Introduction to Decision Trees

Jarlath Quinn

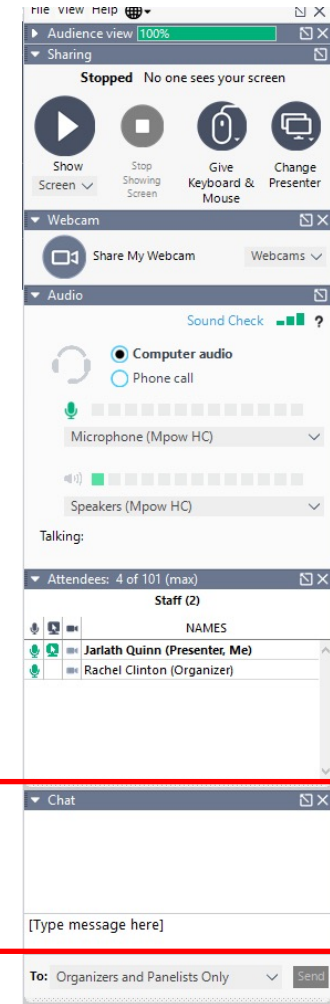A SELECT INTERNATIONAL COMPANY

FAQ's

- Is this session being recorded? Yes

- Can I get a copy of the slides? Yes, we'll email a PDF copy to you after the session has ended.

- Can we arrange a re-run for colleagues? Yes, just ask us.

- How can I ask questions? All lines are muted so please use the chat facility – if we run out of time we will follow up with you.

- Premier accredited partner to IBM and Predictive Solutions specialising in advanced analytics & big data technologies

- Work with open source technologies (R, Python, Spark etc.)

- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
  - Retail
  - Gaming
  - Utilities
 - Insurance
  - Telecommunications
  - Media
  - FMCG
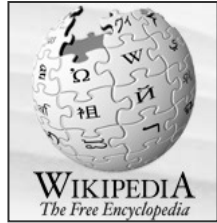
"**Predictive analytics** encompasses a variety of [statistical](#) techniques from [data mining](#), [predictive modelling](#), and [machine learning](#), that analyze current and historical facts to make [predictions](#) about future or otherwise unknown events."

What do we mean by 'Predictive Analytics'?

- It's different from Business Intelligence or MI reporting

- Actually, it's not *always* about prediction

- However, Predictive Analytics *does* creates important new data

- These data take the form of estimates, probabilities, forecasts, recommendations, propensity scores, classifications or likelihood values

- Which in turn can be incorporated into key operational and/or insight systems
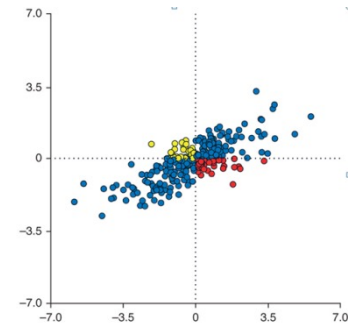
Types of Predictive Analytics

- **Classification / Propensity**
  - Who is most likely to respond / upgrade/recommend/defect based on the historical behavioural data we have about them?

- **Clustering**
  - How can I divide my customers into meaningful and usable groups as a framework for marketing communications?

- **Association & Sequence**
  - What combinations of product purchases or events co-occur more often than normal? What natural affinities exist within the data?

- **Time Series**
  - What will product demand/revenue/website hits/visitor numbers look like in the next hour/day/month/quarter/ year?

Where do Decision Trees fit within Predictive Analytics?

- Decision trees are used *extensively and widely* within Predictive Analytics
- Decision trees can be used to
  - Build profiles of customers/employees/clients
  - Find key behavioural segments
  - Generate predictive models
- Decision Trees can be expressed as a series of hierarchical rules which means that they can be converted in languages like SQL for database scoring
- Decision Trees are especially popular because
  - they are fairly visual representations of models
  - relatively easy to understand

Understanding Decision Trees – a worked example

- What were the most important factors determining survival during the sinking of the RMS Titanic?

Gender?

Survival on the RMS Titanic

		Count	Percent %
survive	Did not survive	1490	68%
	Survived	711	32%
	Total	2201	100%

Age?

Class?

