Proceedings of the Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2023)
IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2

A Study of Collocations in Sentiment Analysis

1[st] Raj Kishor Bisht
School of Computing
Graphic Era Hill University
Dehradun, India
bishtrk@gmail.com

2[nd] Sarthak Sharma
Department of Mathematics
Graphic Era Hill University
Dehradun, India
sarthakparashar1408@gmail.com

3[rd] Ashna Gusain
Department of Mathematics
Graphic Era Hill University
Dehradun, India
ashnagusain1811@gmail.com

4[th] Nisha Thakur
Department of Mathematics
Graphic Era Hill University
Dehradun, India
nishathakur11398@gmail.com

***Abstract*—Collocations are not merely frequently appearing word combinations (n-grams). Words in collocations have some kind of strong association among them. Collocations play an important role in various natural language processing (NLP) applications. Sentiment analysis is one of the growing areas of research in NLP because of its utilization in various business strategies. The present paper investigates collocations in positive and negative sentiments and their usefulness in sentiment analysis. We considered Amazon Products Review dataset for the purpose and analyzed positive and negative reviews separately. Different statistical techniques; Pointwise Mutual information (PMI), Chi Square test (Chi2), t-test, and likelihood ratio (LH) have been used to extract collocations from these texts and the common collocations have been extracted and analyzed. We found that collocation may be a potential feature for sentiment analysis.**

***Index Terms*—Collocation, sentiment analysis, Pointwise Mutual information, Chi Square test, t-test, likelihood ratio test**

I. Introduction

Natural Language Processing (NLP) is the processing of natural languages with the help of machines, that is, computers so that a machine can deal with natural language as a human being. This includes various tasks, for example, text-to-speech conversion, speech-to-text conversion, machine translation, sentiment analysis etc. NLP is a combination of machine learning, deep learning, and statistics which make computers enable to understand and process human language. These techniques allow computers to perform tasks such as language translation, text summarization, sentiment analysis, and more. Sentiment analysis is a technique to determine the emotion or sentiment behind a piece of text with the help of natural language processing. Sentiment analysis detects customer's feedback to gain insights into customer perceptions, brand reputation, and market trends, social media posts, product reviews, and other forms of text data. The goal of sentiment analysis is to understand how people feel about a particular topic or product, which can be used by businesses to improve products, services, and marketing strategies. Sentiment analysis can be applied to various forms of text data such as social media, reviews, and customer feedback to help companies understand customer needs, gauge brand reputation and improve customer service.

Collocations are multi-word phrases or statements that are quite likely to occur together and some kind of association exists among them. Collocations are groups of words that express a particular thing, for example, 'strong coffee' has a special meaning where we cannot replace 'strong' with any other similar word. Context is crucial in many tasks of natural language processing, for example, word sense disambiguation, machine translation etc. In terms of collocations, context is crucial. Collocations usually take place in the context of a document, that is, a list of words. Finding collocations in this collection of terms is looking for frequent phrases that appear in the text and represent some special meaning. Collocation identification has a variety of applications, including 'keyword extraction' locating the most pertinent keywords in documents to determine what topics are most frequently discussed; Bigrams/Trigrams may produce more significant features for predictive models in NLP issues. Many previous researchers utilized bigrams and trigrams for sentiment analysis [1]] but collocations are something different from bigrams and trigrams. Our objective of the present study is to study collocations in negative and positive sentiment text separately and to analyze whether collocation can be a potential feature for sentiment analysis.

II. Literature Review

Sentiment analysis is nowadays quite popular research topic due to its applicability in business and organizations [2]. Different online platforms are available for people to express their opinions. Public review is quite important for organizations to look at the shorting comings and work accordingly to improve them. Sentiment analysis can also be termed as opinion mining. A number of authors, in their work, provided the detailed survey of previous research work in the direction of sentiment analysis [3]–[7]. Due to the growing nature of online texts and different issues related to natural languages, still too many dimensions are to be explored in sentiment analysis. Utilization of text classification in extracting the sentiment of reviews in Greek language related to an electronic product is discussed in [8]. A new approach for processing emotion detection by looking into the relationships within emotion words in the given corpus is elaborated in [9]. Emotion words play a very important role in analyzing sentiments as they help to capture emotional perceptions. Emotion words can be further grouped into emotion-describing and emotion-inducing words. Sentiment word and its context provides useful information for sentiment analysis [10]. A review may contain different aspects; thus, there is a need to look at different aspects of the review also [11]–[13]. Different aspects should be evaluated in [14] or summarized in [15] for getting the information regarding the sentiment expressed by a text. A sentence may represent a particular aspect; thus, sentiment analysis is also performed at sentence level [16]. Identification of different features of a text is also an important task as features may play a crucial role in the classification of a text [17]. Feature selection methods include lexicon-based and statistical methods [18], [19].

