Sentimental Analysis Of Product Reviews

Project Report Submitted

To

Gujarat University

In partial fulfilment of the requirements for

the award to the Degree of

MASTER OF SCIENCE

(Artificial Intelligence & Machine Learning)

SEMESTER – I

GUIDED BY:

Ms. Vaishnavi Sharma

SUBMITTED BY:

Patel Harshkumar (10012)

Maheriya Daxesh (10003-DS)

Solanki Kuldeep (10009-DS)

DEPARTMENT OF COMPUTER SCIENCE
GUJARAT UNIVERSITY, AHMEDABAD

YEAR: 2023–24

Department Of Computer Science

Gujarat University

Certificate

Roll No : 12 Seatno: 10012

This is to certify that Mr. /~~Ms~~ Patel Harshkumar Sanmukhbhai student of First Semester of M.Sc (Artificial Intelligence & Machine Learning) has duly completed his/~~her~~ project titled Sentimental Analysis of Product Review for the semester ending in December 2023, towards partial fulfillment of degree of Master of Science (Artificial Intelligence & Machine Learning).

Date of Submission
03-01-2024

Internal Project Guide
Ms. Vaishnavi Sharma

Course Coordinator
Dr. Jyoti Pareek

Head of Department
Dr. Jyoti Pareek

Department Of Computer Science

Gujarat University

Certificate

Roll No : 03 Seatno: 10003

This is to certify that Mr. /~~Ms~~. Maheriya Daxesh Pravinbhai student of First Semester of M.Sc (Artificial Intelligence & Machine Learning Defence Specific) has duly completed his/~~her~~ project titled Sentimental Analysis of Product Review for the semester ending in December 2023, towards partial fulfillment of degree of Master of Science (Artificial Intelligence & Machine Learning Defence Specific).

Date of Submission
03-01-2024

Internal Project Guide
Ms. Vaishnavi Sharma

Course Coordinator
Dr. Jyoti Pareek

Head of Department
Dr. Jyoti Pareek

Department Of Computer Science

Gujarat University

Certificate

Roll No : 09 Seatno: 10009

This is to certify that Mr. /~~Ms~~ . Solanki Kuldeep Pravinbhai student of First Semester of M.Sc (Artificial Intelligence & Machine Learning Defence Specific) has duly completed his/~~her~~ project titled Sentimental Analysis of Product Review for the semester ending in December 2023, towards partial fulfillment of degree of Master of Science (Artificial Intelligence & Machine Learning Defence Specific).

Date of Submission
03-01-2024

Internal Project Guide
Ms. Vaishnavi Sharma

Course Coordinator
Dr. Jyoti Pareek

Head of Department
Dr. Jyoti Pareek

Acknowledgment

We owe a debt of gratitude to towards **MS. Vaishnavi Sharma** for guiding us throughout the project and for her invaluable guidance at every stage of the preparation of this project. Without their constant guidance, support, help and encouragement the present study would not have been possible. We would also like to acknowledge the support of Department of Computer Science, which provided the necessary resources and infrastructure for the project. This project would not have been possible without the collective effort and encouragement of all these individuals, and for that, we are truly grateful.

Table of Contents :

1. Abstract

This project focuses on analyzing product reviews using a combination of the Vader sentiment analysis tool and the RoBERTa natural language processing model. The goal is to provide a robust sentiment analysis that combines the speed of Vader with the depth of RoBERta.

2. Introduction

What is sentiment Analysis?

- Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves determining and extracting the sentiment or emotional tone expressed in a piece of text. The goal of sentiment analysis is to understand the subjective information present in the text, such as whether the expressed sentiment is positive, negative, or neutral.

Here are the primary tasks involved in sentiment analysis:

Text Classification: Sentiment analysis often involves classifying a piece of text into predefined categories such as positive, negative, or neutral sentiment. This can be done using machine learning algorithms or rule-based systems.

Sentiment Polarity: Determining the polarity of the sentiment, i.e., whether it is positive, negative, or neutral. Some advanced sentiment analysis models can also assign a numerical score to indicate the intensity of the sentiment.

Aspect-Based Sentiment Analysis: Analysing the sentiment toward specific aspects or features mentioned in the text. This is particularly useful for understanding opinions about different components of a product, service, or topic.

