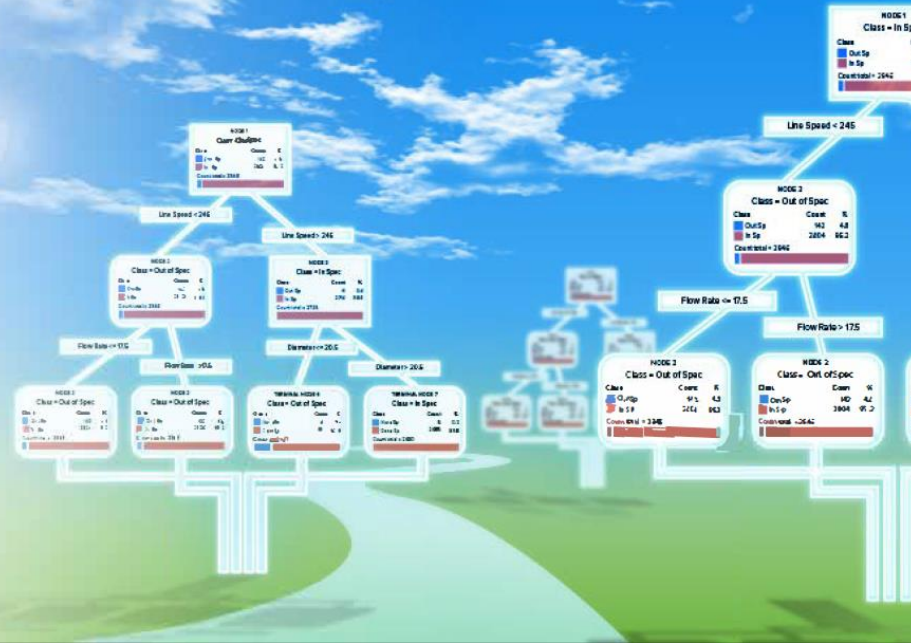




Boost Your Predictive Analytics Skillset with Minitab Statistical Software



Jenn Atlas
*Global Market
Development
Manager*



Shelby Anderson
*Marketing Content
Manager*

Welcome. The webinar will begin shortly.

AUDIO WEB: Please ensure you have your computer audio system activated and your speakers turned up.





Boost Your Predictive Analytics Skillset with Minitab Statistical Software



Meet the Presenter:

Jenn Atlas

Global Market Development Manager

Jenn studies how organizations solve analytical problems to ensure that Minitab delivers on our commitment to make data analytics easier.

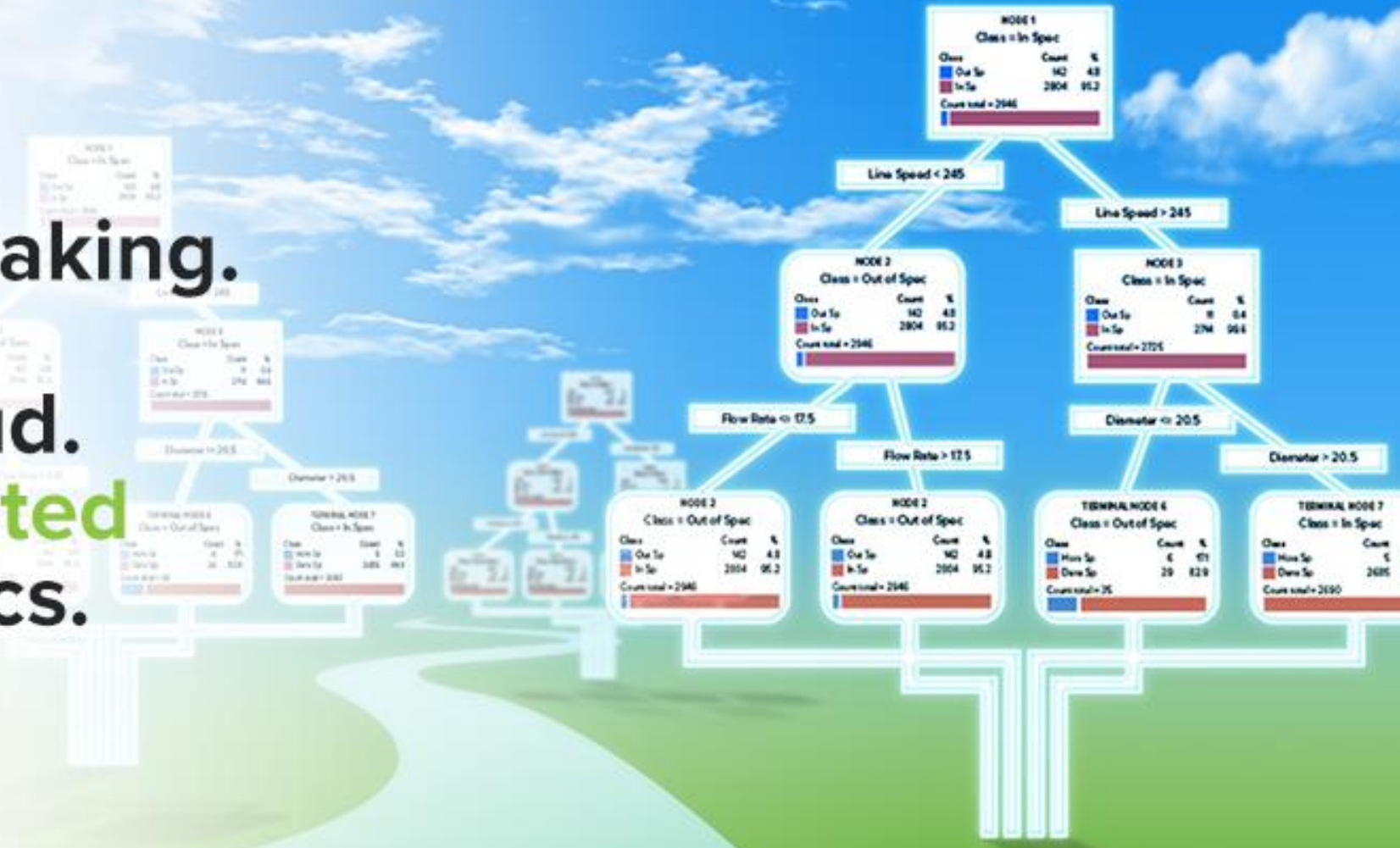
As a classically trained statistician, Jenn has worked, trained and consulted in the field of analytics for over 25 years.





