MineTool User Manual

SciberQuest, Inc.

MineTool is a data mining tool for building classification models from static and time series data. MineTool is easy to use and does not require prior data mining expertise. In its simplest form, MineTool works by reading in the dataset and pressing the Run button with the default parameters. From a user perspective, the steps in data mining using MineTool can be illustrated as follows (Figure 1):
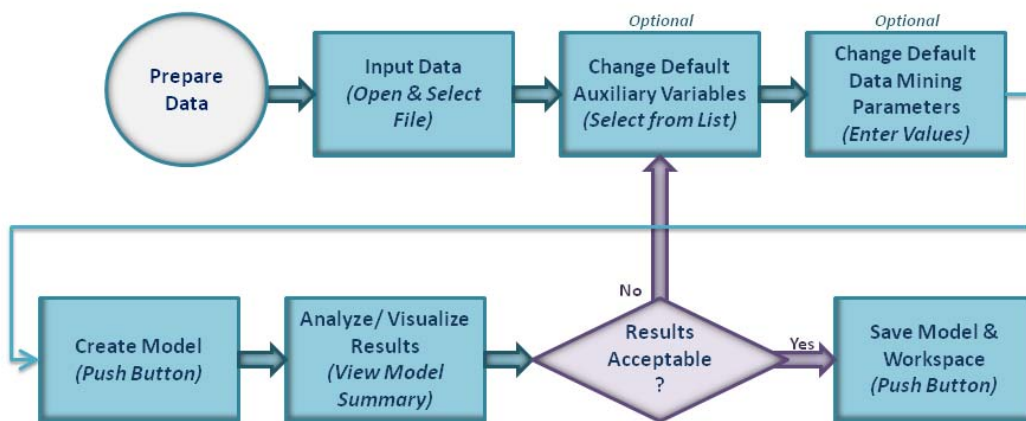
Figure 1. User steps to modeling in MineTool

Using the default parameters, the user can create a static model of the input data simply by reading in a file and clicking on a button. To create a time series model, the user needs to define how many time observations there were in a time series, press then

Convert Time Series button, and press the *Run* button. MineTool also offers the user the ability to change the default settings to experiment with model accuracy.

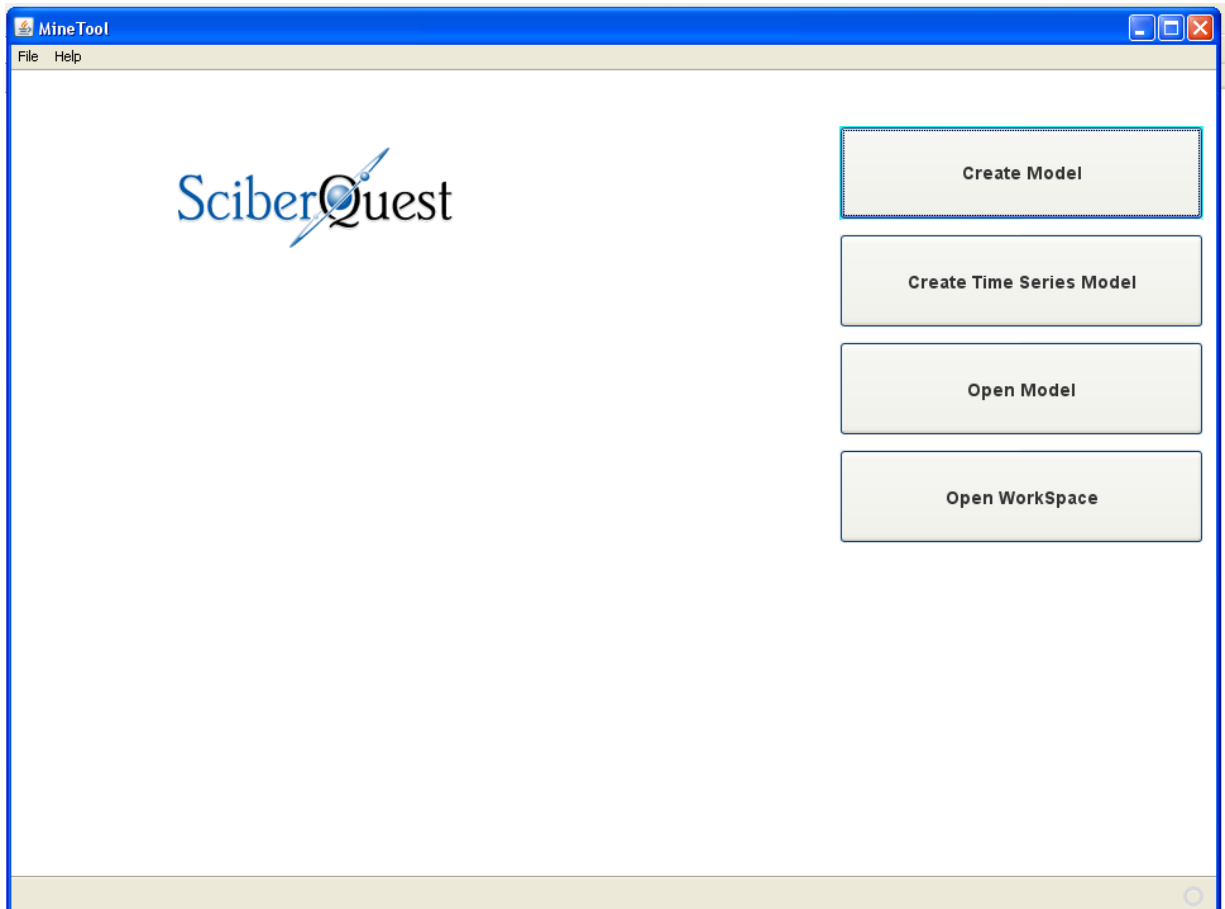The MineTool suite of tools consists of the following four applications:

Figure 2. The main application suite window of MineTool.

