

# **Data Mining Applications for Small and Medium Enterprises**

KOH Hian Chye

Nanyang Business School

Nanyang Technological University

## About the Author

Dr Koh Hian Chye is an Associate Professor and Vice-dean of the Nanyang Business School, Nanyang Technological University (Singapore). He is also a Certified Public Accountant. Dr Koh has published widely in international and regional journals and conferences on topics related to accounting and auditing, business and management, statistical modelling and applications, and data mining. His current interest is in the business applications of data mining. Dr Koh frequently acts as a consultant for private companies, large organisations and government agencies. His recent consultancy projects involve data mining applications. He also serves as a trainer in workshops and executive programmes.

## Foreword

All businesses depend on data to plan their strategies. The data could be on the external market or their own internal production. Large companies do spend a great deal of their resources in analysing the available data – that is what data mining is all about.

Small and Medium Enterprises (SMEs) do not have the fortune to own such resources. Yet they need to benefit from data mining applications in order to remain profitable and to grow.

This book serves to show the SMEs on how to approach the issue intuitively without the same level of resources needed. The examples, illustrations and case studies are a practical and application-oriented guide to inform entrepreneurs on both the usefulness and the limitations of data mining.

I congratulate the author on this meaningful contribution to research on small enterprise development.

Professor Cham Tao Soon  
Director  
Centre for Research on Small Enterprise Development  
Nanyang Technological University

## Preface

Three observations have motivated me to write this book. Firstly, from my interactions with decision makers in commercial and non-commercial organisations, I can see that there is a consistent and persistent desire to know more about data mining. However, what these decision makers need is not so much a technical treatise of data mining but an intuitive introduction to what data mining is, the steps involved, the tools that are used, examples of applications, and some ideas about the limitations and future directions of data mining. That is, what is needed is a practical and application-oriented book that does not require much technical or quantitative knowledge upon which to grasp the fundamentals of data mining. There are, indeed, many books on data mining. There are, however, very few books that meet the needs of this group of decision makers for a good, concise and intuitive introduction to data mining.

Secondly, I have observed a gap in the data mining scene. Big organisations (e.g., banks) have the resources, expertise and ability to develop their own data mining applications, including very sophisticated ones. Many of these big organisations are, in fact, already very advanced in harnessing the benefits of data mining in their business. At the other extreme, there are very small organisations that probably do not need sophisticated data mining applications. Simple data analyses (and perhaps online analytical processing or OLAP) are often sufficient to address their business problems and issues. Small and medium enterprises (or SMEs), however, are squeezed in between. They can certainly benefit from data mining applications but they may not have the same level of resources that

## *Preface*

big organisations have to develop data mining applications. Hence, there is a need for a book that focuses on data mining applications for SMEs.

Finally, my students have been a constant source of encouragement for me to write this book. They find data mining very fascinating and useful and have always wanted to know more about it. Being in the Nanyang Business School, these students approach data mining not from a technical or quantitative perspective but from an application perspective. I have introduced data mining as part of a course that I teach and I have substantial difficulty in sourcing for an appropriate book for business students that has a good balance between technical and practical aspects of data mining. I have then decided that one solution to this problem is to write such a book myself.

Hence, this book is born. I have deliberately made the book as intuitive and practical as I can. This means that I have avoided unnecessary statistics and mathematics and have instead attempted to explain the concepts behind them and show intuitively how they work in data mining. Still, some statistics and mathematics are not avoidable to ensure that the essence of data mining is captured in the book. I have also included many examples, illustrations and case studies. Further, I have included screenshots of the data mining results derived from a very user-friendly data mining software (namely, SPSS Clementine) to familiarise readers with how data mining results look like. Finally, I have focused on illustrations and case studies that are most relevant to SMEs and business students.

My primary objective of writing this book would have been achieved if readers can gain a good understanding of what data mining is after reading the book. It is my wish too that readers will go beyond just understanding what data mining is to developing data mining applications in their organisations. I have no doubt that data mining can be a very powerful technology and methodology for generating information from raw data to address business and other problems. This usefulness, however, will not be realised unless the knowledge of data mining is put to good use.

## **Acknowledgements**

Without the support and encouragement of several individuals and organisations, this book would not have been possible. Hence, I wish to express my sincere thanks to the following individuals and organisations (not in any particular order) who have contributed greatly to this book.

I would like to thank SPSS Singapore for all the software and technical support that it has provided to me during the writing of the book. In addition, I would like to thank Gerald Tan, Vincent Toh and Sean Tan for freely sharing their data mining and other expertise with me and for working with me in data mining projects.

Goh Chwee Peng and Tan Wei Chin have also contributed tremendously to the book, especially to the three chapters on the potential data mining applications for a retailer, service provider and manufacturer. They have also added a lot of knowledge to the book. Hence, my special thanks to them. I also appreciate the great experience working with Chwee Peng and Wei Chin in data mining projects.

I would like to express my sincere gratitude to Distinguished Professor Cham Tao Soon (Director, Centre for Research on Small Enterprise Development, Nanyang Technological University) for the financial and other support that he has provided to me to make this book a reality. I have greatly benefited from his guidance, wisdom and experience.

Further, I would like to thank the Nanyang Technological University and the Nanyang Business School for providing a conducive environment for me to write the book.