Collocations are some special word combinations having some association among them. Thus, collocations may be quite useful for the purpose of sentiment analysis. Detailed information regarding collocations and methods for collocation extraction can be found in [20]. Some new measures for association of words in collocations have been studied by [21], [22]. Collocations for sentiment analysis of Russian text were studied by [23], [24]. Collocations for sentiment analysis of Russian text were studied by [25]. The basic unigram strategy was employed by [26] to collect the sentiment scores related to the words and analyze the customer review with the corresponding rating. The neural network approach has been utilized by [27] for sentiment collocation extraction.

In the previous research works, words, and n-grams have been considered and reviews are converted into vectors utilizing the frequencies, tf.idf etc. of n-grams in the text. Collocations have not been given due weightage so far. Collocations are also n-grams but they show a stronger association between the words. In the present work, as a first step in the direction of utilizing collocations for sentiment analysis, we have made a study of collocations in positive and negative reviews and checked their suitability for becoming a feature for sentiment analysis.

III. Methodology

Our objective is to look at collocations in positive and negative reviews. We have considered the Amazon Products Review dataset available on the Kaggle website [28] for the purpose and analyzed positive and negative reviews. The Amazon Product Review dataset contains 28332 reviews with 24 attributes for diverse products, including information about the products and the evaluations written by customers. Python programming and NLTK library have been used for analysis purpose. The dataset is pre-processed for removal of punctuation marks, stop words (excluding 'not') and conversion of all words to lower case. The dataset is divided into two parts positive sentiment dataset and negative sentiment dataset. Datasets where the rating of sentiments was given in the form of 1,2,3,4,5 in place of positive and negative categories, ratings 1 and 2 have been converted into negative sentiments and 4 and 5 ratings have been converted into positive sentiments categories. Rating 3 is converted into neutral and has not been utilized for the study. We have applied four different collocation extraction methods: Pointwise Mutual information (PMI), Chi-Square (Chi2) test, t-test, and likelihood ratio (LH) test [12] to extract collocations from each of the datasets. In PMI we go for a high score as a high PMI score indicates a high association between words. In other tests, we decide based on a particular significance level. Here we have taken 5% level of significance for our study. After extracting collocations using these methods, we have assigned part of speech tags to extracted collocations and filtered out collocations with tags noun, determiner, and conjunction etc. Words with tags like adjective and adverb are considered as adjective and adverb are more effective for sentiment analysis. Finally, we have merged (inner join) the four different lists of collocations to find common collocations among collocations extracted by four methods. The top fifteen common collocations have been shown here in different tables corresponding to two sentiments positive and negative for demonstration purpose. The process is summarized in Algorithm 1 assuming a dataset D containing reviews in text and their ratings in numbers.

IV. Collocations analysis in different sentiments

Collocation is not merely a bigram having high frequency. In the previous research work, it has shown that two different bigrams having same frequencies may not be equally good candidates for collocations [20]. Table I shows top fifteen frequent bigrams and trigrams from positive and negative reviews of Amazon product review. Table II shows the top fifteen extracted common collocations (bigrams) from Amazon product review positive sentiments using four different methods. In order to extract collocations from word combinations, we can apply some frequency filter to consider only those n-grams whose frequency is above the threshold. Though it is not necessary, but we have considered a random frequency threshold five for positive review and three for negative review as for the frequency filter five, we were not getting enough trigrams due to small dataset of negative review in comparison to the positive review. Table III shows the top fifteen extracted collocations (bigrams) from Amazon product review negative sentiment , Table IV shows trigrams in positive sentiment reviews and Table V shows trigrams in negative sentiment reviews.

From Table II, we observe that many of the extracted bigrams are showing positive sentiments like, 'user friendly', 'reasonably priced', 'highly recommended', 'would recommend', 'extremely beneficial', 'definitely recommend' etc. From Table III, we observe that some of the bigrams for example, 'stopped working', 'stay away', 'not recommend', 'not worth', 'never buy', etc. are showing negative sentiments. Since these bigrams are among top fifteen bigrams, thus they have high PMI scores also and hence indicating that these quite informative for positive/negative sentiments. We also find some bigrams common in both tables, Table II and Table III like 'user friendly', 'fully charged', 'stopped working etc. This indicates that these collocations are commonly used for both of the purposes. Further analysis reveals that negation of such collocations is used for negative sense like 'not user friendly' as shown in the Table IV in negative review. This indicates that the common bigram collocations can be deleted for the both the lists as they are not capable of discriminating positive and negative review, however these bigrams would be a part of trigram collocations in any one of the positive and

TABLE I
HIGHEST FREQUENCY BIGRAMS AND TRIGRAMS IN AMAZON PRODUCT REVIEW DATASETS

In Positive Reviews		**In Negative Reviews**	
bigram	***trigram***	***bigram***	***trigram***
easy use	name brand batteries	last long	not last long
great price	last long time	not last	worst batteries ever
year old	bought year old	batteries last	batteries not last
last long	seem last long	batteries not	would not recommend
kindle fire	batteries last long	would not	got bad batch
great tablet	batteries great price	not buy	name brand batteries
good price	last long name	amazon batteries	batteries last long
name brand	year old loves	waste money	would not buy
great product	tablet year old	name brand	seem last long
battery life	google play store	not work	leaked battery acid
works great	year old son	not good	dont last long
great value	long name brand	aa batteries	brand name batteries
work well	amazon fire hd	batteries ever	batteries ever purchased
great batteries	batteries work well	brand batteries	buy amazon batteries
amazon fire	tablet easy use	bad batch	spend extra money

TABLE II
COLLOCATIONS (BIGRAMS) IN AMAZON POSITIVE SENTIMENT CORPUS

Bigram	**POS tag**	**Chi2 Score**	**PMI Score**	**LH Score**	**t-Score**
user friendly	RB JJ	159701.00	9.76	2348.17	13.55
pleasantly surprised	RB VBD	110452.24	12.11	409.21	5.00
reasonably priced	RB VBN	109940.16	11.70	512.05	5.74
highly recommend	RB VB	103973.69	8.76	2734.97	15.49
would recommend	MD VB	56579.34	7.16	3452.52	19.86
highly recommended	RB JJ	45129.93	9.61	695.74	7.61
going strong	VBG JJ	32798.72	9.51	529.44	6.70
fully charged	RB VBN	32184.19	10.65	265.23	4.47
extremely beneficial	RB JJ	29797.02	12.28	102.41	2.45
stopped working	VBD VBG	26251.72	10.22	284.72	4.69
higher priced	JJR VBN	25246.44	10.17	275.81	4.69
individually wrapped	RB VBD	24956.10	12.02	92.17	2.45
definitely recommend	RB VB	24627.20	7.51	1208.58	11.60
younger sibling	JJR VBG	23724.04	12.21	77.12	2.24
short lived	JJ VBD	21699.01	12.08	77.58	2.24

TABLE III
COLLOCATIONS (BIGRAMS) IN AMAZON NEGATIVE SENTIMENT CORPUS

Bigram	**POS tag**	**Chi2 Score**	**PMI Score**	**LH Score**	**t-Score**
stopped working	VBD VBG	16553.89	9.85	240.18	4.24
fully charged	RB VBN	13645.74	10.74	113.93	2.83
user friendly	RB JJ	12210.11	10.41	122.57	3.00
stay away	VB RB	4958.73	9.96	62.66	2.23
could repossess	MD VB	4723.81	9.89	68.90	2.23
really fast	RB JJ	2059.25	7.44	104.50	3.44
going back	VBG RB	2020.83	7.19	117.28	3.72
last long	JJ JJ	1996.45	5.12	332.96	7.52
low powered	JJ VBD	1860.07	8.28	59.55	2.44
worked fine	VBN JJ	1814.13	7.67	81.00	2.99
started leaking	VBN VBG	1591.62	8.32	49.39	2.23
not recommend	RB VB	1568.68	5.61	222.48	5.63
not worth	RB JJ	1422.37	5.70	194.67	5.19
never buy	RB VB	1173.48	5.72	144.42	4.70
would not	MD RB	1082.63	4.30	263.90	7.35

