

DTREG

Predictive Modeling Software

www.dtreg.com



Using Models to Predict Values

Phil Sherrod
phil@philsherrod.com

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

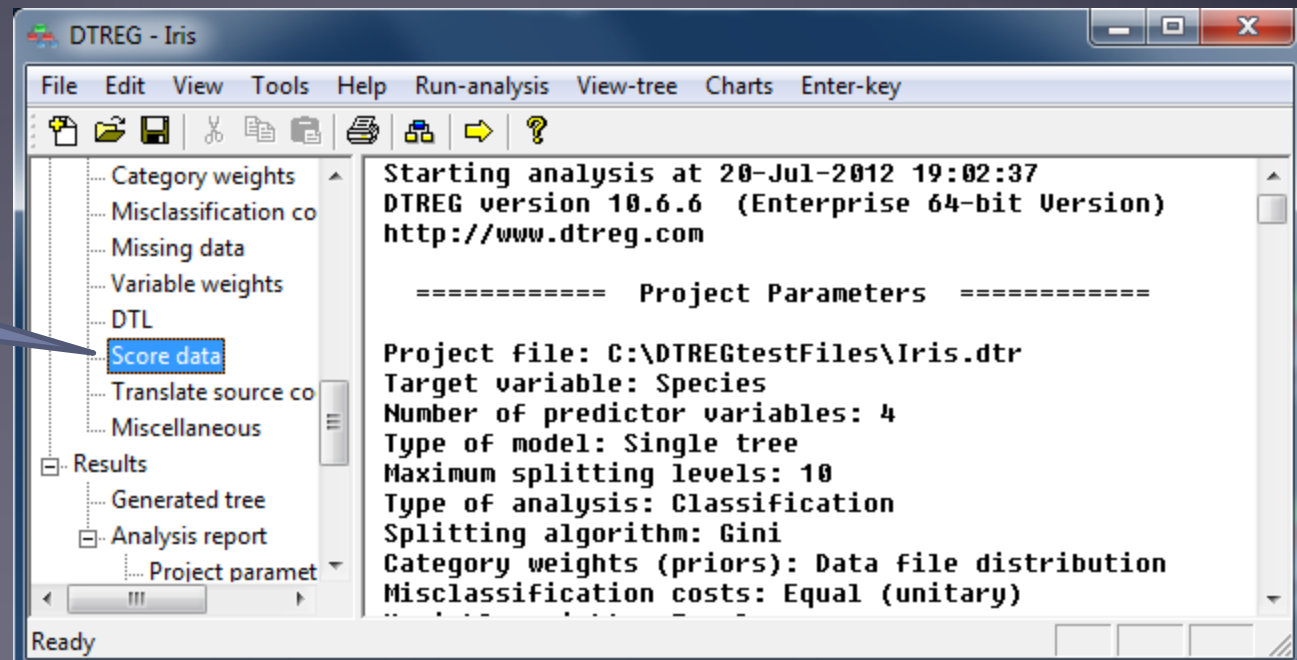
Introduction to Applying Models

- Once a model has been created, it can be applied to new data to predict target variable values.
- This process is known as “scoring” the new data.
- The scoring process reads an input data file in csv format, uses the predictor values to compute predicted target values and writes the predictions to an output file.
- If the input file contains known target values, then DTREG compares the predicted and actual values and displays accuracy measures.

Starting the Scoring Process

- First, train the model, then click “Score data” in the left panel.
- You also can open a previously-trained model.

Click “Score data” to perform scoring



Specifying Scoring Parameters

Input file to be scored

Output file for results

Select which variables should be written to the output file

Select generated values to be written to output

Perform scoring

"Scoring" runs a dataset through a model and generates an output dataset showing the predicted values of the rows.

Input file whose data is to be scored:

Output file where scored results are to be written with predicted values from the model:

Select variables to be written to the output scoring file

Variable	Output
Species	<input checked="" type="checkbox"/>
Sepal length	<input checked="" type="checkbox"/>
Sepal width	<input checked="" type="checkbox"/>
Petal length	<input checked="" type="checkbox"/>
Petal width	<input checked="" type="checkbox"/>

Select variables DTREG should add to output records

Predicted target value

Residual (Actual - Predicted)

Misclassification indicator

Row number

Terminal node number

Probability scores for each category of the target

K-means cluster number

Write variable names to the first row of the score file

Time series forecast

Forecast rows for time series:

Report Generated by Scoring

Scoring report

Scoring was performed 1-Jan-2014 12:10:38

Input file = C:\DTREGtestFiles\Iris.csv
Output file = C:\DTREGtestFiles\IrisScore.csv

Number of observations scored = 150

Category	Actual Count	--Misclassified-- Count	Percent
Setosa	50	0	0.000
Versicolor	50	3	6.000
Virginica	50	1	2.000
Total	150	4	2.667

Overall accuracy = 97.33%

----- Confusion Matrix for Scored Data -----

Actual Category	-----Predicted Category-----		
Category	Setosa	Versicolor	Virginica
Setosa:	50	0	0
Versicolor:	0	47	3
Virginica:	0	1	49

Copy report into clipboard

Close

Output File Generated by Scoring

- Output file has selected input variables and also a column with predicted target variable values.

Predicted target values

```
0 10 20 30 40 50 60 70 80
1 Species, "Sepal length", "Sepal width", "Petal length", "Petal width", PredictedValue
2 Setosa, 5.1, 3.5, 1.4, 0.2, Setosa
3 Setosa, 4.9, 3, 1.4, 0.2, Setosa
4 Setosa, 4.7, 3.2, 1.3, 0.2, Setosa
5 Setosa, 4.6, 3.1, 1.5, 0.2, Setosa
6 Setosa, 5, 3.6, 1.4, 0.2, Setosa
7 Setosa, 5.4, 3.9, 1.7, 0.4, Setosa
8 Setosa, 4.6, 3.4, 1.4, 0.3, Setosa
```

End of Tutorial

- This completes the scoring DTREG training tutorial