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		sex			
		female		male	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	126	26.8%	1364	78.8%
	Survived	344	73.2%	367	21.2%
	Total	470	100.0%	1731	100.0%

Pearson Chi-Square Tests

		sex
survive	Chi-square	456.874
	df	1
	Sig.	.000*

SMART VISION
E u r o p e

A SELECT INTERNATIONAL COMPANY

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		age			
		adult		child	
		Count	Column Percent %	Count	Column Percent %
survive	Did not survive	1438	68.7%	52	47.7%
	Survived	654	31.3%	57	52.3%
	Total	2092	100.0%	109	100.0%

Pearson Chi-Square Tests

		age
survive	Chi-square	20.956
	df	1
	Sig.	.000*

SMART VISION
E u r o p e

Statistical Tests Like Chi Square help to answer this

Survival on the RMS Titanic

		class							
		1st		2nd		3rd		crew	
		Count	Column Percent %	Count	Column Percent %	Count	Column Percent %	Count	Column Percent %
survive	Did not survive	122	37.5%	167	58.6%	528	74.8%	673	76.0%
	Survived	203	62.5%	118	41.4%	178	25.2%	212	24.0%
	Total	325	100.0%	285	100.0%	706	100.0%	885	100.0%

Pearson Chi-Square Tests

		class
survive	Chi-square	190.401
	df	3
	Sig.	.000*

SMARTVISION
Europe

Gender is most important

…and a CHAID Decision tree will reflect this….

Dependent/Target Field

Baseline proportions

Independent/Predictor Field

Node Number

Predicted Survival Outcome
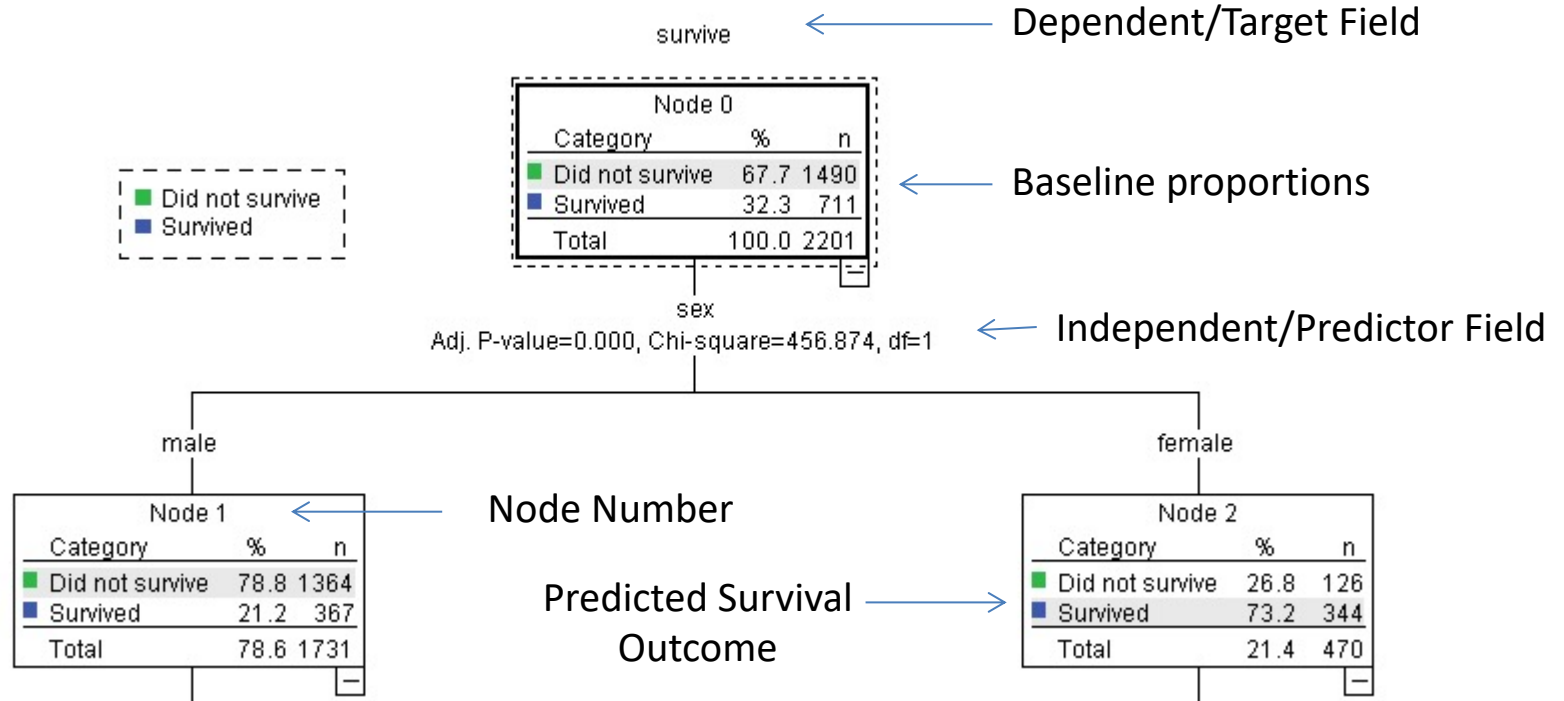
Full CHAID Decision Tree

