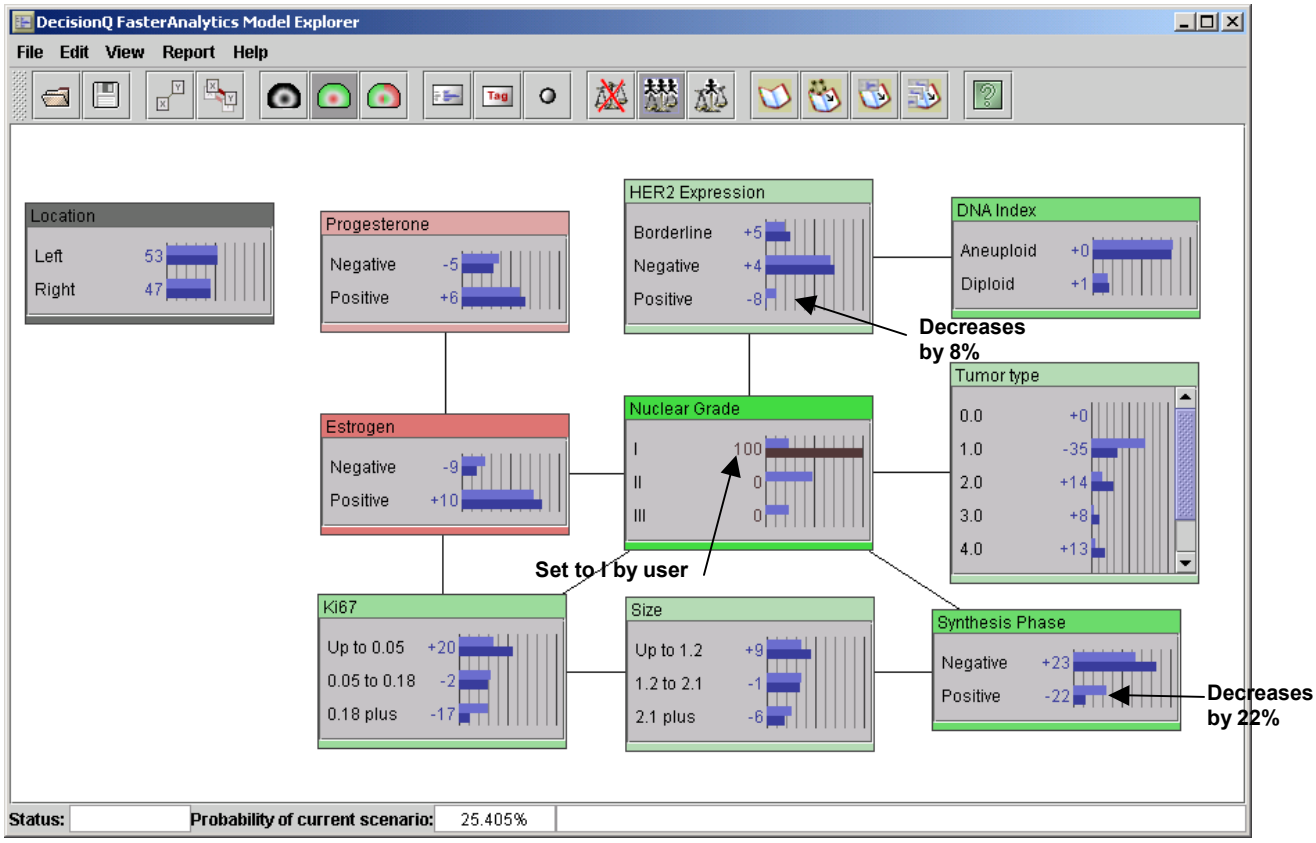


Figure 4: Nuclear Grade set to I (low)



