Contents

1 Data as the Basis of Models . . . 1
1.1 Data-Based Modeling . . . 2
1.1.1 The Concept of Model . . . 2
1.1.2 White-Box, Grey-Box, and Black-Box Models . . . 3
1.1.3 Criticism of Data-Based Modeling . . . 4
1.2 Classification of Data . . . 4
1.2.1 Historical Classification of the Term Data . . . 4
1.2.2 Data in Our Present Time . . . 5
1.2.3 Structural Classification . . . 6
1.2.4 Quantitative Categorization of Data . . . 7
1.2.5 Qualitative Categorization of Data . . . 9
1.2.6 Time Series . . . 10
1.2.7 Scales . . . 10
1.3 Data Mining as a Systematic Process . . . 13
1.3.1 What is Data Mining? . . . 13
1.3.2 Steps to Perform Data Mining . . . 15
1.3.3 Drawing Conclusions from Data . . . 15
1.3.4 Industrial Data Mining and the CRISP-DM . . . 17
1.4 Practical Handling of Data in Python . . . 19
1.4.1 Programs in this Book . . . 19
1.4.2 Libraries for Python . . . 20
1.4.3 Explainable Data in Python . . . 20
Tasks . . . 23
References . . . 23

2 Mathematical Description of Data . . . 25
2.1 Basics of Stochastics . . . 26
2.1.1 Probability . . . 26
2.1.2 Sets . . . 27
2.1.3 Probability Definition According to Laplace . . . 28
2.1.4 Event, Outcome, and Probability Spaces . . . 29
2.1.5 Axioms of Kolmogorov . . . 29
2.1.6 Conditional Probability and Bayes' Theorem . . . 31

2.1.7 Stochastic Process . . . 33
2.1.8 Considering the Probability Density in the Data . . . 35
2.2 Stochastics of Data . . . 36
2.2.1 Stochastics of the Sampling Process . . . 36
2.2.2 The Nyquist Theorem . . . 37
2.2.3 Distinguishing Uncertainties . . . 38
2.3 Statistical Tools for Handling Data . . . 40
2.3.1 Vector Representation of a Single Data Series . . . 40
2.3.2 Expected Value / Expectation Value . . . 40
2.3.3 Variance . . . 41
2.3.4 Empirical Variance . . . 42
2.3.5 Moments of a Probability Distribution . . . 42
2.3.6 Selected Probability Distributions . . . 43
2.3.7 Matrix Representation of Multiple Measurement Series . . . 45
2.3.8 Covariance and Covariance Matrix . . . 46
2.3.9 Correlation and Correlation Matrix . . . 47
Tasks . . . 49
References . . . 50

3 Data Preprocessing . . . 53
3.1 Goals of Preprocessing . . . 54
3.2 Data Cleaning . . . 55
3.2.1 Removal of Faulty Data Points . . . 55
3.2.2 Missing Data . . . 58
3.2.3 Tracking and Eliminating Duplicates . . . 58
3.3 Normalization . . . 59
3.3.1 Reasons for Normalizing Data . . . 59
3.3.2 Types of Normalizations . . . 59
3.3.3 Application of Normalization . . . 60
3.4 Filtering of Data . . . 63
3.4.1 Moving Average . . . 63
3.4.2 Convolution . . . 65
3.4.3 The Median Filter . . . 66
3.5 Triggering . . . 67
3.5.1 Time series and repetitive data . . . 67
3.5.2 Implementation of the trigger . . . 67
3.6 Transformations . . . 69
3.6.1 Differentiation of Data . . . 69
3.6.2 Functional Transformation . . . 70
3.6.3 Fourier Transformation . . . 72
3.6.4 Wavelet Transformation . . . 76
3.7 Quantification of Stochastic Properties . . . 79
3.7.1 Histograms . . . 79
3.7.2 Identification of the Probability Distribution Using Kernel Density Estimators . . . 80

3.8 Determining Key Figures from Data . . . 82
3.8.1 Statistical Key Figures . . . 82
3.8.2 Process-Oriented Key Figures . . . 83
Tasks . . . 85
References . . . 86

