

UNIVERSITY OF TECHNOLOGY, SYDNEY
FACULTY OF INFORMATION TECHNOLOGY

A Strategic Analytics Methodology

By Marcel van Rooyen

A dissertation submitted in fulfillment of the requirements for the degree of Master in
Computing Sciences

2005

Acknowledgements

Associate Professor Simeon Simoff for his vision of what we had to achieve together in this novel area of research, his insightful guidance about how to achieve it, and his encouragement to perseverance at times of frustration. Professor Ken Miller for his direction about marketing subject matter.

Family and friends for their faith in me during a mid-life career change, and for the neglect they endured for it. Yong Mei Tang (Grace) for her unquestionable support during unquestionable neglect, and for supporting with graphic design.

The University of Technology for an RTS Scholarship and the fantastic research infrastructure.

SAS Institute Australia Pty. Ltd. and key individuals there, for access to their software, training support, professional networking opportunities, access to proprietary research literature, and the introduction to the industrial research partner.

An Australian mobile phone company for the industrial research partnership. In particular their Retention and Business Intelligence Managers for their support and interest in the project, and two analysts who endured me while transferring invaluable knowledge to me.

This work is a celebration of the pleasant surprises life presents to the fortuitous and adventurous

Table of contents

Acknowledgements	i
Table of contents	ii
Certificate of Authorship and Originality	xv
List of figures	xvi
List of tables	xviii
Abstract	xix
1 Chapter 1	1
1.1 Context of the research problem	1
1.2 Research problem, goal, objectives, scope, and significance	7
1.2.1 Research problem defined.....	7
1.2.2 Research goal and objectives	7
1.2.3 Research scope	8
1.3 Research methods.....	9
1.3.1 Description of methods	9
1.3.2 Research outcomes and benefits	15
1.3.2.1 For commercial practitioners	15
1.3.2.2 For software vendors and consultants	15
1.3.2.3 For the industry partner.....	15
1.4 Thesis structure	15
1.5 Chapter summary	17
2 CHAPTER 2 – Contextual literature research	19
2.1 Information and Knowledge Management.....	19
2.2 Cognitive Psychology Theory.....	22
2.2.1 Perception process.....	22

2.2.1.1	Awareness response stage	23
2.2.1.2	Perceptive meaning appraisal stage	25
2.2.2	Cognition process	26
2.3	Strategic Planning Model	27
2.3.1	Mission and circumstantial profile	29
2.3.2	Strategic analysis	31
2.3.3	Strategic choice	32
2.3.4	New Strategic and Operating objectives and strategies	33
2.3.5	Implement and deploy new strategies (Execute)	34
2.3.6	Monitor and control	35
2.4	Strategic Planning Model as a Knowledge Management tool	36
2.5	A practical pre-project schema	38
2.6	Relevant sections of Telco ABC's Corporate Strategy	39
2.6.1	Mission (SPM 1.1)	39
2.6.2	Controllable factors profile (<i>SPM 1.2</i>)	40
2.6.3	Uncontrollable factors profile (<i>SPM 1.3</i>)	41
2.6.4	Strategic analysis (SPM 2)	41
2.6.5	Strategic choice (SPM 3)	42
2.6.6	Strategic objectives (SPM 4.1)	42
2.6.7	Grand strategy (SPM 4.2)	43
2.6.8	Operating objectives (SPM 4.3)	43
2.6.9	Operating strategies (SPM 4.4)	44
2.6.10	Implement and deploy new strategies (Execute) (SPM 5)	45
2.6.11	Monitor and control (SPM 6)	45
2.7	Total Quality Management	45
2.8	Lessons from the ERP environment	47

