

DTREG

Predictive Modeling Software

www.dtreg.com



Getting Started with DTREG

Phil Sherrod
phil@philsherrod.com

Copyright © 2014, all rights reserved
This material may not be reproduced without permission

Basic Features of DTREG

- Creates models to predict a *target* variable based on the values of *predictor* variables.
- There may be only one target variable, but there may be many (possibly thousands) predictor variables.
- Target and predictor variables may have continuous (numeric) values or categorical (discrete) values.
- With DTREG you can easily try many types of models such as neural networks and decision trees.
- Validation of the model using cross-validation and other methods is an inherent part of DTREG.

Input, “Training” Data

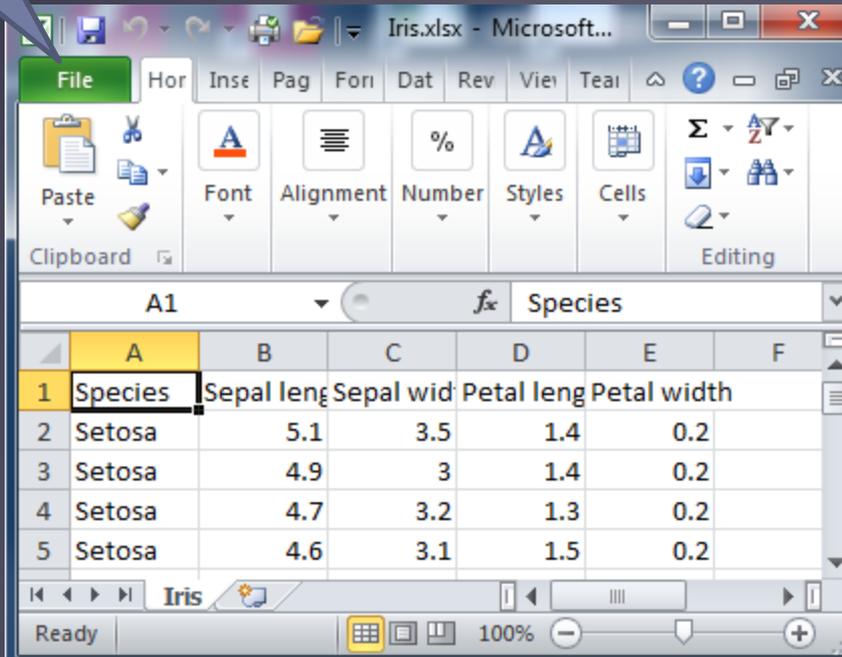
- The input data file used to “train” a model must be presented as a Comma Separated Value (.csv) file.
- SQL and Excel can export data to csv file format.
- The first row of the file must have the variable names
- Comma, space, semicolon or tab may separate columns. You may use quote marks around values.

```
Species,"Sepal length","Sepal width","Petal length","Petal width"  
Setosa,5.1,3.5,1.4,0.2  
Setosa,4.9,3,1.4,0.2  
Setosa,4.7,3.2,1.3,0.2  
Setosa,4.6,3.1,1.5,0.2  
Setosa,5,3.6,1.4,0.2  
Setosa,5.4,3.9,1.7,0.4  
Setosa,4.6,3.4,1.4,0.3  
Setosa,5,3.4,1.5,0.2  
Setosa,4.4,2.9,1.4,0.2
```

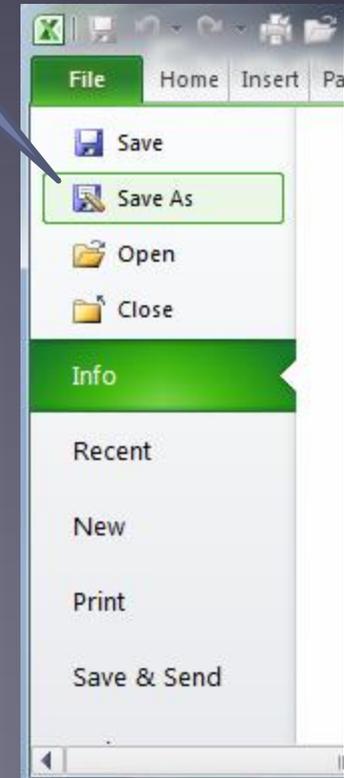
Exporting Data From Excel to a CSV File

- Click “File” then “Save as”.