4 Supervised Learning . . . 87
4.1 Learning . . . 87
4.1.1 What does Learning Actually Mean? . . . 87
4.1.2 Various Forms of Machine Learning . . . 89
4.2 Strategy of Supervised Learning . . . 89
4.2.1 Algorithmic Consideration . . . 89
4.2.2 Classification and Regression . . . 91
4.2.3 Training Data and Test Data . . . 91
4.3 Evolutionary Learning . . . 93
4.3.1 Idea of Evolutionary Change . . . 93
4.3.2 Implementation of a Simple Evolutionary Algorithm . . . 93
4.3.3 Caching the Best Change . . . 95
4.4 LMS Adaptive Filter . . . 97
4.4.1 Theory of the LMS Algorithm . . . 97
4.4.2 Implementation of the LMS Algorithm . . . 98
4.4.3 NLMS Algorithm by Normalizing the Input Variables . . . 99
4.4.4 Training . . . 99
4.4.5 Applications and Properties of the Adaptive Filter . . . 100
4.4.6 Hyperparameters of the LMS Adaptive Filter . . . 100
4.5 Neural Networks . . . 101
4.5.1 The Neuron . . . 101
4.5.2 Networking of Neurons . . . 102
4.5.3 Activation Functions . . . 104
4.5.4 Training of Neural Networks . . . 106
4.5.5 Optimization as Hyperparameter . . . 107
4.5.6 Simple Classification Network with Scikit-Learn . . . 107
4.5.7 Classification Network in Keras . . . 109
4.5.8 Regression network in Keras . . . 112
4.5.9 Overfitting and Cross-Validation . . . 115
4.6 Recurrent Neural Networks . . . 117
4.6.1 Feedback in Neural Network Structures . . . 117
4.6.2 Gates for Controlling the Flow of Information within the Neuron . . . 119
4.6.3 Neural Network with Long Short-Term Memory (LSTM) . . . 119
4.6.4 Implementation of an LSTM Network . . . 121
4.7 Decision Trees . . . 123
4.7.1 Basic Idea of the Decision Tree . . . 123
4.7.2 Entropy and Information Gain . . . 125

4.7.3 Classification and Regression Trees . . . 130
4.7.4 Application Example with Scikit-Learn . . . 131
4.7.5 Implementation of a Simple Decision Tree . . . 132
4.7.6 The Gini Coefficient for Decision Trees . . . 136
Tasks . . . 138
References . . . 139

5 Unsupervised Learning . . . 141
5.1 Unsupervised Learning Paradigms . . . 142
5.2 Principal Component Analysis (PCA) . . . 143
5.2.1 Properties of the Covariance Matrix . . . 143
5.2.2 Eigenspace . . . 144
5.2.3 Eigenspace of the Covariance Matrix . . . 145
5.2.4 Eigenspace Reduction . . . 146
5.2.5 Implementation via Scikit-Learn . . . 147
5.2.6 Discussion of the PCA . . . 149
5.3 K-Means Clustering Method . . . 150
5.3.1 Finding Cluster Centers . . . 150
5.3.2 Implementation with Scikit-Learn . . . 151
5.3.3 Batch K-Means . . . 153
5.4 The t-Distributed Stochastic Neighbour Embedding Algorithm (t-SNE) . . . 153
5.4.1 Idea behind t-SNE . . . 153
5.4.2 Application of the t-SNE Implementation in Scikit-Learn . . . 155
5.4.3 Advantages and Disadvantages of t-SNE . . . 156
5.5 Autoencoder . . . 156
5.5.1 Topology of an Autoencoder . . . 156
5.5.2 Latent Space . . . 158
5.5.3 Anomaly Detection . . . 159
5.5.4 Implementation in Keras . . . 159
5.5.5 Classification in the Latent Space . . . 163
5.5.6 Uncertainty in the AE . . . 164
5.5.7 Detection of an Unknown Anomaly . . . 164
5.5.8 Combination of Methods . . . 169
Tasks . . . 171
References . . . 171

6 Physics-Informed Learning . . . 173
6.1 Introduction . . . 173
6.1.1 What is Physics-Informed Learning? . . . 174
6.1.2 A Simple, Motivating Example . . . 175
6.1.3 Continuous Change in the Process Chain . . . 177
6.1.4 Statistical Balance . . . 177
6.1.5 Reasons for Physics-Informed Machine Learning . . . 178

6.2 Data Enrichment 179
6.2.1 Optimized Choice of Preprocessing 179
6.2.2 Expert Knowledge on Data Objects 180
6.2.3 Automatic Optimization of Data Enrichment in Binary Classification 181
6.3 Embedding of Analytical Expressions in Neural Networks 182
6.3.1 Automatic Differentiation 182
6.3.2 Incorporating a Function as an Optimization Objective 183
6.3.3 Integration of a Function 186
6.3.4 Integration of a First-Order Ordinary Differential Equation 189
6.4 Stochastic Methods for Integrating Uncertainty into the Learning Process 191
6.4.1 Considering Uncertainty in Input Data 191
6.4.2 Quantification of Uncertainty from the Learned Results . . . 194
6.4.3 Stochastic Enrichment of Target Variables 196
6.4.4 Mixture-Density Networks (MDN) 199
6.5 Process Corridors 203
6.5.1 Probability Densities through 2D Histograms 203
Tasks 207
References 208

7 Explainability 209
7.1 Semantic Ordering Structures for the Digitization of Meaning 209
7.1.1 Semantics 210
7.1.2 Overview of Semantic Concepts 211
7.1.3 Alphabets and Words 212
7.1.4 Vocabulary 212
7.1.5 Synonym Ring 213
7.1.6 Taxonomies 215
7.1.7 Ontologies 218
7.2 Sensitivity Analysis 221
7.2.1 Perturbation Theoretical Approach 221
7.2.2 Perturbation of a Decision Tree 222
7.3 Explainable Machine Learning 225
7.3.1 Levels of Explainability 225
7.3.2 Practical Implementation of Explainability 226
7.3.3 Causality in the Technical Process Chain 230
7.3.4 Business Understanding 232
Tasks 232
References 233

Appendix A: Basics in Python 235