Create Model allows the user to read in a static data, perform data preparation steps, select variable transformations, perform modeling, visualize the data, and finally, save the model and/or the workspace.

Create Time Series Model allows the user to read in a time series dataset, define the time series parameters, convert the time series data into a static dataset that contains all the important time-varying information, and then proceed with the modeling just as in the Create Model application.

Open Model enables the user to open a previously save model and score it or run it (i.e. apply it) on a novel or unseen data.

Open Workspace facilitates opening a previously saved workspace, where the input file, selected variables and all the other options of the GUI are preserved. This allows the user to effectively remember the steps they used in the predictive modeling, and continue their investigation and analysis of the data.

The Create Model Application

To start the modeling process, the user Selects an Input Dataset by clicking on the Browse button and selecting an input file. The file needs to be a white space (tab or space), or comma separated file. The variable characteristics and statistics are displayed immediately. The use has the ability to select which variables are to be used in modeling, and which are only illustrative (such as ID, instanceNo etc.). The user also may select some variables for preprocessing, Pressing the Preprocess Selected Variables button makes a menu of data filters appear. The following preprocessing methods are available: Normalize, Interpolate, Smooth, Remove Outliers and NominaltoBinary. The latest file operation can be undone by clicking on Undo, or saved by pressing Save File.
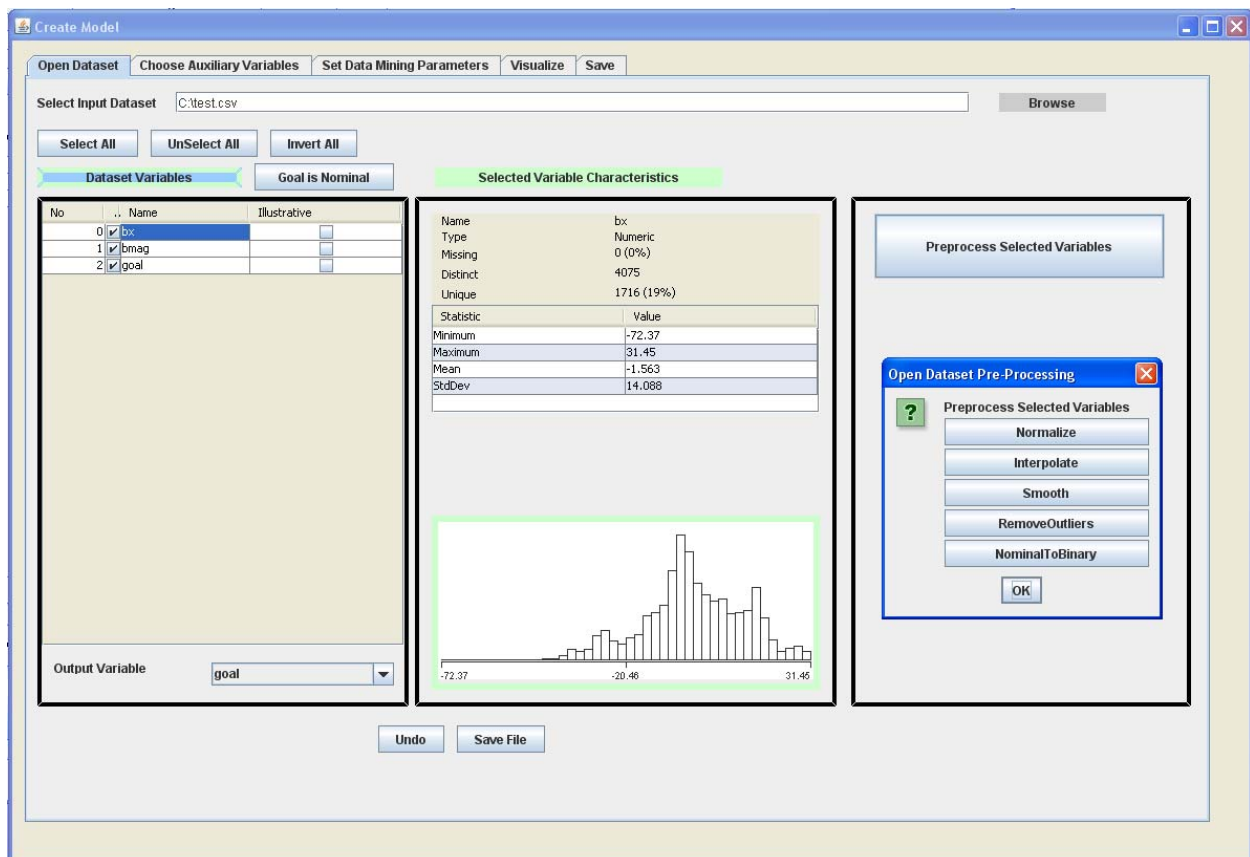

Figure 3. Open Dataset Tab in Create Model application.

