

# Indice

<b>1</b>	<b>Data Understanding</b>	<b>3</b>
1.1	Data understanding . . . . .	3
1.2	Data semantics . . . . .	3
1.3	Distribution of the variables and statistics . . . . .	3
1.4	Assessing data quality . . . . .	3
1.5	Variable transformations . . . . .	3
1.6	Pairwise correlations and eventual elimination of redundant variable	3
<b>2</b>	<b>Clustering analysis</b>	<b>4</b>
2.1	Clustering Analysis by K-means . . . . .	4
2.1.1	Choice of attributes and distance function . . . . .	4
2.1.2	Identification of the best value of k . . . . .	4
2.1.3	Characterization of the obtained clusters by using both analysis of the k centroids and comparison of the distribution of variables within the clusters and that in the whole dataset . .	4
<b>3</b>	<b>Association Rules</b>	<b>5</b>
<b>4</b>	<b>Classification</b>	<b>6</b>

# Elenco delle figure

# Capitolo 1

## Data Understanding

- 1.1 Data understanding
- 1.2 Data semantics
- 1.3 Distribution of the variables and statistics
- 1.4 Assessing data quality
- 1.5 Variable transformations
- 1.6 Pairwise correlations and eventual elimination of redundant variable

# Capitolo 2

## Clustering analysis

### 2.1 Clustering Analysis by K-means

#### 2.1.1 Choice of attributes and distance function

#### 2.1.2 Identification of the best value of $k$

#### 2.1.3 Characterization of the obtained clusters by using both analysis of the $k$ centroids and comparison of the distribution of variables within the clusters and that in the whole dataset

## Capitolo 3

### Association Rules

# Capitolo 4

## Classification