

Progetto Metodi Informatici per la Gestione Aziendale

Dati e Raccomandazioni: Un Viaggio tra Collaborative
Filtering, Content Based e Sentiment Analysis

Alessandro Isceri & Luca Bolis

Dati e importanza delle Review

Per tutti i business le review sono di fondamentale importanza.

Le recensioni analizzate in questo progetto riguardano i prodotti software in vendita su Amazon.



Bhatta's Reviews

★★★★★ **Full Five Stars!**

Reviewed in the United States on May 27, 2012

Verified Purchase

This is a very addictive fun and fast paced game. I must confess that I had side loaded this game several months ago when it was available on Google play and I loved it.

The game play is superb and has very good fast graphics on Kindle fire and hours of fun & excitement. Once you start playing this game I can assure you that you will never put it down - Beautifully done!

While I was very happy to see that it was finally released for Amazon App store. Understandably I could not replace my existing Google play application or update it. I do not want to start all over again (I have scores in the millions with various characters unlocked with hours of play) so I prevented from the APP store replacing my game and kept the old version which I got free from Google play.

This game has the best rating of any Android game ever. Glad it is officially available via Amazon APP store.

Thanks to the developer for making a fantastic free game.

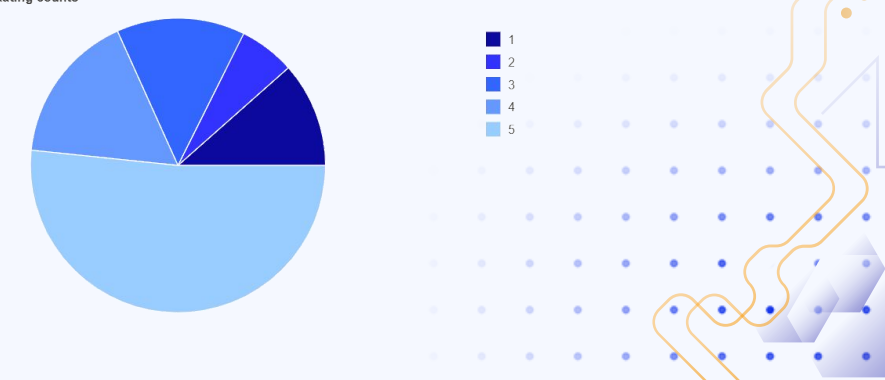
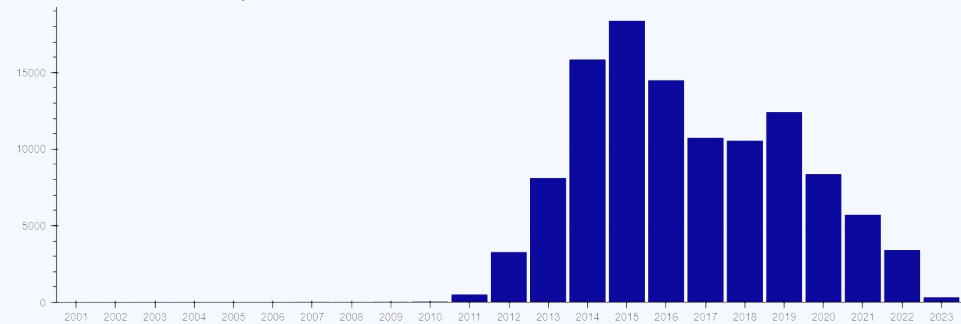
Download it now. It is superb game!

Cheers!

6,150 people found this helpful

[illegible]

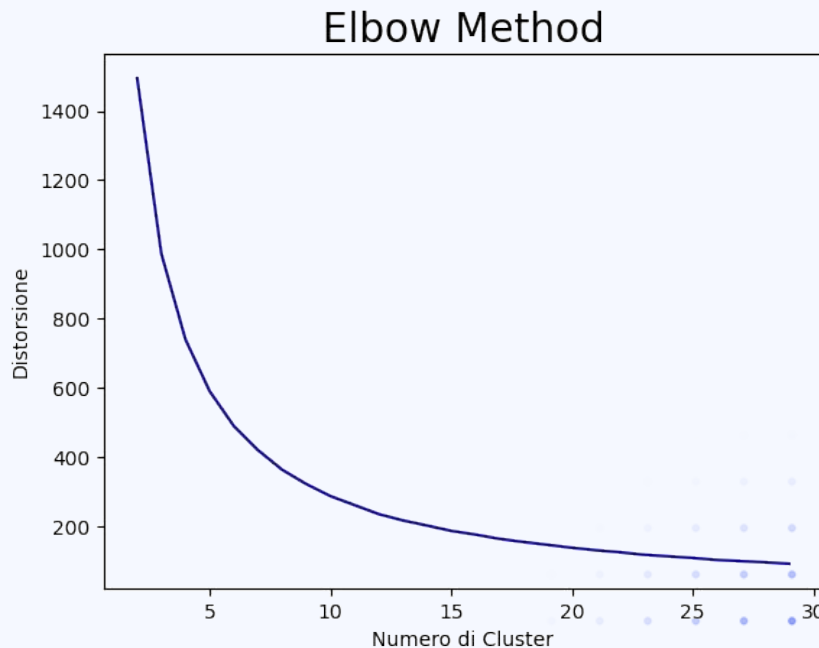
Year	Number of Publications
2001	0
2002	0
2003	0
2004	0
2005	0
2006	0
2007	0
2008	0
2009	0
2010	0
2011	500
2012	3500
2013	8000
2014	16000
2015	18000
2016	14500
2017	10500
2018	10500
2019	12500
2020	8500
2021	5500
2022	3500
2023	500



