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
2.8.1 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 (SPM 1.2) .................................................. 50
  2.8.2.3 Uncontrollable factors profile (SPM 1.3) .............................................. 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 (SPM 4.1) .............................. 53
  2.8.2.8 Company A’s new Grand strategy (SPM 4.2) ....................................... 54
  2.8.2.9 Company A’s new Operating objectives (SPM 4.3) ............................. 54
  2.8.2.10 Company A’s new Operating strategies (SPM 4.4) ......................... 55
  2.8.2.11 Implement and deploy new strategies (SPM 5)(KM 3) ....................... 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
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 Business problem is .............................................................. 155
    5.8.1.2 Business problem’s strategic purpose ............................................. 156
    5.8.1.3 How Business problem supports ................................................... 156
  5.8.2 Potential solutions ................................................................................. 157
    5.8.2.1 What Potential solutions is ............................................................. 157
    5.8.2.2 Potential solutions’s strategic purpose .......................................... 158
5.8.2.3 How Potential solutions supports ......................................................... 159
5.8.3 Develop project mission ........................................................................... 160
  5.8.3.1 What Develop project mission is ......................................................... 160
  5.8.3.2 Develop project mission’s strategic purpose .................................... 162
  5.8.3.3 How Develop project mission supports .............................................. 163
5.8.4 Expert collaboration ................................................................................. 165
  5.8.4.1 What Expert collaboration is .............................................................. 165
  5.8.4.2 Expert collaboration’s strategic purpose ....................................... 167
  5.8.4.3 How Expert collaboration supports .................................................. 167
5.8.5 Identify, assemble, prepare useful data .................................................... 168
  5.8.5.1 What Identify, assemble, prepare useful data is ................................ 168
  5.8.5.2 Identify, assemble, prepare useful data’s support ......................... 169
  5.8.5.3 How Identify, assemble, prepare useful data supports ................... 171
5.8.6 Data mining discovery ............................................................................ 180
  5.8.6.1 What Data mining discovery is .......................................................... 180
  5.8.6.2 Data mining discovery’s support ....................................................... 181
  5.8.6.3 How Data mining discovery supports .............................................. 181
5.8.7 Develop circumstantial knowledge ........................................................... 188
  5.8.7.1 What Develop circumstantial knowledge is ..................................... 188
  5.8.7.2 Develop circumstantial knowledge’s support ................................... 189
  5.8.7.3 How Develop circumstantial knowledge supports ......................... 189
5.8.8 Strategic analysis ...................................................................................... 189
  5.8.8.1 What Strategic analysis is ................................................................. 189
  5.8.8.2 Strategic analysis’ strategic support ............................................... 190
  5.8.8.3 How Strategic analysis supports ....................................................... 190
5.8.9 Strategic choice ......................................................................................... 190
5.8.9.1 What Strategic choice is ................................................................. 190
5.8.9.2 Strategic choice’s strategic support ............................................. 190
5.8.9.3 How Strategic choice supports ..................................................... 191
5.8.10 Define new business objectives and strategies ................................ 192
  5.8.10.1 What Define new business objectives and strategies is ............. 192
  5.8.10.2 Define new business objectives and strategies’ support .......... 192
  5.8.10.3 How Define new business objectives and strategies supports .... 192
5.8.11 Develop data mining plan ............................................................ 192
  5.8.11.1 What Develop data mining plan is ......................................... 192
  5.8.11.2 Develop data mining plan’s support ......................................... 193
  5.8.11.3 How Develop data mining plan supports ................................... 194
5.8.12 Model, evaluate, choose best model(s) ......................................... 194
  5.8.12.1 What Model, evaluate, choose best model(s) is ...................... 195
  5.8.12.2 Model, evaluate, choose best model(s)’s support .................... 195
  5.8.12.3 How Model, evaluate, choose best model(s) supports ................ 195
5.8.13 Operationalise model(s) .............................................................. 196
  5.8.13.1 What Operationalise model(s) is ........................................... 196
  5.8.13.2 Operationalise model(s)’ support ........................................... 196
  5.8.13.3 How Operationalise model(s) supports .................................. 196
5.8.14 Deploy outputs into business ....................................................... 196
  5.8.14.1 What Deploy outputs into business is .................................... 196
  5.8.14.2 Deploy outputs into business’s support .................................. 196
  5.8.14.3 How Deploy outputs into business supports ........................... 196
5.8.15 Execute new business strategies .................................................. 197
  5.8.15.1 What Execute new business strategies is ............................... 197
  5.8.15.2 Execute new business strategies’ support ............................... 197
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 \( \text{CDR}_{\text{modplus}} \) ..................................................... 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 Strategic objective .....................................................298
      7.2.1.2 New Grand strategy ..........................................................299
7.2.1.3 New Operating objective ................................................................. 299
7.2.1.4 New Operating strategies ................................................................. 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