2.8.1	ERP and data mining compared.....	48
2.8.2	An example of an ERP project.....	49
2.8.2.1	Mission (SPM 1.1).....	49
2.8.2.2	Controllable factors profile (<i>SPM 1.2</i>).....	50
2.8.2.3	Uncontrollable factors profile (<i>SPM 1.3</i>).....	52
2.8.2.4	Strategic analysis (SPM 2) (KM 1.3).....	52
2.8.2.5	Strategic choice (SPM 3)(KM 2.1)	53
2.8.2.6	New Strategic and Operating strategies and objectives (SPM 4.1 – 4.4)(KM2.2)	53
2.8.2.7	Company A’s new Strategic objectives (<i>SPM 4.1</i>).....	53
2.8.2.8	Company A’s new Grand strategy (<i>SPM 4.2</i>).....	54
2.8.2.9	Company A’s new Operating objectives (<i>SPM 4.3</i>)	54
2.8.2.10	Company A’s new Operating strategies (<i>SPM 4.4</i>)	55
2.8.2.11	Implement and deploy new strategies (<i>SPM 5</i>)(<i>KM 3</i>).....	56
2.8.3	Role of new knowledge and collaborative expert teamwork.....	57
2.8.4	Role of Strategic Planning Model in the ERP solution.....	57
2.9	Risk Management.....	58
2.10	New marketing subject matter	60
2.10.1	Target	63
2.10.2	Segment.....	64
2.10.3	Position.....	66
2.10.4	Algorithmic innovations: PROMIX.....	68
2.10.5	Concluding about new domain knowledge	73
2.11	CRISP-DM data mining standard for Business Intelligence application.....	74
2.11.1	Business understanding.....	74
2.11.2	Data understanding.....	75

2.11.3	Data preparation	75
2.11.4	Modeling	76
2.11.5	Evaluation	76
2.11.6	Deployment	77
2.12	Chapter summary	77
3	Chapter 3 - Concept drift detection methodology in data mining.....	79
3.1	Context of concept drift	79
3.1.1	Practical applications of concept drift.....	82
3.1.2	The implications of a drifting concept	82
3.2	Concept drift detection methodology.....	83
3.2.1	Summary of concept drift in the automated solution environment.....	94
3.3	Making the case for using concept drift in Strategic Planning Cycle	96
3.3.1	Concept drift as a Knowledge Management process	99
3.4	Chapter summary	103
4	CHAPTER 4 – Evaluation of CRISP-DM.....	105
4.1	Diagnostic Technique for Defining Business Deliverables	106
4.1.1	Analysis of CRISP-DM for diagnostic technique.....	108
4.1.2	Research findings	112
4.1.3	Reflection-in-action.....	113
4.1.4	Evaluative reflection and reframing.....	115
4.2	Introducing new subject matter expertise	116
4.2.1	Analysis of CRISP-DM for introducing new subject matter	116
4.2.2	Research findings	119
4.2.3	Reflection-in-action.....	119
4.2.4	Evaluative reflection and reframing.....	119
4.3	Mapping technique between business deliverables and data mining plan....	119

4.3.1	Analysis of CRISP-DM for mapping technique	120
4.3.2	Research findings	122
4.3.3	Reflection in action	122
4.3.4	Evaluative reflection and reframing.....	124
4.4	Knowledge management activities	125
4.4.1	Analysis of CRISP-DM for knowledge management activities	126
4.4.2	Research findings	129
4.4.3	Reflection in action	129
4.4.4	Evaluative reflection and reframing.....	131
4.5	Monitor and control plan.....	131
4.5.1	Analysis of CRISP-DM for a monitor and control plan	132
4.5.2	Research findings	133
4.5.3	Reflection in action	133
4.5.4	Evaluative reflection and reframing.....	134
4.6	General and soft issues.....	134
4.6.1	General issues.....	134
4.6.1.1	Artificial separation of model and information evaluation	134
4.6.1.2	Consideration for TQM principles	135
4.6.1.3	About data mining discovering knowledge.....	135
4.6.1.4	Lack of content about the open business environment	135
4.6.1.5	Lack of feedback loop.....	136
4.6.2	Soft issues – the impact of the human factor	136
4.6.2.1	Role of subject matter expertise in discovery	136
4.6.2.2	Role of collaborative teamwork.....	136
4.6.2.3	Role of professional circumstances.....	137
4.7	Chapter summary	138

