

---

## Contents

List of Figures		xiii
List of Tables		xxix
Preface		xxxvii
	Structure of the Book	xxxviii
	How to Use this Book	xxxix
	Online Resources	xl
	Acknowledgements	xl
Notation		xliii
	Notational Conventions	xliii
	Notational Conventions for Probabilities	xliv
<b>1</b>	<b>Machine Learning for Predictive Data Analytics</b>	<b>1</b>
	1.1 What Is Predictive Data Analytics?	1
	1.2 What Is Machine Learning?	3
	1.3 How Does Machine Learning Work?	5
	1.4 What Can Go Wrong with Machine Learning?	11
	1.5 The Predictive Data Analytics Project Lifecycle: CRISP-DM	12
	1.6 Predictive Data Analytics Tools	15
	1.7 The Road Ahead	17
	1.8 Exercises	19
<b>2</b>	<b>Data to Insights to Decisions</b>	<b>21</b>
	2.1 Converting Business Problems into Analytics Solutions	21
	2.1.1 Case Study: Motor Insurance Fraud	23
	2.2 Assessing Feasibility	24
	2.2.1 Case Study: Motor Insurance Fraud	26
	2.3 Designing the Analytics Base Table	27
	2.3.1 Case Study: Motor Insurance Fraud	31
	2.4 Designing & Implementing Features	32
	2.4.1 Different Types of Data	34
	2.4.2 Different Types of Features	34
	2.4.3 Handling Time	37
	2.4.4 Legal Issues	41
	2.4.5 Implementing Features	43
	2.4.6 Case Study: Motor Insurance Fraud	43
	2.5 Summary	46
	2.6 Further Reading	49
	2.7 Exercises	50

<b>3</b>	<b>Data Exploration</b>	55
3.1	The Data Quality Report	56
3.1.1	Case Study: Motor Insurance Fraud	57
3.2	Getting to Know the Data	61
3.2.1	The Normal Distribution	64
3.2.2	Case Study: Motor Insurance Fraud	65
3.3	Identifying Data Quality Issues	66
3.3.1	Missing Values	67
3.3.2	Irregular Cardinality	68
3.3.3	Outliers	69
3.3.4	Case Study: Motor Insurance Fraud	70
3.4	Handling Data Quality Issues	73
3.4.1	Handling Missing Values	73
3.4.2	Handling Outliers	74
3.4.3	Case Study: Motor Insurance Fraud	76
3.5	Advanced Data Exploration	77
3.5.1	Visualizing Relationships Between Features	77
3.5.2	Measuring Covariance & Correlation	86
3.6	Data Preparation	92
3.6.1	Normalization	92
3.6.2	Binning	94
3.6.3	Sampling	98
3.7	Summary	100
3.8	Further Reading	101
3.9	Exercises	103
<b>4</b>	<b>Information-based Learning</b>	111
4.1	Big Idea	111
4.2	Fundamentals	114
4.2.1	Decision Trees	115
4.2.2	Shannon's Entropy Model	118
4.2.3	Information Gain	122
4.3	Standard Approach: The ID3 Algorithm	128
4.3.1	A Worked Example: Predicting Vegetation Distributions	131
4.4	Extensions & Variations	138
4.4.1	Alternative Feature Selection & Impurity Metrics	138

4.4.2	Handling Continuous Descriptive Features	143
4.4.3	Predicting Continuous Targets	147
4.4.4	Tree Pruning	152
4.4.5	Model Ensembles	157
4.5	Summary	161
4.6	Further Reading	163
4.7	Exercises	164
<b>5</b>	<b>Similarity-based Learning</b>	<b>169</b>
5.1	Big Idea	169
5.2	Fundamentals	170
5.2.1	Feature Space	171
5.2.2	Measuring Similarity Using Distance Metrics	173
5.3	Standard Approach: The Nearest Neighbor Algorithm	176
5.3.1	A Worked Example	176
5.4	Extensions & Variations	180
5.4.1	Handling Noisy Data	180
5.4.2	Efficient Memory Search	185
5.4.3	Data Normalization	194
5.4.4	Predicting Continuous Targets	199
5.4.5	Other Measures of Similarity	202
5.4.6	Feature Selection	216
5.5	Summary	224
5.6	Further Reading	227
5.7	Epilogue	228
5.8	Exercises	230
<b>6</b>	<b>Probability-based Learning</b>	<b>237</b>
6.1	Big Idea	237
6.2	Fundamentals	239
6.2.1	Bayes' Theorem	242
6.2.2	Bayesian Prediction	246
6.2.3	Conditional Independence & Factorization	251
6.3	Standard Approach: The Naive Bayes Model	257
6.3.1	A Worked Example	259
6.4	Extensions & Variations	262
6.4.1	Smoothing	262