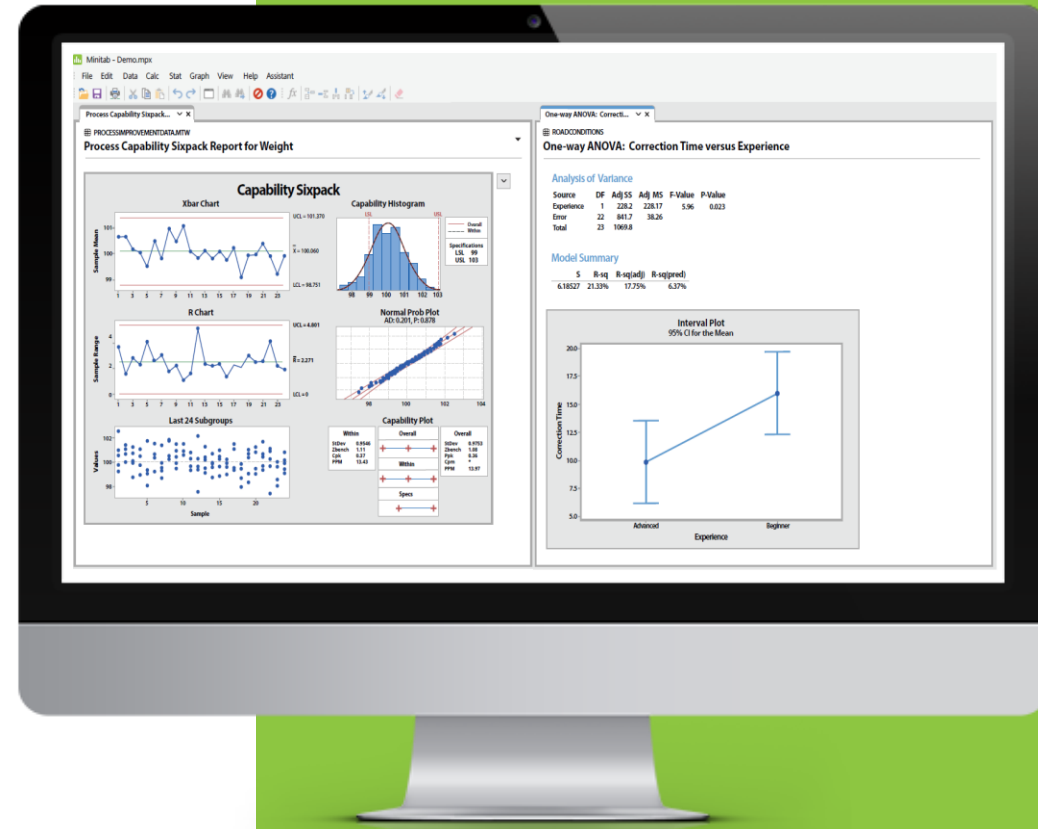
Minitab®

Better decision making.
Faster insights.
Easier on the cloud.
Your path to boosted predictive analytics.



Everyone in an organization, regardless of statistical background, is empowered when they have access to easy-to-use analytics tools.

Minitab Statistical Software delivers business and predictive analytics, paired with essential visualizations to enable data-driven value creating decision making.



Why is an analytics skillset important?

- ▶ Data helps you solve problems that give us (and your business!) a competitive advantage
- ▶ The science of data analytics, paired with your experience guarantees better decision making



Analytics is the science of learning from data

Descriptive

- What happened?

Diagnostic

- Why did it happen?

Predictive

- What might happen?

Prescriptive

- What should I do to make it happen/not happen?

Machine Learning

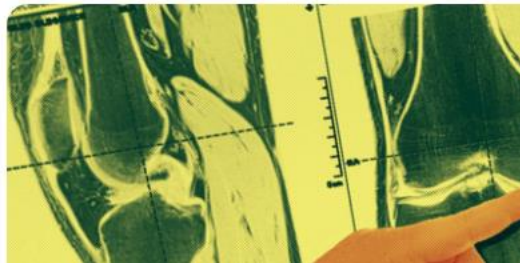
machine learning

All News Images Books Videos

About 2,380,000,000 results (0.72 seconds)

Taking Orthopedic Image Analysis One Step Further with Machine Learning

Yana Yelina - April 26, 2021



ML Ops and the Promise of Machine Learning at Scale

UNLIMITED DATA | BY [JAMES KULICH](#) | 5 MIN READ



Apple will spend \$1 billion to open 3,000-employee campus in North Carolina

PUBLISHED MON, APR 26 2021-7:30 AM EDT | UPDATED 2 HOURS AGO

Kif Leswing
@KIFLESWING

SHARE f t in ✉

KEY POINTS

- Apple announced plans to open a new campus in the Raleigh, North Carolina, area.

on the campus, and it will employ 3,000 people in software engineering and machine learning.



Minitab Machine Learning Methods

Unsupervised Learning in Minitab

- Clustering Algorithms
- Cluster Observations
- Cluster Variables
- Cluster K-Means

Data Reduction Algorithms

- Principal Components Analysis
- Factor Analysis

Supervised learning in Minitab

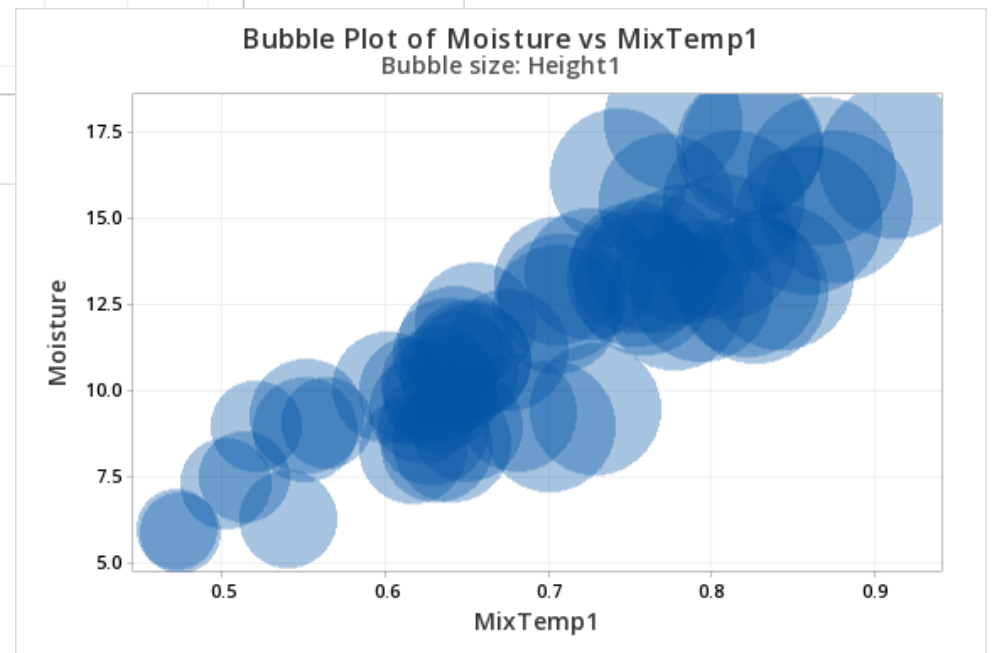
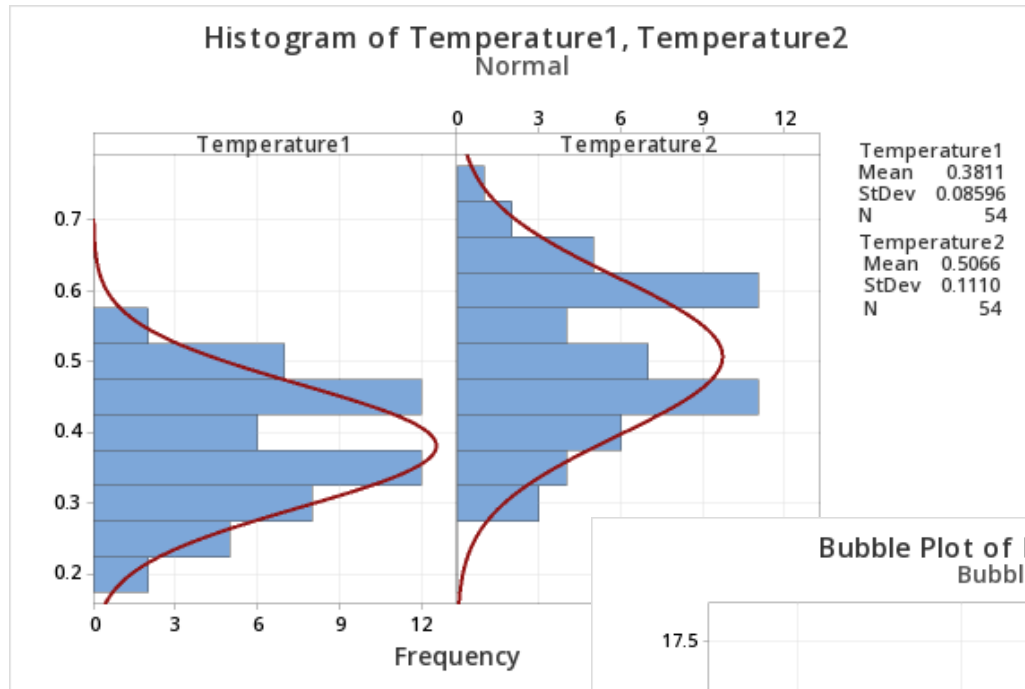
Classification Algorithms

- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)
- Logistic Regression
- **CART Classification Trees**
- **Random Forests**
- **TreeNet**

Regression Algorithms

- Simple Regression
- Polynomial Regression
- Multiple Regression
- Nonlinear Regression
- Partial Least Squares
- **CART Regression Trees**
- **Random Forests**
- **TreeNet**

VISUALIZATIONS

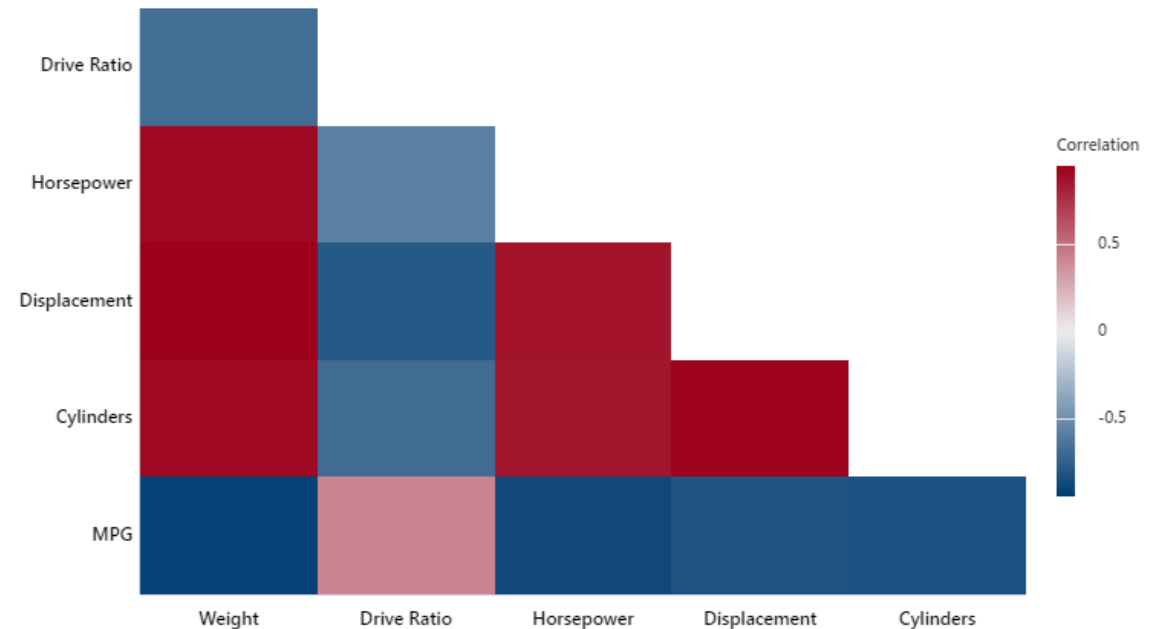


New Visualization: Correlogram

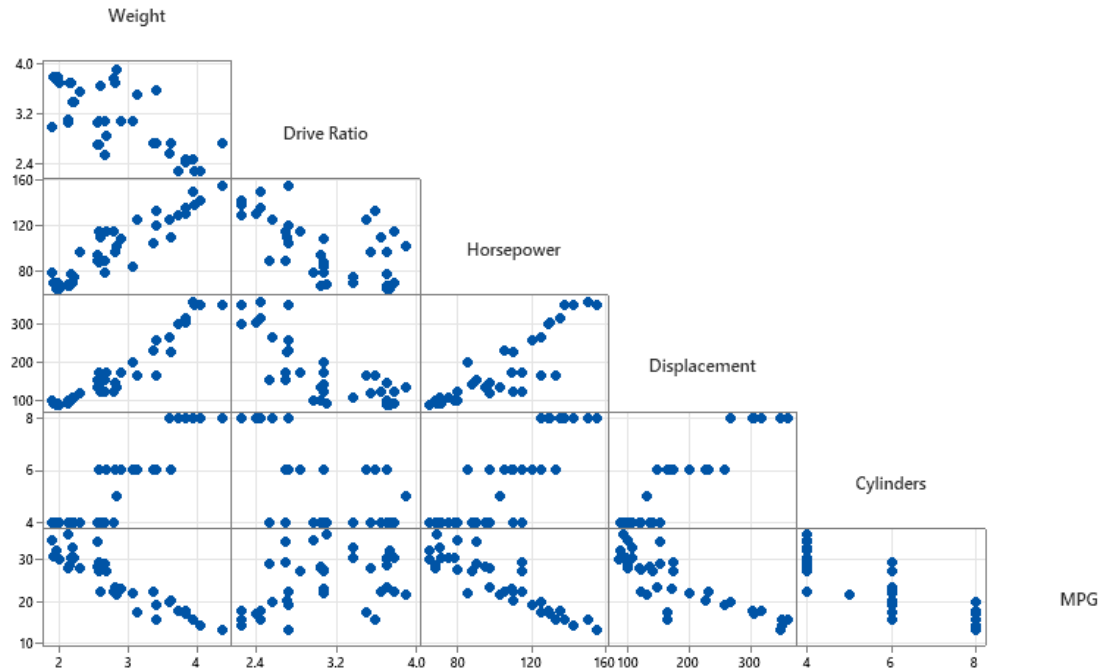
Improvement: A correlogram is a visualization that displays correlation coefficients with variable names defining the rows and columns.

User Benefit: Correlograms are useful for finding important correlations when faced with many variables.

