

Amazon Product Sentiment Analysis using Machine Learning Techniques

Sobia Wassan^a, Xi Chen^b, Tian Shen^{c*}, Muhammad Waqar^d, NZ Jhanjhi^e

Abstract

Online stores like Amazon provide a website for consumers to express their opinions about different items. Since then, it has been established that buying online, 90% of consumers are testing different websites channels to determine the quality of their purchase. To evaluate the text data and then extract the sentiment element from that the field of sentiment analysis is frequently used. From user ratings, suggestions, recommendations and messages, online business websites produce a massive volume of textual data every day. One of the most emerging technological trends in web development is the emergence of social web sites. It helps to communicate the peoples and gather knowledge. It simplifies the user contribution via podcasts, blogs, folksonomies, tagging, and wikis using online Social Networks (OSN). Social Network Analysis (SNA) methods can cover extensive networks from thousands or millions of nodes and links of a graph. It characterizes the network structure in the form of nodes (specific actors, people, or things) that is a network used or within the network and the edges or links between nodes (relationships) that connected the network. In this research study, we presented a novel approach that uses sentimental aspects focused on the characteristics of the item. Amazon consumer reviews have been introduced and validated. We collected the dataset from the data world centre, where opinion rates are first detected in each analysis. The system performs pre-processing operations like stone-coating, tokenization, boxing, deletion of stop-words from the datasets to extract meaningful information like positivity or negativity. The primary study aims to analyze this data's aspect level is a huge benefit to marketers to grasp the preferences of consumers better and then develop their behaviour accordingly. Lastly, we also give some perception into their future text classification work.

Keywords: Machine learning, social network analysis (SNA), amazon, polarity, user reviews.

1. Introduction

Sentimental analysis is one of the machines learning processing methods that help identify sentiments that enable entrepreneurs to obtain information about views of their clients via different online media such as social media, surveys, and e-commerce website reviews.

This knowledge will help us understand the triggers and facets of the degradation of the commodity. The scope of Opinion research expanded in the early 2000s. Author (Haque et al., 2018) presented In the field of sentiment analysis, different fields, different kinds of sentiment classification can be conducted, that one can perform fine-grained sentiments classification via focusing on polarity ranging since extremely negative to relatively positive, a different method can be actual intent or emotions detecting, and data can also be evaluated on the aspect level sentiment. Author (Garanayak et al., 2019) described that We love the android phone for privacy; however, there are a few limitations. Even in the usage of technology within this case, the security features and easiness of applications and the polarization of full sentences could not be forecast. The Author (Alsaeedi & Khan, 2019) has

^aBusiness school Nanjing university china, 210000, China, sobiaali614@gmail.com

^bBusiness school Nanjing university china, 210000, China, chenx@nju.edu.cn

^cSchool of International Education, Nanjing University of CM, China, 210000, shentian025@aliyun.com

^dDepartment of Computer Science COMSATS University Islamabad, Sahiwal Campus, 57000, Pakistan, Waqarrajpoot19@gmail.com

^eSchool of Computer Science and Engineering SCE, Taylor's University, Malaysia

NoorZaman.Jhanjhi@taylors.edu.my

*Corresponding Author: Tian Shen. Email: shentian025@aliyun.com

already studied sentiment classification strategies on tweets. They described multiple techniques to sentiment classification such as paper level, and document category methodology. They based their studies on the study of sentiments and sentiment classification. The proposed machine learning analysis followed by a model of feelings and other methods. The documentation properly considered only for survey seem to be either DH or mathematical linguistic.

Text mining has been in progress for a while, but much work directly has taken place in recent years to consider and classify consumer feedback (Tan et al., 2019). The author analyzed consumer opinion on Amazon.com. They have developed a model using Recurrent Neural Networks (RNNs) with a Gated Recurrent Unit (GRU) that has studied lower-dimensional vectors symbol utilizing text vector and content integrating. (Shrestha & Nasoz, 2019), The methods used in this research is a set of 5.3 million product ratings obtained from Amazon.com. Term vectors much more influence paragraph vector (PVs). The (PVs) module develops vector by forecasting the next word given multiple samples context in a paragraph. Computation of the review encoding built through paragraph vector and GRU-derived product integration is used to train the SVM classifier (SVM) to distinguish feelings. With just the embedding of the review, the predicted classifier achieved 81.29 overall accuracies. Inclusion of the product integrating increased accuracy to 81.82 complete. Authors belief that a comparable methodology should be used to learn consumer knowledge (Shrestha & Nasoz, 2019).

