

```
"""
```

```
https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products
```

---

Dott. Ing. Nicola Lombardi

---

#### Bibliography:

ResearchGate: [https://www.researchgate.net/publication/382851188\\_A\\_whole\\_method\\_to\\_do\\_Data\\_Analysis](https://www.researchgate.net/publication/382851188_A_whole_method_to_do_Data_Analysis)

Kaggle: <https://www.kaggle.com/datasets/bittlingmayer/amazonreviews>

Dataset: <https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products>

```
=====
| Author of the Project - Engineer Nicola Lombardi |
=====
| MSc. in Telecommunications and Internet of Things |
|Dottore Magistrale | Telecomunicazioni | HW ENGINEER Telit Cinterion SpA, Trieste, Sgonico, TS
| Friuli Venezia Giulia |
|Professional Profile | it.linkedin.com/in/nicola-lombardi-09046b205
```

```
====-----
Università degli Studi di Cagliari (UniCa, Sardegna, Cagliari, CA)
Dipartimento di Ingegneria Elettrica ed Elettronica
Piazza d'Armi, 09123 Cagliari
```

```
====-----
"""
```

```
"""
```

L'analisi del sentiment, o opinion mining, è il processo di analisi di grandi volumi di testo per determinare se il tono emotivo del testo è positivo, negativo o neutro.

---

#### About Dataset

---

This dataset consists of a few million Amazon customer reviews (input text) and star ratings (output labels) for to train fastText for sentiment analysis.

The idea here is a dataset is more than a toy - real business data on a reasonable scale - but can be trained in modest laptop.  
Content

The fastText supervised learning tutorial requires data in the following format:

```
__label__<X> __label__<Y> ... <Text>
```

where X and Y are the class names. No quotes, all on one line.

In this case, the classes are \_\_label\_\_1 and \_\_label\_\_2, and there is only one class per row.

MOS = Mean Opinion Score:  $MOS = 1/N \sum(x_i)$  where  $x_i$  is in {1,2,3,4,5}

```
__label__1 corresponds to 1- and 2-star reviews, and
__label__2 corresponds to 4- and 5-star reviews.
```

```
Esse saranno considerate come r'^__label__(\d)\' con la mappatura:
train_data['Sentiment'].map({'1': 'Negative', '2': 'Positive'})
```

(3-star reviews i.e. reviews with neutral sentiment were not included in the original),

The review titles, followed by ':' and a space, are prepended to the text.

Most of the reviews are in English, but there are a few in other languages, like Spanish.

```
=====
DOCUMENTATION ABOUT BUILTINS
=====
```

```

--- pandas: apply ---

df = pd.DataFrame([[4, 9]] * 3, columns=['A', 'B'])

df
   A  B
0  4  9
1  4  9
2  4  9

df.apply(np.sqrt)
   A  B
0  2.0  3.0
1  2.0  3.0
2  2.0  3.0

def clean_text(text):
    text = re.sub(r'^^[^ ]+ ', '', text) # Rimuovi label iniziale
    text = re.sub(r'^\w\s]', '', text) # Rimuovi punteggiatura
    text = text.lower().strip() # Trasforma in lowercase e rimuovi spazi extra
    return text

# Pulizia del testo
train_data['Cleaned_Review'] = train_data['Review'].apply(clean_text)
test_data['Cleaned_Review'] = test_data['Review'].apply(clean_text)

"""

```

In [1]:

```

# Template del codice Python basato sul dataset descritto

import pandas as pd
import re
import nltk.classify.util
from nltk.classify import NaiveBayesClassifier
from sklearn.model_selection import train_test_split

# STEP 1: IMPORTAZIONE E CREAZIONE DEL DATAFRAME
# Simulazione di lettura del CSV dal link Kaggle
csv_filename = "Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products.csv"
data = pd.read_csv(csv_filename, usecols=["reviews.text", "reviews.rating"])

# STEP 2: FEATURES e PREPROCESSING
# Funzione per pulire il testo
def clean_text(text):
    """
    Pulisce una singola recensione rimuovendo etichette iniziali, punteggiatura,
    spazi extra e trasformando il testo in minuscolo per non dar problemi e non
    dover gestire sia Lowercase che Uppercase separatamente.

    Args:
        text (str): La recensione da pulire.

    Returns:
        str: La recensione pulita.
    """
    text = re.sub(r"^[^\w\s]", "", text) # Rimuovi punteggiatura
    text = text.lower().strip() # Trasforma in lowercase e rimuovi spazi extra
    return text

# Pulizia delle recensioni
data["Cleaned_Review"] = data["reviews.text"].apply(clean_text)

# STEP 3: BINARIZZAZIONE DEI LABEL
# Recensioni con rating > 3 sono positive (1), altrimenti negative (0)
data["Sentiment"] = data["reviews.rating"].apply(lambda x: 1 if x > 3 else 0)

# STEP 4: DIVISIONE DEI DATI IN TRAINING E TEST SET
# Separazione dei dati in training (80%) e test (20%)
train_data, test_data = train_test_split(data, test_size=0.2, random_state=42)

# STEP 5: FEATURE EXTRACTION
def extract_features(review):
    """
    Estrae le features da un testo come un dizionario di parole.
    """
    words = review.split()
    return {word: True for word in words}

# Creazione delle features per training e test
train_features = [(extract_features(row["Cleaned_Review"]), row["Sentiment"]) for _, row in train_data.iterrows()]
test_features = [(extract_features(row["Cleaned_Review"]), row["Sentiment"]) for _, row in test_data.iterrows()]

```

```
# STEP 6: ADDESTRAMENTO DEL CLASSIFICATORE NAIVE BAYES
classifier = NaiveBayesClassifier.train(train_features)

# Valutazione delle prestazioni
accuracy = nltk.classify.util.accuracy(classifier, test_features)
print(f"Accuracy del classificatore: {accuracy:.2f}")

# STEP 7: PAROLE PIÙ INFORMATIVE
print("\nTop 10 parole più informative:")
classifier.show_most_informative_features(10)

# STEP 8: TEST SU INPUT MANUALE
input_reviews = [
    "It is an amazing product",
    "This is a terrible product. I would never recommend it to anyone.",
    "The product works great and has excellent features",
    "The quality was disappointing and the design is awful"
]

# STEP 9: PREDIZIONE SU INPUT MANUALE
print("\nPredictions on sample input:")
for review in input_reviews:
    probdist = classifier.prob_classify(extract_features(clean_text(review)))
    pred_sentiment = probdist.max()
    print(f"Review: {review}")
    print(f"Predicted Sentiment: {pred_sentiment}")
    print(f"Probability: {round(probdist.prob(pred_sentiment), 2)}\n")
```

Accuracy del classificatore: 0.61

Top 10 parole più informative:

Most Informative Features

terrible = True	0 : 1	=	74.4 : 1.0
2015 = True	0 : 1	=	44.6 : 1.0
2016 = True	0 : 1	=	44.6 : 1.0
chrome = True	0 : 1	=	44.6 : 1.0
october = True	0 : 1	=	44.6 : 1.0
poor = True	0 : 1	=	44.6 : 1.0
useless = True	0 : 1	=	44.6 : 1.0
worthless = True	0 : 1	=	44.6 : 1.0
fix = True	0 : 1	=	38.7 : 1.0
advertising = True	0 : 1	=	34.7 : 1.0

Predictions on sample input:

Review: It is an amazing product

Predicted Sentiment: 1

Probability: 0.94

Review: This is a terrible product. I would never recommend it to anyone.