To continue to modeling process, the user then moves to the Choose Auxiliary Variables tab. Auxiliary variables are additional variable transformations that can be added to the input dataset. Level-2 transformations create new variable such as $a*b$, a/b, a^2 and similar, out of any two given variables in the dataset. . Level-2 transformations create new variable such as a^2*b, a/b^2, $1/a*b^2$, a^3 and similar, out of any two given variables in the dataset. Additionally, the user can apply exponential, logarithmic, hyperbolic tangent, sine and cosine functions to any of the level-2 or level-3 transformations. Finally, these transformations and the original input variables can be non-linearly transformed using neural network like functions, such as logistic, radial-basis and ridgelets functions. The user may also save this expanded dataset. By default, the expanded input dataset is saved in fTransfromations.txt file.
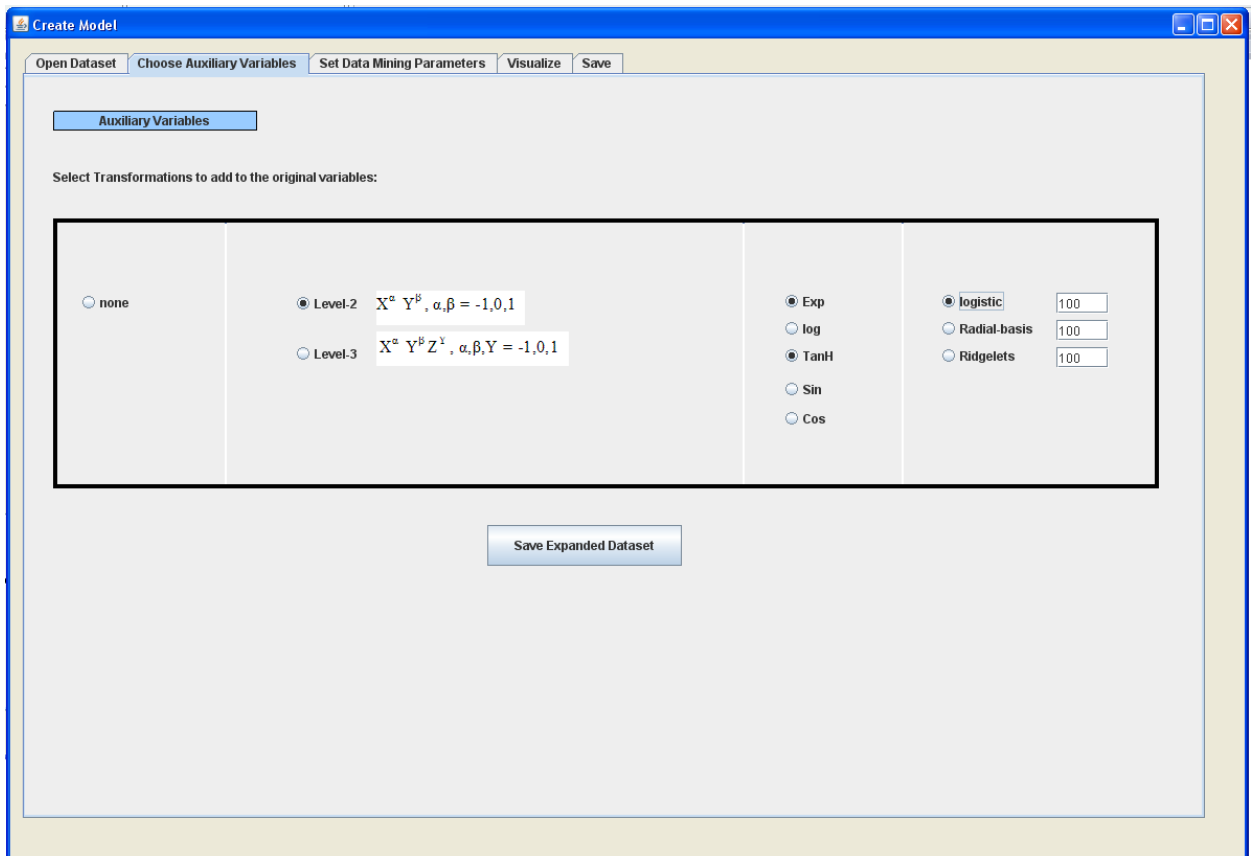

Figure 4. Choose Auxiliary Variables Tab in Create Model application.

