



UNIVERSITÀ DI PISA

RELAZIONE DATA MINING

Il Gentmo Team alla scoperta dei lavoratori che abbandonano l'IBM

Erica Cau

Alfonso Ferraro

Simona Mazzarino

Federico Mazzoni

1 Introduzione

1.1 Data Understanding

2 Data Quality

2.1 Errori nei dati

2.2 Valori mancanti

2.3 Outliers

2.4 Eliminazione data attributes

2.5 Problemi nei valori: calcolo di attributi usando la moda, valori mancanti, valori incongruenti

3 Clustering