(Dong et al., 2018), described The Based on the use of big data for product strategy, Big data analytical methods are pondered to become the utmost important research field of the twenty-first generation. In (Dong et al., 2018) article, sentiment analysis on online product feedback was highlighted. Based on this research work, an innovative system is designed to evaluate various online customer reviews. The fundamental goal of this strategy is to use machine-provide data to assess the preferences of customers. Author (Ireland & Liu, 2018) presented the effect of web reviews on an item for consumption sale. This approach is primarily a collaborative semantic research process. Throughout this work, the author emphasized the significance of evaluating the industry's development by the impact of online feedback. Identify how the comments and ratings of the analysis are interconnected when forecasting the success of the company.

This article analyzed the notion of 'aboutness'

through an essential type of semantic processing task 'subject simulations. All texts are usually composed of several subjects, i.e., topics being spoken about throughout the text. E.g., suppose a project leader at Amazon knows the aspects of a newly launched product (say Amazon Echo) consumers are writing about in their feedback. State they can recognize that 50 % talking for appliances, talking for instrumentals features, and talking for product packaging. In a similar situation where a vendor on Amazon gets much bad feedback and needs to clarify its purpose, this will be incredibly beneficial. We have studied the cases where consumers are hardly misled by overall collective ratings which Amazon without even any polarization between that business evaluation and product review. Hence, we trained our model also to characterize the feedback as positively, negatively and neutrally.

The remainder of the article would be organized as follows. Part 1 is the introduction; part two is a literature review. Part 3 is about data collection and methodology. Part 4 is a result and discussions. The conclusion is presented in Part 5. Finally, future study is addressed in part 6.

2. Literature Review

E-commerce is an essential platform for establishing business contacts between merchants and customers without surpassing geographical barriers. In addition to providing customers with the opportunity to order products online, the e-commerce platform also provides customers with the opportunity to post reviews and ratings about products. This has a significant influence on future customers who want to buy the same product. Therefore, it enables companies to view opinion mining in sentiment analysis of reviews and ratings to monitor product sales and market value. At the sentencing stage, sentiment analysis was conducted, and a complex vocabulary was used with predefined positive and negative expressions to help find the polarity of sentiment. The literature research (Hasan et al., 2018) considered sentiment analysis in sentiment dictionary or machine learning methods. Calculate the semantic direction of sentences, words or phrases for sentences in vocabulary-based sentiment analysis.

Deep learning is developing in the field of emotion analysis. Much current research has been performed. Sentiment analysis can classify and categorize sentiments based on text. Most of the analysis is different reviews from various e-commerce sites, such as finding customer reviews (positive or negative), media posts or many other

commonly used apps. Nevertheless, most researchers use deep learning techniques to study sentiment analysis (Shrestha & Nasoz, 2019). The author performed appropriate textured sentiment analysis of the financial newscast article using the lexicon and machine learning-dependent methodology. Nine studies have been conducted to find more reliable findings. The results of the two methods are often compared (Tammina & Annareddy, 2020). Text categorization accuracy is also a task for text classification. The most suitable method for a straightforward sentiment classification task will be a challenge because the data set's design continuously improves the media data set of all texts in various formats, ratings and blogs. It improves the text data classification model (Mai & Le, 2020). Positive feelings about iPhone privacy and negative feelings about iPhone usage make iPhone predictions more complicated. Author (Rintyarna et al., 2020) described Sentence level sentiment analysis is a method of in-depth analysis of data, which may be valuable for providers and consumers to understand the critical aspects of products that may affect their sales. Author (Patil et al., 2019), presented the "sense of advertising sharing" client work. They proposed a hypothesis that sentiment analysis can explain more deeply the intention of customers to share online advertising. Author (Tang et al., 2019) described, To start the extraction phase with an algorithm, the seed opinion lexicon, the set of general words, data analysis and extraction rules are introduced into the proposed algorithm. Author (Cheng et al., 2018) described the learning is a type of cognitive intelligence. (Vasan, Alazab, Wassan, Safaei, et al., 2020) and (Vasan, Alazab, Wassan, Naeem, et al., 2020) described the Pre-data of test data accompanies the research, feature extraction, feature collection, validation and interpretation of testing data. Optimization of cloud systems has been discussed in (Shafiq et al., 2019). Author (Yu et al., 2019) presented Analyze emotions by using attention-based neural networks. Two-way gated loop unit. In recent years, the application of sentiment classification has been expanded. For example, sentiment analysis techniques used to filter feedback on science papers (Keith Norambuena et al., 2019). Sentiment classification is applied to the Twitter text presented in (Ramanathan & Meyyappan, 2019). to scrutinize visitors' feedback. Among other works, sentiment analysis applications for product reviews were introduced in (Jagdale et al., 2019). (Deshmukh & Tripathy, 2018) described Machine learning methods rely on well-known machine learning (ML)

algorithms to solve sentiment analysis into ordinary text classification problems using syntax and language features). Yang et al. has developed a domain sentiment dictionary in (Kai et al., 2017).

