**Progetto #5: TripAdvisor**

**Motivazione**

Le recensioni scritte dagli utenti circa beni acquistati o servizi usufruiti hanno un elevato impatto sui fornitori di tali beni e servizi in termini di reputazione. Negli ultimi anni, la cosiddetta Brand Reputation, ha assunto un ruolo sempre più rilevante a livello decisionale per permettere di affrontare criticità relative a beni e servizi commercializzati, ridurre la perdita di potenziale clientela ed infine mantenere un elevato potere competitivo all’interno del mercato. Date quindi le opinioni espresse dai clienti sui vari social networks (Twitter, Facebook, TripAdvisor, etc…), è necessario quantificare tali opinioni per fornire indicatori di performance relativi ai beni/servizi offerti, per permettere decisioni correttive/strategiche.

**Dati**

* Recensioni scritte in linguaggio naturale derivanti da TripAdvisor
* Metadati: rating associati a singoli aspetti (rooms, location, service etc…)
* Target: overall rating

Review ID = 1

<**Author**>moonkwean

<**Content**>Everything you need Stayed 9 days in Phoenix but only three nights at this hotel. Wished we had booked our whole stay here. Staff very friendly and available. Any time we requested something it was attended to right away. Rooms clean. Bonus is the breakfast, social time and the 3 computers in the lobby. Would definitely stay here again. Thanks David, Pamela and Mila.

<**Date**>Jan 18, 2006

<**No. Reader**>17

<**No. Helpful**>14

<**Overall**>4

<**Value**>5

<**Rooms**>4

<**Location**>-1

<**Cleanliness**>5

<**Check in / front desk**>-1

<**Service**>5

<**Business service**>-1

**Il modello di sentiment: Naive Bayesian Network**

Date le opinioni di diversi clineti circa le strutture alblerghiere presso le quali hanno soggiornato, si dovrà definire un modello basato su **Reti Bayesiane** (**Naive)** che sia in grado di prevedere il rating associato ad una nuova recensione.

Il sentimento positivo/negativo relativo alle opinioni sui social network può essere inferito definendo il seguente modello Naive Bayes:

**Obiettivi del progetto**

1. Elaborare il testo delle recensioni per derivarne una rappresentazione vettoriale per la successiva fase di definizione del modello Bayesiano

Es: Term/Document Representation (+ metadati + target)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TERMS | | | | | METADATI | | | TARGET |
| Review ID | staff | rooms | clean | shops | breakfast | Value | Rooms | Location | Overall |
| 1 | 1 | 1 | 1 | 0 | 1 | 5 | 4 | 1 | 1 |
| 2 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 5 | 5 |

1. Definire la struttura della rete
2. Stimare i relativi parametri (tabelle di probabilità)
3. Effettuare il task di inferenza. Data una nuova recensione:
   1. Ricavare la relativa rappresentazione vettoriale

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TERMS | | | | | METADATI | | | TARGET |
| Review ID | staff | rooms | clean | shops | breakfast | Value | Rooms | Location | Overall |
| new | 0 | 0 | 0 | 1 | 1 | 1 | 5 | 2 | ? |

* 1. Inferire il valore di rating
  2. Aggregare i rating relativi ai singoli hotel per fornire una valutazione complessiva basata su tutte le recensioni



**Software**

Software utilizzabili:

1. Python: <http://pythonhosted.org/libpgm/>
2. Matlab: <https://code.google.com/archive/p/bnt/>
3. Java:
   1. <https://sourceforge.net/projects/unbbayes/>
   2. <https://sourceforge.net/projects/bnj/>
   3. <http://www.cs.waikato.ac.nz/ml/weka/>
4. R: <http://www.bnlearn.com/>