5	Chapter 5 – Developing Strategic Analytics Methodology	141
5.1	Supporting strategic alignment	141
5.2	Dimensions of Strategic Analytics Methodology	143
5.3	The role of SPM in SAM	143
5.4	Strategic progression.....	144
5.5	SAM as a Knowledge Management cycle	146
5.5.1	Create	147
5.5.2	Legitimise.....	149
5.5.3	Share.....	149
5.5.4	Monitor and control.....	150
5.6	Strategic Planning Cycle phases of SAM	150
5.6.1	Prepare.....	151
5.6.2	Analyse.....	151
5.6.3	Choose.....	152
5.6.4	Define	152
5.6.5	Realise	153
5.6.6	Monitor and control.....	153
5.7	SAM as a reframing of CRISP-DM.....	154
5.8	Functioning of the SAM elements	154
5.8.1	Business problem	155
5.8.1.1	What <i>Business problem</i> is	155
5.8.1.2	<i>Business problem's</i> strategic purpose	156
5.8.1.3	How <i>Business problem</i> supports	156
5.8.2	Potential solutions	157
5.8.2.1	What <i>Potential solutions</i> is	157
5.8.2.2	<i>Potential solutions's</i> strategic purpose.....	158

5.8.2.3	How <i>Potential solutions</i> supports	159
5.8.3	Develop project mission.....	160
5.8.3.1	What <i>Develop project mission</i> is.....	160
5.8.3.2	<i>Develop project mission's</i> strategic purpose.....	162
5.8.3.3	How <i>Develop project mission</i> supports	163
5.8.4	Expert collaboration	165
5.8.4.1	What <i>Expert collaboration</i> is	165
5.8.4.2	<i>Expert collaboration's</i> strategic purpose	167
5.8.4.3	How <i>Expert collaboration</i> supports	167
5.8.5	Identify, assemble, prepare useful data	168
5.8.5.1	What <i>Identify, assemble, prepare useful data</i> is	168
5.8.5.2	<i>Identify, assemble, prepare useful data's</i> support.....	169
5.8.5.3	How <i>Identify, assemble, prepare useful data</i> supports	171
5.8.6	Data mining discovery	180
5.8.6.1	What <i>Data mining discovery</i> is	180
5.8.6.2	<i>Data mining discovery's</i> support	181
5.8.6.3	How <i>Data mining discovery</i> supports	181
5.8.7	Develop circumstantial knowledge	188
5.8.7.1	What <i>Develop circumstantial knowledge</i> is.....	188
5.8.7.2	<i>Develop circumstantial knowledge's</i> support	189
5.8.7.3	How <i>Develop circumstantial knowledge</i> supports.....	189
5.8.8	Strategic analysis.....	189
5.8.8.1	What <i>Strategic analysis</i> is.....	189
5.8.8.2	<i>Strategic analysis's</i> strategic support	190
5.8.8.3	How <i>Strategic analysis</i> supports.....	190
5.8.9	Strategic choice	190

5.8.9.1	What <i>Strategic choice</i> is.....	190
5.8.9.2	<i>Strategic choice's</i> strategic support	190
5.8.9.3	How <i>Strategic choice</i> supports.....	191
5.8.10	Define new business objectives and strategies.....	192
5.8.10.1	What <i>Define new business objectives and strategies</i> is	192
5.8.10.2	<i>Define new business objectives and strategies'</i> support	192
5.8.10.3	How <i>Define new business objectives and strategies</i> supports	192
5.8.11	Develop data mining plan	192
5.8.11.1	What <i>Develop data mining plan</i> is.....	192
5.8.11.2	<i>Develop data mining plan'</i> support.....	193
5.8.11.3	How <i>Develop data mining plan</i> supports.....	194
5.8.12	Model, evaluate, choose best model(s)	194
5.8.12.1	What <i>Model, evaluate, choose best model(s)</i> is	195
5.8.12.2	<i>Model, evaluate, choose best model(s)'s</i> support.....	195
5.8.12.3	How <i>Model, evaluate, choose best model(s)</i> supports	195
5.8.13	Operationalise model(s)	196
5.8.13.1	What <i>Operationalise model(s)</i> is	196
5.8.13.2	<i>Operationalise model(s)'</i> support	196
5.8.13.3	How <i>Operationalise model(s)</i> supports	196
5.8.14	Deploy outputs into business	196
5.8.14.1	What <i>Deploy outputs into business</i> is	196
5.8.14.2	<i>Deploy outputs into business's</i> support.....	196
5.8.14.3	How <i>Deploy outputs into business</i> supports	196
5.8.15	Execute new business strategies	197
5.8.15.1	What <i>Execute new business strategies</i> is	197
5.8.15.2	<i>Execute new business strategies'</i> support	197

