

UNIVERSITÀ DEGLI STUDI DI
MILANO-BICOCCA

DATA ANALYTICS

Amazon Fine Food Reviews

Raffaele Cerizza - 845512



Indice

1	Introduzione	5
2	Analisi esplorativa	7
2.1	Dataset	7
2.2	Preprocessing	7
2.2.1	Preprocessing del dataset	7
2.2.2	Preprocessing del testo	10
2.3	Analisi sulle recensioni	11
2.3.1	Distribuzione delle valutazioni	11
2.3.2	Distribuzione delle recensioni per anno	13
2.3.3	Distribuzione percentuale delle opinioni per anno	13
2.3.4	Sintesi	14
2.4	Analisi sui prodotti	14
2.4.1	Distribuzione del voto medio per prodotto	15
2.4.2	Prodotti con il voto medio più alto	16
2.4.3	Prodotti con il voto medio più basso	16
2.4.4	Sintesi	18
2.5	Analisi sui prezzi	18
2.5.1	Distribuzione delle recensioni per prezzo	20
2.5.2	Correlazione tra valutazioni e prezzi	21
2.5.3	Sintesi	22
2.6	Analisi sulle parole delle recensioni	22
2.6.1	Individuazione delle parole più frequenti	22
2.6.2	Distribuzione delle opinioni sulle recensioni che contengono le parole più frequenti	23
2.6.3	Sintesi	25
3	Sentiment analysis	26
3.1	Approcci basati su lessici	26
3.1.1	AFINN	26
3.1.2	VADER	27
3.2	Approccio basato sul Machine Learning	28
3.2.1	Modello di regressione logistica	29
3.3	Confronto degli approcci di sentiment analysis	30
3.4	Analisi sulle parole	32
3.5	Analisi delle misclassificazioni	33

3.5.1	Esempi di misclassificazione	33
3.5.2	Sintesi	36
3.6	Analisi sui prodotti	36
3.6.1	Prodotti con il sentimento medio più alto	36
3.6.2	Prodotti con il sentimento medio più basso	37
3.6.3	Sintesi	39
3.7	Analisi sui prezzi	41
3.7.1	Correlazione tra sentimenti e prezzi	41
3.7.2	Sintesi	41
3.8	Analisi del rapporto tra testo e sommario delle recensioni . . .	42
3.8.1	Risultati della sentiment analysis sui sommari	42
3.8.2	Sintesi	44
4	Aspect-based sentiment analysis	47
4.1	Approcci per l'aspect-based sentiment analysis	47
4.1.1	Approccio con ASUM	47
4.1.2	Approccio con LCF-ATEPC	49
4.2	Analisi sugli aspetti	50
4.2.1	Individuazione degli aspetti più frequenti	50
4.2.2	Aspetti più apprezzati	51
4.2.3	Aspetti meno apprezzati	53
4.2.4	Sintesi	55
4.3	Analisi sui prezzi	56
4.3.1	Analisi sui sentimenti dei singoli aspetti	57
4.3.2	Analisi temporale	57
4.3.3	Sintesi	58
5	Dashboard	59
6	Conclusioni e sviluppi futuri	61

Elenco delle figure

1	Distribuzione delle valutazioni degli utenti.	12
2	Distribuzione delle opinioni degli utenti.	12
3	Distribuzione delle recensioni per anno.	13
4	Distribuzione percentuale delle opinioni per anno.	14
5	Distribuzione del voto medio per prodotto.	15
6	Migliori 20 prodotti per voto medio.	18
7	Peggiori 20 prodotti per voto medio.	20
8	Distribuzione delle recensioni per prezzo.	21
9	<i>Scatter plot</i> della relazione tra <i>rating</i> e prezzi.	22
10	<i>Wordcloud</i> delle parole più frequenti nel testo delle recensioni.	23
11	Distribuzione delle opinioni sulle recensioni per le parole più frequenti.	24
12	Distribuzione dei sentimenti sulle recensioni per le parole più frequenti.	32
13	Migliori 20 prodotti per sentimento medio.	37
14	Peggiori 20 prodotti per sentimento medio.	39
15	<i>Scatter plot</i> della relazione tra sentimenti e prezzi.	42
16	Distribuzione della lunghezza delle recensioni in termini di numero di parole.	44
17	Distribuzione della lunghezza dei sommari in termini di numero di parole.	45
18	Matrice di confusione per la classificazione dei sentimenti dei sommari con AFINN.	45
19	Matrice di confusione per la classificazione dei sentimenti dei sommari con VADER.	46
20	Aspetti più frequenti identificati con LCF-ATEPC.	51
21	Aspetti più apprezzati identificati con LCF-ATEPC.	53
22	Aspetti meno apprezzati identificati con LCF-ATEPC.	55
23	Sentimento medio per anno dei principali aspetti relativi al prezzo identificati con LCF-ATEPC.	58

Elenco delle tabelle

1	Descrizione degli attributi delle recensioni	8
2	Descrizione degli attributi dei prodotti	8
3	Descrizione del <i>dataset</i> dopo il <i>preprocessing</i>	10
4	Descrizione dei prodotti migliori per voto medio	17
5	Descrizione dei prodotti peggiori per voto medio	19
6	Confronto delle <i>performance</i> dei modelli impiegati per la <i>sentiment analysis</i> sul testo delle recensioni	31
7	Descrizione dei prodotti migliori per sentimento medio	38
8	Descrizione dei prodotti peggiori per sentimento medio	40
9	Confronto delle <i>performance</i> dei modelli impiegati per la <i>sentiment analysis</i> sul testo e sul sommario delle recensioni	43
10	Aspetti positivi identificati con ASUM	52
11	Aspetti negativi identificati con ASUM	54
12	Sentimento medio degli aspetti relativi al prezzo individuati con LCF-ATEPC	57

1 Introduzione

Il presente lavoro ha per oggetto l'analisi di un *dataset* composto da recensioni di cibi e bevande pubblicate sul sito di Amazon.

Questa analisi è stata svolta al fine di rispondere a una serie di quesiti. La finalità di questi quesiti è duplice. Da un lato si vogliono individuare i prodotti più graditi e quelli meno graditi dagli utenti. Dall'altro lato si vuole studiare l'impatto dei prezzi sul gradimento degli utenti. In questo modo si ritiene di poter definire politiche commerciali in grado di massimizzare la soddisfazione degli utenti e il fatturato di piattaforme di *e-commerce*.

In particolare i quesiti considerati sono i seguenti:

- **quesiti esplorativi sulle recensioni degli utenti:**
 - qual è il trend temporale di soddisfazione degli utenti?
 - quali sono i prodotti con la valutazione media più alta?
 - quali sono i prodotti con la valutazione media più bassa?
 - quali sono le fasce di prezzo più recensite?
 - come varia la valutazione dei prodotti in base al prezzo?
 - quali sono le parole più frequenti nelle recensioni degli utenti?
- **quesiti sul sentimento degli utenti espresso nelle recensioni:**
 - quali sono i prodotti più apprezzati dagli utenti?
 - quali sono i prodotti meno apprezzati dagli utenti?
 - quali sono le cause dell'eventuale divergenza tra valutazione e sentimento degli utenti?
 - come varia il sentimento espresso dagli utenti nelle recensioni rispetto al prezzo?
 - il sentimento espresso tramite il sommario diverge da quello espresso tramite il testo della recensione?
- **quesiti sugli argomenti trattati nelle recensioni:**
 - quali sono gli argomenti più frequenti nelle recensioni?
 - quali sono gli aspetti dei prodotti che gli utenti apprezzano di più?

- quali sono gli aspetti dei prodotti che gli utenti apprezzano di meno?
- qual è il sentimento degli utenti rispetto al prezzo dei prodotti?

In particolare le analisi svolte per rispondere a questi quesiti possono essere riassunte come segue: (i) per i quesiti esplorativi è stata svolta un'**analisi esplorativa** del *dataset*; (ii) per i quesiti sul sentimento degli utenti è stata svolta una ***sentiment analysis***; e (iii) per gli argomenti delle recensioni è stata svolta un'***aspect-based sentiment analysis***.

L'esposizione del presente lavoro seguirà questo ordine. Anzitutto verrà presentata l'analisi esplorativa svolta. Poi verranno descritte le tecniche utilizzate per la *sentiment analysis* e i risultati ottenuti. Successivamente verranno illustrate le analisi svolte sul sentimento espresso dagli utenti rispetto agli argomenti discussi nelle recensioni. Dopodiché verrà presentata brevemente la *dashboard* implementata per la visualizzazione dei risultati delle analisi. Infine verranno rassegnate le conclusioni di questo lavoro.

2 Analisi esplorativa

In questo capitolo verrà descritta l’analisi esplorativa svolta. Anzitutto verrà presentato il *dataset* oggetto di analisi. Dopodiché verranno illustrati i risultati dell’analisi esplorativa con questo ordine: (i) prima l’analisi sulle recensioni; (ii) poi l’analisi sui prodotti; (iii) successivamente l’analisi sui prezzi; e (iv) infine l’analisi sulle parole delle recensioni.

2.1 Dataset

Il *dataset* oggetto di studio riguarda recensioni di cibi e bevande pubblicate sul sito di Amazon. Le recensioni sono scritte in lingua inglese. Questo *dataset* fa parte di una più ampia collezione di *dataset* di recensioni pubblicata nel 2013 [7] e aggiornata nel 2019 [8]. Questa collezione è disponibile al seguente link: <https://nijianmo.github.io/amazon/index.html>.

In particolare per questo lavoro è stato utilizzato il *dataset* “Grocery and Gourmet Food” nella versione 5-*core*. In questa versione ogni utente e ogni prodotto presenta almeno 5 recensioni (rispettivamente come autore e oggetto). La scelta di questa versione è stata fatta in quanto l’analisi della versione completa avrebbe richiesto tempi computazionali eccessivamente onerosi.

Il *dataset* utilizzato è a sua volta composto da due parti. Una prima parte rappresenta il vero e proprio *dataset* di recensioni. La seconda parte invece contiene informazioni di dettaglio sui prodotti recensiti. In particolare il *dataset* delle recensioni presenta 12 attributi descritti nella Tabella 1. Il *dataset* contenente le informazioni di dettaglio sui prodotti presenta invece gli attributi descritti nella Tabella 2.

2.2 Preprocessing

2.2.1 Preprocessing del dataset

Già si è detto che il *dataset* utilizzato si divide in (i) *dataset* delle recensioni e (ii) *dataset* dei prodotti. Su questi *dataset* sono state svolte alcune operazioni di *preprocessing*. In particolare:

- il *dataset* delle recensioni e quello dei prodotti sono stati fusi in base al codice ASIN dei prodotti¹;

¹In altri termini i dati di ogni specifico prodotto sono stati aggiunti alle recensioni relative a quel prodotto.

Tabella 1: Descrizione degli attributi delle recensioni

Attributo	Descrizione
asin	codice alfanumerico identificativo del prodotto
reviewerID	codice alfanumerico identificativo del recensore
reviewerName	nominativo alias del recensore
reviewTime	data della recensione con giorno, mese e anno
unixReviewTime	data della recensione in formato unix
reviewText	testo della recensione
summary	sommario della recensione
overall	valutazione del prodotto come numero intero tra 1 e 5
vote	numero di valutazioni positive ricevute dalla recensione
verified	acquisto verificato
style	eventuali dettagli sul prodotto (es. colore, quantità)
image	eventuali URL di immagini della recensione

Tabella 2: Descrizione degli attributi dei prodotti

Attributo	Descrizione
asin	codice alfanumerico identificativo del prodotto
title	titolo del prodotto
description	descrizione del prodotto
price	prezzo del prodotto in USD
imageURL	eventuali URL di immagini del prodotto
imageURLHighRes	eventuali URL di immagini del prodotto ad alta risoluzione
brand	marca del prodotto
category	categorie del prodotto (es. “Grocery and Gourmet Food”, “Jams”)
main_cat	categoria principale del prodotto
rank	rango del prodotto nella classifica delle vendite per categoria
date	data di inserimento dei dati con giorno, mese, anno
details	dettagli del prodotto (es. peso)
similar_item	elenco di prodotti simili
also_view	elenco di prodotti visualizzati insieme a quello corrente
also_buy	elenco di prodotti acquistati insieme a quello corrente
feature	dettagli ulteriori del prodotto (es. specie ed età di piante)
tech1	primo dettaglio tecnico sul <i>dataframe</i> dei prodotti
tech2	secondo dettaglio tecnico sul <i>dataframe</i> dei prodotti
fit	terzo dettaglio tecnico sul <i>dataframe</i> dei prodotti

- sono state rimosse le recensioni duplicate;
- sono state rimosse le recensioni con testo o sommario vuoti;
- sono state rimosse le recensioni per cui i relativi prodotti non hanno un titolo, una descrizione e un prezzo;
- sono stati aggiunti gli attributi relativi alla lunghezza delle recensioni e dei sommari;
- si è convertito il nome dell'attributo *overall* in *rating* in quanto più rappresentativo del suo contenuto;
- è stato aggiunto un attributo *opinion* che rappresenta l'opinione degli utenti espressa tramite il *rating*². Più precisamente:
 - un *rating* maggiore o uguale a 4 corrisponde a un'opinione positiva;
 - un *rating* pari a 3 corrisponde a un'opinione neutrale;
 - un *rating* inferiore a 3 corrisponde a un'opinione negativa³.

Infine per ogni recensione del *dataset* sono stati conservati i seguenti attributi: *asin*, *reviewerID*, *reviewerName*, *reviewTime*, *unixReviewTime*, *reviewText*, *summary*, *reviewLength*, *summaryLength*, *rating*, *opinion*, *vote*, *verified*, *style*, *image*, *imageURLHighRes*, *title*, *description*, *price*⁴.

²In questo lavoro si userà il termine “opinione” per indicare il sentimento espresso esplicitamente dagli utenti attraverso il *rating*. Si userà invece il termine “sentimento” per indicare il sentimento degli utenti inferito attraverso la *sentiment analysis* e l'*aspect-based sentiment analysis*.

³Questa associazione tra *rating* e opinioni è stata determinata sulla base di una scala Likert a 5 valori. Infatti per questa scala (*i*) i valori 1 e 2 corrispondono a un'opinione negativa; (*ii*) i valori 4 e 5 corrispondono a un'opinione positiva; e (*iii*) il valore 3 corrisponde a un'opinione che non è né negativa né positiva (e quindi è neutrale).

⁴Il dato relativo al *rank* di vendita dei prodotti potrebbe essere utile per inferire la soddisfazione degli utenti. Tuttavia si è scelto di non considerare questo dato per diversi motivi. Il primo motivo è che questo dato non è presente per tutti i prodotti. Il secondo motivo è che anche quando per un prodotto viene specificato il *rank*, questo non si riferisce sempre alla stessa categoria merceologica. In questo quadro la mancanza dei dati non permette di ricostruire correttamente un *ranking* dei prodotti. E conseguentemente non è possibile inferire una relazione d'ordine tra i prodotti che possa essere significativa nell'inferenza della soddisfazione degli utenti.

A seguito del *preprocessing* il *dataset* presenta i dati mostrati nella Tabella 3.

Tabella 3: Descrizione del *dataset* dopo il *preprocessing*

Recensioni	Prodotti	Utenti	Acquisti verificati (%)	Voto medio
752154	25218	126485	88.516	4.462

2.2.2 Preprocessing del testo

In questa fase di analisi esplorativa sono state svolte alcune operazioni di *preprocessing* anche sul testo delle recensioni e dei relativi sommari. Questo *preprocessing* presenta diverse analogie con il *preprocessing* svolto per la *sentiment analysis* e per l'*aspect-based sentiment analysis*. Se ne anticipa qui la descrizione in quanto rilevante anche per l'analisi sulle parole delle recensioni che verrà mostrata nel paragrafo 2.6.