TABLE IV
COLLOCATIONS (TRIGRAMS) IN POSITIVE SENTIMENTS

Bigram	POS tag	Chi2 Score	PMI Score	LH Score	t-Score
still going strong	RB VBG JJ	27400938.58	19.35	1114.81	6.40
lasted pretty darn	VBN RB VB	10454524.18	20.73	289.38	2.45
would definitely recommend	MD RB VB	3561830.53	15.35	2718.76	9.22
would highly recommend	MD RB VB	2588896.55	15.31	3890.92	7.87
keep coming back	VB VBG RB	1106889.19	17.75	161.32	2.24
durable thus far	JJ RB RB	763912.76	16.94	463.68	2.45
extremely user friendly	RB JJ RB	737592.11	16.84	2387.76	2.24
pretty darn well	RB RB RB	685995.27	16.79	231.19	2.45
bit longer lasting	RB RBR VBG	443744.86	16.17	178.53	2.45
not always respond	RB RB VB	440566.05	16.43	132.44	2.24
still getting used	RB VBG VBN	209048.37	14.34	318.40	3.16
really enjoy using	RB VB VBG	174756.76	13.77	822.68	3.46
keep getting better	VB VBG RBR	168987.79	13.91	197.19	3.32
would definitely buy	MD RB VB	121894.93	12.40	1145.02	4.58
last pretty long	JJ RB RB	110152.91	11.77	4788.87	4.00

TABLE V
COLLOCATIONS (TRIGRAMS) IN NEGATIVE SENTIMENTS

Bigram	POS tag	Chi2 Score	PMI Score	LH Score	t-Score
still debating returning	RB VBG VBG	8294659.38	21.40	89.60	1.73
ever constantly screwing	RB RB VBG	7720413.58	21.30	89.20	1.73
permanently horizontal not	RB JJ RB	4593382.53	20.55	91.46	1.73
port often break	RB RB VB	3982749.70	20.34	80.13	1.73
seemed awesome little	VBD JJ JJ	3787369.52	20.27	80.11	1.73
randomly close apps	RB RB JJ	3237589.65	20.04	79.44	1.73
called amazon numerous	VBN RB JJ	1655315.06	18.34	122.30	2.24
user friendly much	RB RB JJ	883261.51	18.16	150.13	1.73
not user friendly	RB RB JJ	395103.83	16.00	171.80	2.45
not strong enough	RB JJ RB	328135.83	16.74	80.90	1.73
could not log	MD RB VB	205889.98	16.07	127.46	1.73
would not recommend	MD RB VB	187770.37	13.12	594.50	4.58
horizontal not really	JJ RB RB	178960.08	15.86	66.11	1.73
user friendly not	RB RB RB	103927.25	15.00	137.71	1.73
could not either	MD RB VB	47875.23	13.54	123.53	2.00

negative sentiments. Further if we compare highest frequency bigrams shown in Table I with collocations (bigrams/trigram) shown in Table II, Table III, Table IV and Table V, it can be observed that collocations are quite different from highest frequency bigrams/trigrams. Thus, collocation can be considered as an additional feature for sentiment analysis. The presence of a positive collocation in a text can add weightage towards positivity of the sentiment, similarly the presence of a negative collocation can add weightage towards negativity of the sentiment. A review can be assigned a weight based on the different scores of collocation measure. For example, PMI scores can be utilized as the weights of the collocations; if a collocation appears in the review text, then the positive or negative score of the text will be increased by the weight of the collocation. Since from the extracted collocations list, we get the collocations with their scores also, thus, collocation can be used as an important feature for sentiment analysis.