Predicted Sentiment: 0

Probability: 0.96

Review: The product works great and has excellent features

Predicted Sentiment: 1

Probability: 1.0

Review: The quality was disappointing and the design is awful

Predicted Sentiment: 0

Probability: 0.86

In [2]: # Ambiguous sentiment

```
In [3]: """
        "Il prodotto funziona abbastanza bene, ma il prezzo è decisamente troppo alto per quello che offre."
        "L'aspetto è fantastico, ma la durata della batteria è deludente."
        "Il servizio clienti è stato disponibile, anche se ci è voluto molto tempo per risolvere il problema."
        "La qualità dei materiali è eccellente, ma il design non è affatto pratico."
        "Adoro la semplicità d'uso, ma mi aspettavo molte più funzionalità."
        "L'imballaggio era perfetto, ma il prodotto all'interno era leggermente danneggiato."
        "Funziona bene per la maggior parte delle operazioni, ma ogni tanto si blocca inspiegabilmente."
        "Il manuale è chiaro e utile, ma ho trovato difficile configurarlo senza assistenza tecnica."
        "L'estetica è davvero piacevole, ma il dispositivo si surriscalda rapidamente."
        "La spedizione è stata rapida, ma il prodotto non corrisponde esattamente alla descrizione."
        "L'audio è cristallino, ma i bassi sono troppo deboli per i miei gusti."
        "La velocità è impressionante, ma consuma molta energia durante l'utilizzo."
        "Mi piace la compatibilità con altri dispositivi, ma l'app associata è molto lenta."
        "È facile da configurare, ma il software ha ancora troppi bug da risolvere."
        "Le prestazioni sono sopra la media, ma il supporto tecnico lascia molto a desiderare."
        "La confezione regalo è stata un bel tocco, ma il prodotto non ha soddisfatto le aspettative."
        "L'interfaccia è intuitiva, ma alcune funzionalità sono nascoste e difficili da trovare."
        "La luminosità dello schermo è ottima, ma la batteria si scarica rapidamente a quel livello."
        "Ha molte caratteristiche utili, ma alcune non funzionano come descritto."
        """
```

```
"Il prezzo è competitivo, ma dopo alcune settimane di utilizzo non sono sicuro che valga la pena."
""
```

```
Out[3]: '\nIl prodotto funziona abbastanza bene, ma il prezzo è decisamente troppo alto per quello che offre.\nL\'aspe
tto è fantastico, ma la durata della batteria è deludente.\nIl servizio clienti è stato disponibile, anche se c
i è voluto molto tempo per risolvere il problema.\nLa qualità dei materiali è eccellente, ma il design non è af
fatto pratico.\nAdoro la semplicità d\'uso, ma mi aspettavo molte più funzionalità.\nL\'imballaggio era perfe
tto, ma il prodotto all\'interno era leggermente danneggiato.\nFunziona bene per la maggior parte delle operazi
oni, ma ogni tanto si blocca inspiegabilmente.\nIl manuale è chiaro e utile, ma ho trovato difficile configurar
lo senza assistenza tecnica.\nL\'estetica è davvero piacevole, ma il dispositivo si surriscalda rapidamente.\n
La spedizione è stata rapida, ma il prodotto non corrisponde esattamente alla descrizione.\nL\'audio è cristal
lino, ma i bassi sono troppo deboli per i miei gusti.\nLa velocità è impressionante, ma consuma molta energia d
urante l\'utilizzo.\nMi piace la compatibilità con altri dispositivi, ma l\'app associata è molto lenta.\nÈ f
acile da configurare, ma il software ha ancora troppi bug da risolvere.\nLe prestazioni sono sopra la media, ma
il supporto tecnico lascia molto a desiderare.\nLa confezione regalo è stata un bel tocco, ma il prodotto non h
a soddisfatto le aspettative.\nL\'interfaccia è intuitiva, ma alcune funzionalità sono nascoste e difficili da
trovare.\nLa luminosità dello schermo è ottima, ma la batteria si scarica rapidamente a quel livello.\nHa mol
te caratteristiche utili, ma alcune non funzionano come descritto.\nIl prezzo è competitivo, ma dopo alcune set
timane di utilizzo non sono sicuro che valga la pena.\n'
```

```
In [4]: # Recensioni simulate come input ambiguo per il classificatore
ambiguous_reviews = [
    "Il prodotto funziona abbastanza bene, ma il prezzo è decisamente troppo alto per quello che offre.",
    "L'aspetto è fantastico, ma la durata della batteria è deludente.",
    "Il servizio clienti è stato disponibile, anche se ci è voluto molto tempo per risolvere il problema.",
    "La qualità dei materiali è eccellente, ma il design non è affatto pratico.",
    "Adoro la semplicità d'uso, ma mi aspettavo molte più funzionalità.",
    "L'imballaggio era perfetto, ma il prodotto all'interno era leggermente danneggiato.",
    "Funziona bene per la maggior parte delle operazioni, ma ogni tanto si blocca inspiegabilmente.",
    "Il manuale è chiaro e utile, ma ho trovato difficile configurarlo senza assistenza tecnica.",
    "L'estetica è davvero piacevole, ma il dispositivo si surriscalda rapidamente.",
    "La spedizione è stata rapida, ma il prodotto non corrisponde esattamente alla descrizione.",
    "L'audio è cristallino, ma i bassi sono troppo deboli per i miei gusti.",
    "La velocità è impressionante, ma consuma molta energia durante l'utilizzo.",
    "Mi piace la compatibilità con altri dispositivi, ma l'app associata è molto lenta.",
    "È facile da configurare, ma il software ha ancora troppi bug da risolvere.",
    "Le prestazioni sono sopra la media, ma il supporto tecnico lascia molto a desiderare.",
    "La confezione regalo è stata un bel tocco, ma il prodotto non ha soddisfatto le aspettative.",
    "L'interfaccia è intuitiva, ma alcune funzionalità sono nascoste e difficili da trovare.",
    "La luminosità dello schermo è ottima, ma la batteria si scarica rapidamente a quel livello.",
    "Ha molte caratteristiche utili, ma alcune non funzionano come descritto.",
    "Il prezzo è competitivo, ma dopo alcune settimane di utilizzo non sono sicuro che valga la pena."
]

# Classificatore che prevede il sentiment per queste recensioni
print("\nPredictions for ambiguous reviews:")
ambiguous_predictions = []
for review in ambiguous_reviews:
    cleaned_review = clean_text(review) # Pulizia del testo
    probdist = classifier.prob_classify(extract_features(cleaned_review))
    pred_sentiment = probdist.max()
    prob = round(probdist.prob(pred_sentiment), 2)
    ambiguous_predictions.append((review, pred_sentiment, prob))
    print(f"Review: {review}")
    print(f"Predicted Sentiment: {pred_sentiment}, Probability: {prob}\n")

ambiguous_predictions
```