5.8.15.3	How <i>Execute new business strategies</i> supports	197
5.8.16	Monitor and control.....	197
5.8.16.1	What <i>Monitor and control</i> is.....	197
5.8.16.2	<i>Monitor and control's</i> support	198
5.8.16.3	How <i>Monitor and control</i> adds business value.....	198
5.9	Chapter summary	200
6	Chapter 6 – Apply SAM for discovery	204
6.1	Business problem	204
6.1.1	The existing paradigm.....	204
6.1.2	Pre-project schema	205
6.1.3	Estimating the economic magnitude of Business problem	207
6.2	Potential solutions	208
6.3	Develop project mission.....	208
6.3.1	SPC goals and strategies	208
6.3.1.1	Understanding the extent of the problem (SCHEMA 1).....	208
6.3.1.2	Understand the causes of retention problem (SCHEMA 2).....	212
6.3.1.3	Investigating a new solution (SCHEMA 3)	212
6.3.1.4	Solution development (SCHEMA 4)	218
6.3.1.5	Solution support (SCHEMA 5).....	219
6.3.2	The departments affected	220
6.3.3	Frequency and duration.....	220
6.4	Identify, assemble, prepare useful data	221
6.4.1	Profile for relevance	221
6.4.1.1	Profile for business relevance	221
6.4.1.2	Select business relevant data	224
6.4.1.3	Profile for potential technical signal	224

6.4.1.4	Select data with potential technical signal	225
6.4.2	Profile data assembleability	225
6.4.2.1	Profile data assembleability	226
6.4.2.2	Identify assembleable data	228
6.4.3	Develop technical desirability.....	228
6.4.3.1	Extract	228
6.4.3.2	Data assembly	229
6.4.3.3	Label the data for the churn event.....	230
6.4.3.4	Clean	233
6.4.3.5	Create additional features.....	237
6.4.3.6	Segmentation of $CDR_{modplus t}$	238
6.4.3.7	Sampling	241
6.4.3.8	Dealing with data Distribution issues – final data transforms	241
6.4.3.9	Statistical feature selection.....	244
6.4.3.10	Data partitioning.....	246
6.5	Data mining discovery	247
6.5.1	Develop data mining mission.....	247
6.5.1.1	Data mining goal for SPC 2	247
6.5.1.2	Data mining strategies for SPC 2.....	247
6.5.1.3	Determine the where for SPC 2	248
6.5.1.4	Set confidence levels for SPC 2.....	248
6.5.1.5	Data mining goals for SPC 3 - targeting	248
6.5.1.6	Data mining strategies for SPC 3 - targeting	248
6.5.1.7	Data mining goal for SPC 3 – segmenting.....	249
6.5.1.8	Data mining strategies for SPC 3 – segmenting.....	249
6.5.1.9	Data mining goals for SPC 3 – root cause profiling	250