Click File

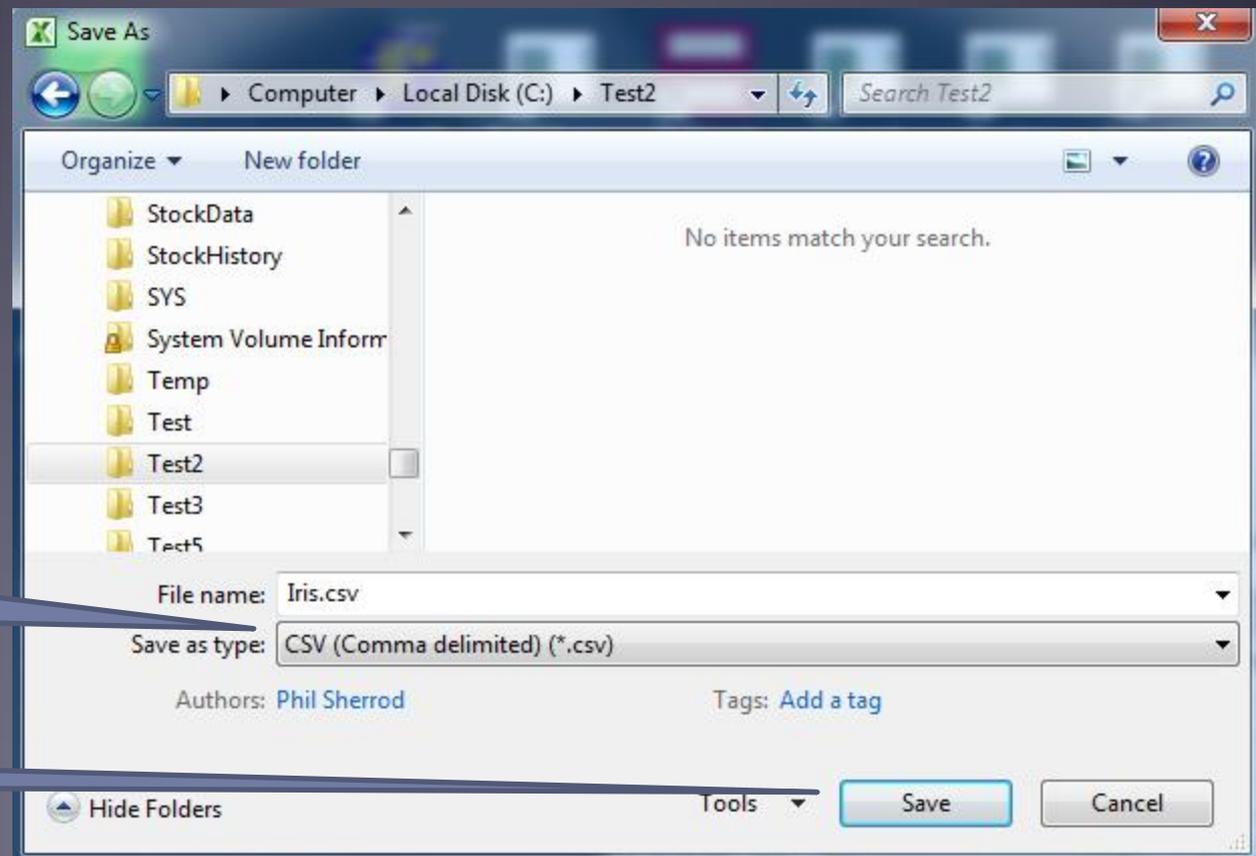


Click Save As



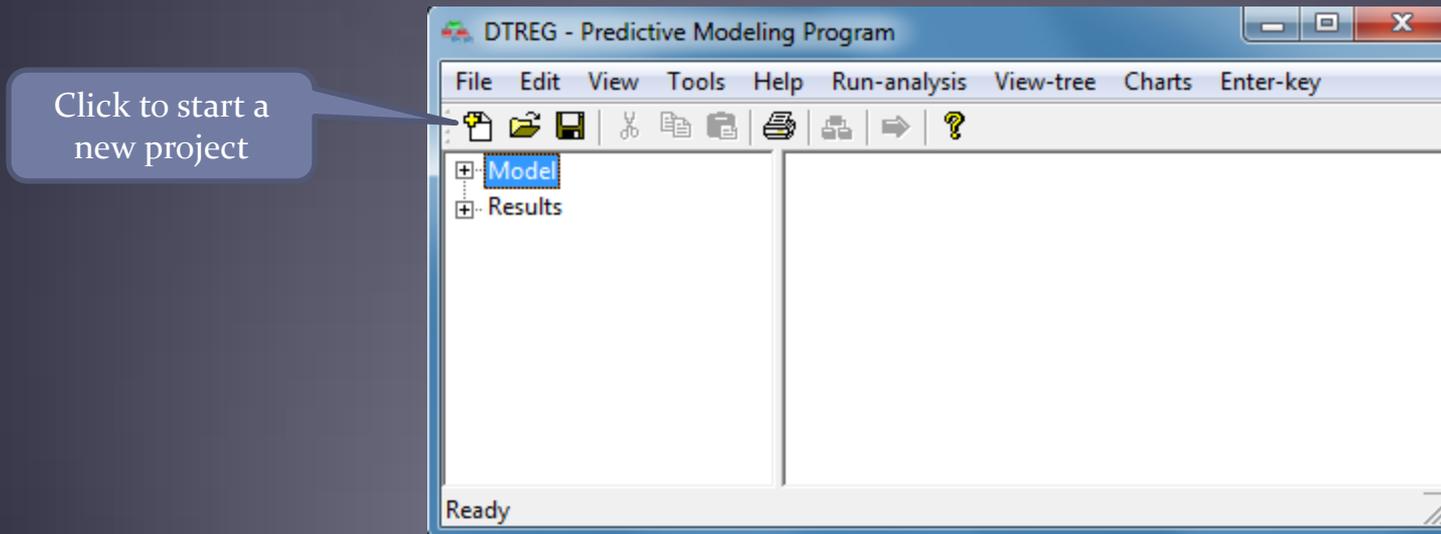
Exporting Data From Excel to a CSV File

- Select “CSV (Comma Delimited)” and click Save.



Creating a New DTREG Project

- Click  icon to start creating a new project.
- Click  icon to open an existing DTREG project.



Specify Input Data and Title

Title of project
Fisher Iris species prediction

Input data file
C:\DTREGtestFiles\Iris.csv

