.

Analysis of the dataset *Credit Card Default*

Gaspare Ferraro (520549)
Javad Khalili (546677)
Mario Matovic (583449)

December 11, 2018

University of Pisa
Exam: Data Mining
Year: 2018/2019
Instructors: Dino Pedreschi, Anna Monreale, Riccardo Guidotti

Contents

Chapter 1

Introduction

Chapter 2

Data Understanding

Chapter 3

Clustering

3.1 K-means

3.1.1 Choice of attributes and distance function

3.1.2 Choise of the best value of k

3.1.3 Cluster analysis

3.2 DBSCAN

3.2.1 Choice of attributes and distance function

3.2.2 Study of the clustering parameters

3.2.3 Characterization and interpretation of the obtained clusters

3.3 Hierarchical clustering

3.3.1 Choice of attributes and distance function

3.3.2 Discussion of dendograms using different algorithms

3.4 Evaluation of clustering approaches and comparison of the clustering obtained

Chapter 4

Association Rules Mining

Chapter 5

Classification

5.1 Choice of attributes for the decision trees

5.2 Decision trees interpretation and validation with test and training set

5.3 Discussion of the best prediction model

Chapter 6

Conclusion