<table>
<thead>
<tr>
<th>Figure number</th>
<th>Figure name</th>
<th>on page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Elements of research problem</td>
<td>6</td>
</tr>
<tr>
<td>1.2</td>
<td>Research goals operationalising</td>
<td>14</td>
</tr>
<tr>
<td>1.3</td>
<td>Thesis structure</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Perception and cognition</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Strategic Planning Model</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Telco ABC's Strategic choice</td>
<td>43</td>
</tr>
<tr>
<td>2.4</td>
<td>Application of SPM in an ERP project</td>
<td>57</td>
</tr>
<tr>
<td>2.5</td>
<td>A retention Risk Management framework for Telco ABC</td>
<td>61</td>
</tr>
<tr>
<td>2.6</td>
<td>CRISP-DM</td>
<td>76</td>
</tr>
<tr>
<td>3.1</td>
<td>Elements of concept drift detection</td>
<td>97</td>
</tr>
<tr>
<td>3.2</td>
<td>Concept drift as an KM tool</td>
<td>102</td>
</tr>
<tr>
<td>4.1</td>
<td>Required Knowledge Management activities</td>
<td>130</td>
</tr>
<tr>
<td>5.1</td>
<td>Strategic Analytics Methodology</td>
<td>145</td>
</tr>
<tr>
<td>5.2</td>
<td>Knowledge management in SAM</td>
<td>150</td>
</tr>
<tr>
<td>5.3</td>
<td>SAM's reframing of CRISP-DM</td>
<td>157</td>
</tr>
<tr>
<td>5.4</td>
<td>Data preparation in SAM</td>
<td>175</td>
</tr>
<tr>
<td>6.1</td>
<td>The three data domains within Telco ABC</td>
<td>225</td>
</tr>
<tr>
<td>6.2</td>
<td>Identify, assemble, prepare data</td>
<td>234</td>
</tr>
<tr>
<td>6.3</td>
<td>Binomial distribution of the label feature</td>
<td>246</td>
</tr>
<tr>
<td>6.4</td>
<td>Distribution of LoyP (standardised)</td>
<td>247</td>
</tr>
<tr>
<td>6.5</td>
<td>Distribution of LoyP after binning</td>
<td>247</td>
</tr>
<tr>
<td>6.6</td>
<td>Relative feature significance by $\chi^2$</td>
<td>249</td>
</tr>
<tr>
<td>6.7</td>
<td>Lift on Validation$_t$</td>
<td>258</td>
</tr>
<tr>
<td>6.8</td>
<td>Cumulative lift on Validation$_t$</td>
<td>260</td>
</tr>
<tr>
<td>6.9</td>
<td>Standardised means plot segment 1</td>
<td>272</td>
</tr>
<tr>
<td>6.10</td>
<td>Standardised means plot segment 2</td>
<td>273</td>
</tr>
<tr>
<td>6.11</td>
<td>Relationship between three measures</td>
<td>276</td>
</tr>
<tr>
<td>6.12</td>
<td>Handset Type's relative impact within segment</td>
<td>280</td>
</tr>
<tr>
<td>7.1</td>
<td>Segment 2 - Bulls</td>
<td>288</td>
</tr>
<tr>
<td>7.2</td>
<td>Segment 5 - Cash cows</td>
<td>288</td>
</tr>
<tr>
<td>7.3</td>
<td>Segment 6 - Stingy stodgies</td>
<td>289</td>
</tr>
<tr>
<td>7.4</td>
<td>Segment 1 - Disloyal dogs</td>
<td>290</td>
</tr>
<tr>
<td>7.5</td>
<td>Segment 4 - Loyal friends</td>
<td>291</td>
</tr>
<tr>
<td>7.6</td>
<td>Segment 3 - Wealthy assertive friends</td>
<td>291</td>
</tr>
<tr>
<td>7.7</td>
<td>New classifier confusion matrix</td>
<td>310</td>
</tr>
<tr>
<td>7.8</td>
<td>New classifier lift</td>
<td>310</td>
</tr>
<tr>
<td>7.9</td>
<td>New classifier ROC</td>
<td>311</td>
</tr>
<tr>
<td>Figure number</td>
<td>Figure name</td>
<td>on page</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>7.10</td>
<td>Five-monthly segment profile of most-at-risk segment</td>
<td>317</td>
</tr>
<tr>
<td>7.11</td>
<td>Change in number of nearest clusters</td>
<td>323</td>
</tr>
<tr>
<td>7.12</td>
<td>Change in relative feature importance over time</td>
<td>323</td>
</tr>
<tr>
<td>7.13</td>
<td>Change in captured R2 over time</td>
<td>324</td>
</tr>
<tr>
<td>7.14</td>
<td>Drift in overall segment profiles</td>
<td>324</td>
</tr>
<tr>
<td>7.15</td>
<td>Change in overall inter-segment distinctiveness</td>
<td>325</td>
</tr>
<tr>
<td>7.16</td>
<td>Change in percentage of consumers within a segment</td>
<td>326</td>
</tr>
<tr>
<td>7.17</td>
<td>Classifier predictive accuracy over time</td>
<td>326</td>
</tr>
</tbody>
</table>
# List of tables

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

 o knowledge management activities for developing the discovered information into competitive business solutions which are executable under the organisation’s limited resources and limiting circumstances;

 o substantial qualitative and quantitative technique for developing monitor and control plans for both the analytics and the business solution;

 o 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;

 o 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:

 o 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;

 o formulated and executed a data mining project which produced the information required by the business deliverables;

 o improved the Telco’s calculation of the extent of the problem;

 o developed knowledge from the discovered information which complemented applicable new marketing subject matter;

 o 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;

 o developed a comprehensive monitor and control plan for the campaigns and the operationalised data analytics solution;

 o quantified the project ROI as about 187 times the investment.

May 2006 m_vanrooven@hotmail.com / +61-402-032-059