Segmentazione degli Utenti: K-Means

L'algoritmo K-Means consente di raggruppare utenti con caratteristiche simili.

In questo modo sempre più aziende offrono promozioni mirate aumentando le vendite.



Sistemi di Raccomandazione

Collaborative Filtering



- Si basa solamente sui rating assegnati dagli utenti ai prodotti
- Si basa sulle interazioni e preferenze espresse da un gruppo di utenti
- Semplice da implementare

Content Based



- Sfrutta anche i dati testuali dei prodotti (descrizione, titolo, ...)
- Raccomanda nuovi contenuti che condividono caratteristiche simili con quelli preferiti dall'utente
- Richiede tecniche di Natural Language Processing

KNN VS Matrix Factorization

Preferenze Utente	KNN	Matrix Factorization
<ul style="list-style-type: none">• Dillard• Maps For Google• Smart System Info• The Island Castaway: Lost World• OneDrive	<ul style="list-style-type: none">• Pocketmine server• AirPlayMirror Receiver• Dino Digger• Fraction calculator• RPG Clicker	<ul style="list-style-type: none">• Zoom• Instagram• Kindle for Android• Speedtest by Ookla• Xfinity Stream(live TV)

- L'algoritmo Matrix Factorization, con la sua configurazione ottimale, ha prodotto un RMSE leggermente migliore di quello prodotto da KNN.

Proprietà del RecSys Collaborative Filtering

Novità

Suggerire prodotti che un utente potrebbe non aver ancora visto

Rilevanza

Raccomandare prodotti che un utente ritiene più interessanti

Serendipità

Suggerire prodotti "inaspettati"

Ottimalità

Configurazione ottimale per ridurre al minimo gli errori



Natural Language Processing

Dividere il testo in
base agli spazi

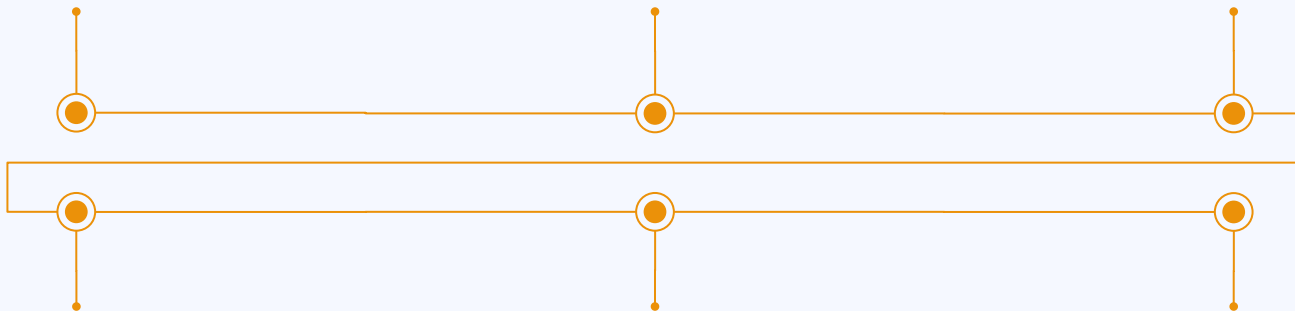
First

Rimuovere le stop
word

Next

Ridurre le parole a dei
token

Next



Next

Creare una
rappresentazione
numerica

Next

Unire i rating alla
rappresentazione
numerica

Last

Sistema di
raccomandazione
content based

Collaborative Filtering VS Content Based

Esistono numerosi software di categorie diverse, con descrizioni decisamente differenti tra loro.

In questo contesto catturare le preferenze dell'utente attraverso l'analisi testuale di titoli e descrizioni dei prodotti può rivelarsi un task non banale.

