

Contents

1	Knowledge Management	1
1.1	The Big Data Era	1
1.2	Business Intelligence	2
1.3	Knowledge Management	3
1.4	Computer Support Systems	3
1.5	Data Mining Forecasting Applications	5
1.6	Data Mining Tools	6
1.7	Summary	7
	References	8
2	Data Sets	11
2.1	Gold	12
2.2	Other Datasets	14
	2.2.1 Financial Index Data	14
	2.2.2 Loan Analysis Data	17
	2.2.3 Job Application Data	17
	2.2.4 Insurance Fraud Data	18
	2.2.5 Expenditure Data	18
2.3	Summary	19
	References	20
3	Basic Forecasting Tools	21
3.1	Moving Average Models	21
3.2	Regression Models	22
3.3	Time Series Error Metrics	26
3.4	Seasonality	27
3.5	Demonstration Data	31

3.6	Software Demonstrations	35
3.6.1	R Software	36
3.6.2	Weka	39
3.7	Summary	42
4	Multiple Regression	45
4.1	Data Series	45
4.2	Correlation	47
4.3	Lags	50
4.4	Summary	55
5	Regression Tree Models	57
5.1	R Regression Trees	57
5.2	Random Forests	61
5.3	WEKA Regression Trees	67
5.3.1	Decision Stump	68
5.3.2	Random Tree Modeling	69
5.3.3	REP Tree Modeling	70
5.3.4	Random Forest	70
5.3.5	M5P Modeling	76
5.4	Summary	77
	Reference	77
6	Autoregressive Models	79
6.1	ARIMA Models	79
6.2	ARIMA Model of Brent Crude	80
6.3	ARMA	82
6.4	GARCH Models	87
6.4.1	ARCH(q)	87
6.4.2	GARCH(p, q)	87
6.4.3	EGARCH	88
6.4.4	GJR(p, q)	88
6.5	Regime Switching Models	89
6.6	Application on Crude Oil Data	89
6.7	Summary	92
	References	93
7	Classification Tools	95
7.1	Bankruptcy Data Set	95
7.2	Logistic Regression	97
7.3	Support Vector Machines	102
7.4	Neural Networks	106
7.5	Decision Trees	107
7.6	Random Forests	110
7.7	Boosting	112

7.8	Comparison	113
7.9	WEKA Classification Modeling	116
7.10	Summary	120
	Reference	121
8	Predictive Models and Big Data	123
	References	125