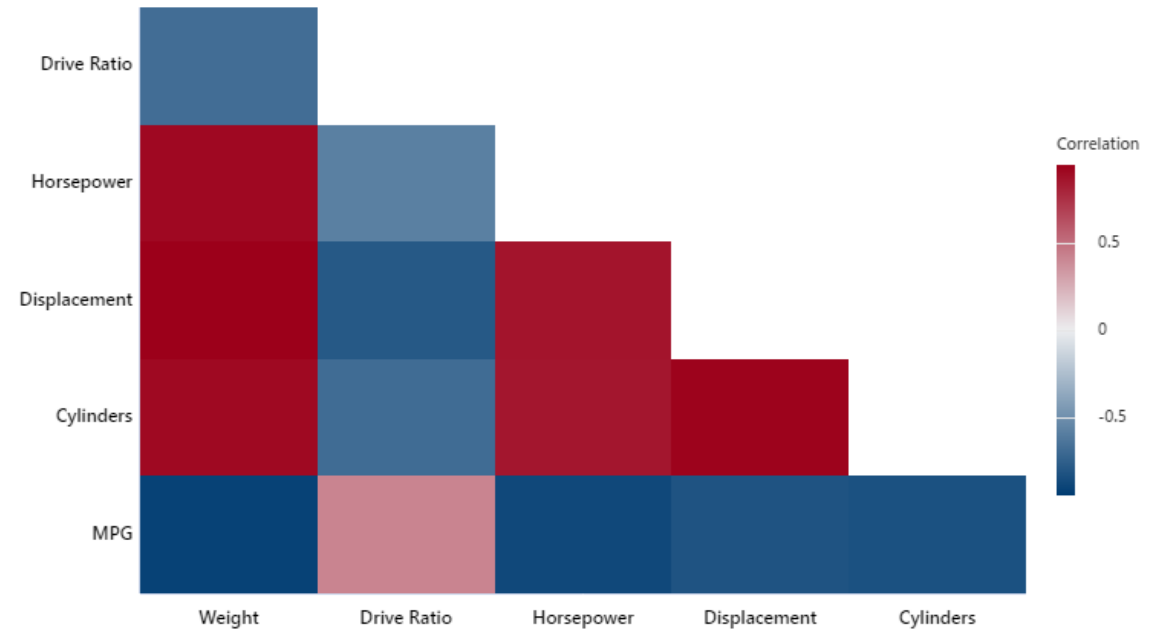
The correlogram makes it easy to visualize a matrix or correlations, particularly when the number of variables is large.



Matrix Plot



Correlogram



Root Cause Analysis



BRAINSTORM
RELEVANT
VARIABLES



COLLECT DATA



PREPARE DATA



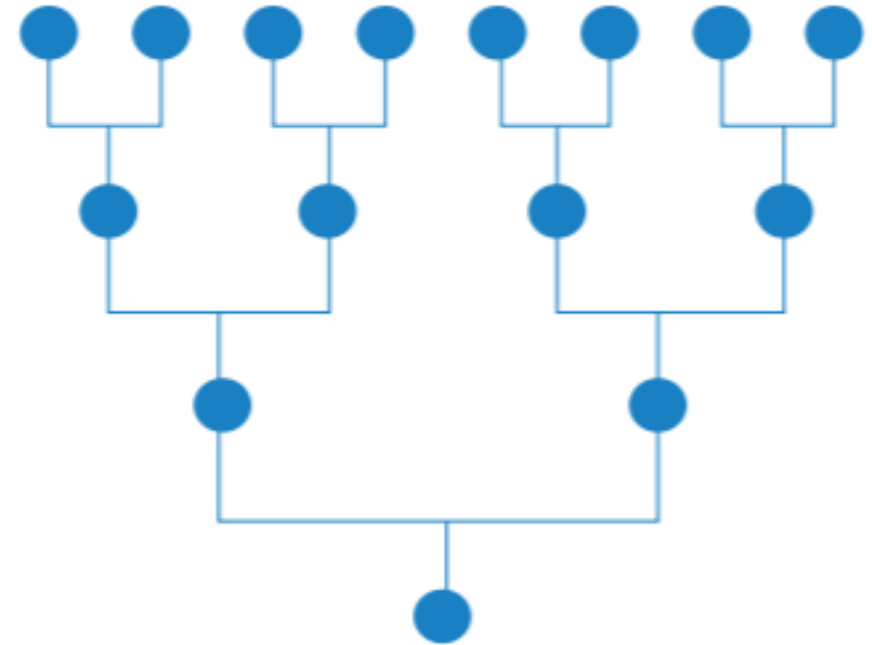
ANALYZE DATA



MAKE
RECOMMENDATIONS

The New Essentials for your Analytics Toolkit: Tree-Based Methods

- Tree-based algorithms utilize a series of if-then rules to create predictions from one or more decision trees
- Tree-based methods empower predictive analytics with not only speed to answer, but also provide remarkable accuracy and ease of interpretation.
- Quickly understand the key drivers of a process.



The New Essentials for your Analytics Toolkit: Tree-Based Methods

Tree-based machine learning algorithms in an easy-to-access module

Our proprietary, best-in-class, tree-based machine learning algorithms provide deeper insights.

- Map non-linear relationships
- Visualize multiple, complex interactions
- Equipped to handle larger data
- Perform well with messy data:
 - missing values
 - outliers



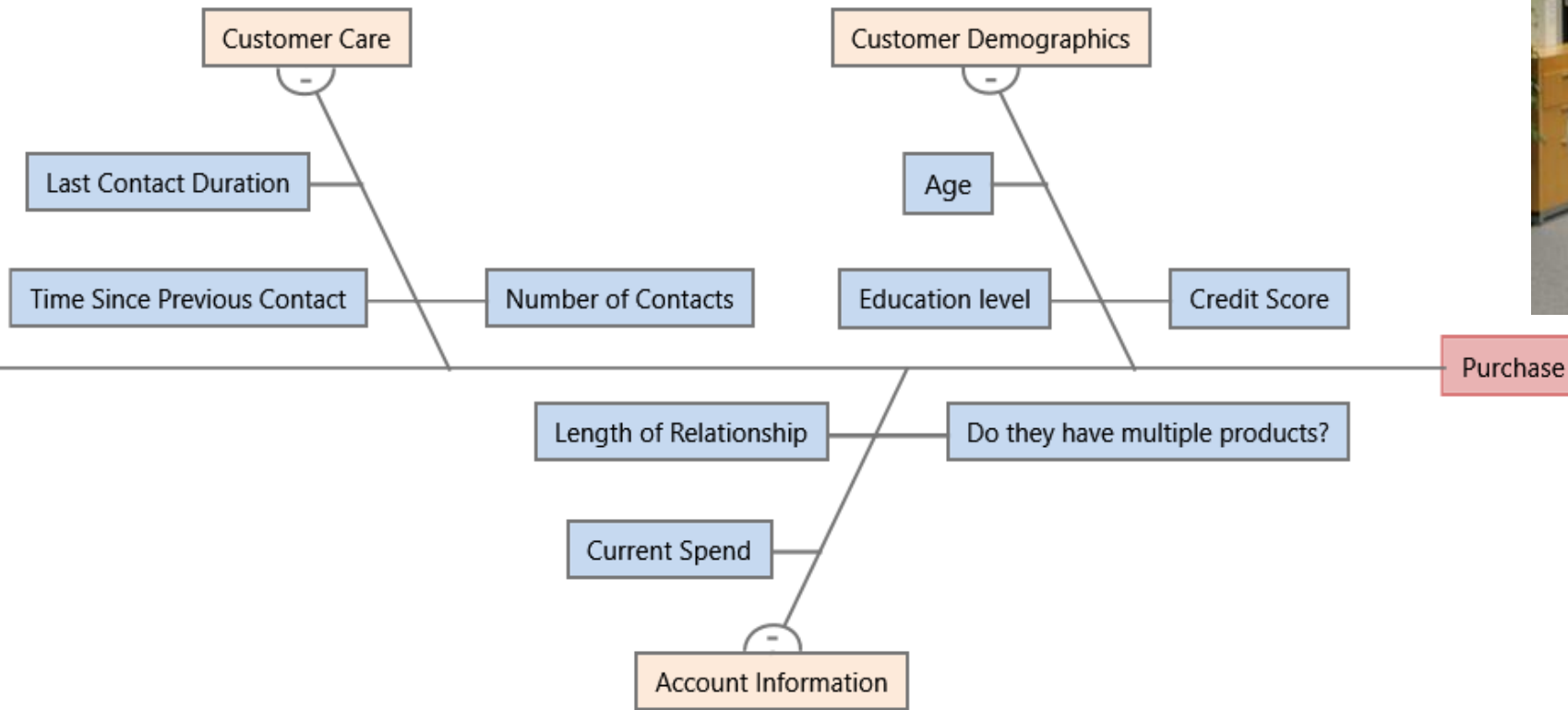


Predicting Customer Behavior

- How can I predict who will purchase my products?
- Who will renew their subscriptions?
- Who will purchase additional products?