6.5.1.10	Data mining strategies for SPC 3 – root cause profiling.....	250
6.5.1.11	Determine the where for SPC 3	250
6.5.1.12	Set the confidence levels for SPC 3	250
6.5.1.13	Comment on SPC 4 and 5	252
6.5.1.14	On executing data mining mission for SPC 2	252
6.5.2	Execute data mining mission for SPC 2 and 3	252
6.5.2.1	Execute SPC 3 strategy one and SPC 2 strategy	252
6.5.2.2	Knowledge development looping for SPC 2 – root cause	256
6.5.2.3	Execute SPC 3 strategy two, three, and four – targeting	257
6.5.2.4	Knowledge development looping for SPC 3 – targeting	260
6.5.2.5	Execute SPC 3 strategies – segmenting	261
6.5.2.6	Knowledge development looping for SPC 3 – segmentation	265
6.5.2.7	Execute SPC 3 strategies – root cause profiling	272
6.5.2.8	Knowledge development looping for SPC 3 – root cause profiling	273
6.6	Chapter summary	276
7	Chapter 7 – Apply SAM for solution development	280
7.1	Knowledge development loop (Develop circumstantial knowledge, Strategic analysis, and Strategic choice)	282
7.1.1	Develop circumstantial knowledge	282
7.1.1.1	Execute SPC 4 strategy one	282
7.1.2	Strategic analysis.....	296
7.1.3	Strategic choice	298
7.2	Define new business objectives and strategies.....	298
7.2.1	Execute SPC 4 SPC strategy two	298
7.2.1.1	New <i>Strategic objective</i>	298
7.2.1.2	New <i>Grand strategy</i>	299

7.2.1.3	New <i>Operating objective</i>	299
7.2.1.4	New <i>Operating strategies</i>	299
7.3	Develop data mining plan	300
7.3.1	Data mining objectives.....	300
7.3.2	Data mining strategies.....	301
7.4	Model, evaluate, choose best model(s)	303
7.4.1	Optimise the classifier.....	304
7.4.1.1	Evaluate and select.....	305
7.4.2	Re-target.....	306
7.4.3	Re-segment.....	307
7.4.3.1	Evaluate and select.....	307
7.5	Operationalise model(s)	308
7.6	Deploy outputs into business	309
7.7	Execute new business strategies	313
7.8	Monitor and control.....	313
7.8.1	Problem understanding.....	314
7.8.2	Business solution relevance	315
7.8.3	Project return on investment	316
7.8.4	Visualising the monitoring.....	317
7.9	Chapter summary	321
8	CHAPTER 8. Conclusions and future research directions	324
8.1	Research contributions	324
8.1.1	Research methods.....	324
8.1.2	Knowledge Discovery and Data Mining.....	324
8.1.1.1	Data mining project methodology.....	324
8.1.1.2	Technical.....	325

8.1.3	Business intelligence practitioners.....	326
8.1.4	Business intelligence software vendors	327
8.1.5	Telco industry and our industrial research partner.....	328
8.2	Ideas for future research.....	328
8.3	Concluding remarks	330
9	Appendix A: Terms, abbreviations, acronyms.....	331
10	Bibliography.....	334

Certificate of Authorship and Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that this thesis has been written by me. Any help that I received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

List of figures

Figure number	Figure name	on page
1.1	Elements of research problem	6
1.2	Research goals operationalising	14
1.3	Thesis structure	16
2.1	Perception and cognition	22
2.2	Strategic Planning Model	29
2.3	Telco ABC's Strategic choice	43
2.4	Application of SPM in an ERP project	57
2.5	A retention Risk Management framework for Telco ABC	61
2.6	CRISP-DM	76
3.1	Elements of concept drift detection	97
3.2	Concept drift as an KM tool	102
4.1	Required Knowledge Management activities	130
5.1	Strategic Analytics Methodology	145
5.2	Knowledge management in SAM	150
5.3	SAM's reframing of CRISP-DM	157
5.4	Data preparation in SAM	175
6.1	The three data domains within Telco ABC	225
6.2	Identify, assemble, prepare data	234
6.3	Binomial distribution of the label feature	246
6.4	Distribution of LoyP (standardised)	247
6.5	Distribution of LoyP after binning	247
6.6	Relative feature significance by χ^2	249
6.7	Lift on Validation _t	258
6.8	Cumulative lift on Validation _t	260
6.9	Standardised means plot segment 1	272
6.10	Standardised means plot segment 2	273
6.11	Relationship between three measures	276
6.12	Handset Type's relative impact within segment	280
7.1	Segment 2 - Bulls	288
7.2	Segment 5 - Cash cows	288
7.3	Segment 6 - Stingy stodgies	289
7.4	Segment 1 - Disloyal dogs	290
7.5	Segment 4 - Loyal friends	291
7.6	Segment 3 - Wealthy assertive friends	291
7.7	New classifier confusion matrix	310
7.8	New classifier lift	310
7.9	New classifier ROC	311

