

RStat: Everyone Can Predict the Future



Rado Kotorov, Ph.D.
 Technical Director of Strategic
 Product Management
 Information Builders

Until recently, predictive analytics was a branch of business intelligence (BI) only used by expert statisticians, who relied on formulas to make predictions. To compete effectively in today's business world, decision-makers at every level of an organization need access to predictive modeling applications. Police officers need to determine where crimes are likely to occur so patrol cars can be in areas where they are most needed. Marketing managers need to predict who is most likely to respond to an e-mail blast or ad campaign. Auto insurance personnel need to create risk profiles based on the likelihood of certain individuals filing claims.

It's time to place predictive models into the hands of operational users to empower proactive decision-making. Most reporting applications do a good job of recording what has happened. But that's just a rear-facing view of the business. They don't provide guidance about future actions. That's where WebFOCUS RStat comes in.

Facing Forward

With RStat, Information Builders bridges the gap between the rear- and forward-facing views of business operations. RStat provides the means to deploy predictive models as scoring applications so operational users can make decisions with confidence instead of relying on their gut instincts. Information Builders now offers the first fully integrated BI and data mining environment for developing predictive models and distributing scoring applications to operational users. Let's look at the fundamental components.

Looking Under the Covers

Data mining extracts historical data and then applies statistical techniques to build a model. Traditionally, highly trained analysts and statisticians built these models. But unless their results are widely deployed, they end up as isolated research products, doing little good for the business.

A scoring application deploys analytic models for repeated use on new data sets by non-technical users to support decision-making. For example, a marketing analyst would use a scoring application to score new mailing lists to screen for the best possible respondents. In simple terms, the scoring application labels a prospect as either good or bad.

Statisticians spend much of their time extracting and querying data. But by working in the same BI environment, developers can create queries that statisticians reuse to create models. The statisticians can compile their models as standard WebFOCUS functions that BI developers turn into WebFOCUS scoring applications, deployable on any platform. There's no need to work with multiple tools or pay for extra licenses. By unifying BI and data mining environments, RStat reduces licensing costs by consolidating software tools. This has the corollary effect of simplifying maintenance and making optimum use of IT resources.

But perhaps more important than streamlining and simplifying these BI processes is being able to generate accurate results – a function of the underlying analysis tools. To that end, RStat is built on the open source R engine.

R is well known as the most powerful and flexible statistical programming language available. It is used by more than a million analysts worldwide, is taught in countless universities, and has more than a thousand packaged extensions for various types of analysis exercises.

Because R is open source, RStat eliminates all statistical software licensing costs. Information Builders customers pay only maintenance and support to use this powerful language. And because RStat is fully integrated within

WebFOCUS Developer Studio, customers can often eliminate software packages that are used primarily for one-off BI functions, like query and analysis.

Explaining the Mystery

Of course, on a practical level, it all comes down to usability. If a tool or solution is beyond the grasp of end users, it simply cannot be effective. R is a powerful scripting environment, but it is designed for

By unifying BI and data mining environments, RStat reduces licensing costs by consolidating software tools.

technical users. RStat integrates R within WebFOCUS Developer Studio. This user interface provides an easy and intuitive workflow. It incorporates the top 10 most commonly used data mining routines, including regression decision trees, neural networks, clustering, association, support vector machine, and other algorithms that are familiar to business and engineering students. A wide variety of users, including statisticians, business analysts, and other professionals, can easily use RStat to develop models.

The unique capability to compile models as native WebFOCUS functions directly in RStat enables an organization to spread data mining benefits to operational

users. Those users will find the scoring application familiar and accessible. With a click on a Web-based form, they can generate scores and predictions to support decision-making without having to know anything about data mining and modeling. This ease of use and scalability will drive adoption by more and more users in the predictive modeling environment.

And that's what it's all about – broadening the reach of predictive analytics to encompass the workers who need it most. You don't need a crystal ball to foresee the benefits: good fortune for the entire organization. 🌟