V. CONCLUSION

The objective of the present study was to know the suitability of collocations for sentiment analysis. Amazon product review dataset was selected for the study of collocations extracted by utilizing different available methods in the literature. Through our study, we found that collocations are subjective, that is related to a particular domain. We found that collocations, in general, are capable of representing positive and negative sentiments for different domains more suitably in comparison to frequency-based bigrams/trigrams. We found that collocations are quite effective in representing sentiments as the scores associated with collocations indicates their intensity also. Thus, collocation can be used as a powerful feature for sentiment analysis. Since collocations are not ordinary word combinations, some associations are there among the words in a collocation, thus utilizing collocations for sentiment analysis will provide additional accuracy to the existing procedures of sentiment analysis which will be a future research work of the present study.

Using this approach of collocation analysis, we found that collocation analysis can be a useful tool in not only analyzing the sentiments of a review but it can also be used to detect the positivity or negativity of different aspects of products also. For example, in Table II, bigram ‘user friendly’ indicating a positive aspect of the product which means that the product is easy to handle and use. Hence a collocation is reflecting a

Algorithm 1 Collocations retrieval, Given the dataset D

```
Create two empty datasets 'positive review' and 'negative review'

for each review in the dataset D do
    Remove punctuation marks.
    Remove stop words excluding 'not'.
    Convert the text into lower case.

    if Rating of a review > 3 then
        Add the review in 'positive review' dataset.
    end if

    if Rating of a review < 3 then
        Add the review in 'negative review' dataset.
    end if
end for

for Each sub dataset 'positive review' and 'negative review' do
    Apply PMI, Chi2 test, t-test, and likelihood ratio (LH) test to extract collocations and their respective scores

    for Each list of collocations extracted by PMI, Chi2 test, t-test, and likelihood ratio (LH) test do

        for Each collocation do
            Assign POS tags
        end for
    end for

    for Each list of tagged collocations extracted by PMI, Chi2 test, t-test, and likelihood ratio (LH) test do

        for Each collocation do

            if POS tag is noun, determiner, or conjunction then
                Remove the collocation from the list to get processed list.
            end if
        end for
    end for
end for
Merge (Inner Join) the four processed lists of collocations.
```

major feature. Similarly, in Table III the bigram collocation 'not worth' is defining another aspect of the product that the price at which the product is bought is disreputable. Therefore, useful aspect based feature extraction can also be done with the help of collocation analysis. The model based on collocations will be more meaningful as it will have some additional and useful information. This will also be a part of future work.

REFERENCES

[1] A. R. Razon and J. A. Barnden, "A new approach to automated text readability classification based on concept indexing with integrated part-of-speech n-gram features," in International Conference Recent Advances in Natural Language Processing, RANLP, 2015, vol. 2015-January.

[2] J. F. Sánchez-Rada and C. A. Iglesias, "Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison," Information Fusion, vol. 52, 2019, doi: 10.1016/j.inffus.2019.05.003.

[3] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," Knowl Based Syst, vol. 226, 2021, doi: 10.1016/j.knosys.2021.107134.

[4] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," in Procedia Computer Science, 2019, vol. 161. doi: 10.1016/j.procs.2019.11.174.

[5] A. Yousif, Z. Niu, J. K. Tarus, and A. Ahmad, "A survey on sentiment analysis of scientific citations," Artificial Intelligence Review, vol. 52, no. 3, 2019, doi: 10.1007/s10462-017-9597-8.

[6] G. Chandrasekaran, T. N. Nguyen, and J. Hemanth D., "Multimodal sentimental analysis for social media applications: A comprehensive review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 11, no. 5. 2021. doi: 10.1002/widm.1415.

[7] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," Artificial Intelligence Review, vol. 54, no. 7, 2021, doi: 10.1007/s10462-021-09973-3.

[8] D. Bilianos, "Experiments in Text Classification: Analyzing the Sentiment of Electronic Product Reviews in Greek," J Quant Linguist, vol. 29, no. 3, 2022, doi: 10.1080/09296174.2021.1885872.

[9] P. Y. Lu, Y. Y. Chang, and S. K. Hsieh, "Causing emotion in collocation: An exploratory data analysis," in Proceedings of the 25th Conference on Computational Linguistics and Speech Processing, ROCLING 2013, 2013.