To move on even further with the modeling process, the user moves to the Set Data Mining Parameters tab, where the colinearity parameter lambda is set, the evaluation method is defined and the methods for choosing the best mode is specified

as well. This tab contains the Run MineTool button that starts the modeling. The modeling results and evaluation scores are presented in the working window to the right of this tab.
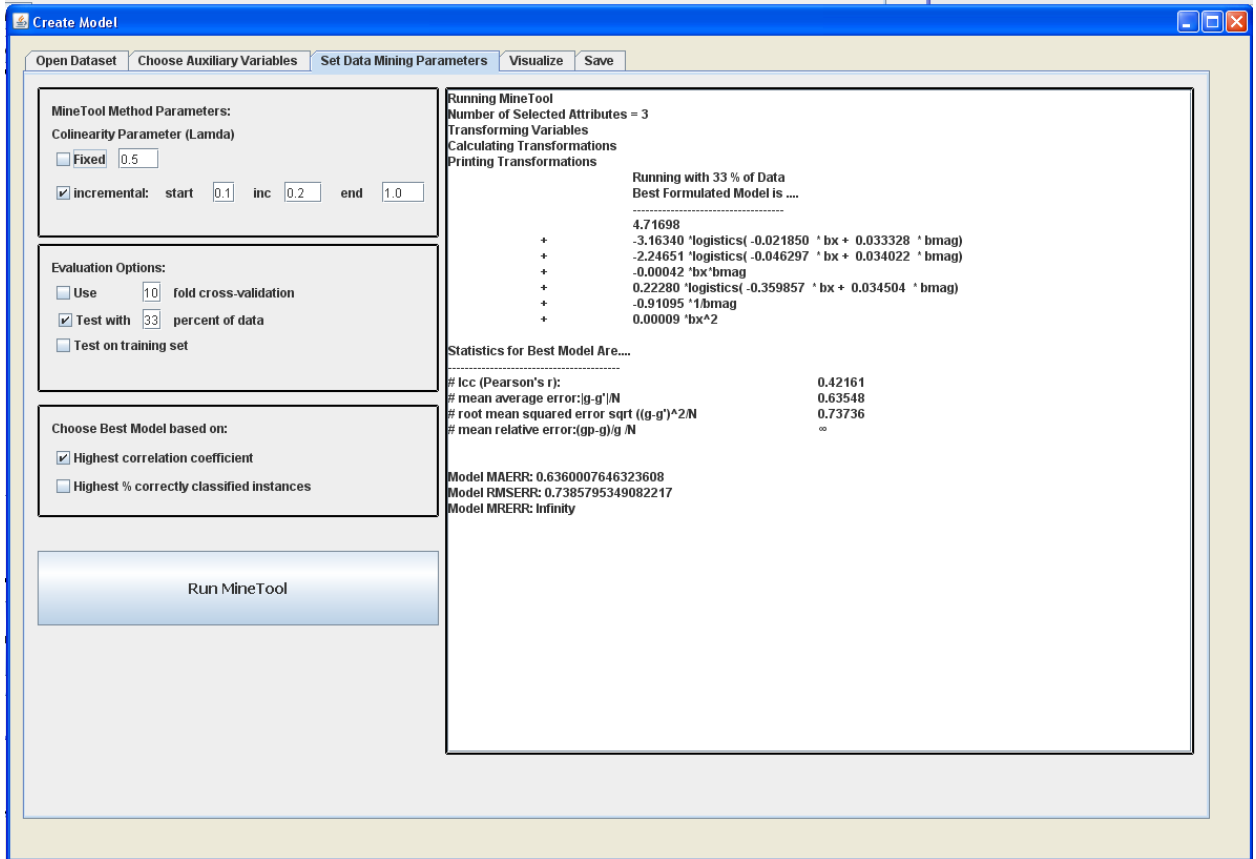

Figure 5. Set Data Mining Parameters Tab in Create Model application.

The user is also able to visualize the variables in our Visualize tab, that plots any two given input variables against each other.