Predictions for ambiguous reviews:

Review: Il prodotto funziona abbastanza bene, ma il prezzo è decisamente troppo alto per quello che offre.  
Predicted Sentiment: 1, Probability: 0.83

Review: L'aspetto è fantastico, ma la durata della batteria è deludente.  
Predicted Sentiment: 1, Probability: 0.94

Review: Il servizio clienti è stato disponibile, anche se ci è voluto molto tempo per risolvere il problema.  
Predicted Sentiment: 1, Probability: 0.83

Review: La qualità dei materiali è eccellente, ma il design non è affatto pratico.  
Predicted Sentiment: 0, Probability: 0.55

Review: Adoro la semplicità d'uso, ma mi aspettavo molte più funzionalità.  
Predicted Sentiment: 1, Probability: 0.94

Review: L'imballaggio era perfetto, ma il prodotto all'interno era leggermente danneggiato.  
Predicted Sentiment: 1, Probability: 0.94

Review: Funziona bene per la maggior parte delle operazioni, ma ogni tanto si blocca inspiegabilmente.  
Predicted Sentiment: 1, Probability: 0.83

Review: Il manuale è chiaro e utile, ma ho trovato difficile configurarlo senza assistenza tecnica.  
Predicted Sentiment: 1, Probability: 0.95

Review: L'estetica è davvero piacevole, ma il dispositivo si surriscalda rapidamente.  
Predicted Sentiment: 1, Probability: 0.94