Note: The first line of the data file must have the names of the variables.

Character used for a decimal point in the input data file
 Period: '.' Comma: ','

Character used to separate columns
 Comma: ',' Semicolon: ';' Space Tab Other:

Data subsetting
 Use all rows in the data file
 Randomly select this percent of the rows:

Store data in a virtual memory disk file. Memory cache (MB):

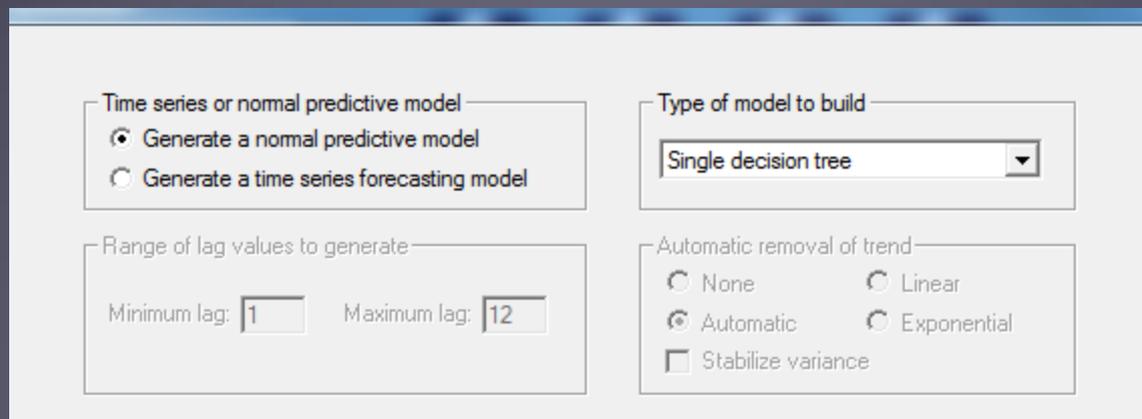
File where information about this project is to be stored
C:\DTREGtestFiles\Iris.dtr