	6.4.2 Handling Continuous Features: Probability Density Functions	266
	6.4.3 Handling Continuous Features: Binning	279
	6.4.4 Bayesian Networks	282
	6.5 Summary	301
	6.6 Further Reading	304
	6.7 Exercises	306
<b>7</b>	<b>Error-based Learning</b>	<b>311</b>
	7.1 Big Idea	311
	7.2 Fundamentals	312
	7.2.1 Simple Linear Regression	312
	7.2.2 Measuring Error	315
	7.2.3 Error Surfaces	318
	7.3 Standard Approach: Multivariable Linear Regression with Gradient Descent	320
	7.3.1 Multivariable Linear Regression	320
	7.3.2 Gradient Descent	322
	7.3.3 Choosing Learning Rates & Initial Weights	329
	7.3.4 A Worked Example	331
	7.4 Extensions & Variations	334
	7.4.1 Interpreting Multivariable Linear Regression Models	335
	7.4.2 Setting the Learning Rate Using Weight Decay	337
	7.4.3 Handling Categorical Descriptive Features	339
	7.4.4 Handling Categorical Target Features: Logistic Regression	341
	7.4.5 Modeling Non-linear Relationships	353
	7.4.6 Multinomial Logistic Regression	361
	7.4.7 Support Vector Machines	364
	7.5 Summary	371
	7.6 Further Reading	374
	7.7 Exercises	376
<b>8</b>	<b>Evaluation</b>	<b>383</b>
	8.1 Big Idea	383
	8.2 Fundamentals	384
	8.3 Standard Approach: Misclassification Rate on a Hold-out Test Set	385
	8.4 Extensions & Variations	391

	8.4.1	Designing Evaluation Experiments	391
	8.4.2	Performance Measures: Categorical Targets	399
	8.4.3	Performance Measures: Prediction Scores	409
	8.4.4	Performance Measures: Multinomial Targets	426
	8.4.5	Performance Measures: Continuous Targets	428
	8.4.6	Evaluating Models after Deployment	433
	8.5	Summary	441
	8.6	Further Reading	442
	8.7	Exercises	443
<b>9</b>		<b>Case Study: Customer Churn</b>	449
	9.1	Business Understanding	449
	9.2	Data Understanding	453
	9.3	Data Preparation	457
	9.4	Modeling	463
	9.5	Evaluation	465
	9.6	Deployment	468
<b>10</b>		<b>Case Study: Galaxy Classification</b>	469
	10.1	Business Understanding	469
		10.1.1 Situational Fluency	472
	10.2	Data Understanding	474
	10.3	Data Preparation	481
	10.4	Modeling	486
		10.4.1 Baseline Models	486
		10.4.2 Feature Selection	489
		10.4.3 The 5-level Model	491
	10.5	Evaluation	494
	10.6	Deployment	495
<b>11</b>		<b>The Art of Machine Learning for Predictive Data Analytics</b>	497
	11.1	Different Perspectives on Prediction Models	499
	11.2	Choosing a Machine Learning Approach	504
		11.2.1 Matching Machine Learning Approaches to Projects	507
		11.2.2 Matching Machine Learning Approaches to Data	508
	11.3	Your Next Steps	509

<b>A</b>	<b>Descriptive Statistics &amp; Data Visualization for Machine Learning</b>	513
	A.1 Descriptive Statistics for Continuous Features	513
	A.1.1 Central Tendency	513
	A.1.2 Variation	515
	A.2 Descriptive Statistics for Categorical Features	518
	A.3 Populations & Samples	520
	A.4 Data Visualization	522
	A.4.1 Bar Plots	522
	A.4.2 Histograms	523
	A.4.3 Box Plots	526
<b>B</b>	<b>Introduction to Probability for Machine Learning</b>	529
	B.1 Probability Basics	529
	B.2 Probability Distributions & Summing Out	534
	B.3 Some Useful Probability Rules	536
	B.4 Summary	538
<b>C</b>	<b>Differentiation Techniques for Machine Learning</b>	539
	C.1 Derivatives of Continuous Functions	540
	C.2 The Chain Rule	542
	C.3 Partial Derivatives	543
	Bibliography	545
	Index	553