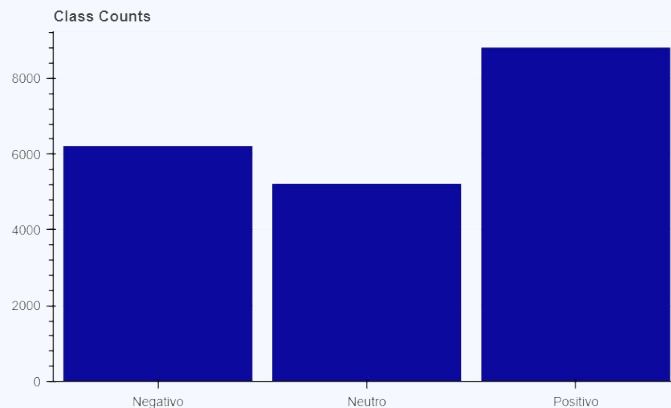
<u>Sistema di Raccomandazione</u>	<u>RMSE</u>
Collaborative Filtering - KNN	1.13
Collaborative Filtering - Matrix Factorization	1.09
Content Based - Transformer	1.12



Sentiment Analysis

La sentiment analysis permette di identificare e classificare automaticamente le emozioni e le opinioni espresse nei testi.

Il dataset è stato ribilanciato per garantire i migliori risultati nella fase di classificazione.



This is a very addictive **fun** and fast paced game. I must confess that I had side loaded this game several months ago when it was available on Google play and I **love** it.

The game play is **superb** and has **very good** fast graphics on Kindle fire and hours of **fun** & **excitement**. Once you start playing this game I can assure you that you will never put it down - **Beautifully** done!

While I was very **happy** to see that it was finally released for Amazon App store. [...]

This game has the best rating of any Android game ever. **Glad** it is officially available via Amazon APP store.

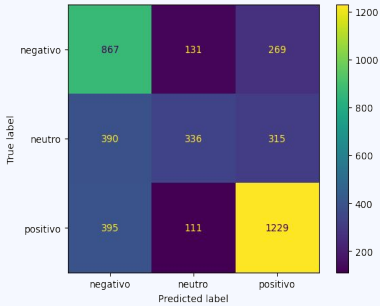
Thanks to the developer for making a **fantastic** free game.

Download it now. It is **superb** game!

Cheers!

Matrici di Confusione e F1-Score

**BoW
Stemmer**

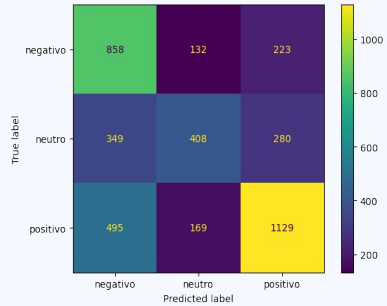


**Negativo
0.59**

**Neutro
0.42**

**Positivo
0.69**

**BoW
Lemmatizer**

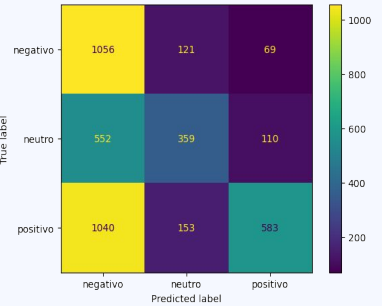


**Negativo
0.59**

**Neutro
0.47**

**Positivo
0.66**

**TF-IDF
Stemmer**

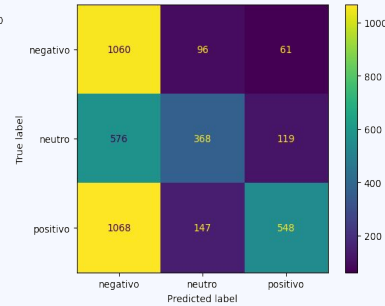


**Negativo
0.54**

**Neutro
0.43**

**Positivo
0.46**

**TF-IDF
Lemmatizer**

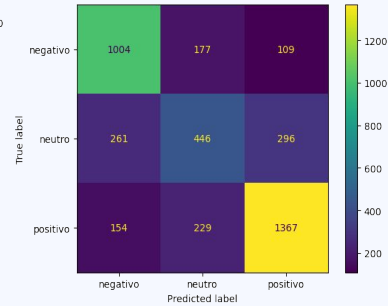


**Negativo
0.54**

**Neutro
0.44**

**Positivo
0.44**

Transformer



**Negativo
0.74**

**Neutro
0.48**

**Positivo
0.78**

Grazie per l'attenzione!

Ci sono domande?

Alessandro Isceri & Luca Bolis