

## Model Overview

### KEY BENEFITS

- ◀ Delivers 10% better predictive accuracy
- ◀ Reduces model development time by 50-75% through automated data attributes
- ◀ Explores hundreds of data attributes to provide better insight
- ◀ Eliminates manual data preparation
- ◀ Updates models on a daily or weekly basis to reflect the most recent customer data

### APPLICATIONS

- ◀ Direct marketing
- ◀ Credit scoring
- ◀ Risk management
- ◀ Fraud detection
- ◀ Cross-sell/up-sell promotion

While the volume of enterprise data is growing exponentially, the capacity to analyze the data has remained relatively unchanged. Semcasting has developed its Model product to break the analytic bottleneck, reduce cycle times and develop predictive models that significantly outperform all others.

In traditional regression-based approaches, data cleaning and variable reduction consume large amounts of time. In the end, a small number of variables are used to build a regression model. While the skill and experience of the analyst can improve the performance of the model, the question remains the same: are you losing predictive power in the data reduction step?

Semcasting Model is a software solution that takes a radically different approach to predictive analytics. Based on patent-pending genetic algorithms, Semcasting Model allows you to create models using hundreds, rather than tens, of variables. Since the software uses more data during the model-building process, there is a greater likelihood that subtle predictors will be found. This results in models that can outperform regression-based approaches by 10% or more.

### What Is a Genetic Algorithm?

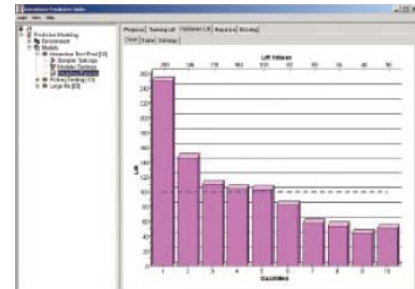
Inspired by the theories of Charles Darwin, genetic algorithms are a new way to accelerate the building of predictive models using a large number of data attributes. This approach begins with the system creating a random set of models. The models then "compete" like animals in a natural environment with the strongest or, in this case, the most predictive, surviving. Models that survive each generation then "breed" and "mutate" just like natural organisms. Each succeeding generation competes on fitness and the best (most fit) models move on. After many generations (but taking only a few hours to process), the model with the greatest predictive power emerges.

*[continued]*

## Model Overview

### How It Works

The first step is to load the data into Semcasting Model. The software can clean and sample the data, or an analyst can do it manually. Next, a dependent variable is selected and model building begins. Over several hours (usually overnight), the software creates and tests thousands of potential models, using the most “fit” ones to breed or create new generations of models. This process continues for thousands of generations and millions of candidate models until a “king” emerges with the best predictive power. The analyst can fine-tune the model and make final adjustments using an understanding of the business problem. If necessary, the software can run additional generations with different parameters. Finally, this model can be applied to the business problem through automated scoring procedures.



Semcasting Model runs on 100% Java and uses familiar tree and tab navigation

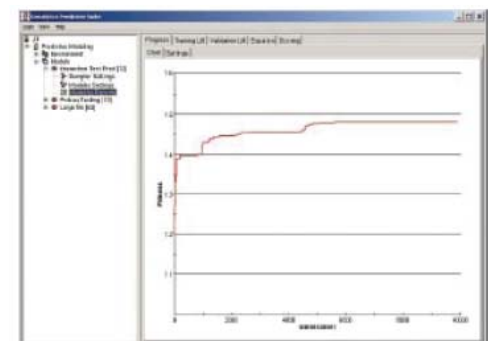
### Architecture

Semcasting Model runs on a 100% Java platform that is easy to install on Linux, Solaris or Windows. The system is scalable and distributed, and can run as a standalone or embedded application within your existing CRM or Data Warehousing infrastructure. A development API is also available. Semcasting Model also provides finished models in SAS or XML code for integration with your current operations. **Semcasting Model has five core components:**

- Database: JDBC relational database
- Sampler: Selects training and validation datasets
- Scheduler: Coordinates the activities of all other components
- Modeler: Builds and evolves models
- Fitness Evaluator: Tests each model for fitness

*You can install these components on one of multiple servers. The user interface can run on any Java-enabled workstation on the network. Additional Fitness Evaluators can be added to increase the overall performance.*

[\[continued\]](#)



Using Semcasting Model interface, you can watch as generations of models improve

## Model Overview

### System Requirements

- ▶ Relational database with a JDBC driver supporting JDBC2.0 (such as Oracle, DB2, Informix, SQL Server, Interbase, MySQL)
- ▶ Minimum configuration: Three servers, each with a dual CPU Pentium 4 or similar processor running at 3+ GHz, 1GB of RAM and at least 4 GB of storage
- ▶ 100 MB network is recommended to connect multiple servers

