

# Software Review

REVIEWED BY PAUL H. LASKY

## Insightful I-Miner 2.0

Insightful Corp.  
1700 Westlake Ave. N., Suite 500  
Seattle, Wash. 98109  
www.insightful.com

**LEVEL:** I-Miner is a platform and its use depends upon how it is loaded. In its basic form (reviewed here) operation is simple. No understanding of statistics beyond the basic concepts of probability, average, standard deviation, correlation and histogram distributions is required.

**OVERALL RATING:** 

**PRICE:** \$9,995

**RECOMMENDED SYSTEM:** Windows 2000/XP/NT 4.0 (service pack 6.0 or later), 512 MB RAM.

**W**hen a software program not only addresses the needs of a current market, but creates new markets among non-experts of its field, it has something going for it. I-Miner is just such a piece of software in that it brings capabilities of advanced data-mining to non-“techies” such as most traders, system designers, trading advisors, analysts, brokers and anyone desiring to unearth patterns in large data sets and build predictive models.

I-Miner uses a clean and intuitive graphical interface and drag-and-drop technology whose basic operations can be completed without reference to any manual. No programming is required to run the platform.





I-Miner’s other big benefit is the use of “pipeline architecture” to handle data. This innovation allows I-Miner to run enormous data files up to

tera-byte size on a personal computer rather than requiring an enterprise-size server. For example, a spreadsheet of 5 million rows by 100 columns (about four gigabytes of data) can be analyzed in a data processing network composed of 100 individual data processing programs or nodes on a typical PC with 512 MB of RAM. All this is achieved without the use of parallel processing.

**OPERATION:** 

The main I-Miner interface contains three windows: a left pane that contains a directory of icons each consisting of a sophisticated data processing program; a larger “worksheet”; and a smaller pane that reports the progress of the various icon/program executions. The user drags and drops the icons/programs onto the worksheet, then connects the icons with clicked-on lines to build a data processing network. “The I-Miner network” (below) shows a portion of a simple I-Miner processing network that reads data, performs preliminary data manipulation, builds three different predictive models, trains and tests each

### RATING SYSTEM:

|   |             |
|---|-------------|
|  | = Excellent |
|  | = Good      |
|  | = Adequate  |
|  | = Poor      |

model, then selects the best model from among the three.

Designing an I-Miner data processing network is something of an art. But the graphical user interface allows the analyst to concentrate on the data processing design and model selection tasks while ignoring the details of parameter selection in the statistical programs standing behind the icons.

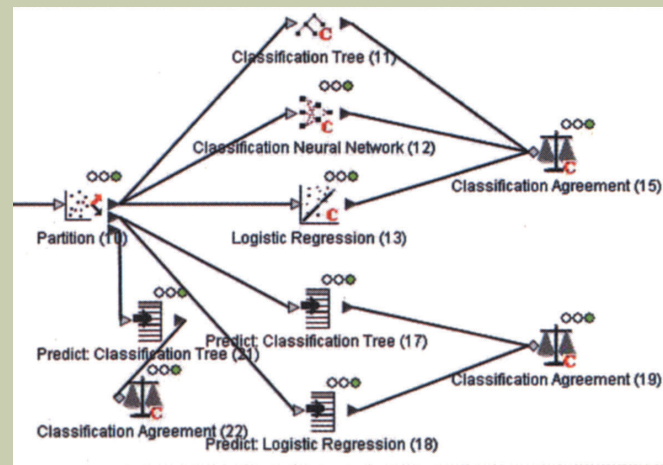
I-Miner runs quickly. The three model example displayed processes 157 MB of data through 27 icon/programs in seven seconds on a 3 Ghz PC with 512 MB RAM. With such speed, a user can tweak the models and filter the inputs to find the optimum models in a few minutes or an hour or two. While the same task could be accomplished using other software, even sophisticated statistical programs such as S-Plus 6.1 or SAS would require days or even weeks due to less-efficient memory management and processing.

**FEATURES:** 

Although the cost of the I-Miner 2.0 is approximately double the competition, some additional features help justify that cost. Icons, which in I-Miner’s graphical-oriented terminology refer to the programs within the package that do the work, are provided for selecting data, preparing data, transforming data, modeling data, summarizing the data in plots and graphs, and most impor-

### THE I-MINER NETWORK

The data is partitioned into train, test and validate sets, and three different models are trained and tested. The model performing best on the test data, a Classification Tree, is selected and validated.



tant for users building trading models, validating the models.

