



From *Applied Data Mining for Forecasting Using SAS*<sup>®</sup>. Full book available for purchase [here](#).

## Contents

|                                                                                   |            |
|-----------------------------------------------------------------------------------|------------|
| <b>Preface .....</b>                                                              | <b>vii</b> |
| <b>Chapter 1 Why Industry Needs Data Mining for Forecasting.....</b>              | <b>1</b>   |
| 1.1 Overview .....                                                                | 1          |
| 1.2 Forecasting Capabilities as a Competitive Advantage .....                     | 2          |
| 1.3 The Explosion of Available Time Series Data.....                              | 2          |
| 1.4 Some Background on Forecasting .....                                          | 3          |
| 1.5 The Limitations of Classical Univariate Forecasting.....                      | 4          |
| 1.6 What is a Time Series Database?.....                                          | 4          |
| 1.7 What is Data Mining for Forecasting? .....                                    | 5          |
| 1.8 Advantages of Integrating Data Mining and Forecasting .....                   | 6          |
| 1.9 Remaining Chapters .....                                                      | 6          |
| <b>Chapter 2 Data Mining for Forecasting Work Process .....</b>                   | <b>7</b>   |
| 2.1 Introduction .....                                                            | 7          |
| 2.2 Work Process Description .....                                                | 7          |
| 2.2.1 Generic Flowchart.....                                                      | 8          |
| 2.2.2 Key Steps .....                                                             | 9          |
| 2.3 Work Process with SAS Tools .....                                             | 16         |
| 2.3.1 Data Preparation Steps with SAS Tools .....                                 | 17         |
| 2.3.2 Variable Reduction and Selection Steps with SAS Tools .....                 | 19         |
| 2.3.3 Forecasting Steps with SAS Tools .....                                      | 20         |
| 2.3.4 Model Deployment Steps with SAS Tools .....                                 | 21         |
| 2.3.5 Model Maintenance Steps with SAS Tools.....                                 | 22         |
| 2.3.6 Guidance for SAS Tool Selection Related to Data Mining in Forecasting ..... | 22         |
| 2.4 Work Process Integration in Six Sigma .....                                   | 23         |
| 2.4.1 Six Sigma in Industry .....                                                 | 23         |
| 2.4.2 The DMAIC Process.....                                                      | 24         |
| 2.4.3 Integration with the DMAIC Process.....                                     | 25         |
| Appendix: Project Charter.....                                                    | 26         |
| <b>Chapter 3 Data Mining for Forecasting Infrastructure.....</b>                  | <b>29</b>  |
| 3.1 Introduction .....                                                            | 29         |
| 3.2 Hardware Infrastructure.....                                                  | 29         |
| 3.2.1 Personal Computers Network Infrastructure.....                              | 30         |
| 3.2.2 Client/Server Infrastructure .....                                          | 30         |
| 3.2.3 Cloud Computing Infrastructure.....                                         | 31         |
| 3.3 Software Infrastructure .....                                                 | 32         |
| 3.3.1 Data Collection Software .....                                              | 32         |
| 3.3.2 Data Preparation Software.....                                              | 32         |
| 3.3.3 Data Mining Software .....                                                  | 32         |
| 3.3.4 Forecasting Software .....                                                  | 33         |