Il *preprocessing* svolto sul testo delle recensioni e dei sommari si articola nei seguenti passi:

1. **Espansione delle contrazioni.** Anzitutto si è proceduto a espandere le contrazioni. Per esempio l'espressione “don't like” è stata convertita in “do not like”. In questo modo (i) si è ridotto il vocabolario di parole utilizzate nelle recensioni e (ii) si è reso più uniforme il testo delle recensioni. Si ritiene di agevolare così gli algoritmi utilizzati per la *sentiment analysis*. Per questa operazione è stata usata la seguente libreria: <https://github.com/kootenpv/contractions>.
2. **Tokenizzazione.** Dopodiché si è proceduto a tokenizzare il testo. In questo modo sono state estratte le singole parole di ogni testo. Per questa attività è stata utilizzata la libreria NLTK [1].
3. **Conversione in minuscolo.** Successivamente ogni token/parola è stato convertito in lettere minuscole. Anche questa attività è stata svolta per ridurre la dimensione del vocabolario.
4. **Semplificazione delle ripetizioni.** Ancora, le parole con molti caratteri ripetuti sono state ridimensionate riducendo le ripetizioni di caratteri a un massimo di due ripetizioni⁵. Infatti nella lingua inglese

⁵Per esempio la parola “tooooo” viene ridotta a “too”.

non esistono parole che presentino lo stesso carattere ripetuto per tre volte di fila.

5. **Rimozione della punteggiatura.** Ancora, è stata rimossa la punteggiatura in quanto non rilevante ai fini dell'analisi esplorativa. La gestione della punteggiatura per la *sentiment analysis* verrà precisata in seguito.
6. **Rimozione dei numeri e dei caratteri singoli.** Per lo stesso motivo della rimozione della punteggiatura sono stati rimossi anche i numeri e i caratteri singoli.
7. **Gestione delle negazioni.** Le negazioni non sono state considerate rilevanti per l'analisi esplorativa del *dataset*. Infatti le negazioni rappresentano termini frequenti che specificano (non tanto il contenuto del testo, quanto) il sentimento degli utenti. In questo quadro le negazioni sono state rimosse dai testi considerati. Assumeranno invece rilevanza per la *sentiment analysis*.
8. **Ulteriori rimozioni.** Infine sono state rimosse le *stop words* di NLTK, i *tag* HTML e gli URL. Si tratta di elementi tipicamente irrilevanti per l'analisi esplorativa. La gestione delle *stop words* per la *sentiment analysis* verrà precisata in seguito.

2.3 Analisi sulle recensioni

In questa sezione si cercherà di rispondere al seguente quesito:

- qual è il trend temporale di soddisfazione degli utenti?

In particolare per rispondere a questa domanda (i) si svolgeranno alcune analisi preliminari sulle valutazioni degli utenti e (ii) poi si concentrerà l'attenzione sull'aspetto temporale.

2.3.1 Distribuzione delle valutazioni

La distribuzione dei *rating* è mostrata in Figura 1. Questa distribuzione mostra un chiaro sbilanciamento. Il *rating* nettamente più frequente è il 5. La sua frequenza è pari a circa il 73% del totale dei *rating*. Il secondo *rating*

per frequenza è il 4, ma ricorre con meno di un quinto della frequenza del *rating* 5. Gli altri *rating* risultano ancora meno frequenti.

La distribuzione delle opinioni segue naturalmente quella delle valutazioni ed è mostrata in Figura 2. L'opinione meno rappresentata è quella neutrale.

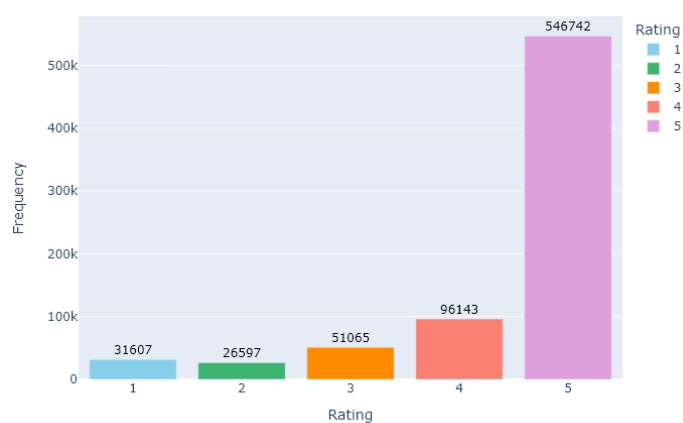


Figura 1: Distribuzione delle valutazioni degli utenti.

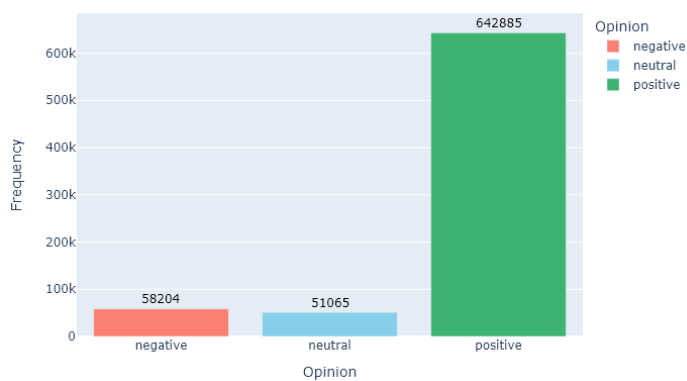


Figura 2: Distribuzione delle opinioni degli utenti.

2.3.2 Distribuzione delle recensioni per anno

La distribuzione del numero di recensioni per anno è mostrata in Figura 3.

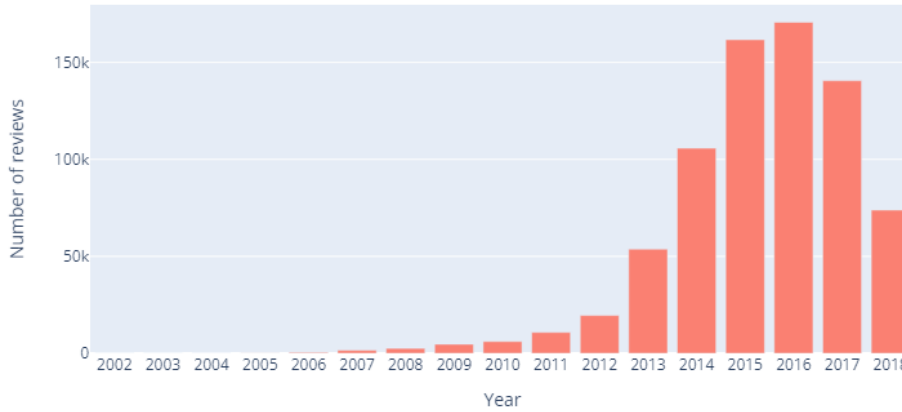


Figura 3: Distribuzione delle recensioni per anno.

La maggior parte delle recensioni ricade tra il 2013 e il 2018. Tuttavia si precisa che questa distribuzione riguarda semplicemente il numero di recensioni presenti nel *dataset* e non riflette necessariamente il reale numero di recensioni pubblicate sul sito di Amazon. In ogni caso l'alto numero di recensioni per gli anni successivi al 2012 consente di ritenere statisticamente più affidabili le analisi svolte per questi anni.

2.3.3 Distribuzione percentuale delle opinioni per anno

La distribuzione delle opinioni (in percentuale) per anno è mostrata in Figura 4. In particolare si nota come per il 2002 e per il 2003 vi siano solo recensioni positive. Nel 2004 compaiono anche recensioni neutrali e negative, con le prime che risultano maggioritarie rispetto alle seconde. Negli anni successivi invece cresce il numero di recensioni negative. E come si è anticipato al paragrafo precedente, i valori per gli ultimi anni risultano più affidabili in ragione del maggior numero di recensioni. In ogni caso le recensioni positive sono le più frequenti in ogni anno. Quindi lo sbilanciamento per le opinioni è costante in tutti gli anni del *dataset*.

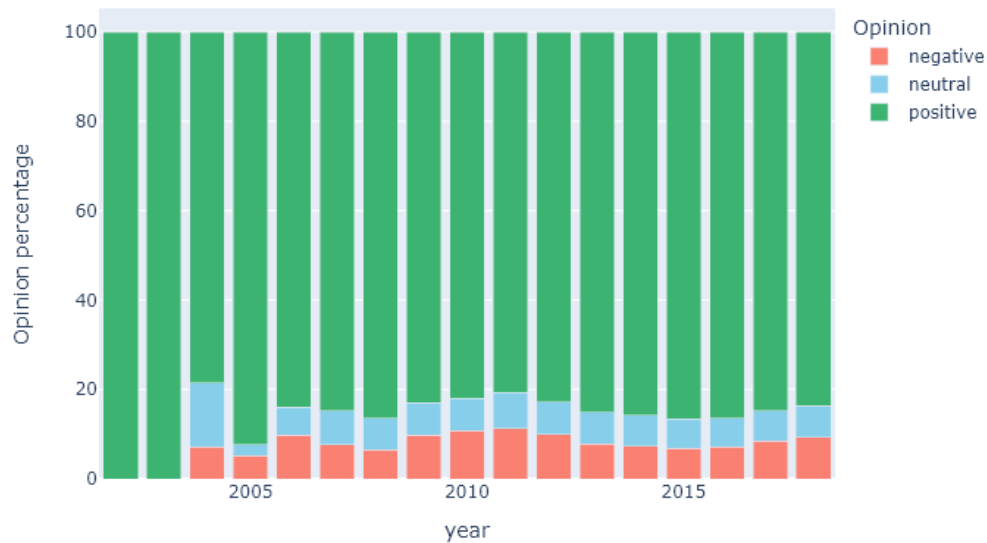


Figura 4: Distribuzione percentuale delle opinioni per anno.

2.3.4 Sintesi

La soddisfazione degli utenti in questa fase esplorativa è stata valutata utilizzando i *rating*. In particolare si è riscontrata una maggior frequenza di recensioni positive in ogni anno. L'opinione neutrale è quella meno rappresentata.

2.4 Analisi sui prodotti

In questa sezione si cercherà di rispondere ai seguenti quesiti:

- quali sono i prodotti con la valutazione media più alta?
- quali sono i prodotti con la valutazione media più bassa?

In particolare per rispondere a queste domande si procederà (i) a calcolare il voto medio dei prodotti e (ii) a individuare una prima descrizione dei prodotti migliori e peggiori.

2.4.1 Distribuzione del voto medio per prodotto

Nella Figura 5 viene mostrata la distribuzione del voto medio per prodotto. Questa distribuzione è stata calcolata in questo modo:

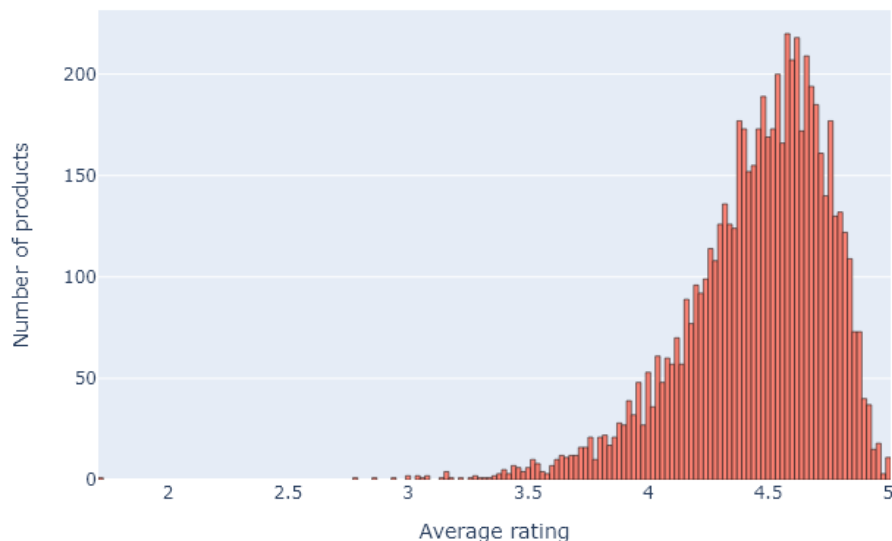


Figura 5: Distribuzione del voto medio per prodotto.

1. Anzitutto si è proceduto a calcolare il numero di recensioni per ogni prodotto.
2. Poi sono stati calcolati i quartili relativi al numero di recensioni. Il secondo quartile corrisponde a 11 recensioni, mentre il terzo quartile corrisponde a 24 recensioni.
3. Successivamente si sono considerati solo i prodotti il cui numero di recensioni fosse maggiore o uguale al terzo quartile appena specificato. Questa scelta è stata fatta al fine di ottenere risultati statisticamente più robusti⁶. Infatti se un prodotto riceve una sola recensione con voto

⁶Si è preferito il terzo quartile al secondo quartile in quanto si è ritenuto che 11 recensioni fossero troppo poche per ottenere risultati statisticamente robusti. In ogni caso i prodotti con un numero di recensioni maggiore o uguale al terzo quartile sono pari a 6496. Si tratta di un numero non piccolo di prodotti giudicato sufficiente per effettuare valutazioni significative.

pari a 5, si potrebbe essere portati a ritenere che il prodotto sia molto buono. Tuttavia questa valutazione positiva potrebbe essere il frutto di un gusto personale ed eclettico del recensore. E quindi potrebbe non riflettere il sentimento generale della comunità rispetto a quel prodotto.

4. Dopodiché si è calcolato il voto medio per ogni prodotto.
5. Infine si è rappresentata graficamente la distribuzione dei voti medi dei prodotti.

Ora, anche il grafico della Figura 5 conferma come la maggior parte dei prodotti abbia un voto alto e più precisamente maggiore di 4. Inoltre si nota come il picco di *rating* medio si aggiri attorno al 4.5. Sono invece pochi i prodotti che hanno raggiunto mediamente il voto massimo pari a 5. E parimenti sono pochi i prodotti che hanno ottenuto mediamente un voto prossimo a 1.

2.4.2 Prodotti con il voto medio più alto

Nel paragrafo precedente si è visto che vi sono pochi prodotti con il massimo voto medio. In particolare i 20 prodotti con il voto medio più alto sono mostrati nella Figura 6. I primi 11 prodotti hanno un voto medio pari a 5. I successivi prodotti hanno un voto medio leggermente inferiore, ma comunque superiore a 4.96.

Nella Tabella 4 vengono riassunti i dettagli di questi prodotti. Le descrizioni dei prodotti migliori suggeriscono che il tè rappresenta un prodotto apprezzato dagli utenti. Infatti 4 prodotti su 20 sono correlati al tè. Inoltre tra i prodotti migliori si trovano anche alimenti correlati al caffè e in generale ai succhi. In questo quadro sembra che i prodotti migliori siano prevalentemente bevande. Tra i prodotti solidi risalta invece la presenza del sale.

2.4.3 Prodotti con il voto medio più basso

I 20 prodotti con il voto medio più basso sono mostrati nella Figura 7. In questo caso il prodotto peggiore ha un voto medio pari a 1.71. Il secondo prodotto peggiore ha un voto medio già più alto e più precisamente pari a 2.79. Il ventesimo prodotto peggiore ha un voto medio pari a 3.28. Quindi i prodotti peggiori hanno complessivamente un voto medio che rispecchia un'opinione

Tabella 4: Descrizione dei prodotti migliori per voto medio

ASIN	Descrizione	Voto medio
B001LO1BF2	succo di pompelmo	5.000
B000NMI7VO	caramelle	5.000
B00JVAGU40	cioccolatini (M&M's)	5.000
B0083COLGG	tè	5.000
B00EYFKM32	acido citrico	5.000
B00FIWD5C0	tè	5.000
B000F013RE	origano	5.000
B004XJXIM6	sale	5.000
B00DMTAV8M	pretzel	5.000
B007L3NVKU	caffè	5.000
B004M8DPQW	condimenti naturali	5.000
B00IGFRY1Y	condimenti	4.979
B009A07K8G	limonata	4.971
B00Y2S0LLO	sale	4.970
B00GKL51PW	tè	4.969
B00FPZCRJW	semi di chia	4.969
B0046CERZ2	riso	4.967
B01FPZ1O0Y	cannella per tè	4.966
B003KRNHZA	caffè	4.963
B005MH0P5Q	curry	4.963

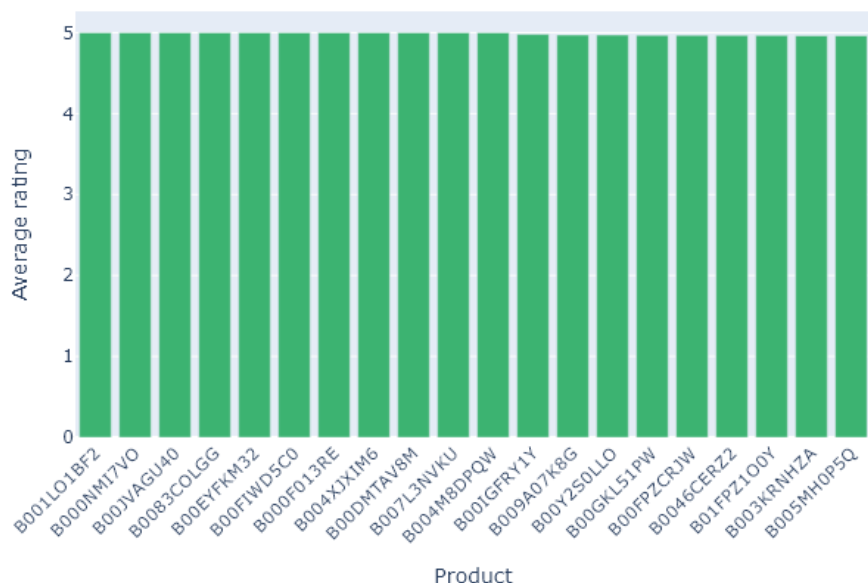


Figura 6: Migliori 20 prodotti per voto medio.

più neutrale che negativa degli utenti. E questo conferma ulteriormente il sentimento generalmente favorevole degli utenti rispetto ai prodotti recensiti.

Nella Tabella 5 vengono riassunti i dettagli di questi prodotti. Queste descrizioni dei prodotti peggiori rivelano una marcata presenza del burro di arachidi. Inoltre si nota la presenza di due prodotti correlati al caffè.

2.4.4 Sintesi

La maggior parte dei prodotti ha registrato un voto medio prossimo a 4.5. I prodotti con il voto medio più alto hanno un voto prossimo a 5. Si tratta principalmente di bevande. Tra questi prodotti spicca il tè. Inoltre si nota la presenza del sale. I prodotti con il voto medio più basso presentano invece un voto più rispondente a un sentimento neutrale che negativo. Tra i prodotti peggiori risalta la presenza del burro di arachidi.

2.5 Analisi sui prezzi

In questa sezione si cercherà di rispondere ai quesiti relativi ai prezzi dei prodotti recensiti. Più precisamente:

Tabella 5: Descrizione dei prodotti peggiori per voto medio

ASIN	Descrizione	Voto medio
B000X1Q1G8	burro di arachidi	1.711
B01H6JO1EI	bevanda energetica	2.789
B0140324OW	cioccolatini (M&M's)	2.857
B003SPO3FC	riso	2.938
B00CGBKIIK	semi di papavero	3.000
B006VXWBTS	burro di arachidi	3.000
B013OV0PTM	primo piatto thailandese	3.032
B00YWL28JO	bustine di avena	3.038
B00N28FBYC	barretta proteica	3.051
B001EJN9KE	brownie con burro di arachidi	3.071
B002O19HWU	pesche affettate	3.073
B0032CJ40U	bevanda di mandorle	3.147
B00FBLKNW8	sciropo di zucca	3.167
B00379SYDQ	cocktail di cetrioli e menta	3.167
B00LKUHCX2	caffè	3.167
B01EB4BM20	panna montata	3.170
B00KO9SH9S	riso	3.185
B005OL1Q8K	caramelle	3.229
B00QX9D36A	biscotti senza glutine	3.264
B00RMX73G2	crema per caffè	3.283

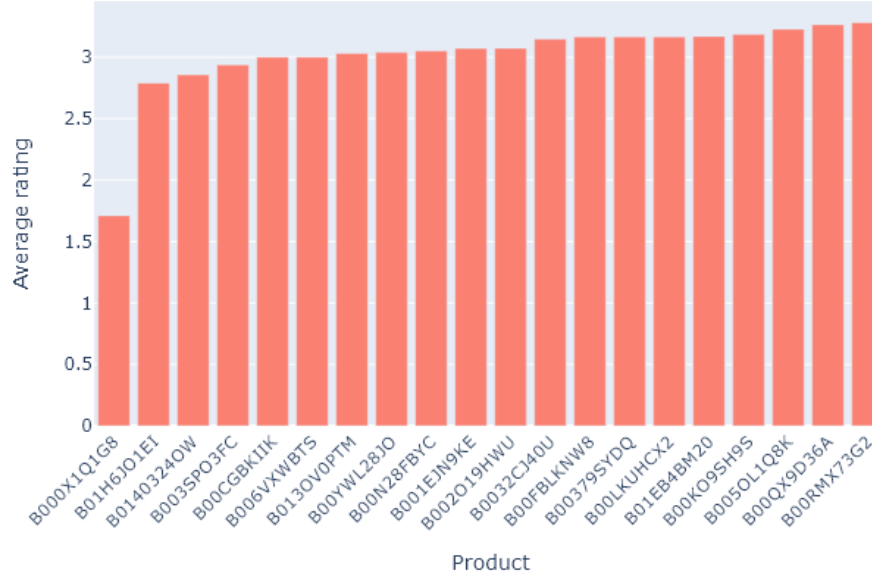


Figura 7: Peggiori 20 prodotti per voto medio.

- quali sono le fasce di prezzo più recensite?
- come varia la valutazione dei prodotti in base al prezzo?

2.5.1 Distribuzione delle recensioni per prezzo

La distribuzione delle recensioni per prezzo è mostrata in Figura 8. In questo caso si è proceduto a discretizzare i prezzi in intervalli per una maggiore chiarezza espositiva. In particolare dal grafico emerge che la maggior parte delle recensioni riguarda prodotti alimentari con un prezzo compreso fra i 12 e i 18 dollari. Le altre fasce di prezzo più popolari sono quelle immediatamente precedenti e successive. I prodotti con un prezzo superiore a 50 dollari sono invece poco recensiti.

Questi dati sono coerenti con il fatto che (i) il prezzo medio che emerge dalle recensioni è di 20.85 dollari e che (ii) il terzo quartile corrisponde a 26 dollari.

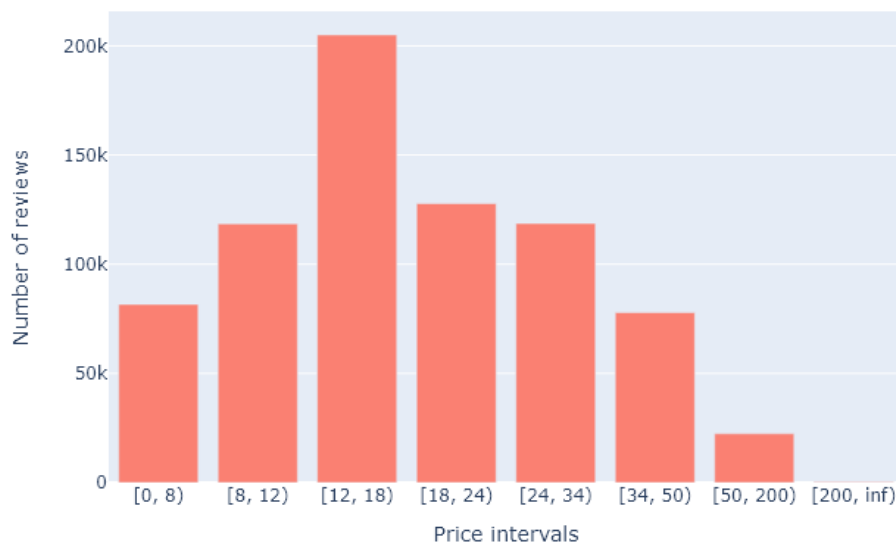


Figura 8: Distribuzione delle recensioni per prezzo.

2.5.2 Correlazione tra valutazioni e prezzi

La correlazione tra le valutazioni degli utenti e i prezzi dei prodotti è stata esaminata sia quantitativamente che qualitativamente.

Per la valutazione quantitativa si è proceduto a calcolare il coefficiente di correlazione di Pearson tra i *rating* e i prezzi. Questo coefficiente ha un valore compreso fra -1 e 1 e misura la correlazione lineare tra due variabili. In questo caso questo coefficiente ha un valore pari a 0.012. Questo significa che le variabili considerate sono poco correlate.

Per confermare questa valutazione quantitativa si è proceduto a effettuare anche una valutazione qualitativa. A questo proposito si mostra in Figura 9 lo *scatter plot* della relazione tra i *rating* e i prezzi. Questo grafico conferma come non sia possibile individuare una relazione lineare tra le due variabili. Infatti i prezzi dei prodotti sembrano equamente distribuiti per ogni *rating*. E questo significa per esempio che vi possono essere sia prodotti costosi apprezzati dagli utenti quanto prodotti costosi non graditi agli utenti.

In questo quadro si può concludere che il prezzo dei prodotti non determina da solo e in modo lineare il gradimento da parte degli utenti.

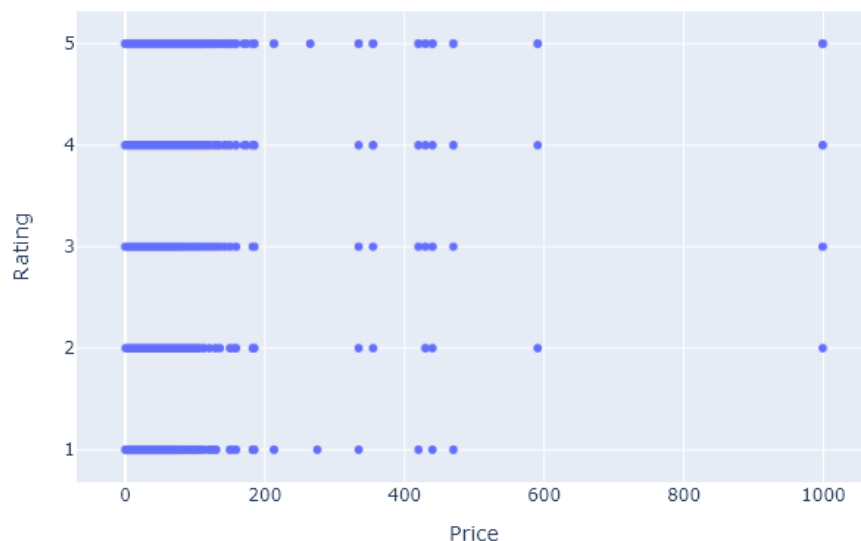


Figura 9: *Scatter plot* della relazione tra *rating* e prezzi.

2.5.3 Sintesi

La fascia di prezzo con più recensioni è compresa fra i 12 e i 18 dollari. Il prezzo medio dei prodotti recensiti è di 20.85 dollari. Non si è riscontrata una correlazione significativa tra il prezzo dei prodotti e il gradimento degli utenti.

2.6 Analisi sulle parole delle recensioni

In questa sezione si cercherà di rispondere alla seguente domanda:

- quali sono le parole più frequenti nelle recensioni degli utenti?

In primo luogo si individueranno le parole più frequenti nelle recensioni. In secondo luogo si analizzerà come queste parole sono distribuite rispetto alle valutazioni degli utenti.

2.6.1 Individuazione delle parole più frequenti

L'analisi *de quo* riguarda solo il testo (e non il sommario) delle recensioni. Il testo utilizzato è stato processato seguendo i passi descritti al paragrafo

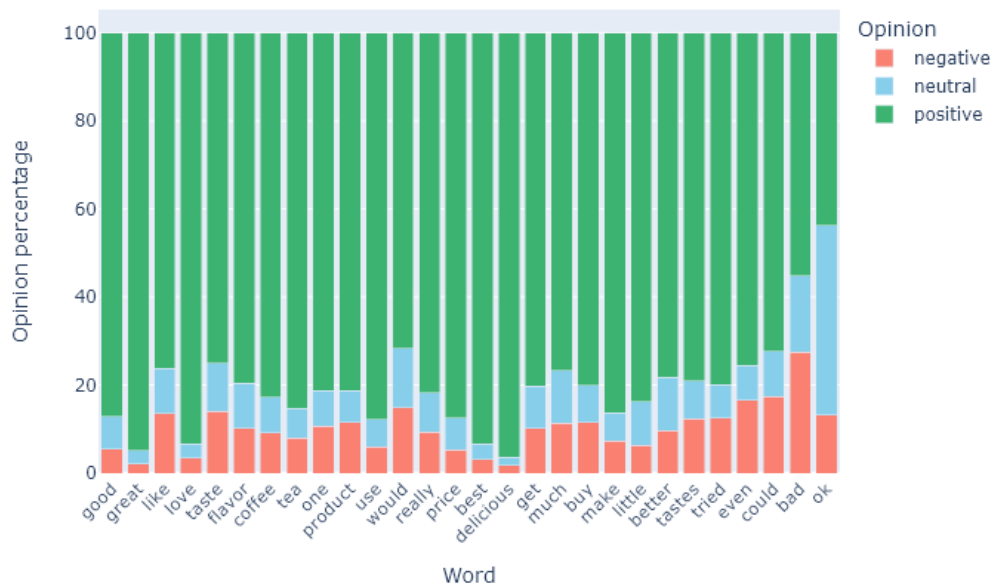


Figura 11: Distribuzione delle opinioni sulle recensioni per le parole più frequenti.

4. poi si è calcolata la percentuale delle opinioni;
5. infine si è rappresentata graficamente la distribuzione delle opinioni per ciascuna parola.

Il risultato è mostrato nella Figura 11. Da questo grafico si possono trarre le seguenti considerazioni:

- Anzitutto l'opinione positiva risulta ancora dominante per la maggior parte delle recensioni che contengono le parole più frequenti. In particolare le recensioni che contengono la parola “delicious” sono quasi sempre positive.
- Inoltre la parola “bad” è spesso associata a recensioni negative, come ci si può banalmente attendere. Tuttavia questa parola è frequente anche in recensioni positive. E questo potrebbe suggerire (*i*) un'ambiguità del contenuto delle recensioni e (*ii*) la descrizione di aspetti diversi con sentimenti diversi nella stessa recensione.

- Infine si nota come la parola “ok” sia associata a recensioni neutrali in quasi la metà dei casi. E questa percentuale è particolarmente significativa se si considera che le recensioni positive sono nettamente più frequenti rispetto a quelle neutrali, come mostrato al paragrafo 2.3.1.

Naturalmente l’analisi svolta in questo paragrafo è solo preliminare. Una disamina più approfondita dei sentimenti espressi dagli utenti rispetto ai singoli aspetti delle recensioni verrà condotta nel capitolo 4.

2.6.3 Sintesi

Le parole più frequenti nelle recensioni riguardano principalmente il giudizio degli utenti sul prodotto. Inoltre sono frequenti anche i richiami ai prodotti del tè e del caffè. E non mancano i riferimenti al prezzo dei prodotti.

Le recensioni che contengono le parole più frequenti sono generalmente positive. Anche parole tipicamente negative sono presenti nelle recensioni positive, come per esempio la parola “bad”. E questo suggerisce un’ambiguità nelle opinioni espresse nei testi delle recensioni. La parola “ok” invece sembra associata in modo significativo alle recensioni neutrali.

3 Sentiment analysis

Sin qui è stata descritta l’analisi esplorativa svolta per rispondere ad alcuni dei quesiti posti. Per rispondere al secondo gruppo di quesiti identificati nel capitolo 1 è stata condotta una *sentiment analysis*.

I risultati verranno presentati con questo ordine. In primo luogo verranno presentati gli approcci di *sentiment analysis* basati sui lessici. In secondo luogo verrà esposto un approccio di *sentiment analysis* basato sul Machine Learning. In terzo luogo verranno individuate le risposte ai quesiti utilizzando l’approccio migliore.

3.1 Approcci basati su lessici

Per la *sentiment analysis* sono stati sperimentati due approcci basati sui lessici. Il primo utilizza AFINN [9]. Il secondo utilizza VADER [3]. Si procederà ora a descrivere separatamente questi approcci.

3.1.1 AFINN

Introduzione. AFINN è uno strumento per la *sentiment analysis* che sfrutta un lessico [9]. Questo significa che specifica una valenza di sentimento per ogni parola di un lessico determinato. Questi valori sono numeri interi compresi fra -5 e +5. Il valore -5 corrisponde al sentimento più negativo, mentre il valore +5 corrisponde al sentimento più positivo.

Il lessico di AFINN comprende parole inglesi relative al contesto dei microblog. Quindi può essere utilizzato anche per le recensioni scritte dagli utenti su un sito come quello di Amazon. Il lessico di AFINN non è stato oggetto di lemmatizzazione né di *stemming*.

AFINN è disponibile come libreria per il linguaggio di programmazione Python: <https://github.com/fnielsen/afinn>. Per gli esperimenti condotti con AFINN è stata utilizzata questa libreria.

Preprocessing del testo. Per il *preprocessing* del testo sono state seguite le attività descritte al paragrafo 2.2.2 per l’analisi esplorativa. Tuttavia rispetto a queste attività è stata operata una modifica in tema di gestione delle negazioni. In particolare non sono state rimosse le seguenti negazioni: “not”, “nor”, “no”, “never”, “against”. Questa scelta è stata fatta in quanto

le negazioni esprimono un sentimento⁷. Pertanto si è ritenuto opportuno conservarle per qualificare il sentimento espresso dagli utenti rispetto ai prodotti recensiti.

Gestione delle negazioni. Le negazioni non sono state rimosse dalle recensioni e dai relativi sommari. Tuttavia AFINN non assegna una valenza a queste negazioni. Pertanto AFINN non coglie l'inversione di polarità del sentimento degli utenti dovuta alle negazioni.

In questo quadro si è proceduto a gestire esplicitamente l'inversione di questa polarità. In particolare si è invertito il segno della valenza di ogni parola immediatamente successiva a una negazione.

Strategia di classificazione del sentimento. Le recensioni e i sommari sono stati classificati sulla base dell'opinione espressa dagli utenti: negativa, neutrale o positiva. Per classificare il sentimento delle recensioni e dei sommari si è proceduto in questo modo. Già si è detto che AFINN assegna una valenza a ogni parola. Pertanto è possibile calcolare la valenza complessiva di un testo come somma delle valenze delle sue parole. In questo quadro il sentimento di un testo è stato calcolato come segue: (i) il sentimento è stato considerato negativo quando la somma delle valenze era negativa; (ii) il sentimento è stato considerato positivo quando la somma delle valenze era positiva; e (iii) il sentimento è stato considerato neutrale quando la somma delle valenze era nulla.

3.1.2 VADER

Introduzione. Anche VADER è uno strumento per la *sentiment analysis* che sfrutta un lessico [3]. A differenza di AFINN, VADER calcola la valenza considerando anche (i) la punteggiatura; (ii) la presenza di lettere maiuscole; (iii) avverbi; (iv) congiunzioni; e (v) negazioni.

Il lessico di VADER comprende parole inglesi relative a *social media*. Queste parole sono applicabili anche al contesto delle recensioni *online*. Infatti gli stessi autori di VADER hanno valutato l'efficacia di questo strumento su recensioni pubblicate sul sito di Amazon. Anche il lessico di VADER non è stato oggetto di lemmatizzazione né di *stemming*.

⁷Per esempio l'espressione "not good" esprime un sentimento opposto rispetto al semplice "good".

VADER è disponibile come libreria per il linguaggio di programmazione Python: <https://github.com/cjhutto/vaderSentiment>. Per gli esperimenti condotti con VADER è stata utilizzata questa libreria.

Preprocessing del testo. Si è detto che VADER calcola il sentimento di un testo considerando una serie di elementi ulteriori rispetto al mero lessico. Pertanto è stato necessario ridurre le operazioni di *preprocessing* del testo in modo da poter sfruttare al meglio le caratteristiche di VADER. In particolare le operazioni di *preprocessing* si sono ridotte a: (i) espansione delle contrazioni; (ii) tokenizzazione; (iii) semplificazione delle ripetizioni; e (iv) rimozione dei *tag* HTML.

Gestione delle negazioni. Le negazioni non sono state rimosse dai testi delle recensioni e dei sommari. In questo modo VADER ha potuto considerare le negazioni per classificare i sentimenti degli utenti. In particolare qui si precisa che VADER inverte la polarità del sentimento del trigramma successivo a una negazione.

Strategia di classificazione del sentimento. La libreria in Python che implementa VADER attribuisce un *compound score* per ogni testo classificato. Questo *compound score* è un numero reale compreso fra -1 e +1. Tipicamente questo valore viene utilizzato per classificare il sentimento come segue: (i) il sentimento è considerato negativo se il *compound score* è minore o uguale a -0.05 ; (ii) il sentimento è considerato positivo se il *compound score* è maggiore o uguale a 0.05 ; e (iii) il sentimento è considerato neutrale altrimenti. Questa strategia di classificazione è suggerita dagli stessi autori di VADER⁸. Ed è stata utilizzata anche per il presente lavoro.

3.2 Approccio basato sul Machine Learning

Oltre agli approcci di *sentiment analysis* basati sui lessici si è utilizzato anche un approccio supervisionato basato sul Machine Learning. In questo paragrafo verrà descritto questo approccio.

⁸A questo proposito si veda quanto scritto in <https://github.com/cjhutto/vaderSentiment> in tema di *compound score*.

3.2.1 Modello di regressione logistica

Introduzione. Il modello di Machine Learning utilizzato per la classificazione supervisionata è un modello di regressione logistica.

Il modello di regressione logistica è stato preferito a Naive Bayes in quanto quest'ultimo assume un bilanciamento delle classi che nel caso di specie non sussiste.

Il modello di regressione logistica è stato preferito anche alle Support Vector Machines. Ora, sia il modello di regressione logistica che le Support Vector Machines nascono per problemi binari. Tuttavia entrambe le tipologie di modelli sono poi state estese anche a problemi multi-classe. In questo caso si è preferito il modello di regressione logistica in quanto computazionalmente meno costoso.

Il costo computazionale ha rappresentato un fattore determinante nella scelta del modello di regressione logistica anche rispetto ad altri modelli come gli alberi decisionali. In questo caso l'elevato numero di recensioni del *dataset* ha in larga parte ristretto le possibilità di scelta del modello di classificazione.

Per il presente lavoro è stata utilizzata l'implementazione offerta da *scikit-learn* [11]. In particolare come algoritmo di ottimizzazione per il modello di regressione logistica è stato scelto l'algoritmo Limited-memory Broyden-Fletcher-Goldfarb-Shanno in quanto efficiente e compatibile anche con problemi multi-classe. La funzione di *loss* utilizzata è invece una *cross-entropy* multinomiale.

Preprocessing del testo. Per il *preprocessing* del testo sono state seguite le operazioni già descritte al paragrafo 2.2.2. Rispetto a queste operazioni è stata modificata la gestione delle negazioni. In questo caso si è seguita la strategia descritta in [10]. In particolare questa strategia prevede di aggiungere il prefisso “NOT_” a ogni parola presente tra una negazione e il primo simbolo di punteggiatura successivo. A questo proposito i termini considerati come negazioni sono stati i seguenti: “not”, “nor”, “no”, “never”, “against”.

Bag of Words. Il modello di regressione logistica è stato addestrato utilizzando Bag of Words come *feature*. In particolare l'approccio Bag of Words rappresenta ogni recensione come un vettore avente una dimensione pari al vocabolario delle parole presenti nel *corpus*. Per ogni parola viene specificata la frequenza nella recensione considerata o nel sommario considerato. Natu-

ralmente questo approccio porta a costruire una matrice sparsa che descrive il *dataset*.

Undersampling. Al paragrafo 2.3.1 si è visto che il *dataset* è sbilanciato rispetto alle opinioni degli utenti. Si è valutata dunque l'opportunità di bilanciare le classi. A questo proposito si è addestrato e testato il modello di regressione logistica (non solo sul *dataset* sbilanciato, ma anche) su un *dataset* bilanciato tramite *random undersampling*. In questo caso le istanze di addestramento (ma non di test) sono state ridotte in modo tale che tutte le classi avessero un numero di istanze pari alla classe con meno istanze. Per l'*undersampling* è stata utilizzata la libreria *imbalanced-learn* [5].

Cross-validation. Per l'addestramento e il test si è adottato un approccio di *cross-validation*. In particolare:

- si è permutato l'ordine delle recensioni in modo casuale;
- si è poi suddiviso il *dataset* in 3 *fold* aventi circa lo stesso numero di istanze. Dato l'elevato numero di recensioni si è ritenuto che la suddivisione in 3 *fold* permettesse un buon compromesso tra efficienza e robustezza dei risultati;
- a ogni passo della *cross-validation* si è considerato un *fold* come *test set* e il resto come *train set*. In questo caso si è considerato che 2 *fold* rappresentanti circa il 67% delle istanze del *dataset* fossero sufficienti come *train set* dato l'elevato numero di recensioni;
- naturalmente le classi considerate per l'addestramento sono state le opinioni degli utenti: negativa, neutrale e positiva.

3.3 Confronto degli approcci di sentiment analysis

I modelli per la *sentiment analysis* descritti nei paragrafi 3.1 e 3.2 sono stati implementati e valutati. In particolare le *performance* di questi modelli sono state valutate con le seguenti metriche: (i) accuratezza; (ii) precisione; (iii) *recall*; e (iv) *f1-score*. Per le ultime tre metriche si è considerato l'approccio di valutazione *macro average* che attribuisce uguale peso a tutte le classi (indipendentemente dalla popolarità). Le metriche sono valutate con l'approccio *one-vs-rest* per i problemi multi-classe. Le *performance* dei modelli

di regressione logistica sono state valutate solo sui *fold* utilizzati come *test set*. Poiché ogni *fold* è stato utilizzato come *test set* esattamente una volta, i risultati ottenuti sono confrontabili con quelli degli approcci basati sui lessici.

I risultati della *sentiment analysis* sul testo delle recensioni sono mostrati nella Tabella 6. I risultati migliori per ogni metrica sono evidenziati in grassetto.

Tabella 6: Confronto delle *performance* dei modelli impiegati per la *sentiment analysis* sul testo delle recensioni

Modello	<i>Undersampling</i>	Accuratezza	Precisione	<i>Recall</i>	<i>F1-score</i>
AFINN	no	0.810	0.514	0.494	0.500
VADER	no	0.823	0.499	0.493	0.496
Reg. log.	no	0.892	0.693	0.582	0.621
Reg. log.	sì	0.789	0.560	0.692	0.591

Questi risultati permettono di svolgere le seguenti considerazioni:

- Le differenze di *performance* tra AFINN e VADER sono poco significative. Infatti si colgono differenze solo a partire dalla seconda cifra decimale.
- Il modello di regressione logistica ha ottenuto *performance* migliori senza *undersampling*⁹. Questo è dovuto al fatto che il *random undersampling* ha portato a una riduzione drastica delle istanze. E questa riduzione ha comportato una perdita di informazione significativa per il modello.
- I modelli di regressione logistica hanno ottenuto *performance* migliori rispetto agli approcci basati sui lessici. In questo quadro per le analisi

⁹Solo per la metrica di *recall* il modello addestrato con *undersampling* ha ottenuto risultati migliori. Il motivo risiede nel fatto che l'*undersampling* ha portato il modello a riconoscere correttamente un maggior numero di istanze relative alle classi di minoranza. E questo giustifica la *recall* più alta in quanto questa metrica misura la corretta identificazione delle istanze *true positive* di una classe. Allo stesso tempo però l'*undersampling* ha portato il modello a commettere più errori di classificazione sulla classe di maggioranza. E questo naturalmente si è tradotto in risultati peggiori su tutte le altre metriche. In ogni caso si è ritenuta più rilevante la metrica *f1-score* rispetto alla *recall* in quanto la prima considera anche la seconda. Infatti la metrica *f1-score* rappresenta una media armonica tra precisione e *recall*.

che verranno presentate nel prosieguo si farà riferimento ai soli risultati ottenuti con i modelli di regressione logistica. E più precisamente si farà riferimento ai risultati ottenuti senza *undersampling* dei dati di addestramento.

3.4 Analisi sulle parole

Prima di procedere con le analisi più attinenti ai quesiti posti occorre soffermarsi brevemente sulle parole utilizzate nelle recensioni. In particolare al paragrafo 2.6.2 si è mostrata la distribuzione delle opinioni rispetto alle recensioni che contengono le parole più frequenti. Ora è opportuno analizzare la stessa distribuzione riguardante i sentimenti predetti. Il procedimento di calcolo dei sentimenti è stato analogo a quello seguito per le opinioni. Il risultato è mostrato nella Figura 12.

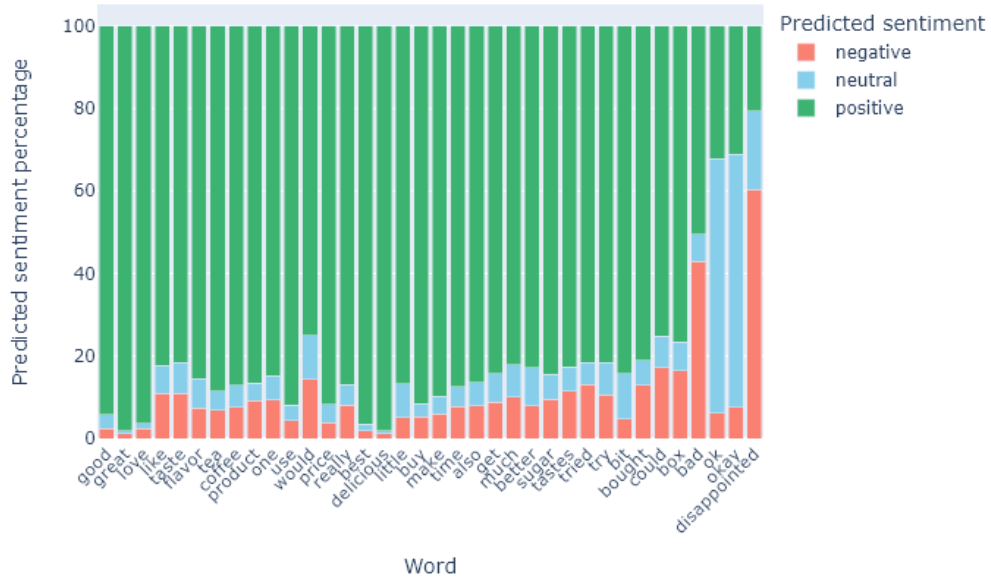


Figura 12: Distribuzione dei sentimenti sulle recensioni per le parole più frequenti.

Da questo grafico si possono trarre le seguenti considerazioni:

- Anzitutto il sentimento positivo risulta ancora dominante per la maggior parte delle recensioni che contengono le parole più frequenti.

- Inoltre le parole “bad” e “disappointed” sono spesso associate al sentimento negativo. Rispetto all’analisi condotta sulle opinioni, il sentimento negativo risulta più forte per la parola “bad”. Permane comunque una percentuale non piccola di sentimento positivo.
- Infine anche in questo caso le parole “ok” e “okay” sono più spesso associate a un sentimento neutrale. Rispetto all’opinione neutrale, l’associazione con il sentimento neutrale è molto più pronunciata.

3.5 Analisi delle misclassificazioni

In questa sezione si proverà a rispondere al seguente quesito:

- quali sono le cause dell’eventuale divergenza tra valutazione e sentimento degli utenti?

Per rispondere a questo quesito si analizzeranno alcuni esempi di errata classificazione dei sentimenti.

3.5.1 Esempi di misclassificazione

Al paragrafo 3.3 si è visto che il miglior modello di *sentiment analysis* implementato ha commesso degli errori di classificazione del sentimento degli utenti. Questo significa che in alcuni casi il modello ha inferito un sentimento dell’utente diverso dalla sua valutazione.

Ora, questa divergenza può avere diverse cause. In parte può essere dovuta alla rappresentazione delle recensioni tramite Bag of Words. In parte può essere dovuta all’inadeguatezza del modello. In parte può essere dovuta all’intrinseca ambiguità del contenuto delle recensioni.

A questo proposito allora si è proceduto a individuare alcuni esempi di misclassificazione al fine di inferirne le possibili cause in modo più preciso. In particolare sono state individuate le seguenti cause di misclassificazione:

- **Sentimenti misti.**

Esempio:

- *Recensione*: “I like the texture and taste of these wraps but they are pricey! They are a bit small when putting things inside so this makes eating a little messy. Otherwise I like them”.

- *Valutazione dell'utente*: 4.
- *Sentimento predetto*: neutrale.

In questo esempio l'utente esprime sentimenti contrastanti. Da un lato apprezza la consistenza e il sapore del prodotto. Dall'altro lato ritiene il prodotto troppo costoso. Questo ha portato il modello a predire un sentimento neutrale. Secondo l'impostazione adottata in questo lavoro la valutazione dell'utente corrisponde invece a un'opinione positiva. Naturalmente occorre anche considerare che non tutti gli utenti utilizzano la stessa scala di valutazione. Quindi un voto pari a 4 può esprimere un'opinione positiva per alcuni e neutrale per altri.

- **Ironia.**

Esempio:

- *Recensione*: “Let me start off by saying... ICK.. Disgusting. These are absolutely terrible. That’s exactly why I rated this product five stars. It was exactly and completely what I was expecting, and thankfully my coworkers were not expecting. It’s a great office game/prank and I will order again soon!”.
- *Valutazione dell'utente*: 5
- *Sentimento predetto*: negativo.

In questo esempio l'utente ha valutato positivamente un prodotto che considera disgustoso. Si tratta dunque di una recensione ironica. Tuttavia il modello addestrato non è riuscito a cogliere questa ironia. Pertanto ha considerato la recensione negativa in ragione della presenza di parole tipicamente associate al sentimento negativo.

- **Sentimenti dipendenti dal contesto.**

Primo esempio:

- *Recensione*: “The nuts are medium in size, not large or ‘premium’, as expected. They do not taste fresh and sweet, seem a bit stale or ‘off’ ”.
- *Valutazione dell'utente*: 3.
- *Sentimento predetto*: positivo.

Secondo esempio:

- *Recensione*: “This is extremely sweet. I can’t even taste the hibiscus its so darn sweet.”.
- *Valutazione dell’utente*: 2.
- *Sentimento predetto*: positivo.

Terzo esempio:

- *Recensione*: “This product has changed. It now has very large tea leaves and tastes a bit bitter. It was my favorite Earl Grey tea of all time. I’m hunting for a new brand. The lid was defective as well. Sigh.....”.
- *Valutazione dell’utente*: 2
- *Sentimento predetto*: neutrale.

Quarto esempio:

- *Recensione*: “This coffee has a very bold taste compared to the original Folgers blend. I prefer this stronger bitter taste and will continue to buy this particular blend.”.
- *Valutazione dell’utente*: 5.
- *Sentimento predetto*: neutrale.

Questi esempi mostrano una problematica legata alla rappresentazione Bag of Words. In particolare questa rappresentazione non coglie la composizionalità delle frasi in quanto assume che ogni parola sia indipendente dalle altre. E non distingue i diversi significati della stessa parola. In questo quadro il modello addestrato su Bag of Words non è riuscito a comprendere che la parola “bitter” può essere positiva per il caffè e negativa per il tè. E allo stesso modo la dolcezza può essere apprezzata in alcuni prodotti e non gradita in altri.

• La parola “ok”.

Esempio:

- *Recensione*: “ok”.
- *Valutazione dell’utente*: 5.

– *Sentimento predetto*: neutrale.

Gli esempi di misclassificazione di recensioni composte dalla sola parola “ok” sono numerosi. In tutti i casi il sentimento predetto è neutrale, mentre la valutazione dell’utente è positiva. Una spiegazione di questo fenomeno è riconducibile alla frequenza della parola “ok” nelle recensioni. In particolare come mostrato in Figura 12 la parola “ok” si trova in recensioni neutrali in più della metà dei casi. Pertanto il modello di classificazione ha considerato questa parola come fattore significativo di un sentimento neutrale. Naturalmente occorre anche considerare che non tutti gli utenti attribuiscono la stessa semantica alla parola “ok”.

3.5.2 Sintesi

In questa sezione si sono indagate le cause della divergenza tra i sentimenti inferiti tramite *sentiment analysis* e le valutazioni degli utenti. Tra le cause individuate vi sono: (i) la presenza di sentimenti contrastanti nelle recensioni; (ii) l’uso di ironia; (iii) l’impiego di parole che esprimono sentimenti diversi in contesti diversi; e (iv) il significato ambiguo della parola “ok”.

3.6 Analisi sui prodotti

In questa sezione si cercherà di rispondere ai seguenti quesiti:

- quali sono i prodotti più apprezzati dagli utenti?
- quali sono i prodotti meno apprezzati dagli utenti?

In particolare per rispondere a queste domande si procederà (i) a calcolare il sentimento medio dei prodotti e (ii) a individuare una prima descrizione dei prodotti migliori e peggiori.

3.6.1 Prodotti con il sentimento medio più alto

Per individuare i prodotti più apprezzati dagli utenti si è proceduto a calcolare il sentimento medio espresso dagli utenti per ciascun prodotto. A questo proposito è stato utilizzato il sentimento inferito tramite *sentiment analysis*. Al fine di calcolare il sentimento medio si è proceduto a convertire i sentimenti in numeri interi: -1 per il sentimento negativo; 0 per il sentimento

neutrale; e 1 per il sentimento positivo. Per questa analisi sono stati considerati solo i prodotti con almeno 24 recensioni per le stesse motivazioni esposte al paragrafo 2.4.1.

I 20 prodotti con il sentimento medio più alto sono mostrati nella Figura 13. In questo caso i primi 20 prodotti hanno tutti un sentimento medio pari a 1. Nella Tabella 7 vengono riassunti i dettagli di questi prodotti.

I prodotti più apprezzati sembrano riguardare principalmente (i) sale, (ii) cioccolato, (iii) tè e (iv) pepe. Qui si ricorda che il sale era già stato individuato come prodotto valutato positivamente dagli utenti nella Tabella 4. Rispetto a questa Tabella le bevande risultano meno presenti. Non si riscontrano corrispondenze tra i codici ASIN dei prodotti delle due Tabelle.

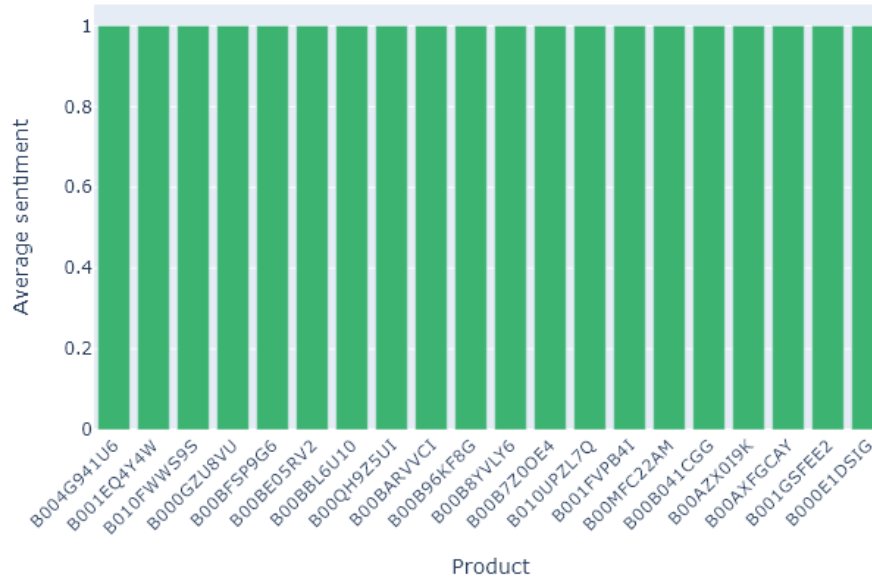


Figura 13: Migliori 20 prodotti per sentimento medio.

3.6.2 Prodotti con il sentimento medio più basso

I 20 prodotti con il sentimento medio più basso sono mostrati nella Figura 14. In questo caso un solo prodotto ha una media negativa. Gli altri prodotti hanno una media compresa fra 0 e 0.3 che esprime un sentimento più neutrale che negativo. Nella Tabella 8 vengono riassunti i dettagli di questi prodotti.

Tabella 7: Descrizione dei prodotti migliori per sentimento medio

ASIN	Descrizione	Sentimento medio
B004G941U6	minestrone	1.000
B001EQ4Y4W	condimento	1.000
B010FWWS9S	cracker	1.000
B000GZU8VU	sale	1.000
B00BFSP9G6	sale	1.000
B00BE05RV2	sale	1.000
B00BBL6U10	tè	1.000
B00QH9Z5UI	cannella	1.000
B00BARVVCI	bevanda energetica	1.000
B00B96KF8G	cioccolato	1.000
B00B8YVLY6	pancake	1.000
B00B7Z0OE4	piante	1.000
B010UPZL7Q	cioccolato	1.000
B001FVPB4I	pepe	1.000
B00MFC22AM	polvere di carruba	1.000
B00B041CGG	pepe	1.000
B00AZX0I9K	cioccolato	1.000
B00AXFGCAY	burro di arachidi e cioccolato	1.000
B001GSFEE2	biscotti	1.000
B000E1DSIG	budino al caramello	1.000

I prodotti meno apprezzati riguardano principalmente il burro di arachidi. Vi sono anche due prodotti legati al caffè. È interessante notare che ben 12 prodotti poco apprezzati per il sentimento medio sono presenti anche nella Tabella 5 relativa ai prodotti con il voto medio peggiore. Questo significa che il modello di *sentiment analysis* adottato riesce a rilevare correttamente i sentimenti fortemente negativi. E per questo risulta coerente con i relativi voti.

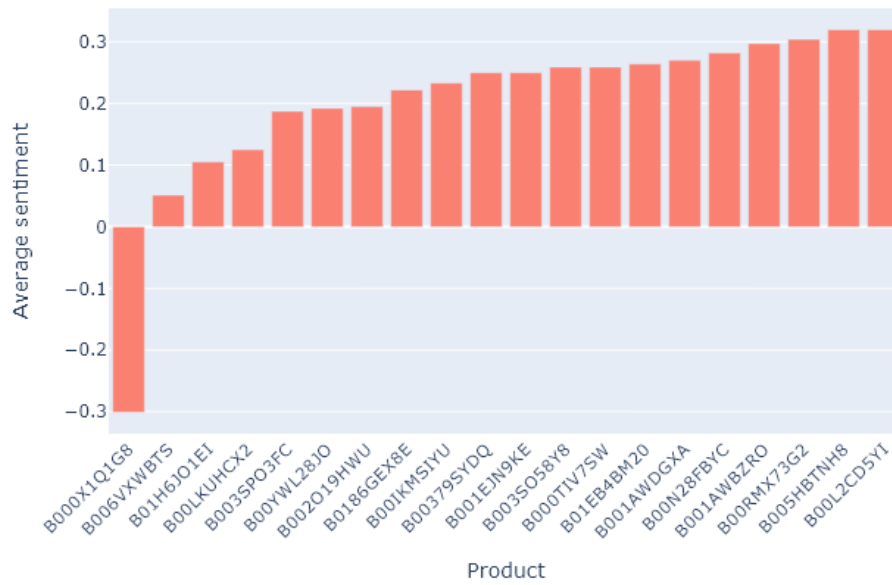


Figura 14: Peggiori 20 prodotti per sentimento medio.

3.6.3 Sintesi

Il sentimento medio dei prodotti più apprezzati è molto alto. In particolare i prodotti più apprezzati dagli utenti riguardano principalmente (i) sale, (ii) cioccolato, (iii) tè e (iii) pepe. Il sentimento medio dei prodotti meno apprezzati è più vicino a un sentimento neutrale che negativo. Tra i prodotti peggiori si nota la presenza del burro di arachidi. I prodotti meno apprezzati corrispondono in buona parte ai prodotti con il voto medio più basso.

Tabella 8: Descrizione dei prodotti peggiori per sentimento medio

ASIN	Descrizione	Sentimento medio
B000X1Q1G8	burro di arachidi	-0.301
B006VXWBTS	burro di arachidi	0.051
B01H6JO1EI	bevanda energetica	0.105
B00LKUHCX2	caffè	0.125
B003SPO3FC	riso	0.188
B00YWL28JO	bustine di avena	0.192
B002O19HWU	pesche affettate	0.195
B0186GEX8E	cereali	0.222
B00IKMSIYU	succo	0.233
B00379SYDQ	cocktail di cetrioli e menta	0.250
B001EJN9KE	brownie con burro di arachidi	0.250
B003SO58Y8	tè	0.259
B000TIV7SW	propoli	0.259
B01EB4BM20	panna montata	0.264
B001AWDGXA	sardine	0.270
B00N28FBYC	barretta proteica	0.282
B001AWBZRO	cibo per gatti	0.297
B00RMX73G2	crema per caffè	0.304
B005HBTNH8	integratore alimentare	0.320
B00L2CD5YI	pasta	0.320

3.7 Analisi sui prezzi

In questa sezione si proverà a rispondere al seguente quesito:

- come varia il sentimento espresso dagli utenti nelle recensioni rispetto al prezzo?

Per rispondere a questo quesito si analizzerà la correlazione tra i sentimenti degli utenti e i prezzi dei prodotti recensiti. Si ricorda qui che il *preprocessing* svolto sul *dataset* ha portato ad associare a ogni recensione il prezzo in dollari del relativo prodotto. Non vi sono recensioni per cui manchi il dato relativo al prezzo.

3.7.1 Correlazione tra sentimenti e prezzi

La correlazione tra i sentimenti degli utenti (inferiti con la *sentiment analysis*) e i prezzi dei prodotti è stata esaminata sia quantitativamente che qualitativamente.

Per la valutazione quantitativa si è proceduto a calcolare il coefficiente di correlazione di Pearson tra i sentimenti e i prezzi. In questo caso questo coefficiente ha un valore pari a 0.002. Pertanto si può affermare che non vi sia correlazione tra le variabili.

Questa valutazione quantitativa è confermata anche da una valutazione qualitativa. A questo proposito si mostra in Figura 15 lo *scatter plot* della relazione tra i sentimenti e i prezzi.

In questo caso i sentimenti sono stati indicati con numeri interi: -1 per il sentimento negativo; 0 per il sentimento neutrale; e 1 per il sentimento positivo. Il grafico conferma come non sia possibile individuare una relazione lineare tra sentimenti e prezzi. In questo quadro il prezzo non sembra da solo sufficiente a determinare il gradimento degli utenti.

3.7.2 Sintesi

I sentimenti espressi dagli utenti nelle recensioni non sembrano correlati linearmente ai prezzi dei prodotti recensiti. Naturalmente questo non significa che il prezzo non rilevi nel giudizio degli utenti. Significa semplicemente che il prezzo del prodotto non è da solo sufficiente a determinare il gradimento degli utenti. Un'analisi più approfondita sul ruolo dei prezzi nel sentimento degli utenti sarà svolta nel capitolo 4 dedicato all'*aspect-based sentiment analysis*.

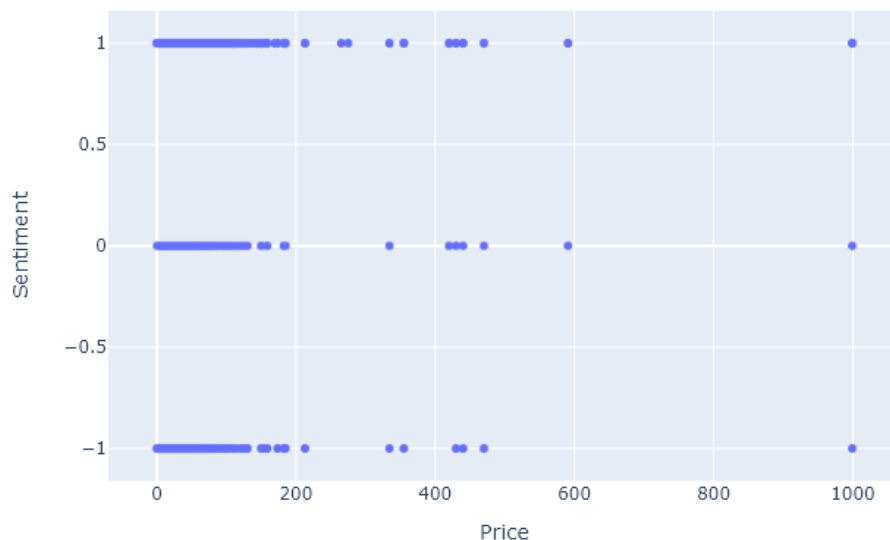


Figura 15: *Scatter plot* della relazione tra sentimenti e prezzi.

3.8 Analisi del rapporto tra testo e sommario delle recensioni

In questa sezione si cercherà di rispondere al seguente quesito:

- il sentimento espresso tramite il sommario diverge da quello espresso tramite il testo della recensione?

A questo proposito verranno presentati i risultati della *sentiment analysis* svolta sul sommario delle recensioni.

3.8.1 Risultati della sentiment analysis sui sommari

I modelli impiegati per la *sentiment analysis* sul testo delle recensioni sono stati utilizzati anche per la *sentiment analysis* sui sommari. Le metriche di *performance* utilizzate sono le stesse menzionate al paragrafo 3.3.

In particolare i risultati della *sentiment analysis* sui sommari sono mostrati nella Tabella 9. Anche in questo caso i risultati migliori per ogni metrica sono evidenziati in grassetto.

Questi risultati permettono di svolgere le seguenti considerazioni:

Tabella 9: Confronto delle *performance* dei modelli impiegati per la *sentiment analysis* sul testo e sul sommario delle recensioni

Modello	Contenuto	<i>Undersampling</i>	Accuratezza	Precisione	<i>Recall</i>	<i>F1-score</i>
AFINN	testo recensione	no	0.810	0.514	0.494	0.500
VADER	testo recensione	no	0.823	0.499	0.493	0.496
Reg. log.	testo recensione	no	0.892	0.693	0.582	0.621
Reg. log.	testo recensione	sì	0.789	0.560	0.692	0.591
AFINN	sommario	no	0.521	0.512	0.471	0.404
VADER	sommario	no	0.526	0.490	0.454	0.406
Reg. log.	sommario	no	0.901	0.807	0.617	0.680
Reg. log.	sommario	sì	0.789	0.549	0.724	0.593

- AFINN e VADER hanno ottenuto *performance* migliori sul testo delle recensioni rispetto al mero sommario. La motivazione è la seguente:
 - i sommari contengono meno parole, come mostrato nelle Figure 16 e 17;
 - spesso AFINN e VADER non attribuiscono una valenza di sentimento alle poche parole presenti nei sommari;
 - questo porta AFINN e VADER a classificare molti sommari come neutrali quando in realtà la maggior parte delle recensioni è positiva, come mostrato al paragrafo 2.3.1;
 - questo emerge chiaramente dalle matrici di confusione relative alle classificazioni di AFINN e VADER mostrate rispettivamente nelle Figure 18 e 19;
 - in questo quadro l’informazione contenuta nei sommari non è sufficiente a classificare le recensioni tramite AFINN e VADER.
- L’accuratezza dei modelli di regressione logistica sul testo delle recensioni e sui sommari è analoga. Le differenze più significative si notano sulle altre metriche e in particolare sulla metrica *f1-score* che risulta particolarmente importante perché (i) rappresenta una combinazione di precisione e *recall* e (ii) identifica meglio gli errori sulle classi di minoranza rispetto all’accuratezza¹⁰. In questo caso i modelli di regressione logistica hanno ottenuto *performance* migliori sui sommari

¹⁰Si supponga di avere un *dataset* composto da due classi. E si supponga che una classe rappresenti il 90% del *dataset*. Un modello che classificasse tutte le istanze come

rispetto al testo delle recensioni. Questo significa che l'informazione contenuta nei sommari è comunque significativa e in alcuni casi è meno ambigua rispetto al testo delle recensioni o comunque più aderente alla valutazione degli utenti.

- Le *performance* complessivamente migliori sono state ottenute dai modelli di regressione logistica sui sommari e senza *undersampling*.

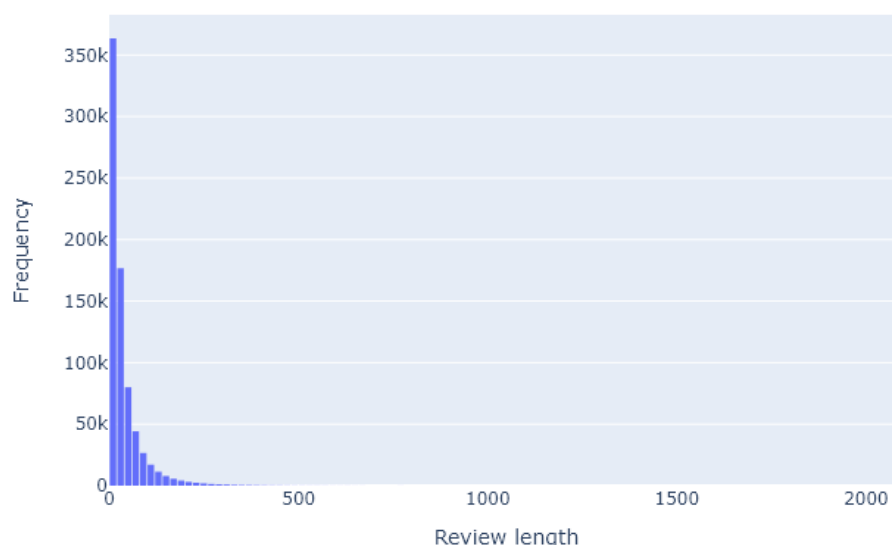


Figura 16: Distribuzione della lunghezza delle recensioni in termini di numero di parole.

3.8.2 Sintesi

Per classificare il sentimento espresso dagli utenti tramite il sommario delle recensioni sono stati utilizzati dei modelli di *sentiment analysis*. I modelli migliori hanno ottenuto *performance* di classificazione superiori rispetto ai

appartenenti alla classe di maggioranza avrebbe un'accuratezza pari a 0.9. Tuttavia questo modello non sarebbe in grado di cogliere alcuna istanza della classe di minoranza. In questo caso la metrica *f1-score* sarebbe pari a 0. In questo modo la metrica *f1-score* coglie meglio il fatto che il modello non riesce a identificare correttamente alcuna istanza della classe di minoranza.

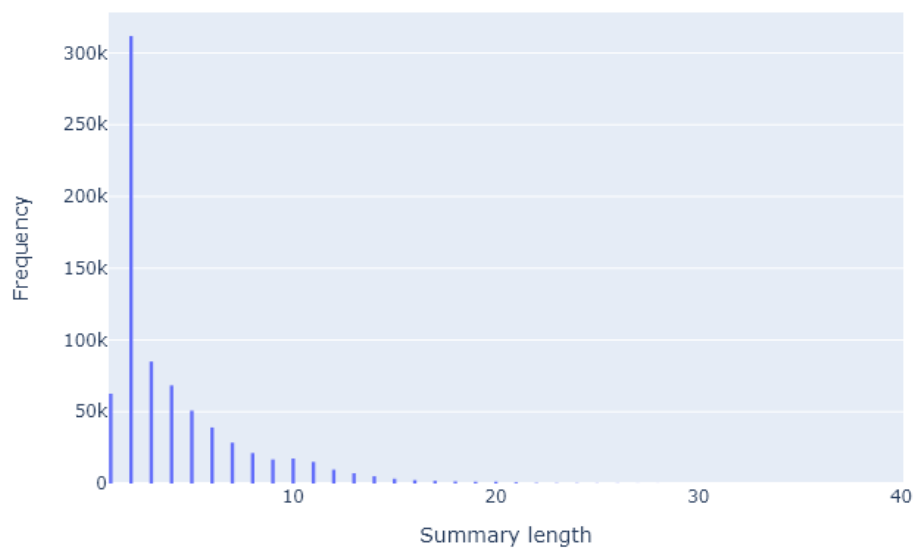


Figura 17: Distribuzione della lunghezza dei sommari in termini di numero di parole.

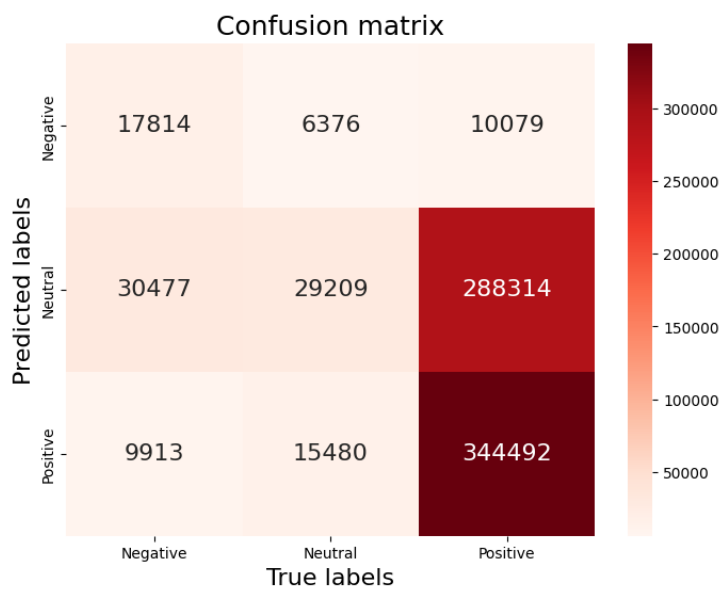


Figura 18: Matrice di confusione per la classificazione dei sentimenti dei sommari con AFINN.

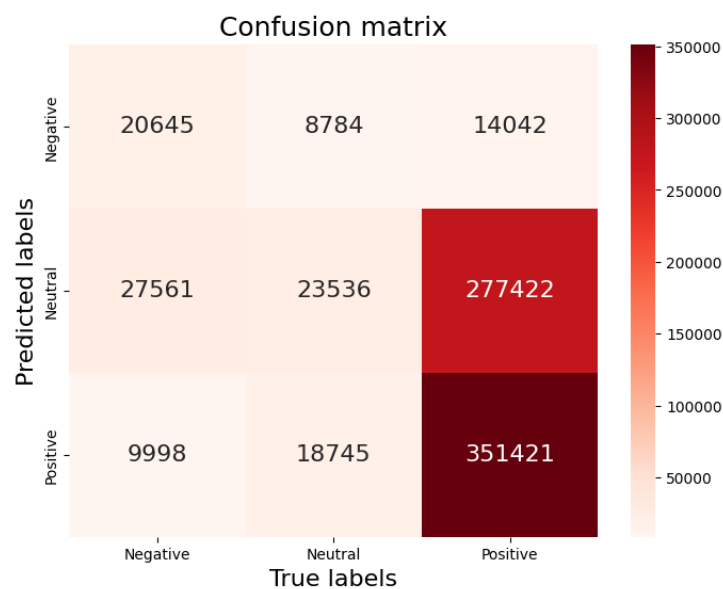


Figura 19: Matrice di confusione per la classificazione dei sentimenti dei sommari con VADER.

modelli addestrati sul testo delle recensioni. Questo suggerisce che l'informazione del sommario è significativa e in alcuni casi è meno ambigua del testo delle recensioni o più aderente alla valutazione degli utenti. In ogni caso si ricorda qui che le ragioni delle misclassificazioni sono diverse e non tutte imputabili a veri e propri errori del modello. Sul punto si rimanda a quanto scritto al paragrafo 3.5.

4 Aspect-based sentiment analysis

Sin qui è stata descritta l'analisi esplorativa del *dataset* e sono stati presentati i risultati della *sentiment analysis*. Per rispondere all'ultimo gruppo di quesiti individuato nel capitolo 1 queste analisi non sono sufficienti. Infatti per individuare gli argomenti trattati nelle recensioni e il relativo sentimento occorre svolgere un'*aspect-based sentiment analysis*. Nel prosieguo si useranno i termini “argomento” e “aspetto” come sinonimi.

In particolare i risultati dell'*aspect-based sentiment analysis* svolta verranno presentati come segue. In primo luogo verranno descritti gli approcci utilizzati. In secondo luogo verranno esposti i risultati relativi agli aspetti in generale. In terzo luogo verranno illustrati i risultati relativi al sentimento sui prezzi in particolare.

4.1 Approcci per l'aspect-based sentiment analysis

4.1.1 Approccio con ASUM

Descrizione. Il primo approccio di *aspect-based sentiment analysis* utilizzato è basato sul modello Aspect and Sentiment Unification Model (ASUM) [4]. ASUM è un *topic model* che estrae *topic* e sentimenti in maniera unificata. Per ASUM gli aspetti sono identificati con i *topic*. In particolare ASUM estende il *topic model* Latent Dirichlet Allocation (LDA) incorporando anche i sentimenti. I *topic* individuati da ASUM sono rappresentati da distribuzioni di probabilità sulle parole. Ogni *topic* viene associato a un sentimento.

ASUM assume che (i) ogni documento sia composto da frasi e che (ii) ogni frase del documento sia relativa a un singolo aspetto e a un singolo sentimento. Questa assunzione risulta pertinente al caso di specie. Infatti una singola recensione può riguardare aspetti diversi a cui vengono associati sentimenti diversi. In questo quadro questa assunzione ha portato a preferire ASUM ad altri modelli che invece assumono che ogni documento sia relativo a un singolo aspetto. Tra questi modelli rientra anche Joint Sentiment/Topic Model (JST) [6].

Un altro vantaggio di ASUM è che si tratta di un approccio non supervisionato all'estrazione di aspetti e sentimenti. Quindi non richiede un *dataset* previamente etichettato. E questo risulta appropriato nel caso di specie in quanto il *dataset* utilizzato per questo lavoro non è stato etichettato per l'*aspect-based sentiment analysis*.

Ancora, si precisa qui che ASUM richiede comunque alcune *sentiment seed words*. Si tratta di parole generiche espressive di sentimenti che non siano specifici per determinati aspetti. Queste parole vengono usate da ASUM per inferire i sentimenti specifici degli utenti sugli aspetti descritti nelle recensioni. Le *sentiment seed words* descritte in [4] sono relative solo al sentimento positivo e a quello negativo. Ai fini di questo lavoro si è ritenuto che fossero sufficienti a rispondere ai quesiti posti. Pertanto non si è provveduto a individuare *sentiment seed words* per il sentimento neutrale.

Infine si evidenziano alcuni limiti dell’approccio basato su ASUM: (i) questo modello non considera il contesto delle parole; (ii) il numero di *topic* è un iperparametro che occorre predeterminare; e (iii) le distribuzioni di probabilità sulle parole prodotte dal modello richiedono un’interpretazione del loro significato più generale.

Per questo lavoro si è fatto uso dell’implementazione di ASUM disponibile al seguente link: <https://yohanjo.github.io/research/WSDM11/>. Si tratta dell’implementazione ufficiale di ASUM scritta in linguaggio Java.

Preprocessing del testo. Le operazioni di *preprocessing* del testo hanno seguito quelle specificate nel paragrafo 2.2.2 per l’analisi esplorativa. Rispetto a queste operazioni sono state apportate due modifiche:

- La prima ha riguardato le negazioni. In particolare ogni parola “not” è stata concatenata alla parola successiva, analogamente a quanto mostrato in [4] per le *sentiment seed words* negative.
- La seconda modifica ha riguardato l’applicazione dello *stemming* a tutte le parole. Questa scelta è stata fatta in analogia a quanto descritto in [4]. In questo caso si è scelto di utilizzare lo *stemmer* Snowball. Si tratta di un’evoluzione dello *stemmer* Porter [12] utilizzato in [4]. E tra l’altro è stato ideato dallo stesso autore. Si è utilizzata l’implementazione offerta da NLTK [1].

Le frasi all’interno delle recensioni sono state individuate utilizzando i seguenti simboli di punteggiatura: “.”, “!” e “?”. Anche questa scelta è stata fatta in aderenza a quanto descritto in [4].

Iperparametri. Per l’addestramento del modello sono stati utilizzati i valori degli iperparametri specificati in [4] con due eccezioni. La prima riguarda

il numero di *topic* per sentimento che è stato definito nel numero di 10 (al posto di 70). Si è scelto un numero basso di *topic* al fine di evitare una sovrapposizione eccessiva. La seconda eccezione riguarda il numero di *thread* utilizzati nell'esecuzione parallela del modello. In questo caso è stato utilizzato il numero massimo di *thread* disponibili.

4.1.2 Approccio con LCF-ATEPC

Descrizione. Il secondo approccio di *aspect-based sentiment analysis* impiegato è basato su LCF-ATEPC [13]. In particolare è stata utilizzata l'implementazione di questo modello messa a disposizione tramite la libreria PyABSA [14].

LCF-ATEPC integra un modello pre-addestrato di BERT [2] e un meccanismo di *local context focus*. In particolare LCF-ATEPC utilizza dei dati annotati con aspetti e sentimenti per imparare a estrarre gli aspetti e a predire i relativi sentimenti. L'estrazione degli aspetti avviene sfruttando il formato BIO (Beginning, Inside, Outside). Per i token considerati come aspetti viene poi classificata la polarità. In questo quadro il modello viene addestrato per risolvere congiuntamente questi due distinti obiettivi.

Gli aspetti identificati da LCF-ATEPC non sono *topic*. Quindi non sono distribuzioni di probabilità sulle parole. In questo caso gli aspetti sono parole specifiche del testo che descrivono caratteristiche del prodotto recensito. E il numero di aspetti non deve essere fissato attraverso un iperparametro.

Per quanto riguarda l'architettura, LCF-ATEPC è un modello neurale che utilizza due *layer* indipendenti basati sul modello pre-addestrato di BERT: (i) uno per estrarre *feature* di contesto locale e (ii) un altro per estrarre *feature* di contesto globale¹¹. Pertanto questo modello utilizza anche il contesto delle recensioni, a differenza di ASUM. E come si è visto al paragrafo 3.5 il contesto può essere determinante per inferire correttamente il sentimento dell'utente.

Per il presente lavoro è stato utilizzato un modello di LCF-ATEPC pre-addestrato su diversi *dataset* in lingua inglese tra cui: (i) *dataset* su cibi e ristoranti; (ii) *dataset* su recensioni di prodotti pubblicate sul sito di Amazon; e (iii) ulteriori *dataset* di *benchmark* per l'*aspect-based sentiment analysis*.

¹¹In questo caso il contesto globale si riferisce alla frase nel suo complesso, mentre il contesto locale si riferisce ad alcune specifiche parole associate a un aspetto sulla base di una distanza semantica [15].

Preprocessing del testo. Si è detto che LCF-ATEPC utilizza informazione di contesto. Per questo motivo si è applicato un *preprocessing* del testo leggero. In particolare sono state eseguite le seguenti operazioni di *preprocessing*: (i) le contrazioni sono state espanse; (ii) le parole sono state convertite in token; (iii) i token sono stati convertiti in caratteri minuscoli¹²; (iv) le ripetizioni sono state semplificate; e (v) parimenti sono stati rimossi anche i *tag* HTML.

4.2 Analisi sugli aspetti

In questa sezione si cercherà di rispondere ai seguenti quesiti:

- quali sono gli argomenti più frequenti nelle recensioni?
- quali sono gli aspetti dei prodotti che gli utenti apprezzano di più?
- quali sono gli aspetti dei prodotti che gli utenti apprezzano di meno?

Per rispondere a questi quesiti si farà riferimento ai risultati dell'*aspect-based sentiment analysis* svolta con ASUM e LCF-ATEPC.

4.2.1 Individuazione degli aspetti più frequenti

Il primo quesito a cui si cercherà di rispondere è quello relativo agli aspetti più frequenti. A questo proposito sono stati utilizzati i risultati ottenuti con LCF-ATEPC.

In particolare LCF-ATEPC ha permesso di individuare per ogni recensione (i) gli aspetti trattati e (ii) i relativi sentimenti. Pertanto è stato possibile calcolare la frequenza degli aspetti. Gli aspetti più frequenti sono mostrati nella Figura 20.

Questi aspetti possono essere divisi in almeno tre categorie:

- **Tipologie di prodotti.** Anzitutto diversi aspetti fanno specifico riferimento a tipologie di prodotti. Per esempio si trovano: “coffee”, “tea”, “chocolate” e “salt”. Si tratta di tipologie di prodotti già menzionate a proposito dei prodotti più apprezzati e meno apprezzati.

¹²Questa scelta è stata dettata dalla necessità di evitare che il modello considerasse come diversi aspetti distinti solo per l'uso di caratteri maiuscoli o minuscoli. Questo avrebbe reso più difficile prendere in considerazione il dato relativo alla frequenza degli aspetti.

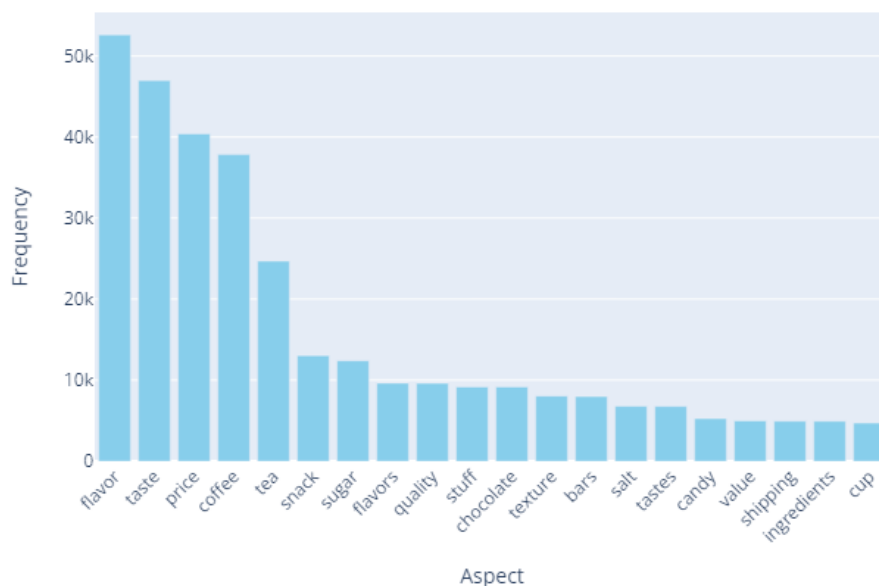


Figura 20: Aspetti più frequenti identificati con LCF-ATEPC.

- **Aspetti commerciali.** Inoltre vi sono aspetti prettamente commerciali. Tra questi si trovano: “price”, “value” e “shipping”. Anche questi aspetti influenzano il giudizio degli utenti.
- **Qualità e ingredienti.** Infine si trovano naturalmente aspetti legati alla qualità dei prodotti. A questo proposito si menzionano: “flavor”, “taste”, “texture” e “ingredients”. Pertanto il sentimento degli utenti è orientato anche dal gusto e dagli ingredienti dei prodotti.

4.2.2 Aspetti più apprezzati

Gli aspetti più apprezzati dagli utenti sono stati individuati con due modalità. Queste modalità riflettono i risultati ottenuti rispettivamente con ASUM e con LCF-ATEPC. Procediamo con ordine.

Aspetti più apprezzati per ASUM. ASUM ha permesso di individuare 10 *topic* associati al sentimento positivo. Questi *topic* corrispondono ad aspetti e sono descritti come distribuzioni di probabilità sulle parole. In questo caso si è proceduto ad associare manualmente un nome a ogni *topic*

per maggiore chiarezza espositiva. Il risultato è mostrato nella Tabella 10. In particolare per ogni *topic* vengono specificate le 10 parole più probabili.

Tabella 10: Aspetti positivi identificati con ASUM

Chocolate	Salt	Coffee	Sauces	Snacks	Price	Tea	Candies	Gluten-free	Quality
tast	use	coffe	flavor	great	good	tea	love	good	great
flavor	oil	flavor	use	snack	price	flavor	good	use	good
good	salt	good	good	good	great	tast	candi	tast	product
love	good	tast	like	bar	product	love	great	great	veri
great	tast	like	tast	love	buy	good	like	make	tast
use	great	cup	great	tast	amazon	like	flavor	love	price
chocol	coconut	love	sauc	eat	love	great	one	like	qualiti
like	love	roast	love	like	order	drink	tast	free	love
milk	like	great	make	veri	veri	veri	gift	gluten	excel
make	popcorn	one	soup	one	best	green	chocol	bread	fresh

Il sentimento positivo sul sale e sul tè riflette le analisi compiute al paragrafo 3.6.1 in merito ai prodotti con il sentimento medio più alto. In questo caso anche l'aspetto commerciale sul prezzo sembra positivo. E a proposito degli ingredienti, la mancanza di glutine sembra giudicata positivamente dagli utenti.

Aspetti più apprezzati per LCF-ATEPC. LCF-ATEPC ha restituito un elenco di aspetti e di relativi sentimenti per ogni recensione. Per individuare gli aspetti più apprezzati si è proceduto come segue:

1. i sentimenti sono stati convertiti in numeri interi: -1 per il sentimento negativo; 0 per il sentimento neutrale; 1 per il sentimento positivo;
2. per ogni aspetto si è calcolata la relativa frequenza intesa come numero di occorrenze;
3. si sono poi considerati solo gli aspetti con una frequenza maggiore di 25. Questo valore rappresenta il novantacinquesimo percentile della frequenza. Questa scelta è stata fatta al fine di ottenere risultati statisticamente robusti;
4. per ogni aspetto si è calcolato il sentimento medio come rapporto tra la somma dei relativi sentimenti e la frequenza;
5. gli aspetti più apprezzati sono quelli con il sentimento medio più alto.

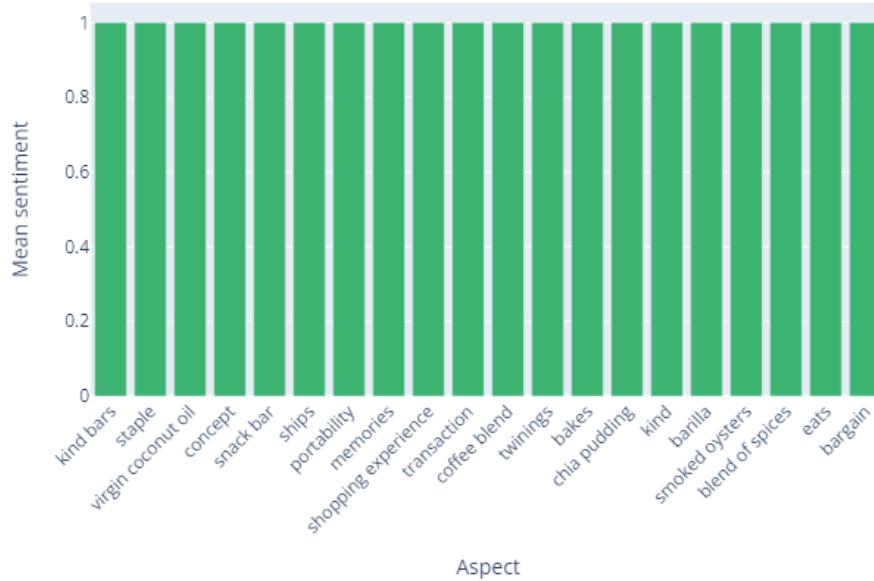


Figura 21: Aspetti più apprezzati identificati con LCF-ATEPC.

I risultati sono mostrati nella Figura 21.

Tutti gli aspetti più apprezzati presentano un sentimento medio pari a 1. L’aspetto “twinings” si riferisce a una nota marca di tè. E conferma l’apprezzamento per questo tipo di prodotto. Gli aspetti “kind bars” e “kind” si riferiscono alla marca di *snacks* Kind. In particolare Kind produce *snacks* che vengono spesso pubblicizzati come molto salutari. Anche questo conferma i risultati di ASUM. Tra gli aspetti commerciali apprezzati emerge lo sconto o l’affare evidenziati con “bargain”.

4.2.3 Aspetti meno apprezzati

Passiamo ora agli aspetti meno apprezzati. Anche in questo caso si analizzeranno i risultati ottenuti sia con ASUM che con LCF-ATEPC.

Aspetti meno apprezzati per ASUM. Gli aspetti meno apprezzati individuati con ASUM sono specificati nella Tabella 11. Anche in questo caso per ogni *topic* vengono mostrate le 10 parole più probabili.

In questo caso tra le tipologie di prodotti con apprezzamento negativo emerge il burro di arachidi. Questo conferma le analisi già condotte ai pa-

Tabella 11: Aspetti negativi identificati con ASUM

Amazon	Drinks	Cooking	Peanut	Sugar	Sweets	Price	Calories	Shipment	Packaging
product	day	water	chocol	tast	tast	store	sugar	order	bag
review	drink	use	like	like	like	buy	ingredi	time	packag
star	use	cup	tast	sugar	flavor	price	fat	buy	box
would	help	make	delici	flavor	veri	local	calori	one	open
tri	eat	add	flavor	use	littl	amazon	contain	bag	one
one	get	cook	peanut	sweet	sweet	groceri	protein	tri	like
like	like	minut	veri	water	smell	get	product	box	use
amazon	one	mix	butter	sweeten	realli	find	serv	would	contain
compani	work	put	sweet	drink	bit	much	high	purchas	veri
becaus	time	time	nut	much	doe	better	sodium	last	plastic

ragrafi 2.4.3 e 3.6.2. Inoltre anche gli alimenti dolci risultano avere una connotazione negativa. In merito agli aspetti commerciali si segnalano problemi legati alle consegne dei prodotti. Il prezzo compare anche tra gli aspetti negativi. Infine risultano poco apprezzati gli ingredienti altamente calorici e legati allo zucchero.

Aspetti meno apprezzati per LCF-ATEPC. Gli aspetti meno apprezzati individuati con LCF-ATEPC sono stati calcolati con lo stesso procedimento descritto per gli aspetti più apprezzati. I risultati sono mostrati nella Figura 22.

Questi risultati consentono di trarre le seguenti considerazioni:

- **Importanza della salute.** In primo luogo emerge chiaramente l’attenzione dei consumatori per la salute. Infatti tra gli aspetti negativi vengono menzionati: *(i)* olio di canola, *(ii)* acidi grassi saturi e *(iii)* sodio. L’olio di canola può comportare rischi per la salute dovuti alla presenza di acido erucico. Per questo motivo per esempio la Commissione europea ha fissato dei tenori massimi per la presenza di acido erucico in alcuni prodotti alimentari¹³. Gli acidi grassi saturi sono legati a malattie cardiovascolari. L’eccesso di sodio nei prodotti salati è un fattore di rischio per l’ipertensione.

L’importanza della salute emerge anche dall’attenzione per la data di scadenza. Questo è dimostrato dal sentimento negativo per “expiration” ed “expiration date”.

- **Importanza del prezzo.** In secondo luogo gli utenti sembrano valutare molto negativamente i prodotti che considerano troppo costosi

¹³Sul punto si veda il regolamento (UE) n. 696/2014 della Commissione europea.

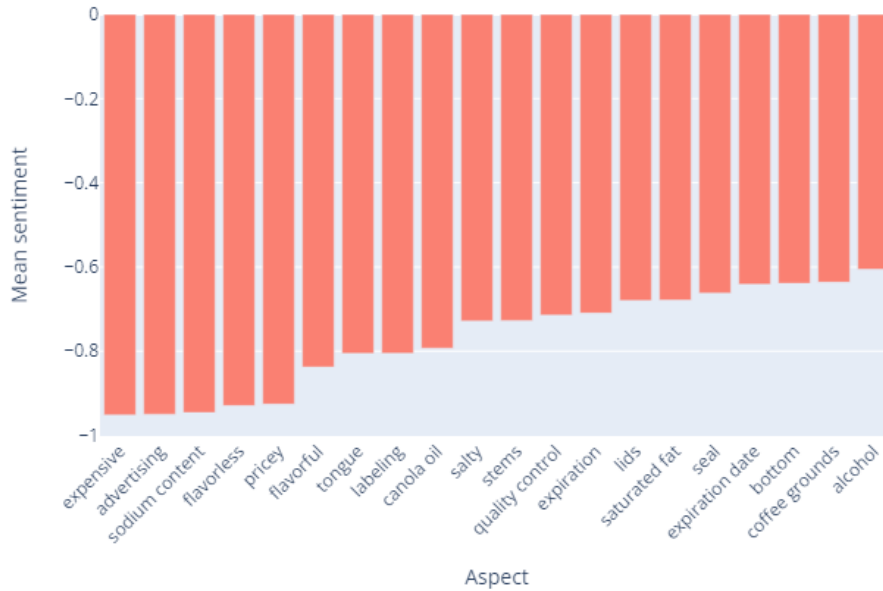


Figura 22: Aspetti meno apprezzati identificati con LCF-ATEPC.

(rispetto alla qualità offerta)¹⁴. A questo proposito è sufficiente considerare che l’aspetto “expensive” è l’aspetto peggiore con un sentimento medio prossimo a -1.

- **Sentimento medio.** Al paragrafo 3.6.2 si è visto che i prodotti peggiori hanno un sentimento più neutrale che negativo. In questo caso invece gli aspetti peggiori riflettono più propriamente un sentimento negativo. Infatti il sentimento medio degli aspetti peggiori è più vicino a -1 che a 0.

4.2.4 Sintesi

Gli aspetti più frequenti menzionati nelle recensioni riguardano principalmente: (i) gli aspetti commerciali; (ii) le tipologie di prodotti; e (iii) gli ingredienti utilizzati.

Gli aspetti più apprezzati possono essere riassunti come segue:

¹⁴Si ricorda qui che al paragrafo 3.7.1 si è mostrato come non vi sia una correlazione tra prezzo e sentimento. Quindi un prezzo elevato non è da solo sufficiente a determinare un giudizio negativo degli utenti.

- **Tipologie di prodotti.** I prodotti più apprezzati sono tè e *snacks*. Per questi prodotti si segnalano come marche gradite rispettivamente Twinings e Kind. Anche il sale è gradito dagli utenti.
- **Aspetti commerciali.** Sul piano commerciale è apprezzata la presenza degli sconti e sono graditi gli affari da un punto di vista del rapporto tra qualità e prezzo.
- **Ingredienti.** I prodotti senza glutine incontrano la soddisfazione degli utenti. Questo si giustifica con il fatto che il glutine può essere fonte di intolleranze alimentari e di malattie come la celiachia.

Gli aspetti meno apprezzati possono essere riassunti come segue:

- **Tipologie di prodotti.** I prodotti meno apprezzati sono principalmente dolci e alimenti altamente calorici. In particolare si segnala la presenza del burro di arachidi.
- **Aspetti commerciali.** Tempi di consegna non rispettati e prezzi eccessivi (rispetto alla qualità offerta) determinano un sentimento negativo per gli utenti.
- **Ingredienti.** La presenza di ingredienti poco salutari influenza negativamente il sentimento degli utenti. Tra questi ingredienti si segnalano gli acidi grassi saturi, l'olio di canola e l'eccessiva presenza di sodio.

4.3 Analisi sui prezzi

In questa sezione si proverà a rispondere al seguente quesito:

- qual è il sentimento degli utenti rispetto al prezzo dei prodotti?

Una risposta parziale a questo quesito è già stata data nei paragrafi precedenti. Infatti si è già visto che gli utenti (*i*) considerano positivamente gli sconti e gli affari e (*ii*) considerano negativamente i prezzi sproporzionati rispetto alla qualità dei prodotti. In questa sezione si svolgeranno ulteriori approfondimenti.

4.3.1 Analisi sui sentimenti dei singoli aspetti

LCF-ATEPC ha individuato diversi aspetti correlati al tema del prezzo. I termini più rilevanti sono i seguenti: “bargain”, “convenience”, “cost”, “costs”, “deal”, “discount”, “expensive”, “money”, “price”, “priced”, “prices”, “pricey” e “pricing”. Il sentimento medio di ciascuno di questi aspetti è mostrato nella seguente Tabella 12.

Tabella 12: Sentimento medio degli aspetti relativi al prezzo individuati con LCF-ATEPC

Aspetto	Media	Dev. st.	Frequenza
bargain	1.000	0.000	72
convenience	0.909	0.396	242
cost	0.004	0.987	1735
costs	-0.217	0.966	152
deal	0.967	0.253	1365
discount	-0.120	0.952	25
expensive	-0.951	0.308	206
money	-0.133	0.985	1060
price	0.695	0.705	40398
priced	0.473	0.880	2252
prices	0.254	0.952	1067
pricey	-0.926	0.378	27
pricing	0.627	0.764	572
<i>Totale</i>	0.630	0.765	49173

Questa Tabella conferma che la convenienza economica è valutata molto positivamente dagli utenti. In questo caso il termine “discount” sembra avere un’accezione negativa. Ma la sua scarsa frequenza non consente di ritenere il risultato robusto. Il costo sproporzionato (“expensive”) è valutato molto negativamente. Gli altri aspetti relativi al prezzo esprimono un sentimento mediamente neutrale o positivo. Ma presentano anche un’alta deviazione standard. Quindi il relativo sentimento varia a seconda del contesto. Questi aspetti sono nettamente la maggioranza.

4.3.2 Analisi temporale

Nella Figura 23 viene mostrato l’andamento temporale del sentimento medio degli aspetti principali relativi al prezzo.

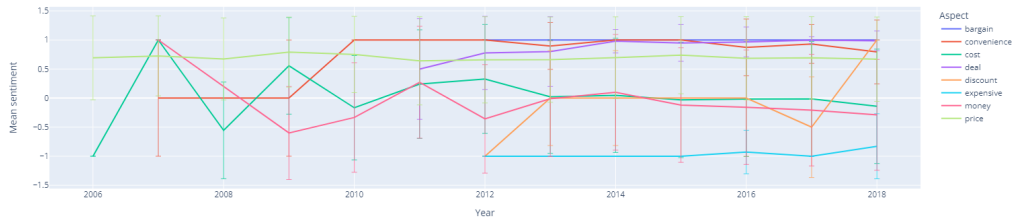


Figura 23: Sentimento medio per anno dei principali aspetti relativi al prezzo identificati con LCF-ATEPC.

Questa Figura conferma le considerazioni precedenti. In questo caso emerge ancora più chiaramente che l'aspetto "expensive" è sempre stato giudicato molto negativamente. E parimenti gli aspetti "bargain" e "deal" sono sempre stati valutati molto positivamente. Il termine "money" sembra tendere verso sentimenti più negativi. Al contrario il termine "discount" è passato da un sentimento medio pari a -1 nel 2012 a un sentimento medio pari a 1 nel 2018. In questo quadro la media complessivamente negativa di questo aspetto potrebbe essere dovuta a un disvalore per gli sconti che è diminuito solo recentemente.

4.3.3 Sintesi

Diversi aspetti sono correlati al tema dei prezzi dei prodotti. Gli aspetti più apprezzati riguardano la convenienza economica e gli sconti/affari. Gli aspetti meno apprezzati riguardano la sproporzione del prezzo rispetto alla qualità del prodotto. Il sentimento degli altri aspetti è invece variabile a seconda dei contesti. E questi aspetti sono i più frequenti.

5 Dashboard

Per rispondere ai quesiti introduttivi è stato necessario valutare quantitativamente e qualitativamente i risultati delle analisi svolte. La valutazione qualitativa è avvenuta tramite diversi grafici. Per agevolare la creazione di alcuni di questi grafici è stata implementata una *dashboard*.

In particolare questa *dashboard* è stata implementata in Python sfruttando la libreria Dash pubblicata al seguente link: <https://github.com/plotly/dash>. La *dashboard* implementata presenta la seguente struttura:

- **Analisi esplorativa.** Una prima sezione riguarda l'analisi esplorativa. In questa sezione possono essere svolte quattro tipologie di analisi:
 - In primo luogo possono essere analizzate le recensioni. In particolare è possibile visualizzare (i) la distribuzione delle opinioni; (ii) la distribuzione delle recensioni per anno; e (iii) la distribuzione in percentuale delle opinioni per ogni anno.
 - In secondo luogo possono essere analizzati i prodotti. In particolare è possibile visualizzare informazioni di dettaglio su uno specifico prodotto. Ed è possibile visualizzare i prodotti con il voto medio più alto e più basso. Il numero di prodotti può essere modificato con uno *slider*.
 - In terzo luogo possono essere analizzati i prezzi dei prodotti. In particolare è possibile visualizzare la distribuzione dei prodotti per intervalli di prezzo. Ed è possibile visualizzare lo *scatter plot* della relazione tra prezzi e voti.
 - In quarto luogo possono essere analizzate le parole delle recensioni. In particolare vengono mostrate le *word cloud* delle parole più frequenti.
- **Sentiment analysis.** Una seconda sezione riguarda la *sentiment analysis*. In questa sezione possono essere svolte tre tipologie di analisi:
 - In primo luogo è possibile visualizzare i risultati della *sentiment analysis*: (i) la distribuzione dei sentimenti; (ii) i prodotti migliori e peggiori per sentimento medio; e (iii) lo *scatter plot* della relazione tra prezzi e sentimenti. Il numero di prodotti migliori e peggiori può essere regolato con uno *slider*.

- In secondo luogo è possibile ricercare recensioni di uno specifico prodotto. In particolare viene mostrato un esempio di recensione del prodotto cercato con associato il relativo sentimento.
 - In terzo luogo è possibile cercare esempi di divergenze tra il sentimento inferito e la vera opinione dell'utente sul prodotto recensito.
- ***Aspect-based sentiment analysis.*** Una terza sezione riguarda l'*aspect-based sentiment analysis*. In questa sezione possono essere svolte due tipologie di analisi:
 - In primo luogo è possibile visualizzare gli aspetti migliori e peggiori per sentimento medio. Il numero di aspetti può essere regolato con uno *slider*.
 - In secondo luogo è possibile (i) visualizzare l'andamento nel tempo del sentimento di uno specifico aspetto; e (ii) cercare recensioni con uno specifico aspetto.

6 Conclusioni e sviluppi futuri

In questo lavoro sono state svolte diverse analisi su un *dataset* di recensioni di cibi e bevande in lingua inglese. In particolare sono state svolte: (a) un'analisi esplorativa del *dataset*; (b) una *sentiment analysis* sul sentimento espresso dagli utenti nelle recensioni; e (c) un'*aspect-based sentiment analysis* sugli aspetti più apprezzati e meno apprezzati dagli utenti. Queste analisi sono state svolte al fine di (i) individuare i prodotti più graditi e meno graditi dagli utenti e (ii) comprendere il ruolo dei prezzi nel loro giudizio.

In base alle analisi svolte si possono rassegnare le seguenti conclusioni:

- I prodotti più apprezzati appartengono alle categorie merceologiche del tè e degli *snacks*. In particolare per i tè è molto apprezzata la marca Twinings. Per gli *snacks* invece si segnala la marca Kind. Anche il sale è gradito dagli utenti.
- I prodotti meno apprezzati sono tipicamente dolci. Inoltre un ruolo importante tra i prodotti meno apprezzati è occupato dal burro di arachidi. Tuttavia anche il sentimento sui prodotti meno apprezzati è più neutrale che negativo.
- La salute è un aspetto decisivo nella valutazione degli utenti. Questo comporta un giudizio positivo per gli alimenti privi di glutine. E comporta un giudizio negativo per gli alimenti che presentano acidi grassi saturi, olio di canola e un alto tasso di sodio.
- Un'ultima considerazione merita il rapporto tra il prezzo e il sentimento degli utenti. In generale il prezzo non è da solo sufficiente a influenzare il giudizio degli utenti. Tuttavia una sproporzione significativa tra il prezzo e la qualità del prodotto è giudicata molto negativamente. E al contrario, la convenienza economica dei prodotti è giudicata molto positivamente.

Alcuni aspetti delle analisi svolte possono essere approfonditi. In primo luogo si potrebbero svolgere analisi (oltre che sui prodotti, anche) sui consumatori. In particolare si potrebbe indagare se vi siano differenze di sentimento per diverse categorie demografiche. In secondo luogo si potrebbe studiare l'influenza e il successo delle raccomandazioni dei prodotti delle piattaforme di *e-commerce*. In terzo luogo si potrebbe indagare il rapporto tra il sentimento degli utenti e il giudizio espresso (non tanto sui prodotti,

quanto) sulle recensioni di altri utenti. Infine si potrebbe approfondire l'analisi delle recensioni sotto il profilo dello studio dell'ironia e dei sentimenti ambigui o contrastanti.

Riferimenti bibliografici

- [1] BIRD, S., KLEIN, E., AND LOPER, E. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.
- [2] DEVLIN, J., CHANG, M., LEE, K., AND TOUTANOVA, K. BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805 abs/1810.04805* (2018).
- [3] HUTTO, C. J., AND GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media* (2014).
- [4] JO, Y., AND OH, A. H. Aspect and sentiment unification model for on-line review analysis. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining* (2011), WSDM '11, Association for Computing Machinery.
- [5] LEMAÎTRE, G., NOGUEIRA, F., AND ARIDAS, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research* (2017).
- [6] LIN, C., AND HE, Y. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (2009), CIKM '09, Association for Computing Machinery.
- [7] MCAULEY, J. J., AND LESKOVEC, J. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd International Conference on World Wide Web* (2013), WWW '13, Association for Computing Machinery.
- [8] NI, J., LI, J., AND MCAULEY, J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (2019), K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., EMNLP '19, Association for Computational Linguistics.
- [9] NIELSEN, F. Å. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903* (2011).

- [10] PANG, B., LEE, L., AND VAITHYANATHAN, S. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing* (2002), EMNLP '02, Association for Computational Linguistics.
- [11] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* (2011).
- [12] WIESE, A., HO, V., AND HILL, E. A comparison of stemmers on source code identifiers for software search. In *27th IEEE International Conference on Software Maintenance* (2011), ICSM '11.
- [13] YANG, H., ZENG, B., YANG, J., SONG, Y., AND XU, R. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction. *arXiv preprint arXiv:1912.07976 abs/1912.07976* (2019).
- [14] YANG, H., ZHANG, C., AND LI, K. Pyabsa: A modularized framework for reproducible aspect-based sentiment analysis. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (2023), CIKM '23, Association for Computing Machinery.
- [15] ZENG, B., YANG, H., XU, R., ZHOU, W., AND HAN, X. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences* (2019).