The authors (Alsaeedi & Khan, 2019) described Twitter sentiment analysis methods like machine learning techniques and word frequency-based techniques. The efficiency of SVM and NB in performing SA has been compared. When bigram models were used, Support vector machine provided the highest accuracy of 85 %. Different methods can also be compared to determine the best way to detect sentiments in tweeting data. Author (Chakraborty et al., 2019) present that information-related writing generally involves overcoming many challenges in achieving goals such as equity, relevance, and accountability.

Author (Kaur & Kumari, 2016) presented that Social networks and shopping websites act as media for users to post product reviews, which can be used for classification. Author (Wahyudi & Kristiyanti, 2016) defined that online shopping has gradually become a developed field due to the development of computer and Internet . When you buy any product on the Internet, the ratings and reviews of existing users significantly impact. It also affects the reputation of companies that use online markets to promote their products.

Main contribution

Our main contribution is producing a novel methodology that proposed utilizing reflective elements based on the item's characteristics. We collected the data from the data world website the system performs pre-processing operations such as tokenization, extracting stop-words from databases to extract meaningful information such as positivity or negativity. As we know, in analyzing sentiments, there are two terms called "polarity" and "subjectivity". Subjectivity applies to the individual's values, opinions, or personal sentiments, whereas polarity relates to thoughts conveyed favourably/positively, negatively, or neutrally. We have evaluated each product reviews and amazon user reviews. We used 28,000 user reviews with over 60 product types in this dataset. The ratio of each product is different from another product. We have analyzed the polarity for measuring the amazon product semantical analysis. Thus, if the bar height in the graph towards 1 means the product reviews are positive, and if the bar height looks down or -1, that means the semantics of that product is negative same as well when we analyze for an amazon user review. The user's review is different from another user. We have

analyzed the polarity for measuring the amazon user's semantics. Thus, if the bar height in the graph towards five means the product reviews are positive, and if the bar height looks down, that means the semantics of that users are negative. The primary study aims to analyze this data's aspect level is a huge benefit to marketers to grasp the preferences of consumers better and then develop their behavior accordingly.

3. Methodology for The Sentiment Analysis of Amazon

Specific work has been performed on finding the analysis of online social networks. They can be organized in three different categories geometrical, statistically, and topological. Mostly analysis systems used these steps, i.e., detection, extraction, selection, and classification. They work accurately to determine the OSN network analysis or graph visualization using mutual sharing work of different algorithms and techniques. This research work also provides a small overview of social network analysis with visualization of graphs. Firstly, a selection of social network websites for analysis such as Amazon. The standard features of social networks sites will summarize by implementing different experiments on social networks such as Amazon. Social Network visualization techniques help find different relationships and features between different entities existing in social network websites users. Different software is open-source and used for visualizing and analyzing different network graphs. We use user reviews from Amazon user reviews to perform sentiment analysis. To perform the different tasks in sequence, let us quickly read the CSV file. When we are solving natural language processing (NLP), we have removed the stop words first. The number of stop word can also calculate some additional information that we have lost. By using the Natural Language Toolkit (NLTK) library, we have imported stop words. We have learned so far how basic features extraction from reviews.

Before text and feature extraction, we have cleaned our dataset for obtaining the best features. We have achieved these features by doing basis pre-processing steps on the training dataset. We use the "textblob" library for spelling correction because this step is more useful in pre-processing to reduce the copies of words to understand those words, which are entirely not understandable in reading. We used tokenization to segment the user reviews into a series of terms or phrases. We have converted user reviews into such a blob and then transformed them into a collection of terms using a

"textblob" library. By using a simple rule-based approach, stemming refers to the deletion of suffices, like "ing," "ly," "s," and so on. We have used a porter stemmer from the NLTK library as illustrated below Fig. 1.