C.H.A.I.D
Chi-Square
Automatic
Interaction
Detector

Terminal nodes generate probability scores
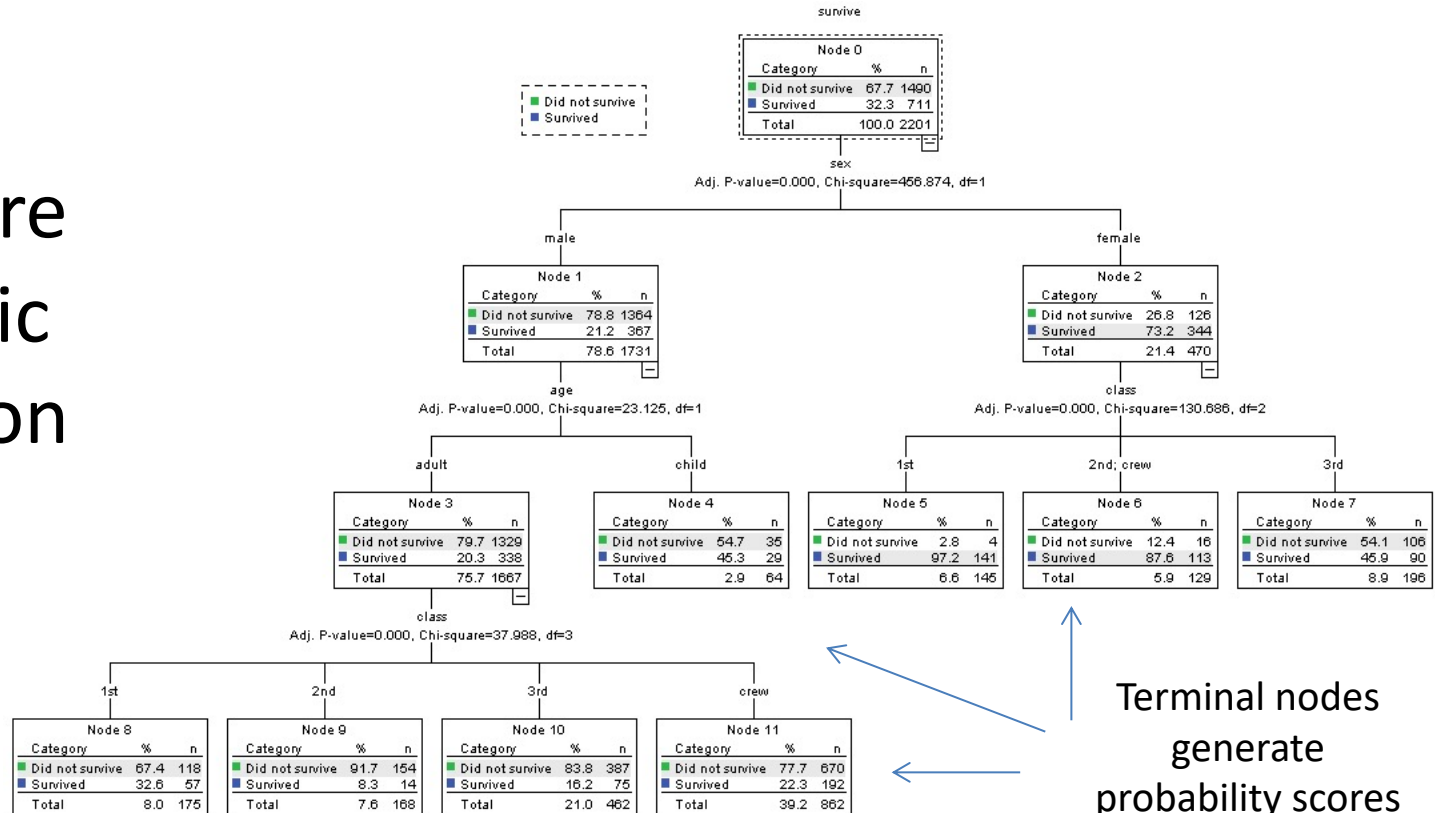
Merging/Splitting in CHAID Trees