Predicting Customer Behavior

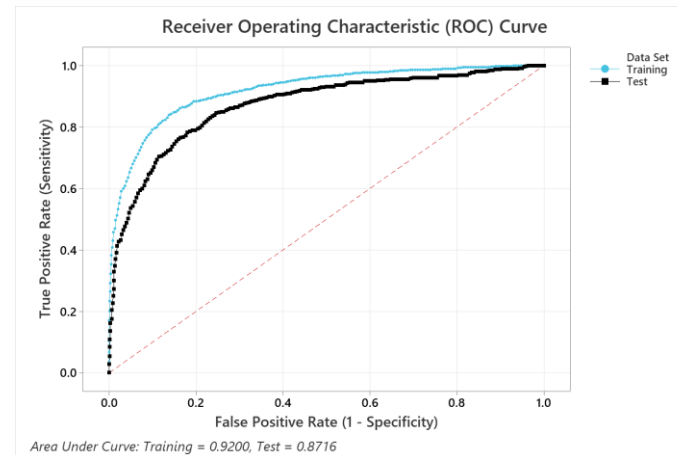
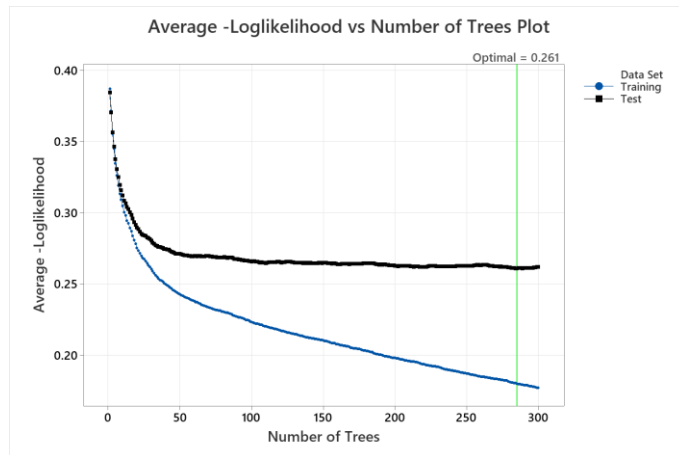


Addition: TreeNet[®] (Gradient Boosting)

New Feature: TreeNet Classification and TreeNet Regression. Includes Fit Model and Discover Key Predictors

Gradient boosting can deliver optimal prediction accuracy, and unique insights.

Available in Minitab's New Predictive Analytics Module



Learn more about TreeNet[®]

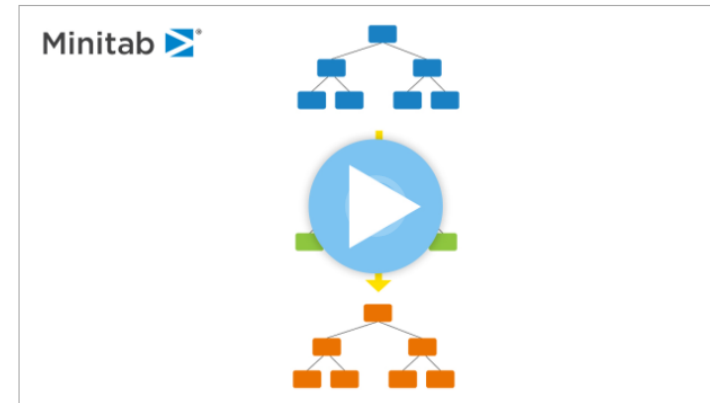
GRADIENT BOOSTING

TreeNet[®]

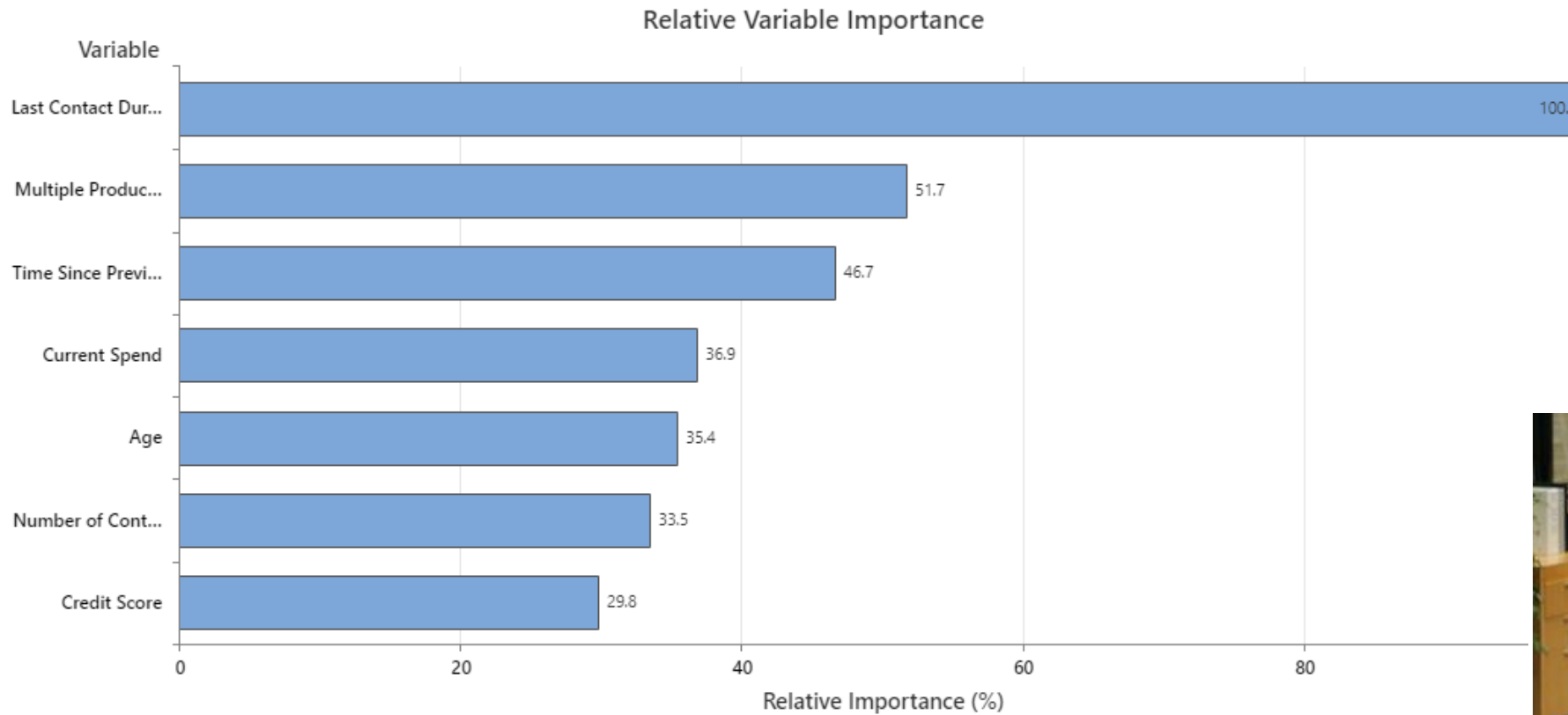
Minitab's most flexible, award-winning and powerful machine learning tool, TreeNet[®] Gradient Boosting, is capable of consistently generating extremely accurate models.

