



UNIVERSITÀ DEGLI STUDI DI
MILANO-BICOCCA

F1801Q127

DATA ANALYTICS

Amazon Reviews Sentiment Analysis

Studenti:

Basso Matteo

Ferri Marco

Matricole:

807628

807130

Luglio 2019

Abstract

Lorem ipsum dolor sit amet.

Indice

1	Introduzione	5
1.1	Dataset	5
1.2	Obiettivi	5
1.3	Ipotesi e assunzioni	5
1.4	Software	5
2	Basic Analysis	6
2.1	Schema	6
2.2	Dimensioni	6
2.3	Distribuzioni	6
2.4	Analisi business-oriented	6
3	Network Analysis	7
3.1	Struttura della rete	7
3.2	Grado dei nodi	7
3.3	Misure di centralità ?	7
4	Sentiment Analysis	8
4.1	Assunzioni	8
4.2	Binarizzazione	8
4.3	Undersampling	8
4.4	Elaborazione del testo	8
4.5	Parole più usate	8
5	Sentiment Prediction	9
5.1	Pesatura dei termini (TF-IDF)	9
5.2	Termini più rilevanti	9
5.3	Modelli di predizione	9
5.3.1	Random Forest	9
5.3.2	Naive Bayes	9
5.3.3	SVM	9
5.4	Pipeline	9
6	Aspect Based Sentiment Analysis	10
6.1	Elaborazione del testo	10
6.2	Estrazione degli aspetti	10
6.3	Identificazione del sentiment	10
6.4	Risultati	10
7	Collaborative Filtering	11

7.1	Funzionamento	11
7.2	Risultati	11
8	Web Demo	12
8.1	Architettura	12
8.2	Sentiment Prediction	12
8.3	Aspect Based Sentiment Analysis	12
9	Conclusioni	13

Elenco delle figure

Elenco delle tabelle

1 Introduzione

1.1 Dataset

1.2 Obiettivi

1.3 Ipotesi e assunzioni

1.4 Software

2 Basic Analysis

2.1 Schema

2.2 Dimensioni

2.3 Distribuzioni

2.4 Analisi business-oriented

3 Network Analysis

3.1 Struttura della rete

3.2 Grado dei nodi

3.3 Misure di centralità ?

4 Sentiment Analysis

4.1 Assunzioni

4.2 Binarizzazione

4.3 Undersampling

4.4 Elaborazione del testo

4.5 Parole più usate

5 Sentiment Prediction

5.1 Pesatura dei termini (TF-IDF)

5.2 Termini più rilevanti

5.3 Modelli di predizione

5.3.1 Random Forest

5.3.2 Naive Bayes

5.3.3 SVM

5.4 Pipeline

6 Aspect Based Sentiment Analysis

6.1 Elaborazione del testo

6.2 Estrazione degli aspetti

6.3 Identificazione del sentiment

6.4 Risultati

7 Collaborative Filtering

7.1 Funzionamento

7.2 Risultati

8 Web Demo

8.1 Architettura

L'interfaccia web è stata sviluppata utilizzando l'architettura a 3 layer, con separazione di frontend, backend e database.

Il database utilizzato in fase di lettura è quello fornito inizialmente, senza alcuna modifica. Esso consiste quindi in un file SQLite interrogabile e modificabile semplicemente tramite un web server. Questo risulta particolarmente utile per fornire i dettagli dei giocatori ed eventualmente dei team così che l'utente possa visualizzarli e sceglierli attraverso l'opportuna interfaccia.

Per lo sviluppo del backend è stato deciso di utilizzare l'engine Javascript tramite il popolare progetto Node.js ?. Esso è in grado di agire come middleware tra il frontend e il database, separando al meglio le logiche di manipolazione del dato. É inoltre incaricato di chiamare adeguatamente lo script R per la predizione del vincitore della partita e per svolgere le inferenze richieste.

Il frontend è invece sviluppato utilizzando la libreria Javascript React.js ?

8.2 Sentiment Prediction

8.3 Aspect Based Sentiment Analysis

9 Conclusioni

Riferimenti bibliografici

Node.js. URL: <https://nodejs.org/it/>.

React. URL: <https://reactjs.org/>.