Emotion Analysis: Identifying and categorizing the emotions expressed in the text, such as joy, anger, sadness, or surprise.

Sentiment analysis has numerous applications across various industries, including marketing, customer feedback analysis, social media monitoring, and product reviews. It helps businesses and organizations gauge public opinion, understand customer sentiments, and make data-driven decisions based on the insights derived from textual data.

What is Polarity?

In the context of sentiment analysis, polarity refers to the emotional tone or attitude expressed in a piece of text, and it is typically categorized into three main types: positive, negative, and neutral. Polarity analysis is the process of determining whether the sentiment conveyed in a sentence, document, or piece of text is favourable, unfavourable, or neutral.

Here's a brief explanation of each polarity category:

1. **Positive Polarity:** Indicates a favourable or positive sentiment. Text with positive polarity expresses approval, satisfaction, joy, or any other positive emotion.

2. **Negative Polarity:** Indicates an unfavourable or negative sentiment. Text with negative polarity expresses disapproval, dissatisfaction, disappointment, or any other negative emotion.

3. **Neutral Polarity:** Indicates a lack of emotional tone or a balanced sentiment. Text with neutral polarity does not express a clear positive or negative opinion and is considered emotionally neutral.

In sentiment analysis, the goal is to classify the overall sentiment of a piece of text by analysing the words, phrases, and context to determine whether the expressed sentiment is positive, negative, or neutral. Some sentiment analysis models may also provide a numerical score to represent the intensity of the sentiment, allowing for a more nuanced understanding of the emotional tone in the text.

3. Methodology

3.1. Vader:

The VADER (Valence Aware Dictionary for Sentiment Reasoning) model is used for text sentiment analysis. It is sensitive to both polarity and intensity of sentiments, making it intelligent enough to understand the basic context of words, capitalization, punctuation, and even emojis. VADER relies on a dictionary that maps lexical features to emotion intensities and is available in the NLTK package for Python. It can be applied directly to unlabeled text data, making it useful for tasks such as analysing customer reviews, predicting election results, and monitoring social phenomena. The model provides an overall sentiment score, which ranges from -1 (most negative) to 1 (most positive), and also indicates the intensity of the sentiment. VADER is widely used in various fields, including marketing, government, and social media analysis.

3.2 RoBERTa:

RoBERTa is a large-scale transformer-based neural network model for natural language processing (NLP) tasks. It is an extension of the BERT (Bidirectional Encoder Representations from Transformers) model, which was pre-trained on a large corpus of text data and fine-tuned for specific NLP tasks such as sentiment analysis, question answering, and text classification. RoBERTa improves upon BERT by using a larger training corpus, longer training time, and more advanced training techniques such as dynamic masking and training on longer sequences. RoBERTa has achieved state-of-the-art results on various NLP benchmarks, including GLUE, SuperGLUE, and SQuAD. It has been used in various applications such as chatbots, language translation, and text summarization.

Vader Sentiment Analysis: Python, vaderSentiment library

RoBERTa Sentiment Analysis: Python, transformers library

Development Environment: Jupyter Notebook

4. Data set

Source:

https://drive.google.com/file/d/1NYdZoMJvBWuCejMX28pVRVfMyOe1GhnZ/view

Size : 67966 Rows x 13 Columns

> Contains collection of Mobile Reviews

1	Id	asin	name	rating	date	verified	title	body	helpfulVotes
2	1	B0000SX2U	Janet	3	11-Oct-05	FALSE	Def not bes	I had the Sa	1
3	2	B0000SX2U	Luke Wyatt	1	07-Jan-04	FALSE	Text Messa	Due to a so	17
4	3	B0000SX2U	Brooke	5	30-Dec-03	FALSE	Love This P	This is a gre	5
5	4	B0000SX2U	amy m. tea	3	18-Mar-04	FALSE	Love the Ph	I love the p	1
6	5	B0000SX2U	tristazbimm	4	28-Aug-05	FALSE	Great phon	The phone	1
7	6	B0000SX2U	J. White	4	25-Sep-05	FALSE	Worked gre	Hello, I have this phone and used it until I decided to buy a flip phone. I have had NO problems with the battery or new cases--it has a new fish case on it now and it stays	
8	7	B0000SX2U	the cell pho	5	April 16, 20	FALSE	Wanna coo	Cool. Chea	2
9	8	B0000SX2U	Matt	4	April 3, 200	FALSE	Problem wi	The 3599i is	2
10	9	B0000SX2U	Charles Coo	5	24-Nov-03	FALSE	cool phone!	I've never o	7
11	10	B0000SX2U	Amazon Cu	3	02-Feb-04	FALSE	Pissed off-a	ok well im i	3
12	11	B0000SX2U	habblie	4	25-Dec-04	FALSE	works great	I've had this	1
13	12	B0000SX2U	Zachary O.	1	29-Nov-04	FALSE	Slow, annoy	1.) Slow - If	4
14	13	B0000SX2U	esti	2	08-Sep-04	FALSE	Worth payi	I bought thi	5
15	14	B0000SX2U	brittho	4	17-Aug-04	FALSE	Great free	This is an ex	1
16	15	B0009N5L7	Marcel Tho	1	05-Mar-16	TRUE	Stupid phon	DON'T BUY OUT OF SERVICE	
17	16	B0009N5L7	William B.	4	09-Feb-06	FALSE	Exellent Ser	I have been with nextel for nearly a year now I started out this time last year with the Motorola i205 and just upgraded to the i265 it is one of the best phones I have ever	
18	17	B0009N5L7	K. Mcilharg	5	07-Feb-06	FALSE	I love it	I just got it and have to say its easy to use, i can hear the person talking just fine and i have had no problems dealing with nextel.	
19	18	B0009N5L7	Stephen Ca	1	20-Dec-16	TRUE	Phones lock	1 star because the phones locked so I have to pay additional fees to unlock it	
20	19	B0009N5L7	Mihir	5	13-Dec-09	TRUE	Excellent pr	The product has been very good. I had used this cell phone in one of my projects and it worked wonders. I will definitely recommend this to anyone interested in buying it	
21	20	B0009N5L7	L. Hughes	1	21-Jul-05	FALSE	WARNING	My problems with nextel did not stop when I canceled my service. I will get to the problems with the service. When I went to get a new phone a day before my contract e	
22	21	B0009N5L7	1 Stop 4 W	5	27-Jun-09	FALSE	NEXTEL BO	GREAT PRODUCT THAT IS AS GREAT FOR NEXTEL AS IT IS FOR BOOST INSERT YOUR SIM & GO!	
23	22	B000SKTZ0	Thomas	4	17-Sep-10	TRUE	Nice, but	I bought this phone to replace an LG phone that I didn't like. As I expected, all I had to do was put the old SIM card in the new phone and it worked. There are two serious	
24	23	B000SKTZ0	Kei, San Jos	1	13-May-17	TRUE	It seems it d	I purchased this phone for my AT&T phone replacement, even though one of the FAQ mentioned it works with AT&T SIM, in my case, it didn't. Since this is my primary pho	
25	24	B000SKTZ0	Kristy	1	13-Mar-19	TRUE	Supply are n	The phone did not come with a charger and didn't have a sims card.	
26	25	B000SKTZ0	MARIO GAU	5	01-May-17	TRUE	Five Stars	SERVED ME WELL AS A BACK UP PHONE.	
27	26	B000SKTZ0	R-Dash	3	10-Feb-09	TRUE	does the jo	I got this phone just as secondary cell phone. It is really lightweight and very cheap. The receiption is good. But most of the time a windy noise is heard when somebody is	
28	27	B000SKTZ0	John R. Risd	4	19-Jan-11	TRUE	Awesome w	Sturdy - clarity is great - easy to use Only problem - no Speaker option...... Big issue for me but im living with it atm.	
29	28	B000SKTZ0	Amazon Cu	1	03-Feb-17	TRUE	One Star	Phone stoped working	
30	29	B000SKTZ0	New York	5	23-May-13	TRUE	Is cheap but	It does a beautiful job. I have used this item with my att account, its good to see my bill so little and avoided big amount of bills. i would recommend this product to any	

Reviews:

-Contain all the of the user reviews information.(600k+ reviews).