Bind PCA transformation to project

Notes about this project
Classic model presented by Sir Ronald Fisher in 1936

Select Standard Model or Time Series model and Initial Model Type

- We will build a normal (not time-series) model using a single decision tree.



The screenshot shows a software interface with four panels for configuring a model:

- Time series or normal predictive model:** Two radio buttons. The first, "Generate a normal predictive model", is selected. The second is "Generate a time series forecasting model".
- Type of model to build:** A dropdown menu currently showing "Single decision tree".
- Range of lag values to generate:** Two input fields. "Minimum lag:" is set to "1" and "Maximum lag:" is set to "12".
- Automatic removal of trend:** Four radio buttons: "None", "Linear", "Automatic" (selected), and "Exponential". There is also a checkbox for "Stabilize variance" which is currently unchecked.

Select Target and Predictor Variables

Variables

| Variable | Target | Predictor | Weight | Categorical | Character |
|--------------|-------------------------------------|-------------------------------------|--------------------------|-------------------------------------|-------------------------------------|
| Species | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| Sepal length | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Sepal width | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Petal length | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Petal width | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Predictor range

All predictors

Predictor coverage

Type range

All categorical

All continuous

All numeric

All character

All reset

Search

Report options

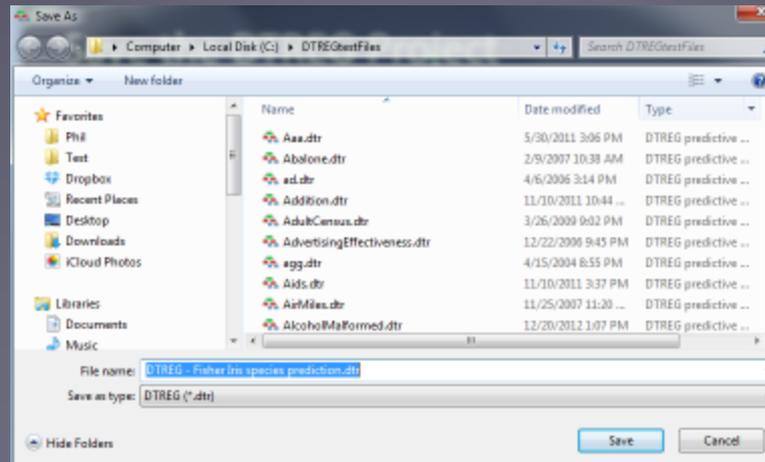
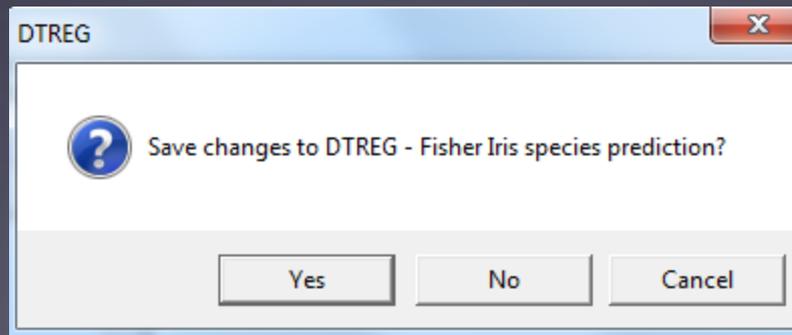
- Report summary of variables
- Report category statistics for categorical variables
- Report category statistics for continuous variables
- Report Min., Max., Mean for continuous variables

Surrogate variables for missing value imputation

Number of surrogates to store: Max. polynomial order:

Minimum surrogate association: Report surrogate variables

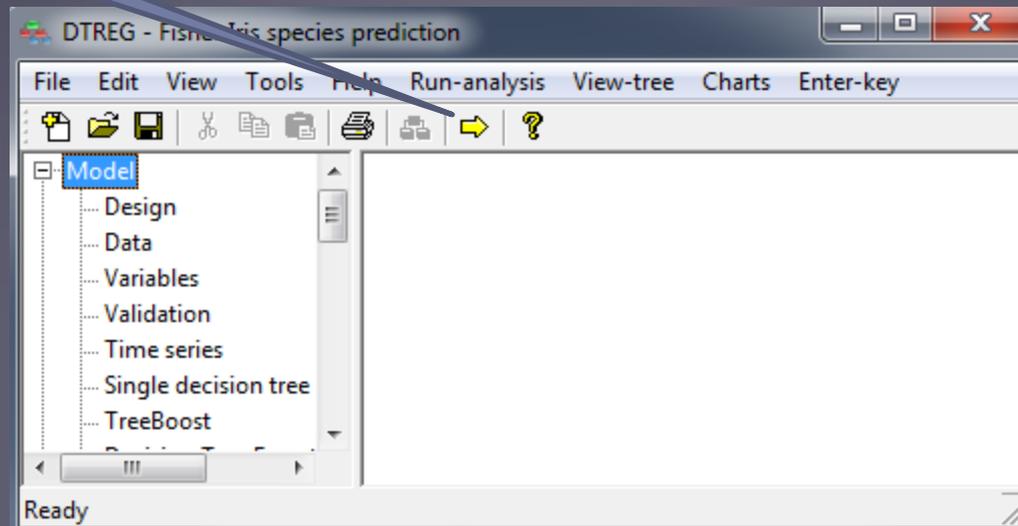
Save the DTREG Project



Tell DTREG to Train the Model

- Click the  icon to start training the model.

Click to start training



View Analysis Results

- Once training is complete, select items in the left panel to scroll to the section in the report.

The screenshot shows the DTREG software window titled "DTREG - Fisher Iris species prediction". The interface includes a menu bar (File, Edit, View, Tools, Help, Run-analysis, View-tree, Charts, Enter-key) and a toolbar. On the left, a tree view under "Results" shows "Analysis report" expanded, with "Project parameters" selected and highlighted in blue. A blue callout bubble points to this selection with the text "Click items to select report sections". The main window displays the text of the "Project Parameters" section, which includes details such as project title, file path, target variable, number of predictor variables, model type, splitting algorithm, and various model parameters.