For those new to TreeNet, it is a powerful implementation of the modern machine learning class of algorithms generally known as Stochastic Gradient Boosting. Developed by Jerome Friedman at Stanford University, the technique is known for its superb predictive accuracy. The secret is in the way a model is built: at each iteration a small tree is added to the current ensemble of trees to correct the combined errors of the ensemble.

Utilizing the variety of the supplied loss functions, the process can be tuned for the specific predictive modeling task, like least squares regression, robust regression, classification, etc. To assist with the model interpretation, TreeNet goes one step further and automatically generates various 2D and 3D plots to explain the nature of dependency of the response variable on the model inputs. The model is flexible enough to automatically discover and incorporate various non-linearities and multi-way interactions. A further set of controls allows the user to fine-tune model interactions to meet specific design objectives.



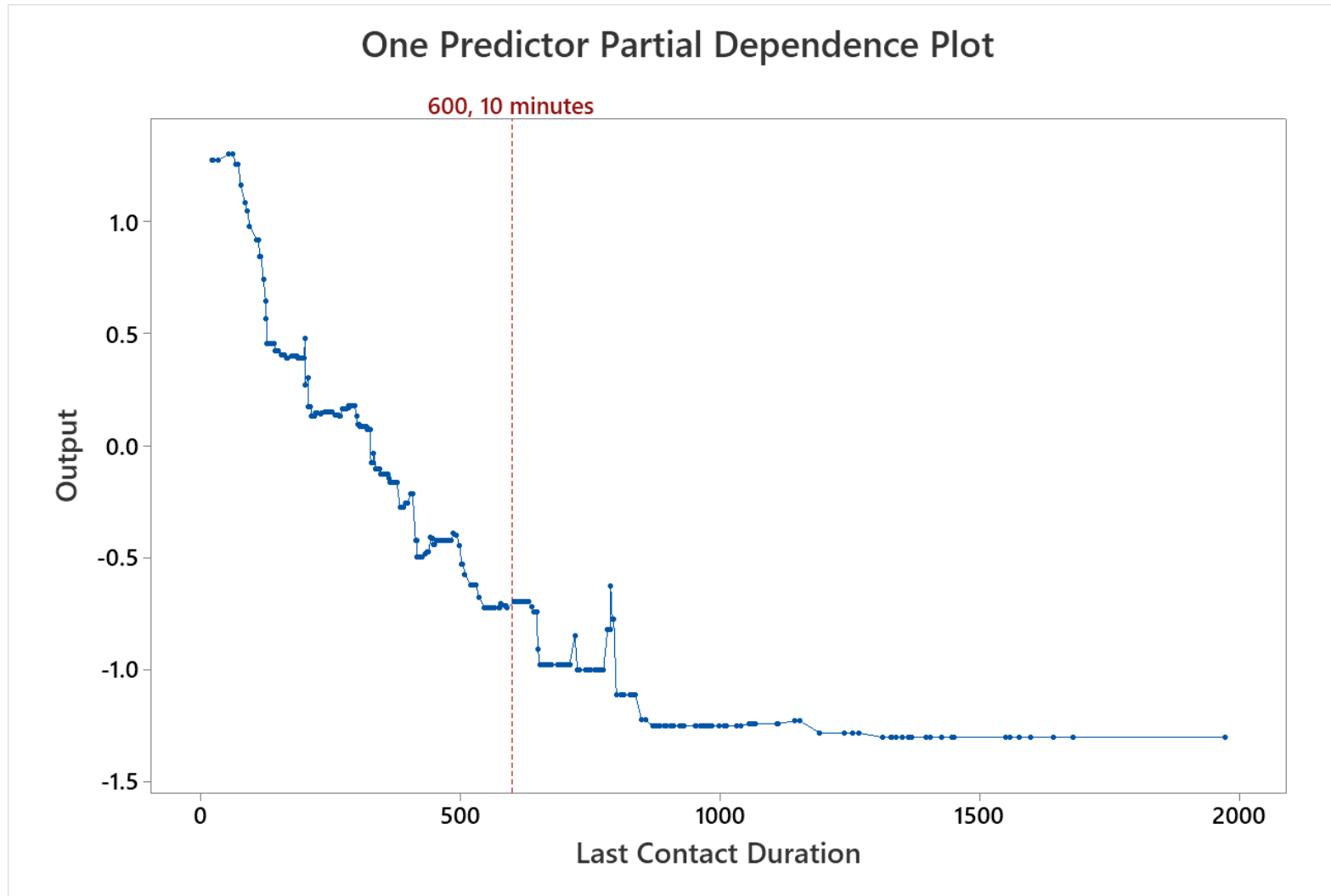
Predicting Customer Behavior using TreeNet®



Variable importance measures model improvement when splits are made on a predictor. Relative importance is defined as % improvement with respect to the top predictor.



Predicting Customer Behavior using TreeNet®



I already have Binary Logistic Regression for this problem

Regression Equation

$$P(\text{Yes}) = \frac{\exp(Y')}{1 + \exp(Y')}$$

Multiple Products

No $Y' = -2.446 + 0.04701 \text{ Age} - 0.000097 \text{ Current Spend} - 0.01103 \text{ Last Contact Duration} + 0.07410 \text{ Number of Contacts} + 0.01677 \text{ Time Since Previous Contact} - 0.000940 \text{ Age*Age} + 0.000005 \text{ Last Contact Duration*Last Contact Duration} - 0.000020 \text{ Time Since Previous Contact*Time Since Previous Contact} + 0.01315 \text{ Age*Number of Contacts} - 0.001140 \text{ Number of Contacts*Time Since Previous Contact}$

Yes $Y' = -2.615 + 0.04701 \text{ Age} - 0.000097 \text{ Current Spend} - 0.01103 \text{ Last Contact Duration} - 0.3861 \text{ Number of Contacts} + 0.01374 \text{ Time Since Previous Contact} - 0.000940 \text{ Age*Age} + 0.000005 \text{ Last Contact Duration*Last Contact Duration} - 0.000020 \text{ Time Since Previous Contact*Time Since Previous Contact} + 0.01315 \text{ Age*Number of Contacts} - 0.001140 \text{ Number of Contacts*Time Since Previous Contact}$

Example

**Engineering:
What factors affect
the process?**

**Response: Moisture
Factors: 88
Rows of data: 54**



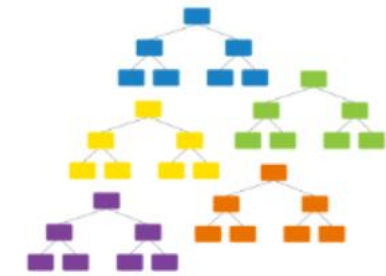
Addition: Random Forests®

New Feature: Random Forests consist of a large number of individual decision trees that operate as an ensemble.