Provided are five different icons for reading data and five more for outputting data to files of numerous formats. Six different icons are available for graphing, tabulating, correlating and summarizing. Data cleaning is important, and I-Miner provides two icons for dealing with the tenebrous problems of outliers and missing data. Sixteen icons are available for data manipulations, such as combining or normalizing it.

The nine modeling icons are the heart of the program. They include multiple regression, regression trees, regression neural nets, logistic regression, classification trees, classification neural nets, bayesian classification, K-means clustering and principle component reduction.

Two different icons are available to aid the analyst in the subjective task of assessing the value of a classification model: one icon displays three “lift charts” including a Receiver Operating Characteristic Chart (ROC), a valuable tool rarely available in software, as well as the standard classification matrix (affectionally termed the “confusion matrix”). These icons should, but do not include valuable objective assessments such as Chi-squared P-values nor the area above the ROC 45 degree line.

I-Miner can be programmed, specialized and extended for advanced tasks with five S-Plus icons, a task best left to experienced programmers.

A negative perception in the data and finance modeling world is that after performing so many complicated data manipulation and modeling steps, analysts don't document in detail what they have done. As a result frequently they can't repeat the steps. I-Miner addresses this problem with a graphical workspace that depicts the processing steps, saves and documents the whole complicated effort. New data can be validated by simply recalling the workspace and running the new data through it.

One feature could be improved. The I-Miner uses a simplistic segmentation procedure to increase the accuracy of classification trees. A true recursive “boosting” procedure that builds increasingly more powerful trees could be included that would increase the power of I-Miner's boosting operation.

### DOCUMENTATION AND HELP:

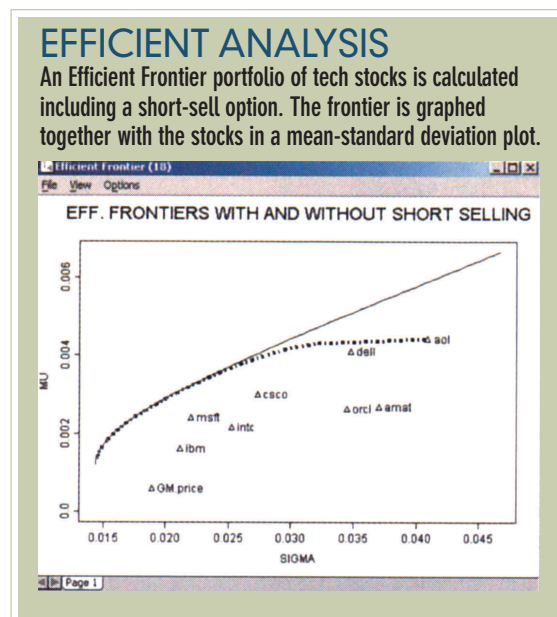


Insightful's manual is available both in hard-copy and as a PDF online file. It includes two detailed tutorials involving moderately large data sets both of which are included. Completing both of these tutorials is adequate training to allow a user to begin to build predictive models.

The manuals include no formulas. Everything is explained in non-technical language free of statistical lingo. Emphasis is placed on teaching a disciplined step-by-step approach to building data processing networks.

The manuals contain no explanation of the important technical concepts that underlie the icons/program. Although most users probably don't care which definition of “deviance” is employed in the classification tree model, some knowledgeable users may wish to have this knowledge to better choose from among the small number of options available. The manual should include a complete glossary containing this kind of information. Once analysts graduate from neophytes to experienced users, most would appreciate a fuller explanation of the pro and cons of icon property choices and trade-offs to be expected.

Finally, the explanation of how to program I-Miner with S-Plus is too truncated to be understandable. Granted, most users will never wish nor need to



program I-Miner, but the topic should be more fully covered for those who wish to customize the software.

### SUMMARY & EXTENSIONS

Insightful recently made available a financial-oriented extension of I-Miner. The suite includes the ability to build processing icons or nodes from other Insightful products, including its non-linear optimizer, NuOpt, and its time-series analyzer Fin-Metrics. Also available is the enormous database of historical intraday security prices from the University of Chicago's Center for Research on Stock Prices (CRSP).

The suite integrates the CRSP database and lets users access it with one icon and use the efficient frontier icon with NuOpt to build optimal portfolios from any subset of listed stocks. Another icon can be linked up to produce a tabbed work-book of price charts.

Using the I-Miner with these newer icons together with the CRSP database greatly extends the platform's usefulness to finance analysis topics.

**Paul H. Lasky, a southern California lawyer and statistical consultant, has traded financials for more than 25 years. E-mail: phlasky@earthlink.net.**