By simply clicking on any of the given visualization squares, a plot of the two specified variables appears.
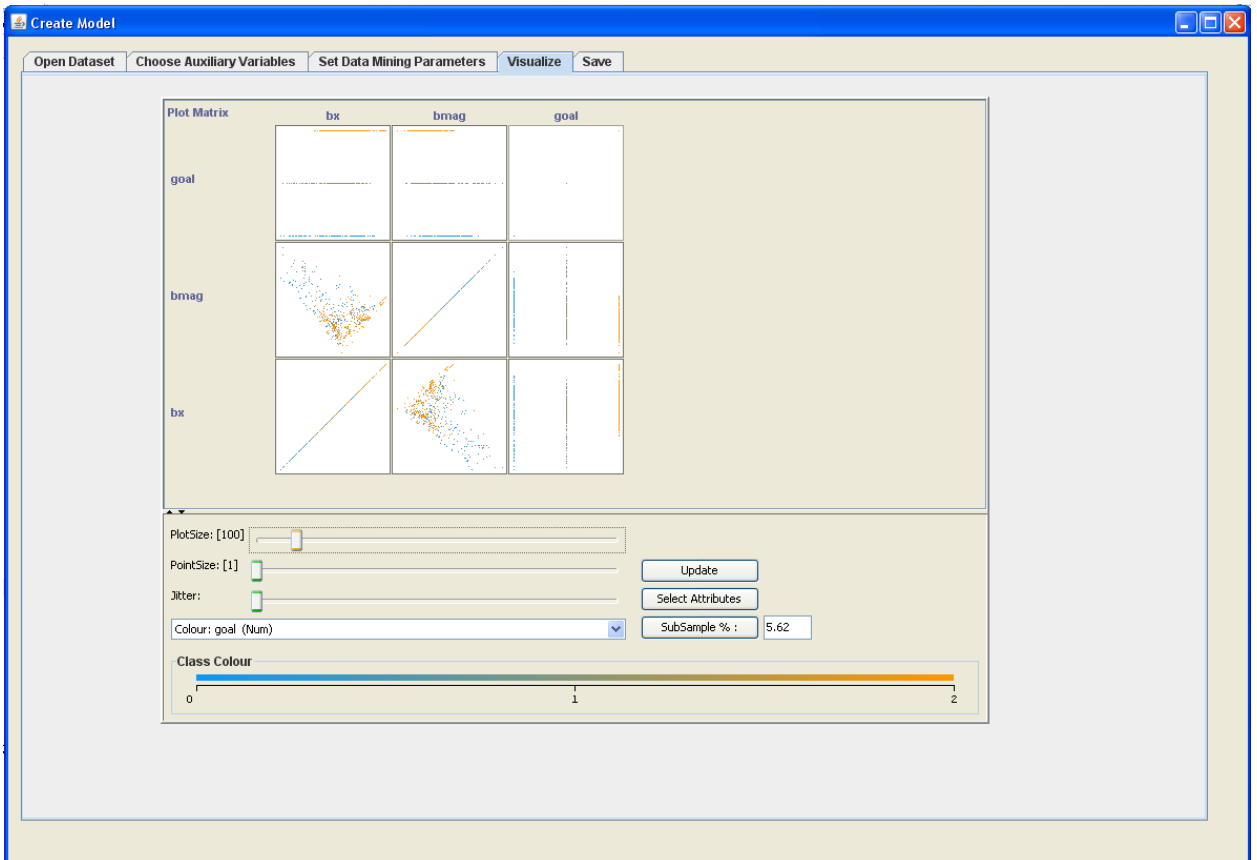
Figure 6. Visualize Tab in Create Model application.

Finally, the user is able to save the best model and the workspace. The save model can then be reopened and used to score or apply on unseen, novel data and collect predictions. A saved workspace allows the user to continue a stopped session, and go on without having to remember which options were used. It also enables the user to come back to the modeling task and perform further experimentation with the MienTool options.
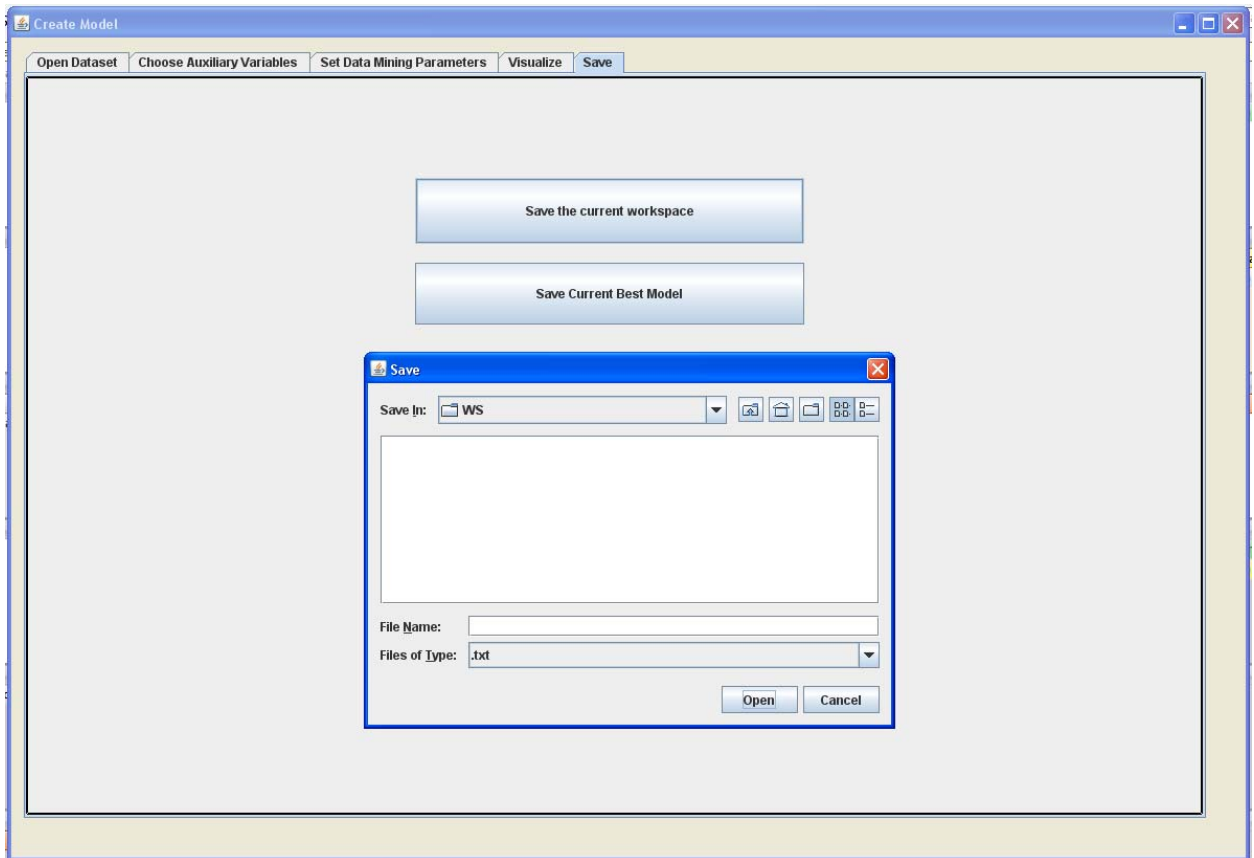
Figure 7. Save Tab in Create Model application.

The Create Time Series Model Application

The Create Time Series application is able to read in a time series file, where multiple rows are assigned to one instance/example of the data. The user needs to preprocess the time series file first, but defining the number of observations in one time series and the number of time series, and define the metafeatures and global features to be collected, by pressing the Change Metafeature Defaults.
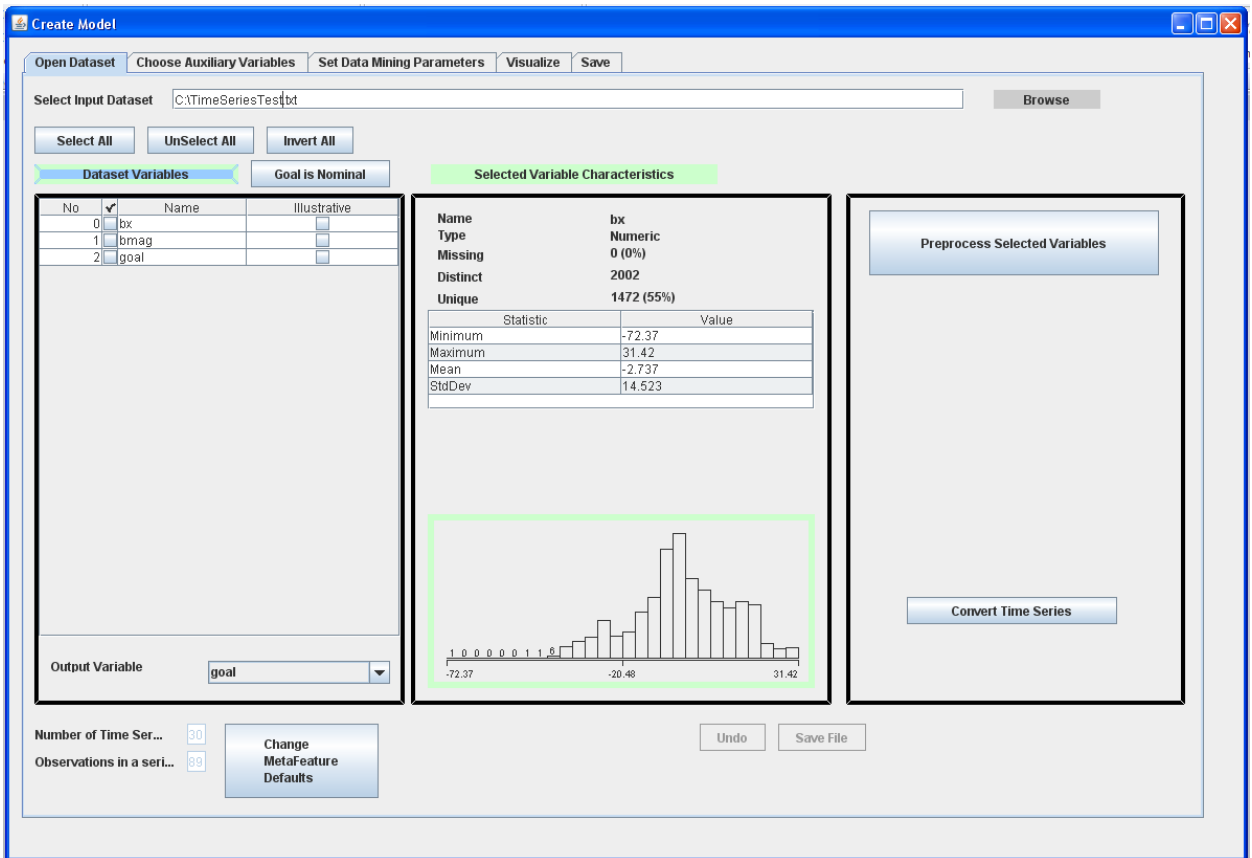
Figure 8. Open Dataset Tab in Create Time Series Model application.

The Change Metafeature Defaults window (Figure 9) allows the user to specify the number of time series and the number of observations in each of the time series. It also allows the user to select which metafeatures (increasing, decreasing and plateau) and which global features (mean, minimum and maximum) they wish collected on which of the time series channels.
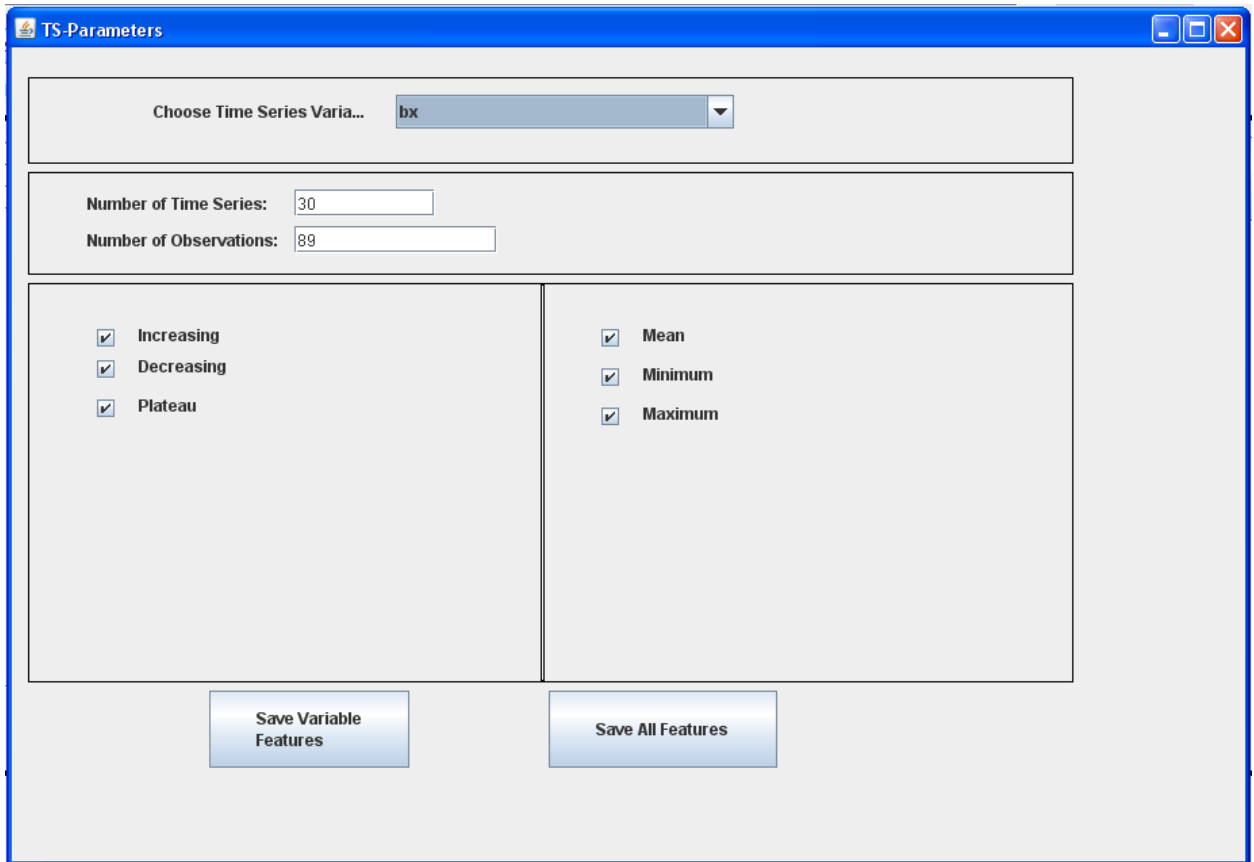
Figure 9. Save Change Metafeature Defaults window.

The Open Model Application

The Open Model application enables the user to open a previously save model and score it or run it (i.e. apply it) on a novel or unseen data. The user then specifies which dataset to apply the model to, and presses the Run button.
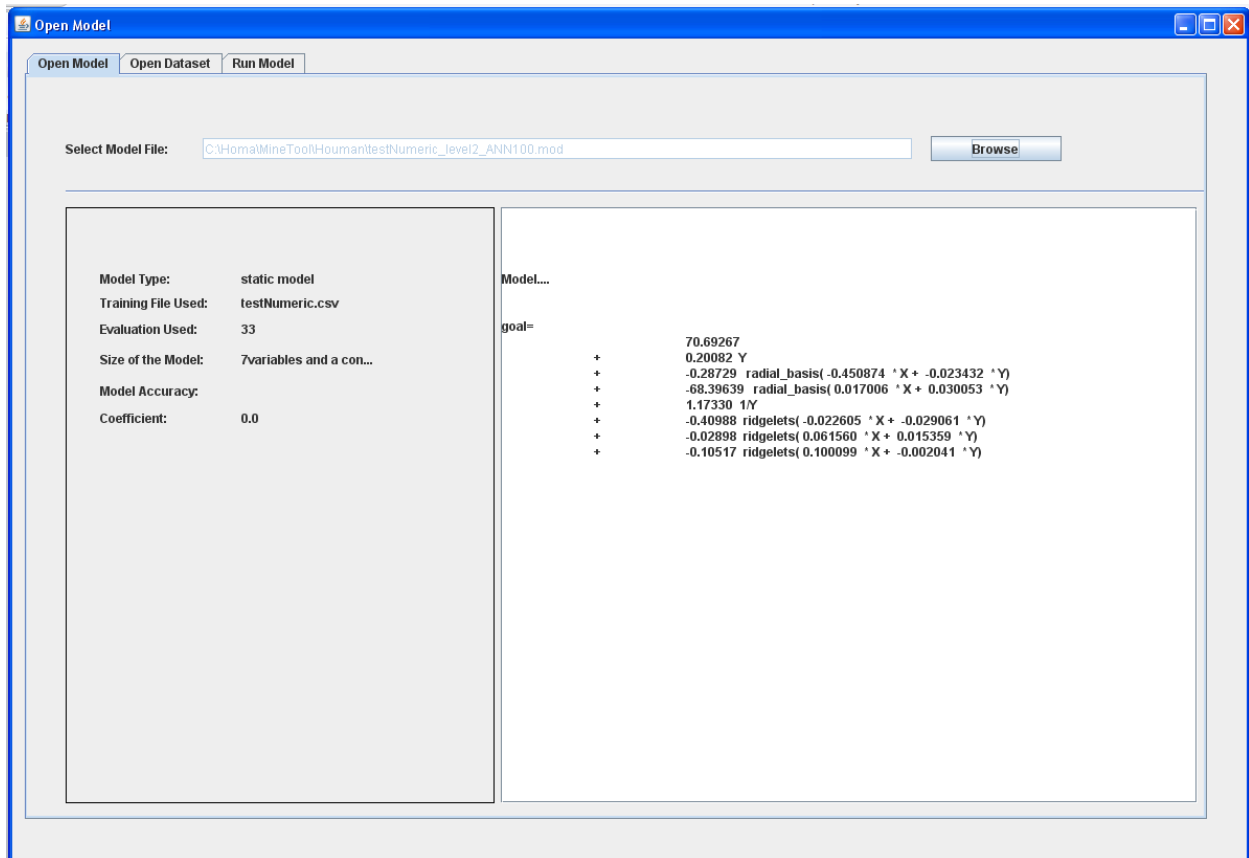
Figure 10. Open Model Application.