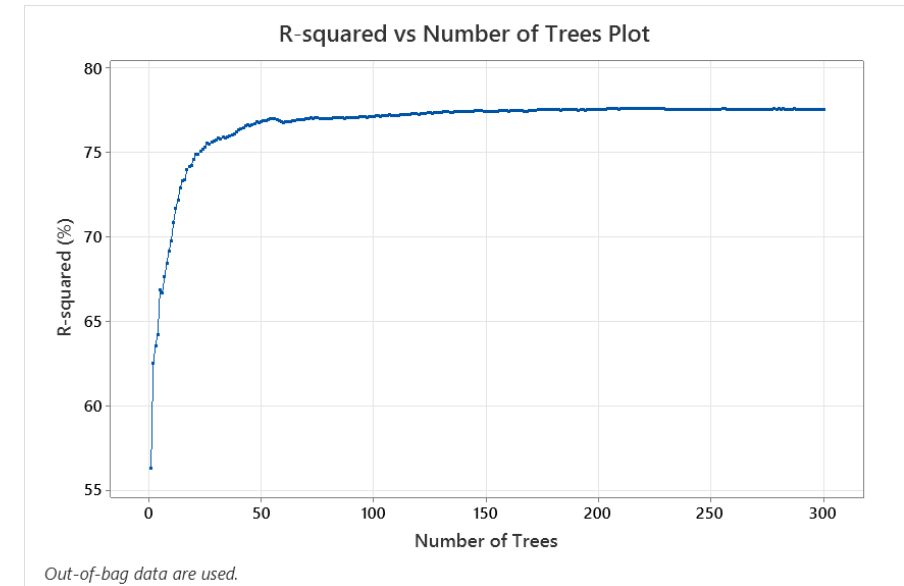
User Benefit: Random Forests generally provide better predictive power than a single decision tree.

Based on a collection of CART Trees, Random Forests leverages repetition, randomization, sampling, and ensemble learning in one convenient place that brings together independent trees and determines the overall prediction of the forest.

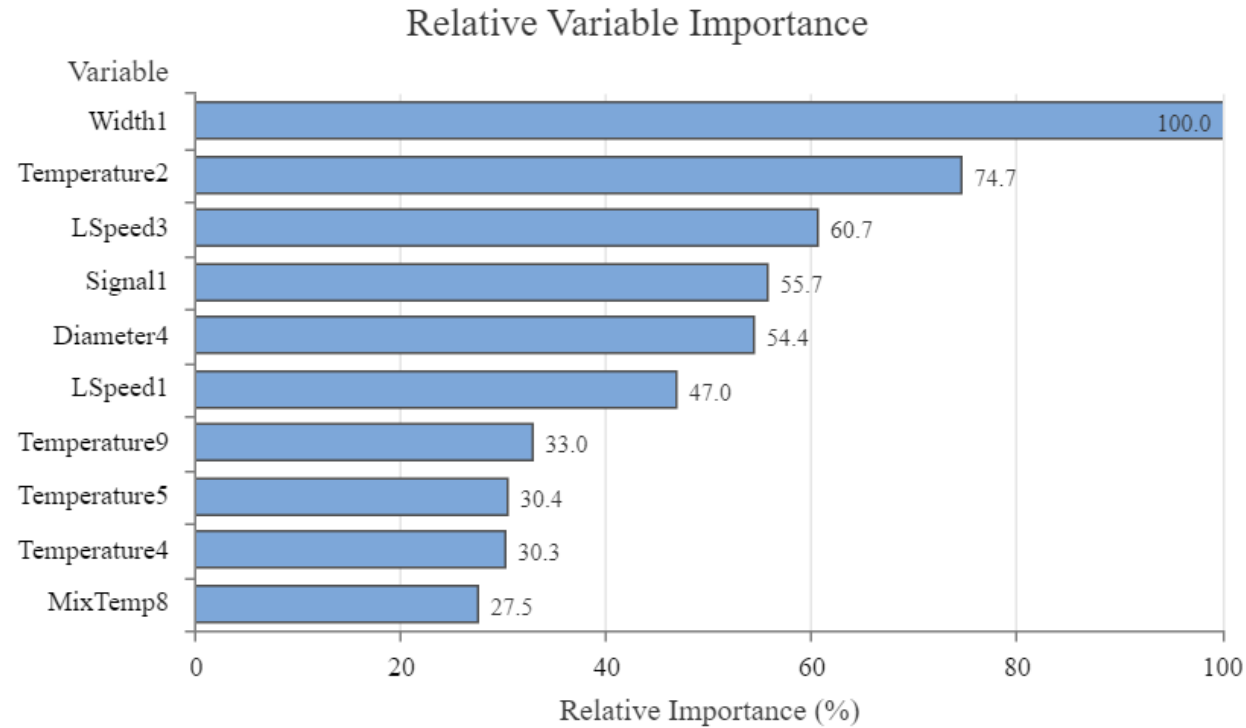
Available in Minitab's Predictive Analytics Module