|                                                                                     |           |
|-------------------------------------------------------------------------------------|-----------|
| 3.3.5 Software Selection Criteria .....                                             | 33        |
| 3.4 Data Infrastructure.....                                                        | 34        |
| 3.4.1 Internal Data Infrastructure.....                                             | 34        |
| 3.4.2 External Data Infrastructure.....                                             | 34        |
| 3.5 Organizational Infrastructure.....                                              | 35        |
| 3.5.1 Developers Infrastructure .....                                               | 35        |
| 3.5.2 Users Infrastructure.....                                                     | 36        |
| 3.5.3 Work Process Implementation .....                                             | 37        |
| 3.5.4 Integration with IT .....                                                     | 37        |
| <b>Chapter 4 Issues with Data Mining for Forecasting Application .....</b>          | <b>39</b> |
| 4.1 Introduction .....                                                              | 39        |
| 4.2 Technical Issues.....                                                           | 39        |
| 4.2.1 Data Quality Issues .....                                                     | 39        |
| 4.2.2 Data Mining Methods Limitations .....                                         | 43        |
| 4.2.3 Forecasting Methods Limitations.....                                          | 45        |
| 4.3 Nontechnical Issues .....                                                       | 46        |
| 4.3.1 Managing Forecasting Expectations .....                                       | 46        |
| 4.3.2 Handling Politics of Forecasting .....                                        | 47        |
| 4.3.3 Avoiding Bad Practices .....                                                  | 49        |
| 4.3.4 Forecasting Aphorisms .....                                                   | 49        |
| 4.4 Checklist “Are We Ready?” .....                                                 | 50        |
| <b>Chapter 5 Data Collection .....</b>                                              | <b>53</b> |
| 5.1 Introduction .....                                                              | 53        |
| 5.2 System Structure and Data Identification .....                                  | 53        |
| 5.2.1 Mind-Mapping .....                                                            | 54        |
| 5.2.2 System Structure Knowledge Acquisition.....                                   | 54        |
| 5.2.3 Data Structure Identification.....                                            | 55        |
| 5.3 Data Definition.....                                                            | 57        |
| 5.3.1 Data Sources .....                                                            | 57        |
| 5.3.2 Metadata .....                                                                | 59        |
| 5.4 Data Extraction.....                                                            | 61        |
| 5.4.1 Internal Data Extraction.....                                                 | 61        |
| 5.4.2 External Data Extraction .....                                                | 61        |
| 5.5 Data Alignment.....                                                             | 62        |
| 5.5.1 Data Alignment to a Business Structure .....                                  | 62        |
| 5.5.2 Data Alignment to Time.....                                                   | 63        |
| 5.6 Data Collection Automation for Model Deployment .....                           | 64        |
| 5.6.1 Differences between Data Collection for Model Development and Deployment..... | 64        |
| 5.6.2 Data Collection Automation for Model Deployment .....                         | 64        |
| <b>Chapter 6 Data Preparation .....</b>                                             | <b>67</b> |
| 6.1 Overview .....                                                                  | 67        |
| 6.2 Transactional Data Versus Time Series Data .....                                | 67        |
| 6.3 Matching Frequencies.....                                                       | 75        |
| 6.3.1 Contracting .....                                                             | 75        |
| 6.3.2 Expanding .....                                                               | 78        |
| 6.4 Merging .....                                                                   | 83        |
| 6.5 Imputation.....                                                                 | 84        |

