

Construct, validate and deploy recursive partitioning models.

Use with both continuous and binary descriptors.

Apply for mining large volumes of data, such as for producing models for virtual screening of compound libraries.

Decision Trees Collection

The Decision Trees Collection provides components for the construction, validation, and deployment of recursive partitioning models. Using ArborPharm™ technology from NovoDynamics® (www.novodynamics.com), these components work with both continuous and binary descriptors, such as molecular structure fingerprints. This method is ideal for mining large discovery data sets like those from high throughput screening.

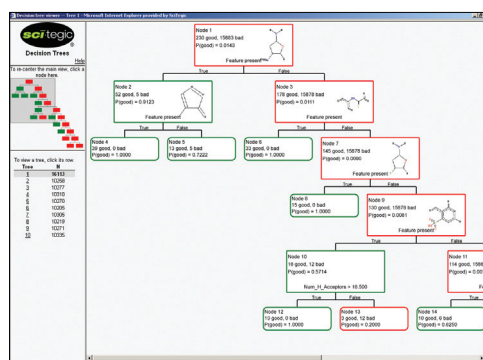
With the Decision Trees Collection you can:

- Model huge volumes of data with massive descriptor sets
- Learn the importance of each descriptor to the response (such as activity)
- Screen compound collections or virtual libraries with predictive models
- Deploy self-contained validated models to a broad audience

observations per node where splitting should cease. A cross-validation component allows the tuning of these parameter to ensure an optimal final model is built. When employed in conjunction with the novel Extended Connectivity Fingerprints (available separately in the Chemistry Collection), it is ideal for building virtual screening models of high throughput screening data.

View Decision Trees

With recursive partitioning, it's easy to interpret the models that are produced. Paths through the decision trees leading to “good” nodes provide sets of rules for “good” data, including structures. For example, these trees make it easy for you to discover rules suggesting that active molecules should have certain structural motifs (and not others) and have physical properties within certain ranges (and not others). The View Recursive Partitioning Tree component allows the trees to be displayed with chemical structural features making it possible to discover these rules.



Learn Decision Trees Models

The Learn Decision Trees model component rapidly produces a two-class decision tree from large amounts of data using a recursive partitioning algorithm from NovoDynamics. It allows control over the underlying algorithm so you can maximize the accuracy and predictive ability of the model. Parameters are available that specify the number of trees to generate in an ensemble analysis, the maximum depth of each tree, and the minimum number of

Validate Decision Trees Models

It's important to validate and assess models for predictive power and accuracy before deploying them. A component is available that generates an enrichment plot for a set of test data, typically data that is withheld from the model building process. With the separately available Modeling Collection, you can calculate an ROC plot to further assess the model's quality.

Use Decision Trees Models

Once generated, a Decision Trees model becomes a standard property calculator component in Pipeline Pilot. Any new data can then be passed through the component and its “goodness” predicted. Use the model component in any Pipeline Pilot protocol, including those deployed via the Web interface to a broad community of potential users.

© 2008 Accelrys Software Inc. All brands or product names may be trademarks of their respective holders.



Accelrys Corporate Headquarters
10188 Telesis Court, Suite 100
San Diego, CA 92121-4779, USA
Tel: +1 858 799 5000

Accelrys European Headquarters
334 Cambridge Science Park
Cambridge, CB4 0WN, UK
Tel: +44 1223 228500

Accelrys Asia Headquarters
Nishi-shimbashi TS Bldg 11F
Nishi-shimbashi 3-3-1, Minato-ku,
Tokyo, 105-0003, Japan
Tel: +81 3 3578 3861