

Alpine Data Labs provides an enterprise solution to the problems of data access, model building, model deployment and model performance monitoring for predictive analytics applications. As data mining and machine learning technologies move out of the realm of esoterica, so organizations will need a full implementation and management environment. I spoke with Steven Hillion, the Chief Product Officer at Alpine Data Labs and he was very aware of the need to satisfy the requirements of business management, as well as those of IT and data scientists. Many analytics technologies focus on the technical aspects with scant regard for the monitoring of model performance and the sharing of information in a collaborative environment. Although this is one of the less glamorous aspects of predictive technologies, in many ways it is one of the most important. Without the means to establish confidence in predictive models the technology will always be underexploited and untrusted.

One of the central philosophies of the Alpine approach concerns the movement, or rather lack of it, of data. Most data mining technologies require that data be extracted from the production or data warehousing environment and set aside for model development. Alpine takes a different approach and executes data exploration, data transformation, modeling, testing and deployment in the native database environment. In practice this means that models are created and execute in the same operating environment and on the same hardware as the database. Hadoop, Greenplum and most relational database systems are supported.

One of Alpine's central messages is a direct outcome of its ability to access data sources directly. Cross functional analytics is certainly possible with Alpine, and in some organizations will be possible. In reality however, highly stovepiped businesses will have a hard time exploiting this capability.

The user interface to Alpine's capabilities is through a the web browser, accessing a server that simply functions as an interface to data resources and as the broadcaster of the web based environment. At the data exploration phase it provides a plethora of visualization tools - frequency diagrams, box plots, scatter charts and so on. It provides the tools for data transformation modeling, testing and scoring.

All the popular mining algorithms are supported and there is ongoing activity aimed at creating fast implementations of these. Logistic regression and support vector machine are two methods which have been subjected to this treatment. Random Forests comes in two flavors. One is essentially exploratory in nature and executes faster than the full blown implementation, which is more likely to be used for creating a final model.

A new feature called Chorus is one of the stronger differentiators. This supports strong collaboration between data scientists, IT and management, and allows information pertaining to models to be freely shared. It's an effective way of building up a knowledge base for those working on and using models, and will eventually give management the visibility they need. It seems likely that Alpine will partner with BI and data visualization tools providers to open up the analysis of data, and performance monitoring to a wide audience.

Alpine is one of a new generation of predictive analytics solution, providing a platform that satisfies all those with a stake in the exploitation of data mining technologies, and specifically IT, data scientists and management.

www.alpinedatalabs.com