|                                                                                                        |            |
|--------------------------------------------------------------------------------------------------------|------------|
| 6.6 Outliers .....                                                                                     | 87         |
| 6.7 Transformations .....                                                                              | 92         |
| 6.8 Summary .....                                                                                      | 94         |
| <b>Chapter 7 A Practitioner's Guide of DMM Methods for Forecasting.....</b>                            | <b>95</b>  |
| 7.1 Overview .....                                                                                     | 95         |
| 7.2 Methods for Variable Reduction.....                                                                | 96         |
| Traditional Data Mining .....                                                                          | 96         |
| Time Series Approach .....                                                                             | 97         |
| 7.3 Methods for Variable Selection .....                                                               | 103        |
| Traditional Data Mining .....                                                                          | 104        |
| Example for Variable Selection.....                                                                    | 111        |
| Variable Selection Based on Pearson Product-Moment Correlation Coefficient .....                       | 112        |
| Variable Selection Based on Stepwise Regression.....                                                   | 114        |
| Variable Selection Based on the SAS Enterprise Miner Variable Selection Node.....                      | 116        |
| Variable Selection Based on the SAS Enterprise Miner Partial Least Squares Node.....                   | 118        |
| Variable Selection Based on Decision Trees .....                                                       | 120        |
| Variable Selection Based on Genetic Programming .....                                                  | 122        |
| Comparison of Data Mining Variable Selection Results .....                                             | 124        |
| 7.4 Time Series Approach .....                                                                         | 125        |
| 7.5 Summary .....                                                                                      | 134        |
| <b>Chapter 8 Model Building: ARMA Models.....</b>                                                      | <b>135</b> |
| Introduction .....                                                                                     | 135        |
| 8.1 ARMA Models.....                                                                                   | 136        |
| 8.1.1 AR Models: Concepts and Application .....                                                        | 136        |
| 8.1.2 Moving Average Models: Concepts and Application .....                                            | 149        |
| 8.1.3 Auto Regressive Moving Average (ARMA) Models.....                                                | 157        |
| Appendix 1: Useful Technical Details.....                                                              | 164        |
| Appendix 2: The "I" in ARIMA .....                                                                     | 166        |
| <b>Chapter 9 Model Building: ARIMAX or Dynamic Regression Modes.....</b>                               | <b>179</b> |
| Introduction .....                                                                                     | 179        |
| 9.1 ARIMAX Concepts .....                                                                              | 180        |
| 9.2 ARIMAX Applications.....                                                                           | 185        |
| Appendix: Prewhitening and Other Topics Associated with Interval-Valued Input Variables .....          | 195        |
| <b>Chapter 10 Model Building: Further Modeling Topics .....</b>                                        | <b>205</b> |
| Introduction .....                                                                                     | 205        |
| 10.1 Creating Time Series Data and Data Hierarchies Using Accumulation<br>and Aggregation Methods..... | 206        |
| Introduction .....                                                                                     | 206        |
| Creating Time Series Data Using Accumulation Methods .....                                             | 206        |
| Creating Data Hierarchies Using Aggregation Methods.....                                               | 210        |
| 10.2 Statistical Forecast Reconciliation .....                                                         | 216        |
| 10.3 Intermittent Demand.....                                                                          | 220        |
| 10.4 High-Frequency Data and Mixed-Frequency Forecasting .....                                         | 224        |
| High-Frequency Data.....                                                                               | 224        |
| Mixed-Interval Forecasting .....                                                                       | 232        |

|                                                                         |            |
|-------------------------------------------------------------------------|------------|
| 10.5 Holdout Samples and Forecast Model Selection in Time Series .....  | 237        |
| Introduction .....                                                      | 237        |
| 10.6 Planning Versus Forecasting and Manual Overrides .....             | 240        |
| 10.7 Scenario-Based Forecasting .....                                   | 241        |
| 10.8 New Product Forecasting.....                                       | 244        |
| <b>Chapter 11 Model Building: Alternative Modeling Approaches .....</b> | <b>251</b> |
| 11.1 Nonlinear Forecasting Models.....                                  | 251        |
| 11.1.1 Nonlinear Modeling Features.....                                 | 251        |
| 11.1.2 Forecasting Models Based on Neural Networks .....                | 252        |
| 11.1.3 Forecasting Models Based on Support Vector Machines .....        | 254        |
| 11.1.4 Forecasting Models Based on Evolutionary Computation .....       | 256        |
| 11.2 More Modeling Alternatives.....                                    | 258        |
| 11.2.1 Multivariate Models.....                                         | 258        |
| 11.2.2 Unobserved Component Models (UCM).....                           | 263        |
| <b>Chapter 12 An Example of Data Mining for Forecasting .....</b>       | <b>273</b> |
| 12.1 The Business Problem.....                                          | 273        |
| 12.2 The Charter .....                                                  | 273        |
| 12.3 The Mind Map.....                                                  | 274        |
| 12.4 Data Sources .....                                                 | 274        |
| 12.5 Data Prep .....                                                    | 275        |
| 12.6 Exploratory Analysis and Data Preprocessing .....                  | 279        |
| 12.7 X Variable Imputation .....                                        | 283        |
| 12.8 Variable Reduction and Selection .....                             | 287        |
| 12.9 Modeling .....                                                     | 291        |
| 12.10 Summary .....                                                     | 307        |
| <b>Appendix A.....</b>                                                  | <b>309</b> |
| <b>Appendix B.....</b>                                                  | <b>311</b> |
| <b>References.....</b>                                                  | <b>313</b> |
| <b>Index.....</b>                                                       | <b>317</b> |

From *Applied Data Mining for Forecasting Using SAS®* by Tim Rey, Arthur Kordon, and Chip Wells. Copyright © 2012, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED.