

# Comparazione di classificatori in Credit Card Approval

Valerio Colitta, Daniele Cominu, Alessio Fiorenza

Aprile 2017

## 1. Descrizione preliminare

Volendo approfondire i metodi e le tecniche di classificazione affrontate nel corso di Metodi Quantitativi per l'Informatica, abbiamo deciso di confrontarli risolvendo il problema presentato in *Credit Approval Data Set* [1]; abbiamo deciso di utilizzare modelli di diverse complessità ed abbiamo osservato che data la conformazione dei dati e la loro dimensione ridotta, si ottengono risultati migliori con modelli meno complessi.

## 2. Il dataset

Il dataset riguarda delle domande di approvazione di carte di credito, ed è composto da 690 campioni, 37 dei quali con valori mancanti. Sono presenti in totale 15 feature, le quali si articolano in 6 continue e 9 categoriche. Tali dati sono poi classificati in due classi  $\{+, -\}$  che rappresentano rispettivamente l'approvazione o meno della carta di credito. Il significato delle 15

feature non è noto per poter mantenere la confidenzialità dei dati.

## 3. Manipolazione dei dati

Come prima cosa sono stati eliminati manualmente dal dataset i 37 campioni di cui non erano stati specificate alcune feature, riducendo ulteriormente la dimensione del dataset a 653 campioni.

Per superare l'eterogeneità nella tipologia delle feature, si è deciso di applicare la tecnica del *One Hot Encoding* [2] per le feature categoriche; l'*One Hot Encoding* è una tecnica utilizzata per trattare feature categoriche in problemi di classificazione e regressione, e consiste nel tradurre una feature categorica che può assumere  $n$  valori distinti in un vettore di  $n$  feature binarie; per ogni campione la feature  $i$ -esima del vettore calcolato assume il valore 1 se e solo se il campione assume il valore  $i$ -esimo per la feature considerata; le restanti  $n-1$  sono dunque settate a 0 per tale campione.

## Riferimenti bibliografici

- [1] John Ross Quinlan  
*Credit Approval Data Set*  
<http://archive.ics.uci.edu/ml/datasets/Credit+Approval>
- [2] Håkon Hapnes Strand  
*One Hot Encoding*  
<https://www.quora.com/What-is-one-hot-encoding-and-when-is-it-used-in-data-science>