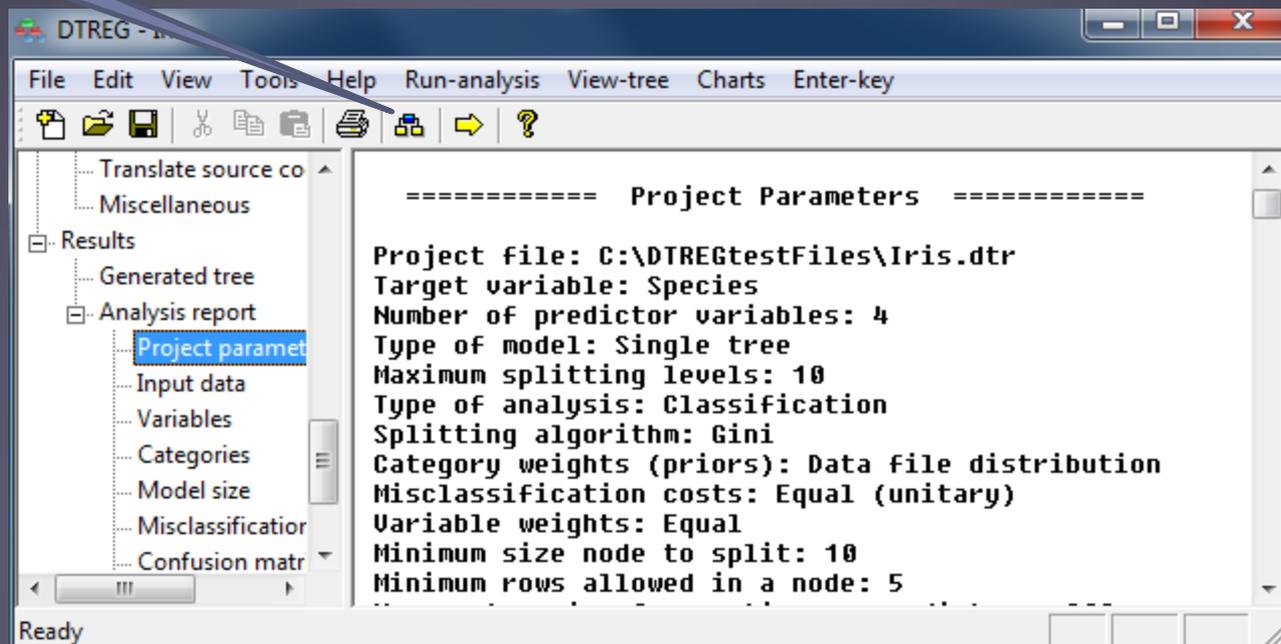
```
===== Project Parameters =====  
  
Project title: Fisher Iris species prediction  
Project file: C:\DTREGtestFiles\DTREG - Fisher Iris species  
prediction.dtr  
Target variable: Species  
Number of predictor variables: 4  
Type of model: Single tree  
Maximum splitting levels: 10  
Type of analysis: Classification  
Splitting algorithm: Gini  
Category weights (priors): Data file distribution  
Misclassification costs: Equal (unitary)  
Variable weights: Equal  
Minimum size node to split: 10  
Minimum rows allowed in a node: 5  
Max. categories for continuous predictors: 1000  
Tree pruning and validation method: Cross validation  
Number of cross-validation folds: 10  
Tree pruning criterion: Minimum cost complexity (0.00 S.E.)  
  
===== Input Data =====
```

Click items to
select report
sections

Viewing the Generated Decision Tree

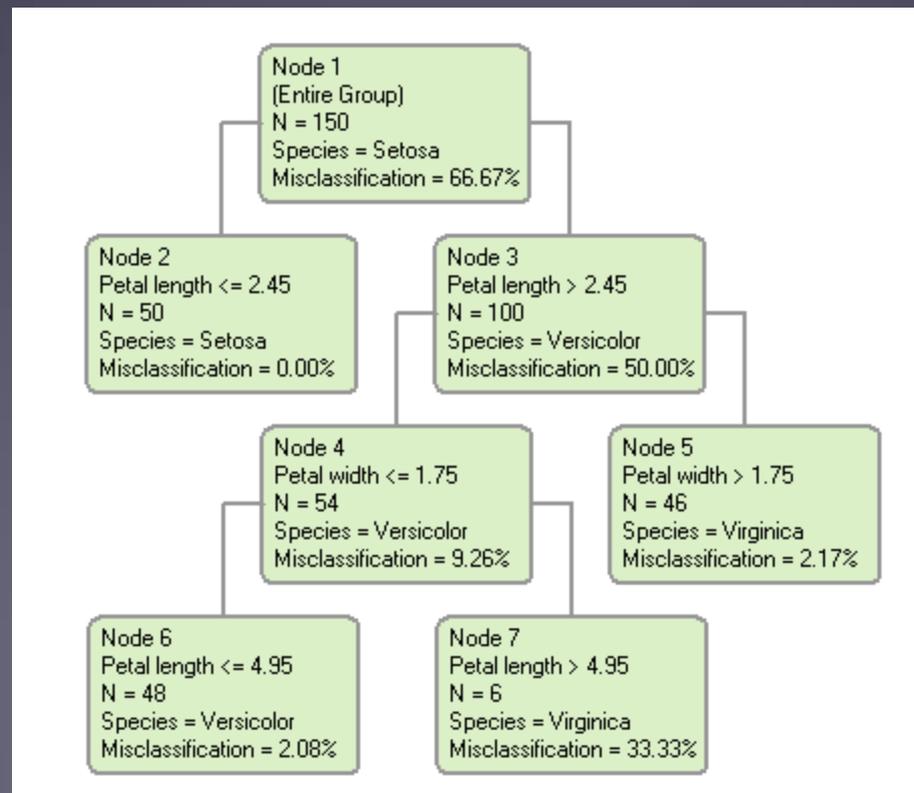
- Click the  icon to display the generated tree.