The Open Workspace Application

The Open Workspace application facilitates opening a previously saved workspace, where the input file, selected variables and all the GUI options are preserved. The user is able to continue with their modeling work and continue with the experimentation without having to remember all the options they chose before they saved the workspace.
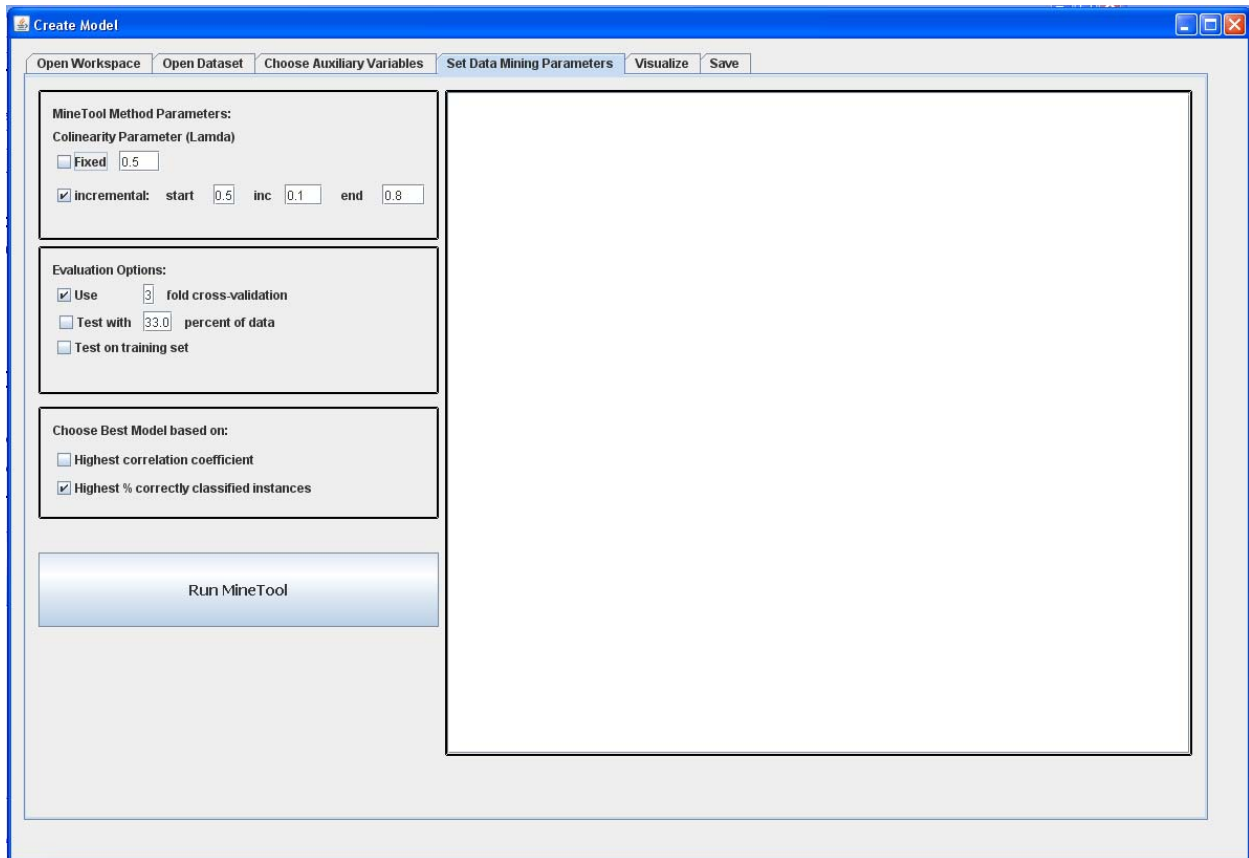
Figure 11. Open Workspace Application.