Review: La spedizione è stata rapida, ma il prodotto non corrisponde esattamente alla descrizione.  
Predicted Sentiment: 1, Probability: 0.5

Review: L'audio è cristallino, ma i bassi sono troppo deboli per i miei gusti.  
Predicted Sentiment: 1, Probability: 0.8

Review: La velocità è impressionante, ma consuma molta energia durante l'utilizzo.  
Predicted Sentiment: 1, Probability: 0.94

Review: Mi piace la compatibilità con altri dispositivi, ma l'app associata è molto lenta.  
Predicted Sentiment: 1, Probability: 0.93

Review: È facile da configurare, ma il software ha ancora troppi bug da risolvere.  
Predicted Sentiment: 0, Probability: 0.89

Review: Le prestazioni sono sopra la media, ma il supporto tecnico lascia molto a desiderare.  
Predicted Sentiment: 1, Probability: 0.96

Review: La confezione regalo è stata un bel tocco, ma il prodotto non ha soddisfatto le aspettative.  
Predicted Sentiment: 0, Probability: 0.75

Review: L'interfaccia è intuitiva, ma alcune funzionalità sono nascoste e difficili da trovare.  
Predicted Sentiment: 1, Probability: 0.95

Review: La luminosità dello schermo è ottima, ma la batteria si scarica rapidamente a quel livello.  
Predicted Sentiment: 1, Probability: 0.92

Review: Ha molte caratteristiche utili, ma alcune non funzionano come descritto.  
Predicted Sentiment: 0, Probability: 0.71

Review: Il prezzo è competitivo, ma dopo alcune settimane di utilizzo non sono sicuro che valga la pena.  
Predicted Sentiment: 1, Probability: 0.5

```
Out[4]: [('Il prodotto funziona abbastanza bene, ma il prezzo è decisamente troppo alto per quello che offre.',
1,
0.83),
("L'aspetto è fantastico, ma la durata della batteria è deludente.", 1, 0.94),
('Il servizio clienti è stato disponibile, anche se ci è voluto molto tempo per risolvere il problema.',
1,
0.83),
('La qualità dei materiali è eccellente, ma il design non è affatto pratico.',
0,
0.55),
("Adoro la semplicità d'uso, ma mi aspettavo molte più funzionalità.",
1,
0.94),
("L'imballaggio era perfetto, ma il prodotto all'interno era leggermente danneggiato.",
1,
0.94),
('Funziona bene per la maggior parte delle operazioni, ma ogni tanto si blocca inspiegabilmente.',
1,
0.83),
('Il manuale è chiaro e utile, ma ho trovato difficile configurarlo senza assistenza tecnica.',
1,
0.95),
("L'estetica è davvero piacevole, ma il dispositivo si surriscalda rapidamente.",
1,
0.94),
('La spedizione è stata rapida, ma il prodotto non corrisponde esattamente alla descrizione.',
1,
0.5),
("L'audio è cristallino, ma i bassi sono troppo deboli per i miei gusti.",
1,
0.8),
("La velocità è impressionante, ma consuma molta energia durante l'utilizzo.",
1,
0.94),
("Mi piace la compatibilità con altri dispositivi, ma l'app associata è molto lenta.",
1,
0.93),
('È facile da configurare, ma il software ha ancora troppi bug da risolvere.',
0,
0.89),
('Le prestazioni sono sopra la media, ma il supporto tecnico lascia molto a desiderare.',
1,
0.96),
('La confezione regalo è stata un bel tocco, ma il prodotto non ha soddisfatto le aspettative.',
0,
0.75),
("L'interfaccia è intuitiva, ma alcune funzionalità sono nascoste e difficili da trovare.",
1,
0.95),
('La luminosità dello schermo è ottima, ma la batteria si scarica rapidamente a quel livello.',
1,
0.92),
('Ha molte caratteristiche utili, ma alcune non funzionano come descritto.',
```

```
0,  
0.71),  
( 'Il prezzo è competitivo, ma dopo alcune settimane di utilizzo non sono sicuro che valga la pena.',  
1,  
0.5)]
```

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js