### *Acknowledgements*

I am indebted to all my students who have contributed to making the book better in one way or another. They have clarified a lot of my thinking and understanding of data mining and have constantly challenged me to greater heights. Their enthusiasm has been a great and constant source of encouragement.

Several of my colleagues at the Nanyang Business School have given me tremendous support, encouragement and valuable help in so many ways. In particular, I wish to thank Professors Ang Kong Beng, Bobby Srinivasan, Chan Yoke Kai, Clement Tan, Fock Siew Tong, Goh Chye Tee, Lee Pui Mun, Low Buen Sin, Mike Leu, Ng Beoy Kui, Ng Eng Juan and Wan Chew Yoong. In addition, Joyce Lee, Lok Siew Leng and Carol Loke have provided me with a lot of clerical, administrative and other support.

My special thanks to Professor Roger Chiang who patiently gave me my first "lecture" on data mining and Professor Low Chan Kee who worked with me on several data mining projects when I first got involved in the data mining area. I would also like to thank Quek Seow Hiah, Karen Goh, Koh Tat Wee and Chua Meow Heam for working with me on some data mining projects. These projects have been great learning experiences.

Most important of all, I would like to thank my wife (Beng Guat) and children (De Wei, Jing Yi and Wan Yi) for their continual support, care, love and encouragement. Writing a book is no small task; it requires a lot of effort, concentration and patience. My family makes it all worthwhile. A special thanks to my lovely wife and children and I dedicate this book to them.

Preface		iii
Acknowledgements		v
<b>Part One</b>	<b>Data Mining Fundamentals</b>	
Chapter 1	<b>Overview of Data Mining</b>	1
	1.1 Background	1
	1.2 Definition	3
	1.3 Methodology	4
	1.3.1 Pre-modelling Stage	5
	1.3.2 Modelling Stage	8
	1.3.3 Post-modelling Stage	10
	1.4 Organisation of the Book	11
Chapter 2	<b>Data Mining Tools</b>	15
	2.1 Introduction	15
	2.2 Description and Visualisation	15
	2.2.1 Illustration of Description and Visualisation	17
	2.3 Association Analysis	24
	2.3.1 Illustration of Association Analysis	28
	2.4 Clustering	33
	2.4.1 Illustration of Clustering	37
	2.5 Summary	40



*Contents*

Chapter 3	<b>Predictive Modelling</b>	47
3.1	Introduction	47
3.2	Regression Models	50
3.2.1	Multiple Regression	50
3.2.2	Logistic Regression	56
3.2.3	Illustration of Regression Model	63
3.3	Neural Networks	67
3.3.1	Kohonen Nets	73
3.3.2	Illustration of Neural Network Model	75
3.4	Decision Trees	79
3.4.1	Chi-square Test of Independence	80
3.4.2	Chi-square Automatic Interaction Detection (CHAID)	84
3.4.3	Other Decision Tree Algorithms	93
3.4.4	Illustration of Decision Tree Model	95
3.5	Summary	99
Chapter 4	<b>Data Mining Issues</b>	103
4.1	Introduction	103
4.2	Model Validation	103
4.3	Optimal Cut-off Points	108
4.4	Evaluation Charts	116
4.5	Financial Assessment of Prediction Models	124
4.6	Illustration of Predictive Modelling – A Revisit	129
4.7	Summary	136
<b>Part Two</b>	<b><u>Data Mining Applications</u></b>	
Chapter 5	<b>Potential Applications for a Retailer</b>	139
5.1	Introduction	139

	5.2	Context	140
	5.3	Background Information for Data Mining Illustrations	141
	5.4	Application 1: Baby Products Promotion	143
	5.5	Application 2: Customer Segmentation	149
	5.6	Application 3: Churn Modelling	155
	5.7	Concluding Remarks	168
Chapter 6		<b>Potential Applications for a Service Provider</b>	171
	6.1	Context	171
	6.2	Background Information for Data Mining Illustrations	172
	6.3	Application 1: Identification of Dissatisfied Customers	175
	6.4	Application 2: Developing Tour Packages	179
	6.5	Application 3: Profiling of Customers	181
	6.6	Application 4: Target Mailing Campaign	188
	6.7	Concluding Remarks	197
Chapter 7		<b>Potential Applications for a Manufacturer</b>	199
	7.1	Introduction	199
	7.2	Background Information for Data Mining Illustrations	201
	7.3	Application 1: Matching of Job and Worker	201
	7.4	Application 2: Determination of Optimal Conditions for Storage	206
	7.5	Application 3: Fault Detection and Development of Operating Guidelines	214
	7.6	Concluding Remarks	223

*Contents*

**Part Three**      **Summary and Conclusion**

Chapter 8	<b>Conclusion</b>	227
	8.1      Summary	227
	8.1.1    Overview of Data Mining	227
	8.1.2    Data Mining Tools and Issues	229
	8.1.3    Lessons from the Cast Studies	233
	8.2      Limitations of Data Mining	236
	8.3      Future Directions	238
	8.3.1    Applications	238
	8.3.2    Data	239
	8.3.3    Software	240
	8.3.4    Hardware	241
	8.4      Some Final Comments	242
Appendix	<b>Introduction to SPSS Clementine</b>	245
Index		257