# Departments of Mathematics and Computer Science Department of Education and Psychology Freie Universität Berlin

# NATURAL LANGUAGE PROCESSING PROF. DR. ARTHUR M. JACOBS SUMMERSEMESTER 2021

# ASPECT BASED SENTIMENT ANALYSIS ON AMAZON PRODUCT REVIEW DATASET

ESRA GÜCÜKBEL 5503113

# **Contents**

1.	God	al and Topic	3
2.	Dat	taset	3
3.	Abs	stract	3
		thodth	
	4.1	Data Reading and Wrangling	4
	4.2	Feature Extraction	5
	4.3	Sentiment Analysis	6
	4.4	Visualization	6
<b>5.</b>	Cod	de	7
6.	Ref	ferences	

### 1. Goal and Topic

When people buy a product, mostly have tendencies to look at the reviews of buyers who have been already bought same product to decide. The aim of this project is to extract Pros and Cons of the products from customer review data and show mostly showed up properties of them.

#### 2. Dataset

In this study, Amazon Arts, Crafts and Sewing category dataset and its metadata will be used.[1]The review dataset includes **overall,vote,verified,reviewTime,reviewerId,asin,style, reviewerName,reviewText,summary,unixReviewTime, imageattributes,** however only **overall,reviewText,summary, asin** will meet the requirements for the purpose of this project. From the metadata, only title,maincat,brand,asin, price attributes will be used. Metadata and review data set is joined on **asin** attribute. (Note: In accordance with memory limits, maincat,brand and price columns can be removed.)

	overall	asin	style	reviewText	summary
0	5.0	0449819906	{'Format:': ' Kindle Edition'}	I've read this book already and I've got plans	A WONDERFUL BOOK
1	5.0	0449819906	{'Format:': ' Kindle Edition'}	Nicely written directions.	Nice
2	5.0	0449819906	{'Format:': ' Paperback'}	love it	Five Stars
3	5.0	0449819906	{'Format:': ' Kindle Edition'}	Good additional knitting reference to have ava	Good Reference in Kindle Edition
4	5.0	0449819906	{'Format:': ' Kindle Edition'}	A gazillion pattern stitches, lucidly explaine	Extremely clear, thorough

Figure 1: First 5 records from data set

	title	main_cat	brand	asin	price
0	You Son of a Bitch! 1987 Embroidered Patch				

Figure 2: First 5 records from meta data

#### 3. Abstract

The whole process is called as "Aspect based sentiment analysis". At the beginning we need to find features and their identifiers such as verbs, adjectives and adverbs. The tagging process can be performed by using spaCy dependency parser. Through the instrument of Vader Sentiment Analysis framework, compound word groups with their descriptors will be linked

and also negations must be taken into account. As a last step, positive and negative first m-feature will be shown with its first n-descriptors.

#### 4. Method

## 4.1 Data Reading and Wrangling

In this study, meta data and review data were used, since review data only consists of asin(unique field) and review related data but to be able to identify which product with its name, we need metadata of it, therefore they were joined on **asin** field. Two datasets can be found in reference [1].

Review dataset has 2875917 and meta dataset has 302988 records(products). Due to memory limits, processor capacities only first 10000 records are used in this work.

	asin	reviewText	summary	overall	title	
0	6665560953	Great quality product, and top notch service from the seller!	Five Stars	5.0	You Son of a Bitch! 1987 Embroidered Patch	
1	6665560953	details on this patch are very good, def a fav gift for the recipient. Would love if they offered a hook & loop option, but won't hold that against the quality of the product.	don't push too many pencils	5.0	You Son of a Bitch! 1987 Embroidered Patch	
2	6665560953	Bought these as a gifts for friends and really impressed with the quality and speed of delivery! I recommend these patches for sure! I picked up 10 of these and plan to share it out. If you buy in the US watch shipping cost. But worth it if you buy in bulk!	Great deal if you buy in bulk!	5.0	You Son of a Bitch! 1987 Embroidered Patch	
3	6665560953	The photo's really didn't do this patch justice. It's made extremely well and looks GREAT! I've seen other similar patches depicting this moment from PREDATOR, but this one was much, much, better than anything I saw on ETSY. It arrived looking great!	It's made extremely well and looks GREAT! I've seen other similar patches depicting this moment	5.0	You Son of a Bitch! 1987 Embroidered Patch	
4	6665560953	Exactly as pictured. My brother loves Predator so I had to get us matching patches!	CIA got you pushing too many pencils?	5.0	You Son of a Bitch! 1987 Embroidered Patch	
2902716	B01HJGIR3E	If you are a crafter or like decorating your home, you can't go wrong with this purchase! So many uses!!!!!	Beautiful sturdy ribbon! So many uses!!!!	5.0	Creative Ideas 3" Inch Solid Grosgrain Ribbon, 25 Yards, Pink, 25 yd	
2902717	B01HJGIR3E	This ribbon was ridged, so I couldn't use it for my project. I needed smooth to attach letters.	Ridged ribbon	2.0	Creative Ideas 3" Inch Solid Grosgrain Ribbon, 25 Yards, Pink, 25 yd	
2902718	B01HJGIR3E	Very pretty ribbon. Good quality and good price.	Very pretty ribbon. Good quality and good price	5.0	Creative Ideas 3" Inch Solid Grosgrain Ribbon, 25 Yards, Pink, 25 yd	

Figure 3: Overview to joined dataset

As a writing language mostly some shortenings, for example won't instead of will not, 've instead of have, is made use of. With this object shortening of these kinds of auxiliary words were enlarged. In addition to that, all of words in the reviews were converted to lower case letter.