Random Forests®

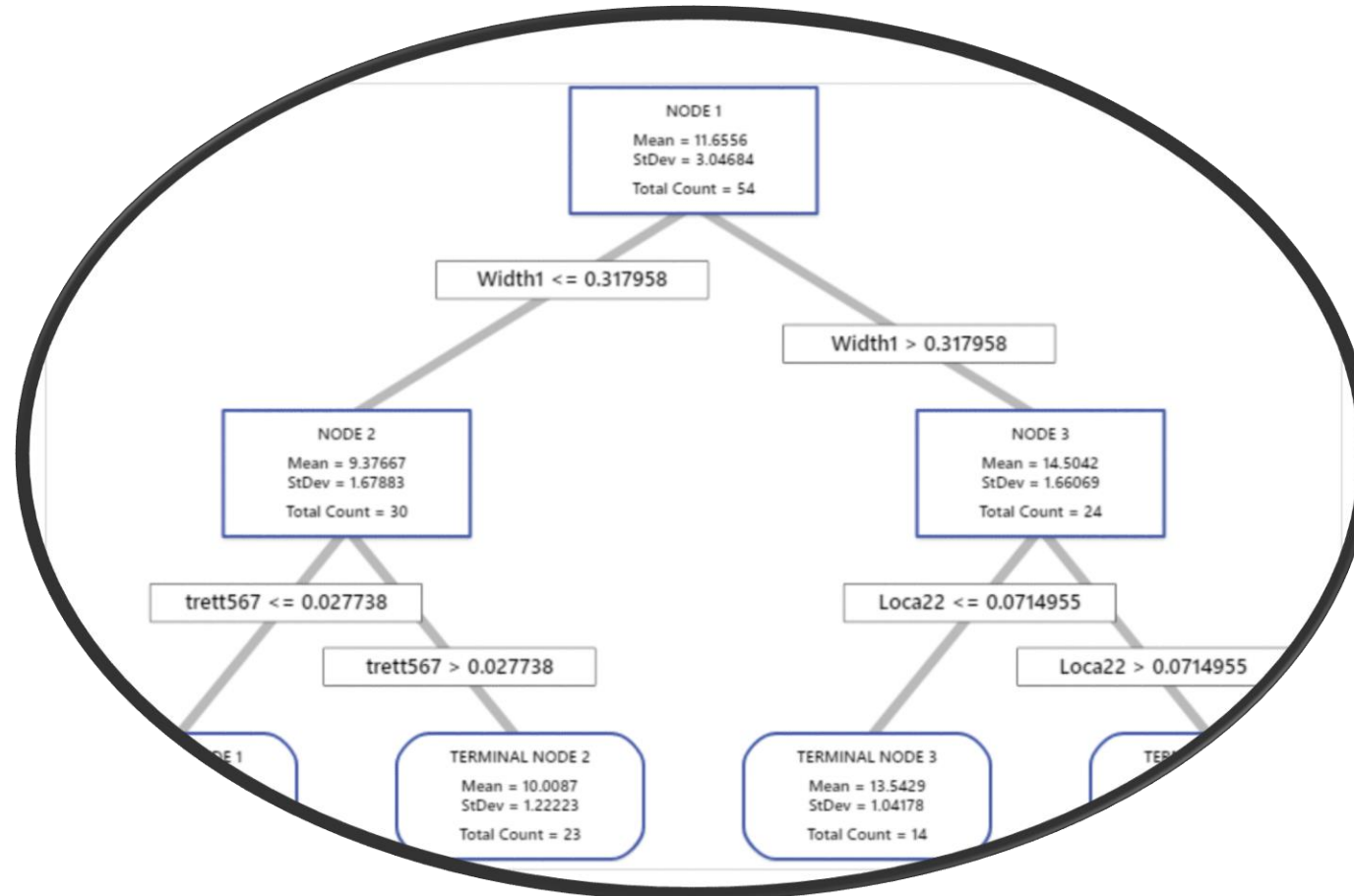


Variable Importance



*Variable importance measures model improvement when splits are made on a predictor.
Relative importance is defined as % improvement with respect to the top predictor.*

CART[®] Tree to share recommendations



Learn more about Random Forests®

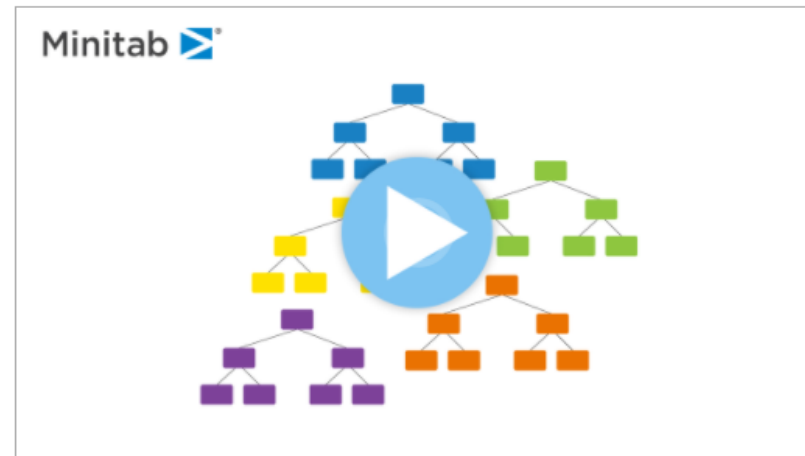
BREIMAN AND CUTLER'S RANDOM FORESTS

Random Forests®

Based on a collection of Classification & Regression Trees (CART®), Random Forests® modeling engine sums the predictions made from each CART tree to determine the overall prediction of the forest, while ensuring the decision trees are not influenced by one another.

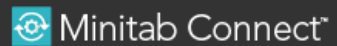
For those new to Random Forests, it is a powerful ensemble technique developed by Leo Breiman and Adele Cutler at the University of California, Berkeley, and is favored by many predictive modeling practitioners. The deceptive simplicity of the algorithm builds hundreds of independent trees and employs lots of sampling from both observations and variables.

Random Forests' unique ability to evaluate unbiased model performance based on the out-of-bag data removes the need to have a separate testing/validation sample. This immediately positions Random Forests as the top predictive modeling tool in the wide data applications where the number of variables exceeds, often many times over, the number of available observations.



Minitab 20.2

ADDITIONAL FUNCTIONALITY



Improvement: Lag

Improvement: Previous versions of Minitab performed lags for 1 variable at a time. Users can now create lags for up to 12 variables (series) at a time. The dialog can also accommodate multiple lags (e.g 1, 2, 3,4 and 5), as shown here.

Benefit: Lagged columns are commonly used in time series modeling and in supervised machine learning such as CART, TreeNet and Random Forests. This improvement allows for faster data preparation.

This command is in **Stat-Time Series-Lag**

Specify root series and lags:

Root Series	Lags
Measures	1:5



C1	C2	C3	C4	C5	C6
Measures	Measures_Lag1	Measures_Lag2	Measures_Lag3	Measures_Lag4	Measures_Lag5
20	*	*	*	*	*
32	20	*	*	*	*
22	32	20	*	*	*
27	22	32	20	*	*
7	27	22	32	20	*
12	7	27	22	32	20
18	12	7	27	22	32
9	18	12	7	27	22
11	9	18	12	7	27
34	11	9	18	12	7
29	34	11	9	18	12
10	29	24	11	9	18

Addition: Open-Source Integration

R

Addition: Call R scripts from Minitab Statistical Software. R is a language and environment for statistical computing and graphics.

R scripts can run in 3 ways:

- Run the **rscr** in the Command Line pane
- Run a Minitab exec that includes the R command
- Customize the interface to run a Minitab exec that includes the R command

Execute external R scripts that use Minitab variables (columns, constants, matrices) as inputs. Results are returned to Minitab and displayed in the Output Navigator and Output Pane.



Takeaways







- Analytics in small, manageable steps
- Analytics + Expertise
- Work Smarter



Our Approach: More Than Business Analytics... Solutions Analytics

Solutions analytics is our integrated approach to providing software and services that enable organizations to make better decisions that drive business excellence.

Software

Data Analysis	Data Transformation	Predictive Modeling	Online Training	Visual Business Tools	Project Oversight
 Minitab®	 Minitab Connect™	 SPM®	 Quality Trainer®	 Minitab Workspace™	 Minitab Engage™
Powerful statistical software everyone can use.	Data access, automation, and governance for comprehensive insights.	Machine learning and predictive analytics software.	Master statistics and Minitab anywhere with online training.	Visual tools to process and product excellence.	Start, track, manage and execute improvement projects in real-time.

Services



Training

Learn first-hand by attending public or customized trainings in your facilities according to your requirements.



Statistical Consulting

Personalized help with statistical challenges from collecting the right data to interpreting analysis more.



Support

Assistance with installation, implementation, version updates and license management.

Flexible, Hands-On Training

Virtual Training

Gain insights into your data with courses guided by expert statisticians.

On-Site Training

Schedule time with our trainers to review your data to address your specific business challenges.

Learn more at:

minitab.com/training



Upcoming Events – Machine Learning Week

Ready for more predictive analytics?

Virtually attend **Machine Learning Week**, the **premier machine learning event** held by Predictive Analytics World (PAW)

Minitab will be presenting and sponsoring this amazing **livestreamed conference**

May 24-28, 2021 for US time zones

June 14-18, 2021 for EMEA time zones



Find out more:

www.minitab.com/en-us/events/

Thank You!

From all of us at

Minitab ®