Decision Trees can merge values of numeric *and* categorical predictors together

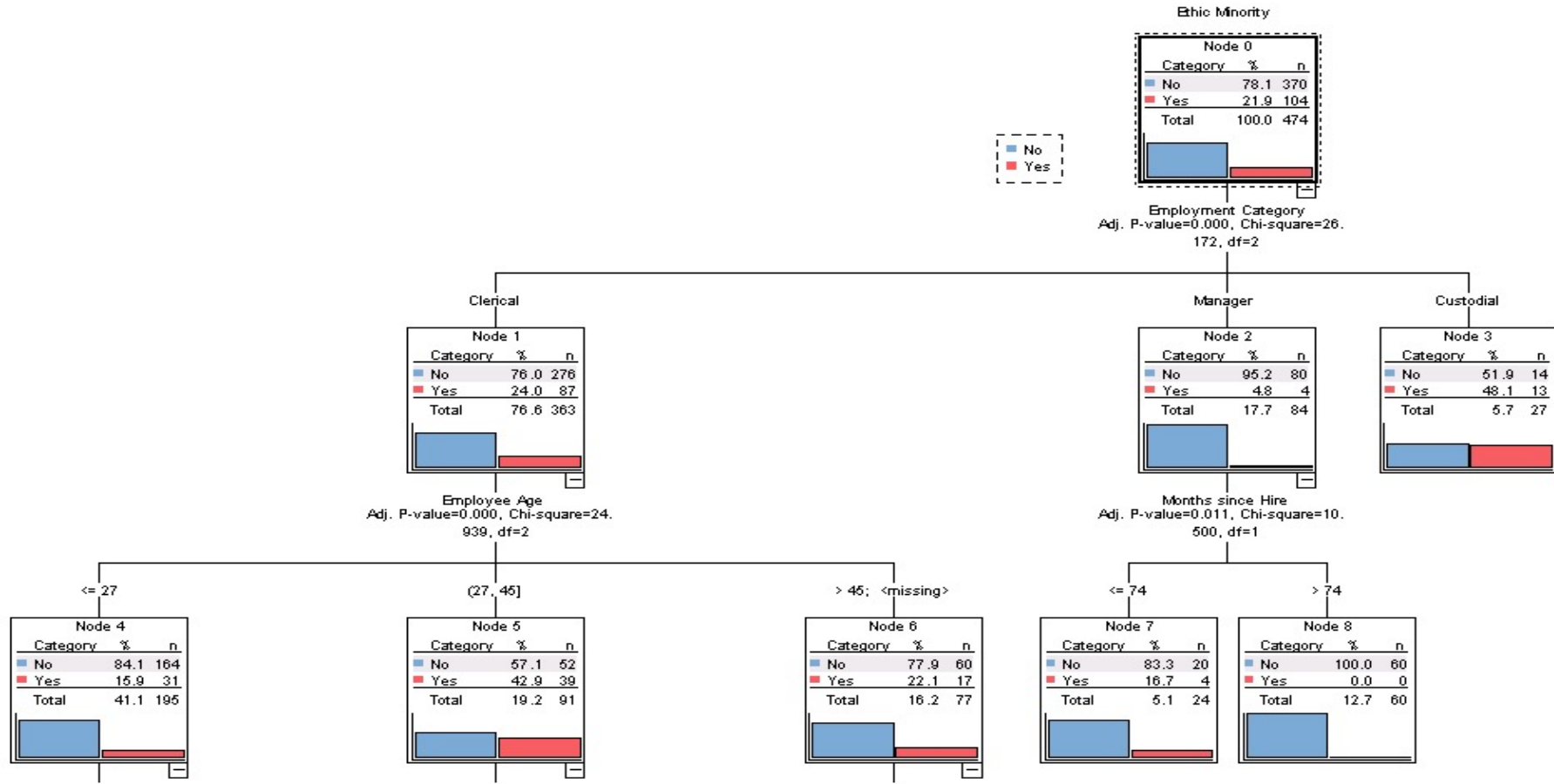