

# Business Analytics

## Methods & Cases for Data-Driven Decisions

Richard Huntsinger

University of California, Berkeley



# Contents

About the Author . . . . .	xix
Preface . . . . .	xx
Acknowledgements . . . . .	xxv
<b>Executive Overview</b>	<b>1</b>
<b>1 Data and Decisions</b>	<b>13</b>
Learning Objectives . . . . .	14
1.1 Introduction . . . . .	15
1.2 Data-to-Decision Process Model . . . . .	20
1.3 Decision Models . . . . .	23
1.4 Sensitivity Analysis . . . . .	32
<b>2 Data Preparation</b>	<b>37</b>
Learning Objectives . . . . .	38
2.1 Data Objects . . . . .	39
2.2 Selection . . . . .	45
2.3 Amalgamation . . . . .	56
2.4 Synthetic Variables . . . . .	63
2.5 Normalization . . . . .	67
2.6 Dummy Variables . . . . .	71
2.7 CASE   High-Tech Stocks . . . . .	74
<b>3 Data Exploration</b>	<b>91</b>
Learning Objectives . . . . .	92
3.1 Descriptive Statistics . . . . .	94
3.2 Similarity . . . . .	99
3.3 Cross-Tabulation . . . . .	102
3.4 Data Visualization . . . . .	108
3.5 Kernel Density Estimation . . . . .	122
3.6 CASE   Fundraising Strategy . . . . .	138
3.7 CASE   Iowa Liquor Sales . . . . .	155

---

<b>4</b>	<b>Data Transformation</b>	<b>163</b>
	Learning Objectives . . . . .	164
4.1	Balance . . . . .	165
4.2	Imputation . . . . .	170
4.3	Alignment . . . . .	174
4.4	Principal Component Analysis . . . . .	179
4.5	CASE   Loan Portfolio . . . . .	197
<b>5</b>	<b>Classification I</b>	<b>211</b>
	Learning Objectives . . . . .	212
5.1	Classification Methodology . . . . .	214
5.2	Classifier Evaluation . . . . .	223
5.3	k-Nearest Neighbors . . . . .	237
5.4	Logistic Regression . . . . .	242
5.5	Decision Tree . . . . .	248
5.6	CASE   Loan Portfolio Revisited . . . . .	257
<b>6</b>	<b>Classification II</b>	<b>267</b>
	Learning Objectives . . . . .	268
6.1	Naive Bayes . . . . .	269
6.2	Support Vector Machine . . . . .	291
6.3	Neural Network . . . . .	308
6.4	CASE   Telecom Customer Churn . . . . .	330
6.5	CASE   Truck Fleet Maintenance . . . . .	345
<b>7</b>	<b>Classification III</b>	<b>361</b>
	Learning Objectives . . . . .	362
7.1	Multinomial Classification . . . . .	363
7.2	CASE   Facial Recognition . . . . .	375
7.3	CASE   Credit Card Fraud . . . . .	383
<b>8</b>	<b>Regression</b>	<b>399</b>
	Learning Objectives . . . . .	400
8.1	Regression Methodology . . . . .	402
8.2	Regressor Evaluation . . . . .	410
8.3	Linear Regression . . . . .	423

---

8.4	Regression Versions . . . . .	444
8.5	CASE   Call Center Scheduling . . . . .	449
<b>9</b>	<b>Ensemble Assembly</b>	<b>469</b>
	Learning Objectives . . . . .	470
9.1	Bagging . . . . .	471
9.2	Boosting . . . . .	475
9.3	Stacking . . . . .	481
<b>10</b>	<b>Cluster Analysis</b>	<b>485</b>
	Learning Objectives . . . . .	486
10.1	Cluster Analysis Methodology . . . . .	488
10.2	Cluster Model Evaluation . . . . .	499
10.3	k-Means . . . . .	507
10.4	Hierarchical Agglomeration . . . . .	514
10.5	Gaussian Mixture . . . . .	525
10.6	CASE   Fortune 500 Diversity . . . . .	555
10.7	CASE   Music Market Segmentation . . . . .	563
<b>11</b>	<b>Special Data Types</b>	<b>577</b>
	Learning Objectives . . . . .	578
11.1	Text Data . . . . .	580
11.2	Time Series Data . . . . .	588
11.3	Network Data . . . . .	607
11.4	PageRank for Network Data . . . . .	614
11.5	Collaborative Filtering for Network Data . . . . .	624
11.6	CASE   Deceptive Hotel Reviews . . . . .	634
11.7	CASE   Targeted Marketing . . . . .	642
	Photo Credits . . . . .	658
	Index . . . . .	660