Data Mining 2020: Contents

0. General 1
1. Introduction 6
1.1 Basic data mining tasks 11
1.2 Approaches and tasks for data mining processes 19
1.3 Data mining issues 25
2. Measurement and data: variables 31
2.1 Outline of measurement 31
2.2 Types of variables or measurements 43
2.3 Continua of variables 51
3. Building data – Dealing with variables 64
3.1 Preparation with variables and their values 64
3.2 Basic computation techniques for data 92
3.3 Distance measures 96
3.4 Binary distance and similarity measures 110
3.5 Distance measures of mixed variables 117
4. Important preprocessing methods 122
4.1 Normalization of variable values 122
4.2 Nearest neighbor searching and classifier 129

4.3 Outliers 140
4.4 Finding outliers effectively 149
4.5 Balancing class distribution 162
5. Variable selection 167
5.1 Simple technique 168
5.2 Principles of variable selection 171
5.3 Variable selection criteria 178
5.4 Search methods 189
5.5 Variable selection methods with criterion functions 193
5.6 Evaluation of variable importance 210
6. Discretization methods 230
6.1 The purpose of discretization 230
6.2 Unsupervised discretization algorithms 237
6.3 Supervised discretization algorithms 240
6.4 Other supervised discretization algorithms 257
7. Variable extraction and dimensionality reduction 266
7.1. Variable extraction in general 269
7.2. Principal component analysis 273
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Testing models built</td>
<td>290</td>
</tr>
<tr>
<td>8.1 Bayesian classification rule</td>
<td>291</td>
</tr>
<tr>
<td>8.2 Testing</td>
<td>295</td>
</tr>
<tr>
<td>8.3 Accuracy values to be computed</td>
<td>302</td>
</tr>
</tbody>
</table>