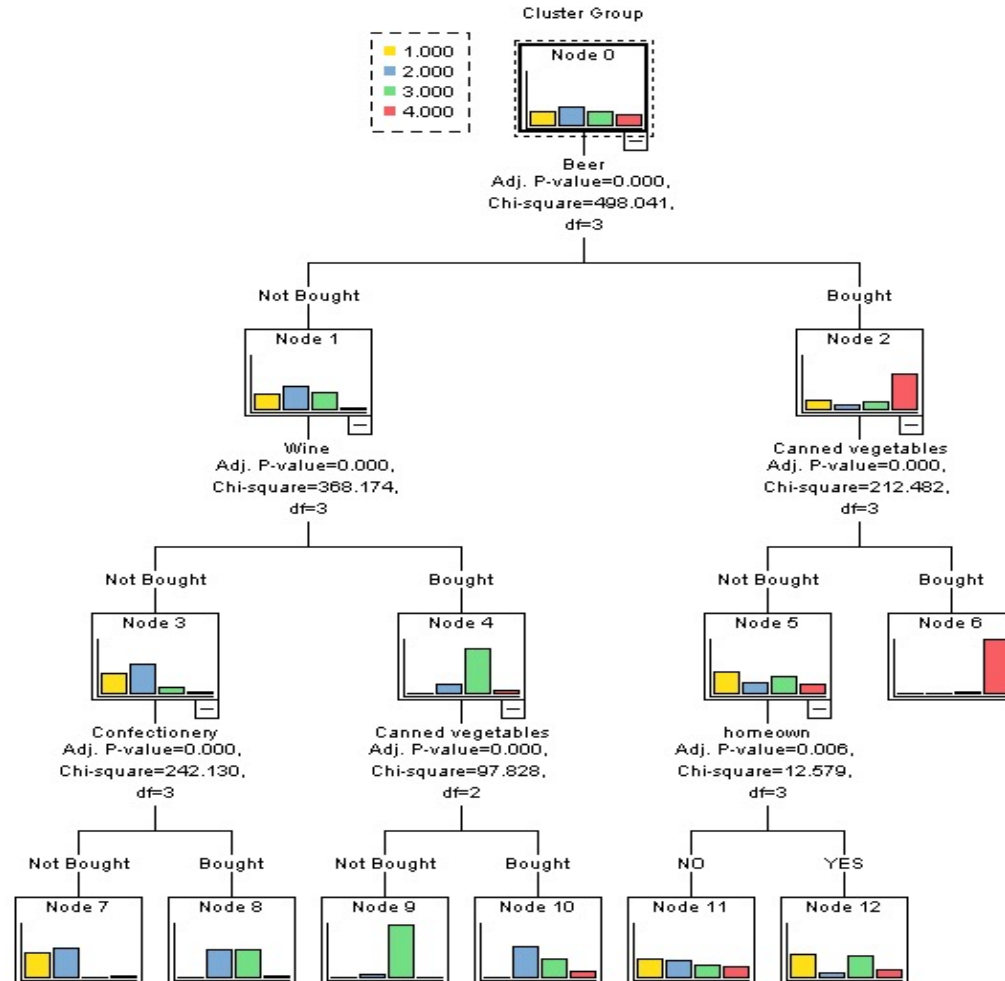This makes the tree more efficient and easier to read