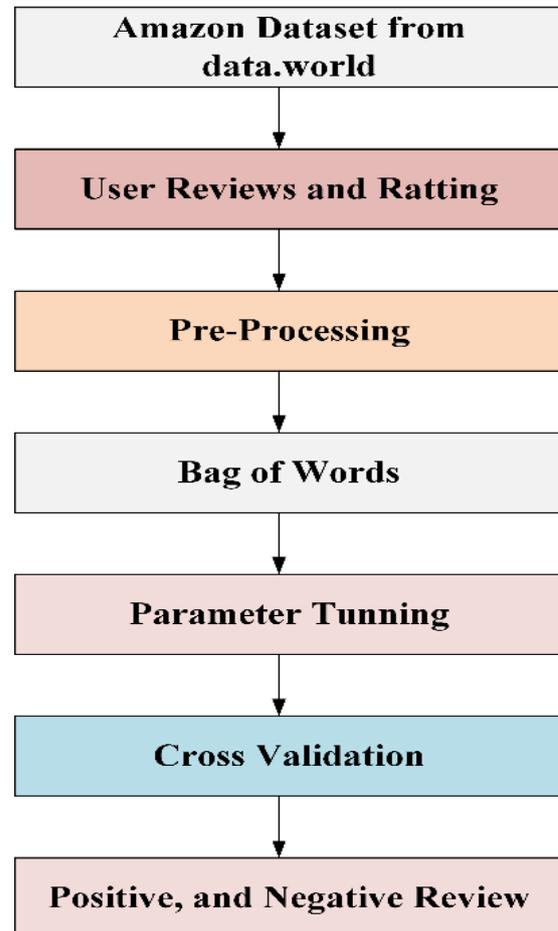


Figure 1. The overall procedure diagram of the sentiment analysis of Amazon user reviews

4. Data Collection

A model for better sentiment analyzes was designed using an ensemble approach to increase the accuracy and efficiency of reviews for keyword trends. First, we must identify the subjective or objective weather of a sentence. If the sentence is subjective, then the SSA decides whether the sentence is positive or negative. Researchers have analyzed no primary variations between the ratings of sentences and document levels because we can say that sentences are small documents only. The data stream has been collected from the data. World. The data collection is a database from over 28,000 consumer reviews of Amazon products. We have used the second category of the dataset in which we have 28,000 user reviews for 60 products of Amazon, as shown in Fig. 2.

id	reviews.rating	reviews.text
2	AVpgNzjwL43Kpxn	3 I order 3 of them and one of the item is bad quality. Is missing backup spring so I have to put a pcs of aluminum to make the battery work.
3	AVpgNzjwL43Kpxn	4 Bulk is always the less expensive way to go for products like these
4	AVpgNzjwL43Kpxn	5 Well they are not Duracell but for the price i am happy.
5	AVpgNzjwL43Kpxn	5 Seem to work as well as name brand batteries at a much better price
6	AVpgNzjwL43Kpxn	5 These batteries are very long lasting the price is great.
7	AVpgNzjwL43Kpxn	5 Bought a lot of batteries for Christmas and the AmazonBasics Cell have been good. I haven't noticed a difference between the brand nam
8	AVpgNzjwL43Kpxn	5 ive not had any problame with these batteries have ordered them in the past been very pleased.
9	AVpgNzjwL43Kpxn	5 Well if you are looking for cheap non-rechargeable batteries that last quite a while then these are perfect. Nothing more to say.
10	AVpgNzjwL43Kpxn	3 These do not hold the amount of high power juice like energizer or duracell, but they are half the price.
11	AVpgNzjwL43Kpxn	4 AmazonBasics AA AAA batteries have done well by me appear to have a good shelf life. I'll buy them again.
12	AVpgNzjwL43Kpxn	5 I find amazon basics batteries to be equal if not superior to name brand ones. Can't believe I didn't start buying them sooner! The package
13	AVpgNzjwL43Kpxn	3 When I first started getting the Amazon basic batteries I really liked them. With recent purchases, they do not seem to last like they had, c
14	AVpgNzjwL43Kpxn	5 Use it for my fish tank's light at night and works great, I love how you can easily switch it off and on if you want it on while guests are there
15	AVpgNzjwL43Kpxn	5 just got em so I can't really comment on how good the do the job, good price, quick delivery but have only put two into one of my keyboar
16	AVpgNzjwL43Kpxn	5 we have many things that need aa battery they are great
17	AVpgNzjwL43Kpxn	5 Thankful that I was able to find on Amazon for a great price and even better shipping. Arrived in perfect condition and did exactly what I n
18	AVpgNzjwL43Kpxn	1 I don't know if I would buy thus brand again seems like they don't last as long as Duracell
19	AVpgNzjwL43Kpxn	2 In my opinion these did not last anywhere near as long as Duracel in things like LED candles (which is crazy) and trail cameras. Cameras we
20	AVpgNzjwL43Kpxn	4 They don't last as long as the brand name but are good enough considering they are much cheaper.