Figure number	Figure name	on page
7.10	Five-monthly segment profile of most-at-risk segment	317
7.11	Change in number of nearest clusters	323
7.12	Change in relative feature importance over time	323
7.13	Change in captured R2 over time	324
7.14	Drift in overall segment profiles	324
7.15	Change in overall inter-segment distinctiveness	325
7.16	Change in percentage of consumers within a segment	326
7.17	Classifier predictive accuracy over time	326

List of tables

Table number	Table name	on page
1.1	Differences between business and structured data mining environments	4
5.1	Components of the project mission	164
5.2	Model measures	188
6.1	Replacement and imputation	240
6.2	Replacement values	241
6.3	Outlier filtering	241
6.4	Clustering base statistics	244
6.5	Selected variable importance	250
6.6	Type III analysis of model effects	259
6.7	Segment base features vital statistics	268
6.8	Technical segmentation measures ($RetSeg_{month\ t+1}$)	268
6.9	Standardised retention segment profile values ($RetSeg_{month\ t+1}$)	271
6.10	Overall retention segment measures ($RetSeg_{month\ t+1}$)	274
6.11	Root cause profile segment 1	278
7.1	Effect A overall impact profile in $SegRet_{month\ t+1}$	292
7.2	Overall Retention Segment Measures ($RetSeg_{Final\ month\ t}$)	312
7.3	Consumer segment membership	315
7.4	Quantitative segment profiles	316
7.5	Monitor and control problem understanding	319
7.6	Monitor and control business solution relevance	321
7.7	Monitor and control project ROI	322

Abstract

Commercial organisations are dependent on generating profit from competitive advantage. Central to this approach, is the Strategic Planning Cycle (SPC). SPC converts new information and new subject matter expertise into competitive knowledge, and then converts that knowledge into executable solutions best suited to the organisation's internal and external circumstances and resources. SPC also maintains the relevance and efficiency of the executed solutions over time.

In order to optimise competitiveness, organisations seek to improve SPC in a number of ways. First, they improve the quality of the informational inputs to SPC. Second, they improve the quality of the knowledge which they develop from that information. Third, they optimise the executability of the solutions, which were based on the knowledge, for the organisation's particular circumstances and resources. Four, they improve the solutions over time, maintaining competitiveness. All four ways of improving SPC are supported by data analytics. It is therefore a necessity ever to improve the integration of data analytics with SPC.

Data mining is an advanced analytics approach, which has been shown to support SPC. Recognising the complexity of integrating data analytics with the business at the turn of the 21st century, the analytics community developed data mining project methodologies to facilitate the integration. The most widely published methodology is CRISP-DM. SAS Institute's SAS Data Mining Projects Methodology (SDMPM) is a second, albeit proprietary, methodology which is also widely used.

Despite the availability of packaged data mining software and project methodologies for more than a decade now, organisations are still finding the integration of data mining with the SPC process complex and daunting. The current situation is that business leaders and data analysts often express the need for better integration of data analytics with SPC and business goals.

The researcher hypothesized that the data mining project methodologies may be a major contributor to the above situation. The researcher therefore formulated the research objective of evaluating data mining methodology for its support of the SPC process. The CRISP-DM methodology was chosen for evaluation because it is in the public domain

and therefore available to other researchers. (The researcher has evaluated SDMPM in a separate paper.)

The research method chosen was Participatory Action Research, specifically that of action science or *expert reflection-in-action*. The research was industry-based, using data from a real-life Telco customer retention management problem. The researcher and the Telco formulated a data analytics project using CRISP-DM. The project was in support of the Telco's strategic initiative drastically to reduce customer churn in their consumer business.

The data mining project would support the initiative in three ways. First, it would predict customer churn behaviour within an upcoming time window. Second, it would segment the most at-risk customers in strategic marketing dimensions. Third, it would profile the segments in dimensions required for retention campaign re-design.