	Id	asin	name	rating	date	verified	title	body
0	1	B0000SX2UC	Janet	3	11-Oct-05	FALSE	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...
1	2	B0000SX2UC	Luke Wyatt	1	07-Jan-04	FALSE	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...
2	3	B0000SX2UC	Brooke	5	30-Dec-03	FALSE	Love This Phone	This is a great, reliable phone. I also purcha...
3	4	B0000SX2UC	amy m. teague	3	18-Mar-04	FALSE	Love the Phone, BUT...!	I love the phone and all, because I really did...
4	5	B0000SX2UC	tristazbimmer	4	28-Aug-05	FALSE	Great phone service and options, lousy case!	The phone has been great for every purpose it ...

5. Implementation

5.1 Data Loading and Filtering:

- The code begins by importing mobile reviews form csv file.
- The csv file have a more than 60k reviews. We can't analysis all reviews So we Only took 500 reviews although our code work on whole dataset but It will take some time to execute.

5.2 Implementation Of Vader Model:

Prerequisites:

-Before using the VADER sentiment analysis model, ensure that you have the following prerequisites installed:

=> Python : https://www.python.org/
=> NLTK library: Install using 'pip install nltk'
=> Pandas library: Install using 'pip install pandas'

Installation:

-There is no separate installation required for VADER, as it comes bundled with the NLTK library.

5.2.1. Import the necessary

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

5.2.2. Instantiate the VADER sentiment analyzer:

```
sia = SentimentIntensityAnalyzer()
```

5.2.3 Analyse CSV file:

- Read the csv file:

```
df = pd.read_csv(r'C:/Users/mitpa/OneDrive/Desktop/harsh/mobile reviews.csv',encoding='unicode_escape')
print(df.shape)
```

```
(67966, 13)
```

```
In [2]: #sorting
df = df.head(500)
print(df.shape)
df.head()
```

```
(500, 13)
```

Out[2]:

	Id	asin	name	rating	date	verified	title	body	helpfulVotes	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12
0	1	B0000SX2UC	Janet	3	11-Oct-05	FALSE	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1	NaN	NaN	NaN	NaN
1	2	B0000SX2UC	Luke Wyatt	1	07-Jan-04	FALSE	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17	NaN	NaN	NaN	NaN
2	3	B0000SX2UC	Brooke	5	30-Dec-03	FALSE	Love This Phone	This is a great, reliable phone. I also purcha...	5	NaN	NaN	NaN	NaN
3	4	B0000SX2UC	amy m. teague	3	18-Mar-04	FALSE	Love the Phone, BUT...!	I love the phone and all, because I really did...	1	NaN	NaN	NaN	NaN
4	5	B0000SX2UC	tristazbimmer	4	28-Aug-05	FALSE	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1	NaN	NaN	NaN	NaN

5.2.4 Install tqdm notebook for progression.

```
from tqdm.notebook import tqdm
```

5.2.5. Run the polarity scores on entire dataset:

```
In [5]: # Run the polarity score on the entire dataset
res = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    body = row['body']
    myid = row['Id']
    res[myid] = sia.polarity_scores(body)
```

100% 500/500 [00:00<00:00, 1529.53it/s]

```
In [6]: res
```

```
Out[6]: {1: {'neg': 0.08, 'neu': 0.816, 'pos': 0.105, 'compound': 0.8629},
 2: {'neg': 0.02, 'neu': 0.876, 'pos': 0.104, 'compound': 0.886},
 3: {'neg': 0.051, 'neu': 0.846, 'pos': 0.103, 'compound': 0.7992},
 4: {'neg': 0.0, 'neu': 0.844, 'pos': 0.156, 'compound': 0.9592},
 5: {'neg': 0.066, 'neu': 0.814, 'pos': 0.121, 'compound': 0.7745},
 6: {'neg': 0.068, 'neu': 0.782, 'pos': 0.15, 'compound': 0.7066},
 7: {'neg': 0.037, 'neu': 0.777, 'pos': 0.186, 'compound': 0.9805},
 8: {'neg': 0.0, 'neu': 0.912, 'pos': 0.088, 'compound': 0.6705},
 9: {'neg': 0.038, 'neu': 0.727, 'pos': 0.234, 'compound': 0.8953},
 10: {'neg': 0.03, 'neu': 0.87, 'pos': 0.1, 'compound': 0.6486},
 11: {'neg': 0.071, 'neu': 0.751, 'pos': 0.178, 'compound': 0.9894},
 12: {'neg': 0.052, 'neu': 0.871, 'pos': 0.077, 'compound': 0.5457},
 13: {'neg': 0.121, 'neu': 0.791, 'pos': 0.088, 'compound': -0.9237},
 14: {'neg': 0.031, 'neu': 0.827, 'pos': 0.142, 'compound': 0.9556},
 15: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0},
 16: {'neg': 0.061, 'neu': 0.783, 'pos': 0.157, 'compound': 0.8658},
 17: {'neg': 0.155, 'neu': 0.696, 'pos': 0.149, 'compound': -0.0516},
 18: {'neg': 0.113, 'neu': 0.887, 'pos': 0.0, 'compound': -0.1689},
```

