UNIVERSITÀ DI PISA

PROGETTO PER IL CORSO DI DATA MINING

Analisi del Dataset Human Resources Analytics

Gianmarco Ricciarelli & Maria Cristina Uccheddu

Indice

Elenco delle figure

Capitolo 1

Data Understanding

1.1 Data semantics

Il progetto viene svolto sul dataset (simulato) **Human Resources Analytics**. Questo dataset contiene le informazioni sui dipendenti di un'azienda fittizia, suddivise in base ai seguenti campi:

satisfaction_level - valutazione *quantitativa* del livello di soddisfazione di ciascun dipendente, ha un valore compreso tra 0 (minimo) e 1 (massimo);

last_evaluation - tempo trascorso, in anni, dall'ultima valutazione delle performace del dipendente;

number_project - numero di progetti completati durante il periodo di lavoro;

average_montly_hours - media delle ore di lavoro in un mese;

time_spend_company - numero di anni trascorsi nell'azienda;

work_accident - esprime con un 1 il coinvolgimento di un dipendente in un incidente sul lavoro, altrimenti viene impostato come 0;

left - se il dipendente ha lasciato l'azienda viene impostato come 1, altrimenti come 0;

promotion_last_5years - se il dipendente è stato promosso negli ultimi 5 anni viene impostato come 1, altrimenti come 0;

sales - definisce il dipartimento nel quale il dipendente lavora;

salary - esprime il livello (*low*, *medium*, *high*), di salario nel quale rientra il dipendente;

1.2 Distribution of the variables and statistics

1.3 Assessing data quality

1.4 Variable transformations

1.5 Pairwise correlations and eventual elimination od 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