Figure 2. The small sample screenshot of the dataset of the amazon product user reviews

The amazon dataset is from data. World website. amazon.com is the most popular websites. The website contains a unique number of product feedback. We selected 28000 feedback randomly from the website. Review rankings vary from 1 to 5 at amazon.com. One-thousand ratings have been chosen for each rating system. In two polarity sentiment classification experimentations, we perceive 1 and 2 to be negative and 4 and 5 to be positive. Neutrally feedback using a rating of the three shall be extracted since the data. For 3, polarity sentimental classification tests, we assign 1

and 2 to be still negatively, 3, to be neutral, 4 and 5 being positive. On behalf of the 5-polarization emotion experimental research, we assign 1 for significantly positively, 2 for still be negative, 3 for neutrally, 4 also be positively, 5 is being strong positively.

4.1 Calculation Metrics

To assess the performance of sentimental classification, several standard metrics, precision, recall and F-Measurement were adopted. These metrics are defined as in Eq. 1 to Eq. 2.

$$R_i = \frac{\text{The belonging documents of class } i}{\text{The belonging documents of class } i \text{ before classification}} \times 100\% \dots\dots\dots(1)$$

$$P_i = \frac{\text{The belonging documents of class } i}{\text{The belonging documents of class } i \text{ before classification}} \times 100\% \dots\dots\dots(2)$$

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{N} \times 100\% \dots\dots\dots(3)$$

$$\bar{P} = \frac{\sum_{i=1}^n P_i}{N} \times 100\% \dots\dots\dots(4)$$

$$F - \text{measure} = \frac{2\bar{P}\bar{R}}{\bar{P} + \bar{R}} \dots\dots\dots(5)$$

Wherever Ri is the polarity retrieval ratio *i*, *p_i* is the precision, *R* is the average recall rate, *P* is the average precision, and (*N*) is the number of classes.

5. Result and Discussion

Sentiment study is one of the most current and demanding fields of science. The presented work

deals with one of the most significant challenges in sentiment analysis in bipolar words. To optimize sentiment analysis, companies need to use the latest tools and approaches. The definition in the background's presence, its effect on the product's overall ranking, and the essential feature of the research were analyzed, and results were excellent. The study was made. Now it takes a few days for

the tool or software to understand the product's behaviour on the market, both for the customer and the developers. This work has been applied to the Python Framework for analysis purposes. The Python Framework for research purposes. We have used Google co-lab as an IDE for the experimentation. In this research work, we have evaluated the average semantical analysis against each amazon product, also the average rating against each product of amazon. We also have analyzed the amazon all reviews semantic analysis.

5.1 Libraries Used in The Sentiment Analysis

```
import pandas as pd
import json
from textblob import TextBlob
import numpy as np
import matplotlib.pyplot as plt
# import nltk
# nltk.download('punkt')
# nltk.download('averaged_perceptron_tagger')
# nltk.download('averaged_perceptron_tagger')
# nltk.download('brown')
```

5.2 Amazon Products Semantical Analysis

In Fig. 3, we evaluand each product reviews that we have used in amazon user reviews. In this dataset, we have used more than 60 types of products. The ratio of each product is different from another product. We have analyzed the polarity for measuring the amazon product semantical analysis. As we know, there are two terms called "polarity" and "subjectivity" that could be describes in the analysis of sentiments. Subjectivity denotes to the individual views, thoughts, or individual emotions, whereas polarization refers only to emotions reflected positively, adversely, or neutrally. Sentiment Classification shall cover the scope of the phrase, the document, and sub-phrase work. In this analysis, we have analyzed the polarity as "plt.ylabel('Polarity -1 to 1 (-1 Negative 1 positive)')". The meaning of that of the height of the bar in the graph towards 1 means the product reviews are positive and if the height of the bar looks downs this means the semantics of that product are negative. The source code of the following graph is:

```
products=dataset['id'].unique()
products_reviews=[]
for product in products:
product_data=dataset[dataset['id'] == product]
products_reviews.append(product_data)
pro_sentiments=[]
each_review_semantics=[]
for single_product in products_reviews:
```

```
sentiments=[]
for eachreview in single_product['reviews.text']:
blob = TextBlob(str(eachreview))
sentiment=blob.sentiment.polarity
sentiments.append(sentiment)
each_review_semantics.append(sentiment)
pro_sentiments.append(np.sum(sentiments)/len(sentiments))
fig = plt.figure()
ax = fig.add_axes([-1,-1,1,1])
langs = np.arange(len(pro_sentiments))
sentiments_rec = pro_sentiments
ax.bar(langs,sentiments_rec)
plt.ylabel('Polarity -1 to 1 (-1 Negative 1 positive)')
plt.xlabel('Average semantical analysis against each product')
plt.title('Amazon products semantical analysis')
#plt.ylim([-1,1])
plt.show()
```

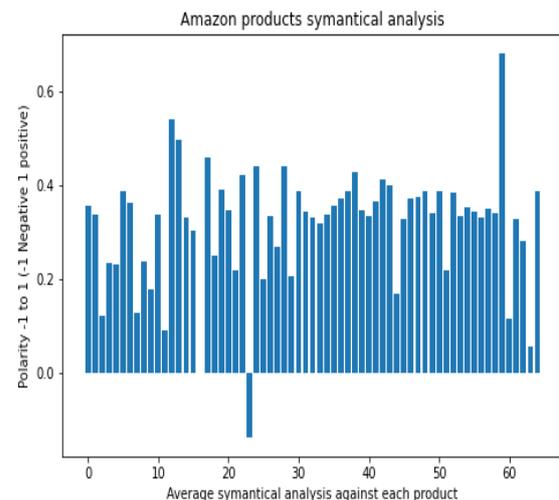


Figure 3. The bar visualization of the amazon product semantical analysis

In Fig. 4, shows the same results of the amazon product semantical analysis of user reviews based on the polarity (-1, 1). In this figure, the visualization of the product looks more precise in the form of dots. The source code of the following graph is:

```
fig = plt.figure()
ax = fig.add_axes([-1,-1,1,1])
langs = np.arange(len(pro_sentiments))
sentiments_rec = pro_sentiments
ax.scatter(langs,sentiments_rec)
plt.ylabel('Polarity -1 to 1 (-1 Negative 1 positive)')
plt.xlabel('Average semantical analysis against each product')
plt.title('Amazon products semantical analysis')
#plt.ylim([-1,1])
plt.show()
```

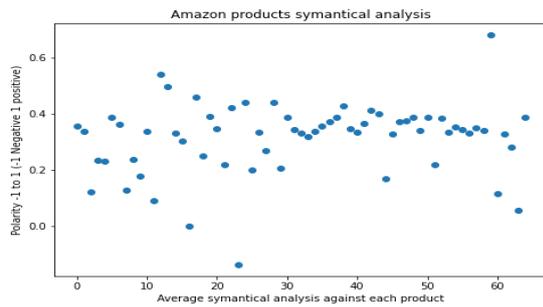


Figure 4. The dot visualization of the amazon product semantical analysis

5.3 Amazon All Reviews Semantical Analysis

In Fig. 5, we evaluand each product reviews that we have used in amazon user reviews. In this dataset, we have used 28000 user reviews. The ratio of reviews for each product is different from another product. We have analyzed the polarity for measuring the amazon product semantical analysis. As we know, there are two terms called "polarity" and "subjectivity" that could be described in the analysis of sentiments. Subjectivity refers to the individual's views, thoughts, or personal emotions, whereas polarity refers only to emotions reflected positively, adversely, or neutrally. Sentiment Classification shall cover the scope of the phrase, the document, and sub-phrase work. The meaning of that of the dots' height in the graph towards 1 means the product reviews are positive, and if the height of the dots is looked downs, this means the semantics of that product is negative. The source code of the following graph is:

```
fig = plt.figure()
ax = fig.add_axes([-1,-1,1,1])
langs = np. range(len(each_review_semantics))
sentiments_rec = each_review_semantics
ax.scatter(langs,sentiments_rec)
plt.xlabel('Each review semantics')
plt. title('Amazon all reviews semantical analysis')
#plt.ylim([-1,1])
plt.show()
```