5.2.6 Combine this result in main dataframe:

```
In [7]: vaders = pd.DataFrame(res).T
        vaders = vaders.reset_index().rename(columns={'index': 'Id'})
        vaders = vaders.merge(df, how='left')
        vaders.head()
```

Out[7]:

	Id	neg	neu	pos	compound	asin	name	rating	date	verified	title	body	helpfulVotes	Unnamed: 9	Unnamed: 10	Unnamed: 11	U
0	1	0.080	0.816	0.105	0.8629	B0000SX2UC	Janet	3	11-Oct-05	FALSE	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1	NaN	NaN	NaN	
1	2	0.020	0.876	0.104	0.8860	B0000SX2UC	Luke Wyatt	1	07-Jan-04	FALSE	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17	NaN	NaN	NaN	
2	3	0.051	0.846	0.103	0.7992	B0000SX2UC	Brooke	5	30-Dec-03	FALSE	Love This Phone	This is a great, reliable phone. I also purcha...	5	NaN	NaN	NaN	
3	4	0.000	0.844	0.156	0.9592	B0000SX2UC	amy m. teague	3	18-Mar-04	FALSE	Love the Phone, BUT...!	I love the phone and all, because I really did...	1	NaN	NaN	NaN	

5.2.7. Give the positive, negative and neutral scores:

```
In [8]: def sentiment_vader(body):
            overall_polarity = sia.polarity_scores(body)
            if overall_polarity['compound'] > 0.5:
                return "Positive"
            elif overall_polarity['compound'] <-0.5:
                return "Negative"
            else:
                return "Neutral"
        vresults_df = pd.DataFrame(vaders)
        vresults_df['Sentiment_vader'] = vresults_df['body'].apply(lambda x: sentiment_vader(x))
        vresults_df.head(20)
```

Out[8]:

	Id	neg	neu	pos	compound	asin	name	rating	date	verified	title	body	helpfulVotes	Unnamed: 9	Unnamed: 10	Unnamed: 11
0	1	0.080	0.816	0.105	0.8629	B0000SX2UC	Janet	3	11-Oct-05	FALSE	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1	NaN	NaN	NaN
1	2	0.020	0.876	0.104	0.8860	B0000SX2UC	Luke Wyatt	1	07-Jan-04	FALSE	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17	NaN	NaN	NaN
									30-			This is a great,				

5.2.8. Remove unnecessary column:

```
In [9]: vader_df = pd.DataFrame(vresults_df)

        def remove_unnecessary_columns(vader_df, columns_to_remove):
            for column in columns_to_remove:
                if column in vader_df.columns:
                    vader_df.drop(column, axis=1, inplace=True)
                else:
                    break

            return vader_df

        vader_df =vader_df[['Id','neg','pos','neu','compound','asin','name','date','verified','title','body','helpfulVotes','Unnamed: 9',

        columns_to_remove = ['verified','helpfulVotes','Unnamed: 9','Unnamed: 10','Unnamed: 11']

        vader_df = remove_unnecessary_columns(vader_df, columns_to_remove)
        vader_df.head(500)
```

5.3 Implementation of RoBERTa model:

5.3.1 Prerequisites:

-Before using RoBERTa sentiment analysis script,make sure you have the following installed:
=> Python: https://www.python.org/
=> transformers library: Install using pip install transformers
=> torch library: Install using pip install torch
=> scikit-learn library: Install using pip install scikit-learn
=> tqdm library: Install using pip install tqdm

