

Contents

1	Introduction to data augmentation in marketing	1
1.1	Research problem and relevance	2
1.1.1	Irrelevant communication and unused chances	2
1.1.2	Academic and practical research context	6
1.1.3	Research question and desired contribution	8
1.2	Research concepts and context	9
1.2.1	Definition of data augmentation and terminology	10
1.2.2	Specific characteristics of external information	14
1.2.3	Independence of source and target variables	17
1.2.4	Delineation of the research subject	19
1.3	Research approach	25
1.3.1	Case study with simulated sampled sources	26
1.3.2	Suitability of the case study method	28
1.4	Structure of the paper	36
2	Strategic motivation for data augmentation	41
2.1	Internal conditions within the company	42
2.1.1	Customer focus as a major marketing goal	42
2.1.2	Targeting in marketing practice	44
2.1.3	The marketing measure: conversion	48
2.2	External conditions around the company	49
2.2.1	Economic environment	50

2.2.2	Technological environment	51
2.2.3	Legal environment	52
2.2.4	Sociological and psychological environment	55
2.3	Sources for data augmentation	56
2.3.1	Public and official sources	57
2.3.2	Company-owned sources	59
2.3.3	Accumulated sources	61
2.3.4	Comparison of sources	64
2.4	Implications for data augmentation	67
3	Literature review on data augmentation	71
3.1	Evolution of data augmentation studies	72
3.1.1	Missing data problems and imputation	73
3.1.2	Statistical matching	78
3.1.3	Data augmentation in marketing	79
3.2	Process of data augmentation	81
3.2.1	Data screening	81
3.2.2	Data preparation	86
3.2.3	Choice of the best data augmentation method	90
3.2.4	Execution and internal evaluation of results	98
3.2.5	External evaluation of augmentation results	102
4	Methodological framework for data augmentation	105
4.1	Data augmentation specifics in marketing	106
4.1.1	The customer database as recipient unit	107
4.1.2	Possible donor units and their characteristics	108
4.1.3	Variable scales, values, and validity	111
4.1.4	Conditional independence of source and target variables	113
4.1.5	Source data mechanism of the donor unit	117
4.2	Data augmentation model	119
4.2.1	Populations and samples	120
4.2.2	Variables	123

4.2.3	Univariate pattern approach	127
4.2.4	Data augmentation target values and uncertainty . . .	129
4.2.5	Ignorability of the source data mechanism	133
5	Test design for evaluating the source characteristics	139
5.1	Research question conceptualization	140
5.1.1	Conceptual model	140
5.1.2	Hypotheses	145
5.2	Dataset generation and case study set up	150
5.2.1	Population characteristics and test design	150
5.2.2	Choice of variables	154
5.2.3	Test samples	166
5.2.4	Methods	172
6	Analysis of data augmentation KPIs	175
6.1	Preserved descriptive data and measures	176
6.2	Quality checks	179
6.2.1	Predictive power and discrimination	179
6.2.2	Managerial implications	183
6.3	Accuracy and precision of results	184
6.3.1	Total correct classification rate	185
6.3.2	Correct classification rate of the target value	193
6.4	Model lifts assessing the results	197
6.4.1	Model lift (chance)	198
6.4.2	Model lift (target)	207
6.4.3	Model lift (uniform)	211
6.4.4	Model lift (source)	215
6.4.5	Comparison of the KPIs	219
6.5	Conversion probability lift	222
7	Analysis of results and test of hypotheses	233
7.1	Validating the source data mechanism	234

7.1.1	Test based conditional dependency calculation	235
7.1.2	Model based conditional dependency calculation	241
7.1.3	Discussion	245
7.2	Influences of source characteristics	248
7.2.1	Representation	248
7.2.2	Overlap	250
7.2.3	Number of donors and surplus	253
7.3	Overall estimation of influences	256
7.3.1	Regression models estimating the model lift	258
7.3.2	Estimating the best method to be chosen	262
7.3.3	Conversion probability and conclusion	266
7.4	Managerial implications	272
8	Limitations of data augmentation and outlook	279
8.1	Limitations and alternatives	280
8.2	Further research opportunities	284
8.2.1	Ignorability of the source data mechanism	285
8.2.2	Deeper exploration of the augmentation process	286
8.2.3	Further augmentation opportunities and use cases . .	287
8.2.4	Uplift models	289
9	Summary	291
	Appendix	303
	Glossary	315
	References	319
	Curriculum Vitae	341
	Abstract	343

Targeting Using Augmented Data in Database
Marketing

Decision Factors for Evaluating External Sources

Hüttenrauch, geb. Krämer, B.

2016, XXII, 343 p. 49 illus., Softcover

ISBN: 978-3-658-14576-7