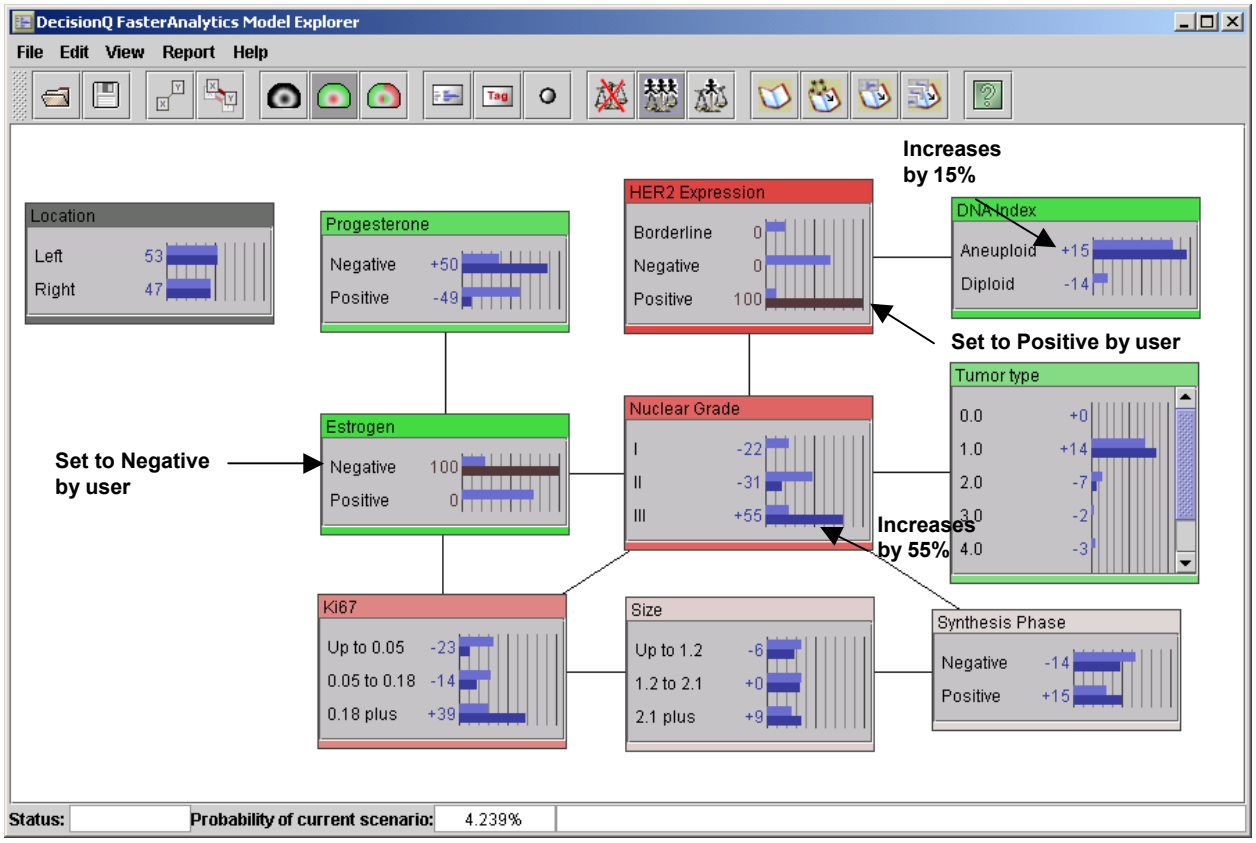
Compare the two models in Figure 3 and 4 above with the base level in Figure 2. While the Nuclear Grade shares conditional dependence with the same nodes, the behavior of those markers changes based upon the aggressiveness of the tumor. The coloring in the graphical model shows the change in population profile quickly and effectively.

It is also possible to select two or more variables simultaneously. The extent to which the HER2 oncoprotein expression and Estrogen Expression affect tumor aggressiveness can be studied together. If we wish to test hypotheses, we can modify any node and see how our hypothesis affects the model. Notice how information flows through the network.

Suppose we are interested in examining how HER2 and Estrogen expression levels affect Nuclear Grade as a prognostic indicator. We first select these nodes and click “Graph” to display the states within these nodes. This can be done for as many variables as we may choose.

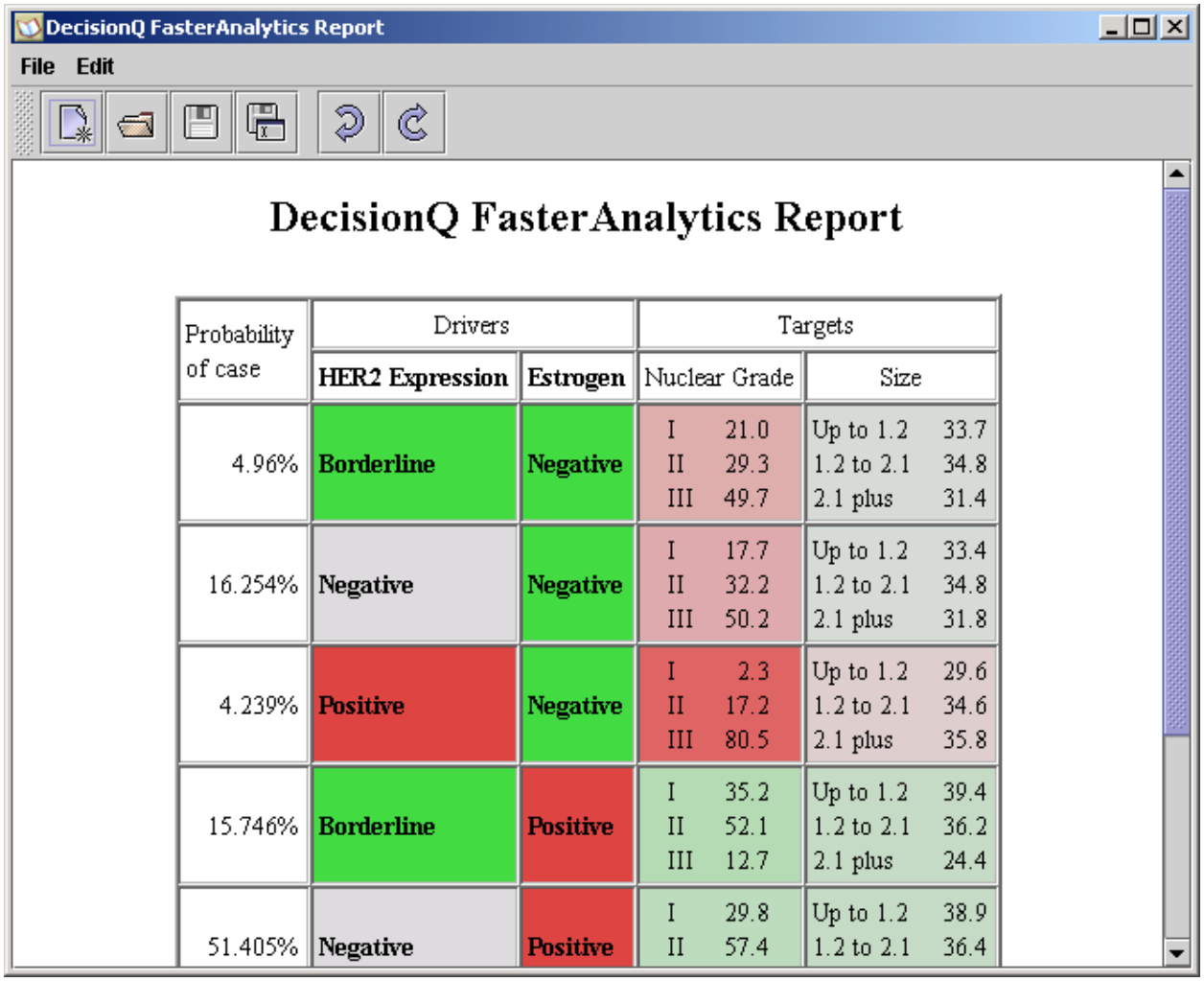
When we change the levels of HER2 and Estrogen expression, we see very clearly the prognostic indication for tumor severity and that Estrogen receptor expression is a much stronger prognostic indicator than HER2 expression. We can see this graphically in Figure 5.

Figure 5: Negative Estrogen receptor expression combined with positive HER2 expression and correlation with Nuclear Grade



The Reporter module can be used to create a report that will show the conditional probabilities (or predicted likelihood) of any target variables, given the expression of any independent variable(s). Any part of the model, visualization can be pasted into Reporter and then transferred into other applications. Figure 6 shows a sample report.

Figure 6: A sample report listing the relationship between the prognostic marker Nuclear Grade with the predictive and prognostic markers of Estrogen receptor and HER2 expression levels



Probability of case	Drivers		Targets	
	HER2 Expression	Estrogen	Nuclear Grade	Size
4.96%	Borderline	Negative	I 21.0 II 29.3 III 49.7	Up to 1.2 33.7 1.2 to 2.1 34.8 2.1 plus 31.4
16.254%	Negative	Negative	I 17.7 II 32.2 III 50.2	Up to 1.2 33.4 1.2 to 2.1 34.8 2.1 plus 31.8
4.239%	Positive	Negative	I 2.3 II 17.2 III 80.5	Up to 1.2 29.6 1.2 to 2.1 34.6 2.1 plus 35.8
15.746%	Borderline	Positive	I 35.2 II 52.1 III 12.7	Up to 1.2 39.4 1.2 to 2.1 36.2 2.1 plus 24.4
51.405%	Negative	Positive	I 29.8 II 57.4	Up to 1.2 38.9 1.2 to 2.1 36.4

DecisionQ sells predictive modeling software and complementary professional services. Alternatively, components from FasterAnalytics can be integrated into third party applications as part of broad data management and analysis platform. If you have any further questions or would like to schedule a more detailed demonstration in person or over the web, please contact us.

DecisionQ Corporation
919 Sir Francis Drake Blvd, Suite 202
Kentfield, CA 94904
www.decisionq.com
Phone: 415-254-7996
Fax : 415-276-6356
Email: info@decisionq.com