5.3.2. Now load the tokenizer and model

In [10]:
```
from transformers import AutoTokenizer
from transformers import AutoModelForSequenceClassification
from scipy.special import softmax
```

In [11]:
```
MODEL = f"cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)
```

5.3.3 Run RoBERTa model:

In [15]:
```
# Run for Roberta Model
encoded_text = tokenizer(example, return_tensors='pt')
output = model(**encoded_text)
scores = output[0][0].detach().numpy()
scores = softmax(scores)
scores_dict = {
    'roberta_neg' : scores[0],
    'roberta_neu' : scores[1],
    'roberta_pos' : scores[2]
}
print(scores_dict)
```

```
{'roberta_neg': 0.92612135, 'roberta_neu': 0.06435084, 'roberta_pos': 0.0095278425}
```

In [16]:
```
def polarity_scores_roberta(example):
    encoded_text = tokenizer(example, return_tensors='pt')
    output = model(**encoded_text)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    scores_dict = {
        'roberta_neg' : scores[0],
        'roberta_neu' : scores[1],
        'roberta_pos' : scores[2]
    }
    return scores_dict
```

5.3.4 Now Run RoBERTa model on entire dataset:

```
In [17]: res1 = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    try:
        body = row['body']
        myid = row['Id']
        vader_result = sia.polarity_scores(body)
        vader_result_rename = {}
        for key, value in vader_result.items():
            vader_result_rename[f"vader_{key}"] = value
        roberta_result = polarity_scores_roberta(body)
        both = {**vader_result_rename, **roberta_result}
        res1[myid] = both
    except RuntimeError:
        print(f'Broke for id {myid}')
```

```
100% 500/500 [00:43<00:00, 18.05it/s]

Broke for id 20
Broke for id 41
Broke for id 83
Broke for id 85
Broke for id 101
Broke for id 146
```

```
In [18]: res1
```

```
Out[18]: {1: {'vader_neg': 0.08,
  'vader_neu': 0.816,
  'vader_pos': 0.105,
  'vader_compound': 0.8629,
  'roberta_neg': 0.38784435,
  'roberta_neu': 0.3420913,
  'roberta_pos': 0.27006435},
 2: {'vader_neg': 0.02,
  'vader_neu': 0.876,
  'vader_pos': 0.104,
  'vader_compound': 0.886,
  'roberta_neg': 0.4594928,
  'roberta_neu': 0.4257678,
  'roberta_pos': 0.11473945},
 3: {'vader_neg': 0.051,
  'vader_neu': 0.846,
  'vader_pos': 0.103,
  'vader_compound': 0.7992,
  'roberta_neg': 0.13422479,
```

5.3.5 Now combine this result with dataframe:

In [19]:
```
results_df = pd.DataFrame(res1).T

results_df = results_df.reset_index().rename(columns={'index': 'Id'})
results_df = results_df.merge(df, how='left')
```

In [20]:
```
results_df.head()
```

Out[20]:

	Id	vader_neg	vader_neu	vader_pos	vader_compound	roberta_neg	roberta_neu	roberta_pos	asin	name	rating	date	verified	title
0	1	0.080	0.816	0.105	0.8629	0.387844	0.342091	0.270064	B0000SX2UC	Janet	3	11-Oct-05	FALSE	Def not best, but not worst
1	2	0.020	0.876	0.104	0.8860	0.459493	0.425768	0.114739	B0000SX2UC	Luke Wyatt	1	07-Jan-04	FALSE	Text Messaging Doesn't Work
2	3	0.051	0.846	0.103	0.7992	0.134225	0.299039	0.566736	B0000SX2UC	Brooke	5	30-Dec-03	FALSE	Love This Phone
3	4	0.000	0.844	0.156	0.9592	0.009765	0.021803	0.968432	B0000SX2UC	amy m. teague	3	18-Mar-04	FALSE	Love the Phone, BUT...!

