

## **Data Mining**

Docente: Annamaria Guolo

### **Prova pratica del 16 giugno 2017**

**ISTRUZIONI:** La durata della prova è di 2 ore e 30 minuti. Redigere una breve relazione che riassume l'analisi dei dati svolta ed i risultati conseguiti e consegnare la stampa in versione cartacea completa di *i*) nome e cognome, *ii*) numero di matricola, *iii*) data. Relazioni senza queste tre informazioni non potranno essere corrette.

Di seguito si riporta la descrizione del file ed una breve traccia dell'analisi da effettuare. È ammesso l'uso del materiale relativo al corso (slides della teoria, dispense del laboratorio, appunti, ...), del libro di testo, ma non di internet.

Dataset *felicità*: i dati si riferiscono alle caratteristiche di alcuni paesi che la rivista Forbes ha classificato in base ad un indice di felicità degli abitanti.

- `salute.indice`: indice di salute
- `salute.spese`: spese in salute pro capite
- `istruzione`: media degli anni di istruzione
- `vita`: aspettativa di vita
- `reddito.distr`: indice di distribuzione del reddito (più alto l'indice peggiore la distribuzione)
- `equatore`: il paese è vicino all'equatore? TRUE/FALSE
- `popden`: densità della popolazione (persone per metro quadro)
- `spese.pagate`: percentuale delle spese in salute pagate dallo stato
- `pilc`: PIL pro capite
- `banca`: indice di efficacia del sistema bancario (più alto l'indice migliore il sistema)
- `democrazia`: indice di democrazia (più alto l'indice maggiore la democratizzazione del paese)
- `oecd`: il paese appartiene all' Organisation for Economic Co-operation and Development? TRUE/FALSE
- `felice`: indice di felicità di Forbes (maggiore l'indice più felice il paese)

Lo scopo dell'analisi è valutare quali variabili siano associate all'indice di felicità del paese.

1. Si consideri il sottoinsieme di dati costituito dalle quattro variabili *felice*, *salute.spese*, *equatore*, *pilc*. Costruire il modello che si ritiene più opportuno per gli scopi dell'analisi. Riportare output e valutazione grafiche del modello / dei modelli che si ritengono adatti per una spiegazione dell'approccio scelto e dei risultati. Un elevato PIL pro capite è associato ad un aumento della felicità del paese?
2. Considerare tutte le variabili del dataset. Procedere alla costruzione del modello che si ritiene più opportuno per gli scopi di analisi. Riportare output e valutazione grafiche del modello / dei modelli che si ritengono adatti per una spiegazione dell'approccio scelto e dei risultati.

\* Precisazione: nel caso in cui si svolgano analisi che richiedano la definizione di un seed, specificare nel testo quale seed viene scelto.