Buy financial service

Node 0		
Category	%	n
■ No	43.6	697
■ Yes	56.4	902
Total	100.0	1599

■ No
■ Yes

Amount in savings or investment
Adj. P-value=0.000, Chi-square=208.199, df=3

<= 4395.000   (4395.000, 6150.000]   (6150.000, 8000.000]   > 8000.000

Node 1		
Category	%	n
■ No	71.2	227
■ Yes	28.8	92
Total	19.9	319

Node 2		
Category	%	n
■ No	51.6	248
■ Yes	48.4	233
Total	30.1	481

Node 3		
Category	%	n
■ No	36.2	119
■ Yes	63.8	210
Total	20.6	329

Node 4		
Category	%	n
■ No	21.9	103
■ Yes	78.1	367
Total	29.4	470

SMARTVISION
Europe

Let's see a demonstration…

A SELECT INTERNATIONAL COMPANY

Demographic Profiling

Interpreting Clusters

Drivers of Satisfaction
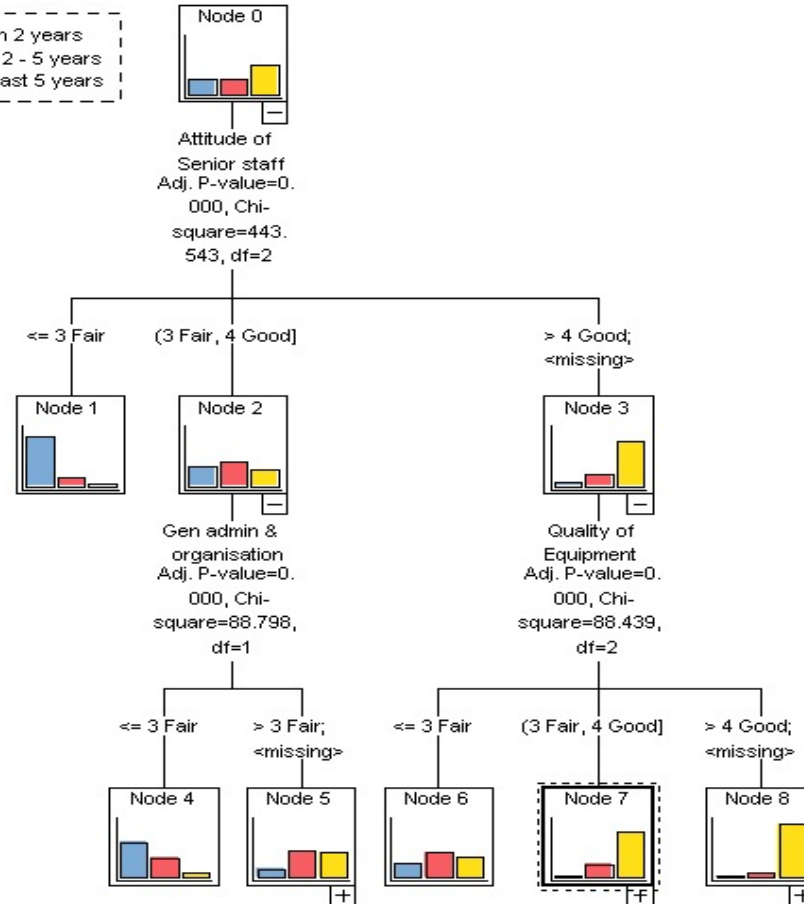
Online training
materials free to
Smart Vision
customers or available
for purchase

Factor and Cluster Analysis with
IBM SPSS Statistics

£75.00
Jarlath Quinn

Introduction to Time Series
Forecasting with IBM SPSS
Statistics

£75.00
Jarlath Quinn

Understanding and applying
logistic regression techniques in
SPSS Statistics

£75.00
Jarlath Quinn

Understanding and Applying
Linear Regression Techniques in
SPSS Statistics

£75.00
Jarlath Quinn

Building predictive models in
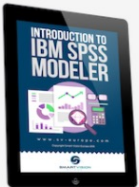SPSS Modeler

£75.00
Jarlath Quinn

Statistical and significance
testing in SPSS Statistics

£75.00
Jarlath Quinn

Working with decision trees in
SPSS Statistics

Introduction to SPSS Modeler
course

Introduction to IBM SPSS
Statistics course

Download our new e-book for free

The insider's guide to predictive analytics

£0.00

- 1 + **Add to basket**

Category: books

Working with Smart Vision Europe Ltd.

- **Sourcing Software**
    - You can buy your analytical software from us often with discounts
    - Assist with selection, pilot, implementation & support of analytical tools
    - http://www.sv-europe.com/buy-spss-online/
- **Training and Consulting Services**
    - Guided consulting & training to develop in house skills
    - Delivery of classroom training courses / side by side training support
    - Identification & recruitment of analytical skills into your organisation
- **Advice and Support**
    - offer 'no strings attached' technical and business advice relating to analytical activities
    - Technical support services around SPSS

SMARTVISION
Europe

A SELECT INTERNATIONAL COMPANY

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you