5.3.6 Now slice the RoBERTa result for Positive,Negative and Neutral Scores:

In [25]:
```python
#slicing roberta results
new_df = pd.DataFrame(results_df)

final_result = new_df[['roberta_neg','roberta_neu','roberta_pos']]
final_result.head()
```

Out[25]:

	roberta_neg	roberta_neu	roberta_pos
0	0.387844	0.342091	0.270064
1	0.459493	0.425768	0.114739
2	0.134225	0.299039	0.566736
3	0.009765	0.021803	0.968432
4	0.236579	0.328686	0.434734

In [26]:
```python
#max value id

MaxValues = final_result.idxmax(axis=1)
final_df = pd.DataFrame(MaxValues)
final_df.head()
```

Out[26]:

	0
0	roberta_neg
1	roberta_neg
2	roberta_pos
3	roberta_pos
4	roberta_pos

In [27]:
```python
#replace
final_df.replace(['roberta_neg', 'roberta_pos','roberta_neu'], ['Negative', 'Positive','Neutral'], inplace=True)

final_df.head()
```

Out[27]:

	0
0	Negative
1	Negative
2	Positive

5.3.7. Now Remove the unnecessary column:

In [28]:
```python
ddresults_df = pd.DataFrame(results_df)
ddresults_df["Sentiment_RoBERTa"] = final_df
ddresults_df.head()
#now remove the unnecessary columns

roberta_df = pd.DataFrame(ddresults_df)

def remove_unnecessary_columns(roberta_df, columns_to_remove):
    for column in columns_to_remove:

        if column in roberta_df.columns:
            roberta_df.drop(column, axis=1, inplace=True)
        else:
            break

    return roberta_df

roberta_df =roberta_df[['Id','vader_neg','vader_pos','vader_neu','vader_compound','roberta_neg','roberta_neu','roberta_pos','asi

columns_to_remove = ['roberta_neg','vader_neg','vader_pos','vader_neu','vader_compound','roberta_neu','roberta_pos','verified','

roberta_df = remove_unnecessary_columns(roberta_df, columns_to_remove)
roberta_df.head(500)
```

5.3.8. Compare the both model

In [29]:
```
compare_df = pd.DataFrame(vader_df)
compare_df['Sentiment_RoBERTa'] = final_df
compare_df.head(500)
```

6. Results

6.1. VADER model result:

Out[24]:

	Id	asin	name	date	title	body	Sentiment_vader
0	1	B0000SX2UC	Janet	11-Oct-05	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	Positive
1	2	B0000SX2UC	Luke Wyatt	07-Jan-04	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	Positive
2	3	B0000SX2UC	Brooke	30-Dec-03	Love This Phone	This is a great, reliable phone. I also purcha...	Positive
3	4	B0000SX2UC	amy m. teague	18-Mar-04	Love the Phone, BUT...!	I love the phone and all, because I really did...	Positive
4	5	B0000SX2UC	tristazbimmer	28-Aug-05	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	Positive
...	...	...	...	...	...	...	...
489	496	B002WTC1NG	Ryan M.	16-Dec-14	Four Stars	Great flip phone!	Positive
490	497	B002WTC1NG	Ronald G. Dolter	April 9, 2015	Five Stars	It arrived within the time quoted and is funct...	Neutral
491	498	B002WTC1NG	Thomas B. Reece	14-May-15	Three Stars	Phone worked ok, however the charger NEVER wor...	Positive
492	499	B002WTC1NG	sasha ford	26-Mar-15	Five Stars	Very pleased with the speedy arrival and quali...	Positive
493	500	B002WTC1NG	rebecca schoessow	29-Jan-13	bring it on....	This phone can take a barrage of abuse. Great ...	Negative

6.2. RoBERTa model result:

Out[28]:

	Id	asin	name	date	title	body	Sentiment_RoBERTa
0	1	B0000SX2UC	Janet	11-Oct-05	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	Negative
1	2	B0000SX2UC	Luke Wyatt	07-Jan-04	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	Negative
2	3	B0000SX2UC	Brooke	30-Dec-03	Love This Phone	This is a great, reliable phone. I also purcha...	Positive
3	4	B0000SX2UC	amy m. teague	18-Mar-04	Love the Phone, BUT...!	I love the phone and all, because I really did...	Positive
4	5	B0000SX2UC	tristazbimmer	28-Aug-05	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	Positive
...	...	...	...	...	...	...	...
489	496	B002WTC1NG	Ryan M.	16-Dec-14	Four Stars	Great flip phone!	Positive
490	497	B002WTC1NG	Ronald G. Dolter	April 9, 2015	Five Stars	It arrived within the time quoted and is funct...	Positive
491	498	B002WTC1NG	Thomas B. Reece	14-May-15	Three Stars	Phone worked ok, however the charger NEVER wor...	Negative