[10] W. Wang, Y. Zhou, A. Yang, J. Zhou, and J. Lin, "Method of Sentiment Analysis for Comment Texts Based on LDA," Shuju Caiji Yu Chuli/Journal of Data Acquisition and Processing, vol. 32, no. 3, 2017, doi: 10.16337/j.1004-9037.2017.03.023.

[11] Y. Noh, S. Park, and S. B. Park, "Aspect-based sentiment analysis using aspect map," Applied Sciences (Switzerland), vol. 9, no. 16, 2019, doi: 10.3390/app9163239

[12] O. Alqaryouti, N. Siyam, A. A. Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," Applied Computing and Informatics, 2019, doi: 10.1016/j.aci.2019.11.003.

[13] A. Firmanto and R. Sarno, "Aspect-based sentiment analysis using grammatical rules, word similarity and SentiCircle," International Journal of Intelligent Engineering and Systems, vol. 12, no. 5, 2019, doi: 10.22266/ijies2019.1031.19.

[14] Y. Zhang, Y. Xie, and J. Sun, "Aspect Level Sentiment Classification Based on Viewpoint Information Unit," in 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2021, 2021. doi: 10.1109/AEECA52519.2021.9574322.

[15] Y. Li, Z. Qin, W. Xu, and J. Guo, "A holistic model of mining product aspects and associated sentiments from online reviews," Multimed Tools Appl, vol. 74, no. 23, 2015, doi: 10.1007/s11042-014-2158-0.

[16] A. Suriya, "A new approach to improve online customer review analysis by a sentence level using vector similarity related text extraction," Journal of Advanced Research in Dynamical and Control Systems, vol. 11, no. 12 Special Issue, 2019, doi: 10.5373/JARDCS/V11SP12/20193322.

[17] C. Pujari, Aiswarya, and N. P. Shetty, "Comparison of classification techniques for feature oriented sentiment analysis of product review data," in Advances in Intelligent Systems and Computing, 2018, vol. 542.

[18] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models," in Decision Support Systems, 2012, vol. 53, no. 4. doi: 10.1016/j.dss.2012.05.023.

[19] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: A survey," in Proceedings - 11th Web Information System and Application Conference, WISA 2014, 2014. doi: 10.1109/WISA.2014.55. C. D. Manning and H. Schütze, Foundations of statistical natural language processing, vol. 1. 1999.

[20] C. D. Manning and H. Schütze, Foundations of statistical natural language processing, vol. 1. 1999.

[21] R. K. Bisht, H. S. Dhami, and N. Tiwari, “An evaluation of different statistical techniques of collocation extraction using a probability measure to word combinations,” Journal of Quantitative Linguistics, vol. 13, no. 2–3, 2006, doi: 10.1080/09296170600850064.

[22] S. Kumova Metin and B. Karaoğlan, “Measuring collocation tendency of words,” Journal of Quantitative Linguistics, vol. 18, no. 2. 2011. doi: 10.1080/09296174.2011.556005.

[23] C. Hung and Y. X. Cao, “Sentiment classification of Chinese cosmetic reviews based on integration of collocations and concepts,” Electronic Library, vol. 38, no. 1, 2020, doi: 10.1108/EL-04-2019-0093.

[24] H. L. Yang and A. F. Y. Chao, “Sentiment analysis for Chinese reviews of movies in multi-genre based on morpheme-based features and collocations,” Information Systems Frontiers, vol. 17, no. 6, 2015, doi: 10.1007/s10796-014-9498-1.

[25] A. Kotelnikova and E. Kotelnikov, “SentiRusColl: Russian collocation lexicon for sentiment analysis,” in Communications in Computer and Information Science, 2019, vol. 1119 CCIS.

[26] A. Dey, M. Jenamani, and J. J. Thakkar, “Senti-N-Gram: An n-gram lexicon for sentiment analysis,” Expert Systems with Applications, vol. 103, 2018, doi: 10.1016/j.eswa.2018.03.004.

[27] Y. Zhao, B. Qin, and T. Liu, “Encoding syntactic representations with a neural network for sentiment collocation extraction,” Science China Information Sciences, vol. 60, no. 11, 2017, doi: 10.1007/s11432-016-9229-y.

[28] Kaggle Datasets, “Consumer Reviews of Amazon Products,” 2023. https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products (accessed Jan. 22, 2023).