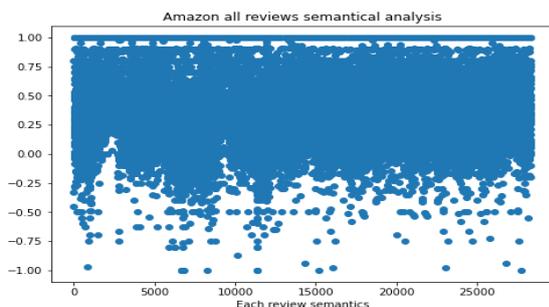


Figure 5. The dot visualization of the amazon all reviews semantical analysis

5.4 Amazon Product User Rating

In Fig. 6, we evaluated each product reviews that we have used in amazon user reviews. In this dataset, we have used 28000 user reviews with more than 60 product types. The ratio of reviews for each product is different from another product. We have analyzed the polarity for measuring the amazon product semantical analysis. The source code of the following graph is:

```
pro_ratings=[]
for single_product in products_reviews:
    rat=[]
    for eachreview in single_product['reviews.rating']:
        rat.append(eachreview)
    pro_ratings.append(np.sum(rat)/len(rat))
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
langs = np.arange(len(pro_ratings))
ratings_rec = pro_ratings
ax.bar(langs,ratings_rec)
plt.ylabel('User ratings (0 for poor 5 for excellent)')
plt.xlabel('Average user ratings against each product')
plt.title('Amazon products average user rating')
#plt.ylim([-1,1])
plt.show() the meaning of that of the height of the bar in the graph towards 5 this means the product reviews are positive rating and if the height of the bar looks downs this means the semantics of that product is negative in rating.
```

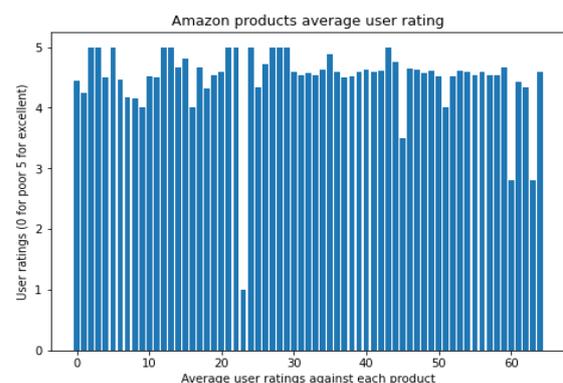


Figure 6. The bar visualization of the amazon product user rating

6. Study Implication

Amazon has become the most popular hub for online purchasing. The useability of each user can be detected by using user comments. The user comments can show the behavior of the buyers satisfactions. The main implication of this research is that we can develop a system that can give us a certain report for any product sold from the Amazon website. We can develop this system by

using different machine learning techniques. We have trained our system on different user reviews and check the polarity of different user feedbacks. After We can then test the reliability of the product and increase the product's productivity.

7. Conclusion

Social Network Analysis (SNA) methods can cover extensive networks from thousands or millions of nodes and links of a graph. It characterizes the network structure in the form of nodes (specific actors, people, or things) that are a network used or within the network and edges or links between nodes that connect the network. In this research article, a novel methodology was proposed utilizing reflective elements based on the item's characteristics. We collected the data from the data world website the system performs pre-processing operations such as tokenization, extracting stop-words from databases to extract meaningful information such as positivity or negativity. As we know, in analyzing sentiments, there are two terms called "polarity" and "subjectivity". Subjectivity applies to the individual's values, opinions, or personal sentiments, whereas polarity relates to thoughts conveyed favourably, negatively, or neutrally. We have evaluated each product reviews and amazon user reviews. We used 28,000 user reviews with over 60 product types in this dataset. The ratio of each product is different from another product. We have analyzed the polarity for measuring the amazon product semantical analysis. Thus, if the bar height in the graph towards 1 means the product reviews are positive, and if the bar height looks down, that means the semantics of that product is negative same as well when we analyze for an amazon user review. The user's review is different from another user. We have analyzed the polarity for measuring the amazon user's semantics. Thus, if the bar height in the graph towards 5, that means the product reviews are positive, and if the bar height looks down, that means the semantics of that users are negative.

8. Future work

This study further can be extended to apply on the rest of the amazon datasets for more directions and aspects related to the review system.

ACKNOWLEDGEMENT

The work was supported in part by the National Natural Science Foundation of China under Grant No. 71771118,72071104, by the Ministry of Education Humanities and Social Sciences

Foundation of China under Grant No.18YJCH146 and by the Key Project of Jiangsu Social Science Foundation under Grant No. 20GLA007.