Click to display tree



The Generated Decision Tree

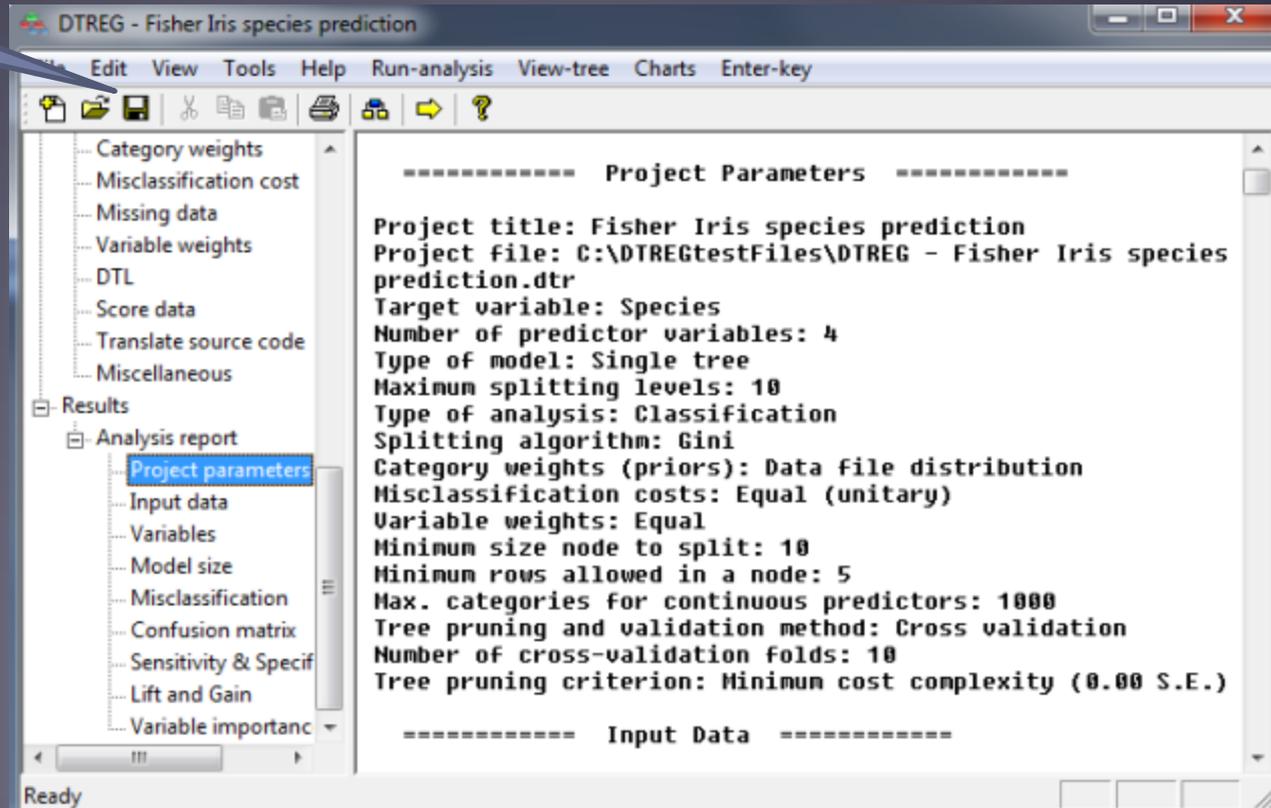
- The decision tree shows how predictor variable values were split to predict the target value.



Save the Trained Project

- Click the  icon to save the generated model project

Click to save
model project



End of Getting-Started Tutorial

- This completes the DTREG training tutorial