

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

<b>1. Goal and Topic .....</b>	<b>3</b>
<b>2. Dataset.....</b>	<b>3</b>
<b>3. Abstract .....</b>	<b>3</b>
<b>4. Method.....</b>	<b>4</b>
<b>4.1 Data Reading and Wrangling.....</b>	<b>4</b>
<b>4.2 Feature Extraction .....</b>	<b>5</b>
<b>4.3 Sentiment Analysis .....</b>	<b>6</b>
<b>4.4 Visualization.....</b>	<b>6</b>
<b>5. Code .....</b>	<b>7</b>
<b>6. References .....</b>	<b>7</b>

## 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				
- [2] Beri, A. (n.d.). Retrieved from Sentiment Analysis Using Vader:  
<https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>