References

- [1] Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2019.0100248>
- [2] Chakraborty, R., Bhavsar, M., Dandapat, S. K., & Chandra, J. (2019). Tweet summarization of news articles: An objective ordering-based perspective. *IEEE Transactions on Computational Social Systems*. <https://doi.org/10.1109/TCSS.2019.2926144>
- [3] Cheng, Z., Ding, Y., Zhu, L., & Kankanhalli, M. (2018). Aspect-aware latent factor model: Rating prediction with ratings and reviews. *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*. <https://doi.org/10.1145/3178876.3186145>
- [4] Deshmukh, J. S., & Tripathy, A. K. (2018). Entropy-based classifier for cross-domain opinion mining. *Applied Computing and Informatics*. <https://doi.org/10.1016/j.aci.2017.03.001>
- [5] Dong, L. yu, Ji, S. juan, Zhang, C. jin, Zhang, Q., Chiu, D. K. W., Qiu, L. qing, & Li, D. (2018). An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2018.07.005>
- [6] Garanayak, M., Mohanty, S. N., Jagadev, A. K., & Sahoo, S. (2019). A recommender system using item-based collaborative filtering (CF) and K-means. *International Journal of Knowledge-Based and Intelligent Engineering Systems*. <https://doi.org/10.3233/KES-190402>
- [7] Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*. <https://doi.org/10.1109/ICIRD.2018.8376299>
- [8] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications*. <https://doi.org/10.3390/mca23010011>
- [9] Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP Journal of Manufacturing Science and Technology*. <https://doi.org/10.1016/j.cirpj.2018.06.003>

- [10] Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-981-13-0617-4_61
- [11] Kai, Y., Cai, Y., Dongping, H., Li, J., Zhou, Z., & Lei, X. (2017). An effective hybrid model for opinion mining and sentiment analysis. *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*. <https://doi.org/10.1109/BIGCOMP.2017.7881759>
- [12] Kaur, B., & Kumari, N. (2016). A Hybrid Approach to Sentiment Analysis of Technical Article Reviews. *International Journal of Education and Management Engineering*. <https://doi.org/10.5815/ijeme.2016.06.01>
- [13] Keith Norambuena, B., Lettura, E. F., & Villegas, C. M. (2019). Sentiment analysis and opinion mining applied to scientific paper reviews. *Intelligent Data Analysis*. <https://doi.org/10.3233/IDA-173807>
- [14] Mai, L., & Le, B. (2020). Joint sentence and aspect-level sentiment analysis of product comments. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-020-03534-7>
- [15] Patil, P. P., Phansalkar, S., & Kryssanov, V. V. (2019). Topic modelling for aspect-level sentiment analysis. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-981-13-1610-4_23
- [16] Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*. <https://doi.org/10.1109/ICBDS.2019.8645596>
- [17] Rintyarna, B. S., Sarno, R., & Fatichah, C. (2020). Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context. *International Journal of Information and Decision Sciences*. <https://doi.org/10.1504/IJIDS.2020.104992>
- [18] Shafiq, D. A., Jhanjhi, N., & Abdullah, A. (2019). Proposing A Load Balancing Algorithm for the Optimization of Cloud Computing Applications. *MACS 2019 - 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics, Proceedings*. <https://doi.org/10.1109/MACS48846.2019.9024785>
- [19] Shrestha, N., & Nasoz, F. (2019). Deep learning sentiment analysis of Amazon.com reviews and ratings. In *arXiv*. <https://doi.org/10.5121/ijsc.2019.8101>
- [20] Tammina, S., & Annareddy, S. (2020). Sentiment analysis on customer reviews using a convolutional neural network. *2020 International Conference on Computer Communication and Informatics, ICCCI 2020*. <https://doi.org/10.1109/ICCCI48352.2020.914086>
- [21] Tan, W., Wang, X., & Xu, X. (2019). Sentiment Analysis for Amazon Reviews. *Reseachgate, March*, 0–9. <https://doi.org/10.13140/RG.2.2.13939.37920>
- [22] Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*. <https://doi.org/10.1016/j.ins.2019.02.064>
- [23] Vasan, D., Alazab, M., Wassan, S., Naeem, H., Safaei, B., & Zheng, Q. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. *Computer Networks*. <https://doi.org/10.1016/j.comnet.2020.107138>
- [24] Vasan, D., Alazab, M., Wassan, S., Safaei, B., & Zheng, Q. (2020). Image-Based malware classification using ensemble of CNN architectures (IMCEC). *Computers and Security*, 92, 101748. <https://doi.org/10.1016/j.cose.2020.101748>
- [25] Wahyudi, M., & Kristiyanti, D. A. (2016). Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization. In *Journal of Theoretical and Applied Information Technology*.
- [26] Yu, Q., Zhao, H., & Wang, Z. (2019). Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3357254.3357262>