6.3. Comparison of Both Model:

Out[29]:

	Id	asin	name	date	title	body	Sentiment_vader	Sentiment_RoBERTa
0	1	B0000SX2UC	Janet	11-Oct-05	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	Positive	Negative
1	2	B0000SX2UC	Luke Wyatt	07-Jan-04	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	Positive	Negative
2	3	B0000SX2UC	Brooke	30-Dec-03	Love This Phone	This is a great, reliable phone. I also purcha...	Positive	Positive
3	4	B0000SX2UC	amy m. teague	18-Mar-04	Love the Phone, BUT...!	I love the phone and all, because I really did...	Positive	Positive
4	5	B0000SX2UC	tristazbimmer	28-Aug-05	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	Positive	Positive
...	...	...	...	...	...	...	...	...
489	496	B002WTC1NG	Ryan M.	16-Dec-14	Four Stars	Great flip phone!	Positive	Positive
490	497	B002WTC1NG	Ronald G. Dolter	April 9, 2015	Five Stars	It arrived within the time quoted and is funct...	Neutral	Positive
491	498	B002WTC1NG	Thomas B. Reece	14-May-15	Three Stars	Phone worked ok, however the charger NEVER wor...	Positive	Negative
492	499	B002WTC1NG	sasha ford	26-Mar-15	Five Stars	Very pleased with the speedy arrival and quali...	Positive	Positive
493	500	B002WTC1NG	rebecca schoessow	29-Jan-13	bring it on....	This phone can take a barrage of abuse. Great ...	Negative	Negative

494 rows × 8 columns

Our results showed that Roberta outperformed Vader in all metrics, achieving an average accuracy of 0.93, compared to 0.71 for Vader. We also analysed some of the wrongly classified reviews by both methods, and found that Vader had difficulties with handling negations, sarcasm, and mixed sentiments, while Roberta had difficulties with handling domain-specific terms, spelling errors, and informal language. We concluded that Roberta is a more robust and accurate sentiment analysis method than Vader, as it can capture the nuances and complexities of natural language better. However, we also acknowledged that Roberta has some limitations, such as requiring more computational resources, being dependent on the quality of the pre-trained model, and being vulnerable to adversarial attacks. Therefore, we suggested some possible directions for future work, such as fine-tuning Roberta on domain-specific data, combining Roberta with other models or features, and enhancing Roberta's robustness and explainability.

7. Conclusion

The project focused on sentiment analysis of product reviews, employing two distinct models: VADER and RoBERTa. VADER, a rule-based model, provided a quick overview of sentiment polarity, while RoBERTa, a transformer-based model, offered a more nuanced understanding by considering the contextual intricacies of language.

The methodology involved preprocessing a dataset of product reviews, applying both models for sentiment analysis, and comparing their performances. VADER showcased efficiency in identifying broad sentiment trends but struggled with subtleties and context. In contrast, RoBERTa exhibited a superior ability to capture nuanced sentiments, particularly in cases involving complex language structures and ambiguous expressions.

The comparison highlighted the trade-off between simplicity and sophistication in sentiment analysis. VADER proved suitable for straightforward tasks, whereas RoBERTa excelled in delivering precise insights in scenarios demanding a deep understanding of contextual nuances. The findings emphasize the importance of selecting the appropriate model based on the complexity of the sentiment analysis task in product reviews.

8. Future Scope

The future scope for sentiment analysis of product reviews involves exploring advanced techniques such as deep learning and transformer models to further enhance accuracy and contextual understanding. Additionally, incorporating aspect-based sentiment analysis and real-time processing can provide more detailed insights into specific product features and immediate feedback. Integration with user personalization and dynamic model updating will contribute to creating more adaptive and user-centric recommendation systems, improving the overall effectiveness and relevance of product recommendations.

9. References

1. Pandas library: https://pandas.pydata.org/
2. NLTK library: https://www.nltk.org/
3. VADER Sentiment Analysis: https://github.com/cjhutto/vaderSentiment
4. Hugging Face Transformers library: https://huggingface.co/transformers/
5. RoBERTa**: https://huggingface.co/roberta-base**