Using expert reflection-in-action, we evaluated the operating and strategic outcome for the Telco, from the project that was formulated using CRISP-DM as the project methodology. The research findings were that the project based on CRISP-DM would be limited in its executibility and strategic impact. This would severely restrict the competitive advantage realisable from the project.

Our research identified six key limitations of CRISP-DM in the SPC environment:

- diagnostic technique for defining the project's business goals or business deliverables. This is about defining the required informational and marketing components required for the strategic initiative;
- introduction of new business and analytics subject matter expertise into the project environment. This relates to increasing the understanding of the business problem and its possible solutions through new marketing and data mining subject matter expertise;
- mapping technique between the project's business deliverables and the supporting data mining plan. This is about assuring that the data analytics best support the project's business deliverables;
- knowledge management activities required by SPC for assessing discovered information against business deliverables, environmental and circumstantial

factors, for adapting the information, and for developing competitive, executable business solutions;

- monitor and control of business and data mining solutions over time for effectiveness and efficiency; and
- a number of soft project and business solution implementation issues.

The main *research goal*, which flowed from the above finding, was to develop a new, more potent data mining project methodology for the SPC environment. In developing this methodology, the researcher used concepts from the Business, Knowledge Discovery, and Data Mining literature, also drawing on his previous corporate management experience and MBA qualification. The researcher called the new method *Strategic Analytics Method (SAM)*.

Essentially SAM is the integration of data analytics project methodology and a proven SPC tool, which is known as Strategic Planning Method (SPM). SPM is a generic decision-making process designed for producing competitive outcomes under conditions of uncertainty and limited resources. SPM is widely used in various guises by business, software engineering, the military, and many other applications.

SAM presents a major departure from CRISP-DM's data centrality, to a project centered on the project's business deliverables. SAM is targeted at data miners and data analysts working in a commercial environment, and at business intelligence practitioners.

Practically SAM contributes the following to data mining projects methodology:

- moving the focus from data-related activities to business deliverables;
- insights about the restrictive impact of the pre-project *status quo* on the results of the project, the dimensions of the status quo which must be defined into a business problem, and how to achieve that definition;
- technique for injecting new business and analytics subject matter into the stale business environment, to enable competitive breakthrough;
- technique for developing business deliverables or goals for the project, which will be competitive. This includes considering the new subject matter, and overcoming the restrictions presented by the current understanding of the *status quo*;
- mapping technique between the project's business deliverables and the data mining plan, which assures the data mining outputs optimally supporting the attainment of the business deliverables;

- technique for assessing discovered information for its relevance to the business deliverables;
- knowledge management activities for developing the discovered information into competitive business solutions which are executable under the organisation's limited resources and limiting circumstances;
- substantial qualitative and quantitative technique for developing monitor and control plans for both the analytics and the business solution;
- activities, which pro-actively manage soft issues before they impact on the project negatively. For instance, we reframe data preparation activities as a process, which gradually reduces project risk associated with the data. This offers more understandable and acceptable justification to the business audience about this resource-intensive part of data mining projects;
- insights for distinguishing between *iteration* and *repetition* of activities on advanced SPC projects, and technique for knowing when to start and stop iterating, or repeating. This distinction provides contextual vocabulary for communicating with the business about required project effort.

The research validates SAM on the same Telco ABC problem, which was used for evaluating CRISP-DM. The validation came through being able to formulate a project using SAM in which we:

- assisted Telco ABC in breaking through their limited pre-project marketing perceptions and expectations, to formulate business deliverables based on new marketing and analytics subject matter, which constituted competitiveness in customer retention management;
- formulated and executed a data mining project which produced the information required by the business deliverables;
- improved the Telco's calculation of the extent of the problem;
- developed knowledge from the discovered information which complemented applicable new marketing subject matter;
- developed the knowledge into a competitive retention management solution executable under the Telco's limiting circumstances and limited campaign resources. We presented the solution as new marketing objectives and strategies, and developed these into a retention campaign strategy with various key components;
- developed a comprehensive monitor and control plan for the campaigns and the operationalised data analytics solution;
- quantified the project ROI as about 187 times the investment.

May 2006

m_vanrooyen@hotmail.com / +61-402-032-059