#### 4.2 Feature Extraction

- 1. Reviews are formed in sentences according to spaCy library nlp method.
- 2. Each line is analyzed independently of others
  - a. If a word is 'NOUN' then it checks the words that located in left side of it.
    - i. If the dependency of word at left is tagged as 'compund' then it gets all of them as word group.
    - ii. If the dependency of word at left is 'amod' (adjectival modifier) and pos tagging is 'adj' (adjective) then it combine them. Moreover if it finds 'advmod' (adverbial modifier) on the left side, it also combine it with adjective and noun words.
  - b. If a word is 'VERB' then it checks the words that located in left and right side of it.
    - i. If the dependency of word at left is 'advmod' (adjectival modifier) and pos tagging is 'adv' (adverb) then it combine them. Moreover if it finds 'advmod' (adverbial modifier) on the left side, it also combine it with adjective and noun words.
    - ii. If a negation word is located in left side of verb, dependency of it 'neg' and pos tagging is 'advmod', then it appends it to left side of verb.
  - c. If a word is 'ADJ' (adjective) then it checks both right and left side of word.
    - i. If the dependency of it is 'xcomp'(open clausal complement) and pos tagging is not 'neg'(negation) then it looks for left side of this word and in case of finding 'aux'(auxialiary) word on left then, it appends to word group.
    - ii. If the dependency of it is 'xcomp'(open clausal complement) and pos tagging is 'neg'(negation) then it checks if the words on left side is 'aux'.If it meets this condition then this word is appended to word group.

At the end of this process, compound words and their descriptors would have been identified.

	asin	reviewText	summary	overall	title	compound_nouns	aspect_keywords
0	6665560953	Great quality product, and top notch service from the seller!	Five Stars	5.0	You Son of a Bitch! 1987 Embroidered Patch	[(quality product, product), (notch service, service)]	[top notch service, great quality product]
1	6665560953	details on this patch are very good, def a fav gift for the recipient. Would love if they offered a hook & loop option, but won't hold that against the quality of the product.	don't push too many pencils	5.0	You Son of a Bitch! 1987 Embroidered Patch	[(fav gift, gift)]	0
2	6665560953	Bought these as a gifts for friends and really impressed with the quality and speed of delivery! I recommend these patches for sure! I picked up 10 of these and plan to share it out. If you buy in the US watch shipping cost. But worth it if you buy in bulk!	Great deal if you buy in bulk!	5.0	You Son of a Bitch! 1987 Embroidered Patch	[(us cost, cost), (watch cost, cost), (shipping cost, cost)]	0
3	6665560953	The photo's really didn't do this patch justice. It's made extremely well and looks GREAT! I've seen other similar patches depicting this moment from PREDATOR, but this one was much, much, better than anything I saw on ETSY. It arrived looking great!	It's made extremely well and looks GREAT! I've seen other similar patches depicting this moment	5.0	You Son of a Bitch! 1987 Embroidered Patch	[(patch justice, justice)]	[similar patches, other patches]
4	6665560953	Exactly as pictured. My brother loves Predator so I had to get us matching patches!	CIA got you pushing too many pencils?	5.0	You Son of a Bitch! 1987 Embroidered Patch	[(matching patches, patches)]	[exactly pictured]

Figure 4: Aspect keywords from reviews

# 4.3 Sentiment Analysis

Sentiment analysis is a text analysis method that detects polarity (e.g. a positive or negative opinion) within the text, whether a whole document, paragraph, sentence, or clause. VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.

SentimentIntensityAnalyzer() produced result for each aspect keywords of every review, it converts word groups into dictionary with scores of polarities for negative, positive and neutral sentiments. Example result of VADER: {'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.4404} Maximum of scores of the review was obtained as output.

As seen on the results, VADER has remarkably true but not always correct for each one, this part is a point to improve output. Textblob and Flair are alternatives using for sentiment analysis.

#### 4.4 Visualization

In order to visualize results and give insight users to identify features easily, Plotly with Dash library was used. Plotly allows developer to create web interface, Jupyter notebook inline or opening a new tab. The interface includes 2 drop down menu and one pie chart to show results. "Product Title" is used for selecting name of product and "Aspect" is used for selecting negative, positive or neutral sentiment. When the user set these criteria up, pie chart will show related 15 word groups with regarding to these criterias.

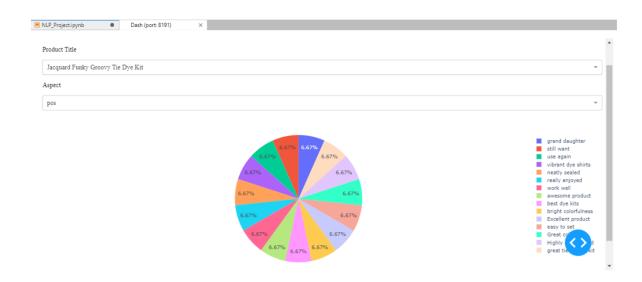


Figure 5: Output of model is presented by using Dash-Plotly

#### 5. Code

The Jupiter notebook which belongs to this project is available on GitHub.

<a href="https://github.com/esragucukbel/NLP\_Term\_Project/blob/main/NLP\_Project.ipynb">https://github.com/esragucukbel/NLP\_Term\_Project/blob/main/NLP\_Project.ipynb</a>

#### 6. References

- [1] Ni, J. (2021, June 18). Retrieved from Amazon Product Reviews: http://deepyeti.ucsd.edu/jianmo/amazon/
- [2] Beri, A. (n.d.). Retrieved from Sentiment Analysis Using Vader: https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664