Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Aspect-Based Analysis of Social Media Post and its Review

P. Rethina Sabapathi [1], Dr. K. P. Kaliyamurthie [2]
[1] Research Scholar, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.
[2] Dean/CSE, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.
[1] rethinasabapathi1@gmail.com, [2] kpkaliyamurthie@gmail.com

Article Info

Page Number: 1017-1030

Publication Issue:

Vol. 71 No. 3s (2022)

Article History

Article Received: 22 April 2022

Revised: 10 May 2022

Accepted: 15 June 2022

Publication: 19 July 2022

Abstract

The exponential growth of the social media provides the platform to users for sharing and generating information. Once the information is shared in the social media it will receive lots of commands, reviews and reply messages. In some cases the message received will not be associated with the post shared. Taking this problem into consideration, the author proposed an aspect based sentimental analysis method for analyzing the content posted and its received messages. The sentimental score output mostly used to identify the positive, negative and neutral sentences with respect to the exacting aspect. Then a total score formulae is used to determine the association between the content posted with its reply messages. Finally, an experiment is designed and processed in different way to visualize the data are provided.

Keywords: Social media content, text mining, sentimental analysis, aspect based sentimental analysis.

1. Introduction

Social media is the interactive technology that allows the user to create and share information by the means of virtual communities and networks. The speedy development and rising fame of social networks have been made in get hold of information from social networking websites for analyzing people's behaviors. Users usually access social media services by the use of desktops, laptops and mobile devices. As users engage with these electronic services, they create highly interactive platforms through which individuals, communities, and organizations can share, co-create, discuss, participate, and modify user-generated content posted online. Social media content post is pieces of information were admin of the group will share, that can be either positive or negative or neutral. Based on the content posted other users will reply through comments that may or may be dependent on the post created.

In the proposed work the author focuses on identifying the association between content posted on social media group and its reply messages. The social media content posted and users review or reply messages are examined using a text mining technique called sentimental analysis. Sentiment Analysis [1] is the procedure of formative whether a script is positive, negative or neutral. Many different sentimental analysis are been introduced, in this methodology the main aim is to focuses on the sentences so it is appropriate to use aspect based sentimental analysis [15]. Aspect – Based sentimental analysis is used to analyze the specific text with different aspects.

Furthermore, our proposed method can be applied on different social media platforms for performance comparisons. The manuscript is structured as follows. Section 2 reviews the related works. Section 3 describes the proposed methodology of sentimental analysis of social media posts and its reply. Section 4 introduces the experiment and result based on the proposed work. Finally, the paper is concluded and suggest for future work.

2. Related Works

Lin Yue et al., [1] presents a survey paper based sentimental analysis on social media. The scope of this survey addressed several essential aspects. And also the paper focuses on the area of sentiment analysis and its three different perspectives where there are compared and categorized in huge quantities. On the added part, they introduced superior tools and unlike type of data, in addition to their boundaries. Based on it the sentimental analysis are discussed and identified. Zoha Rahman et al.,[2] proposed a study that have been analyzed 17 international brands of total 1325 brand posts from Electronics companies. Totally 9 months brand's fan page data is considered for the content analysis. Where, the descriptive statistics of each page is summarized. Fang et al., [3] the papers aim is to handle the sentiment polarity categorization problem. Datasets used for the proposed work are collected from amazon online product reviews. Research is done for sentence level and review level categorization.

Sun, Beiming, and Vincent TY Ng., [4] aim is to analyze the sentimental influence of posts with the results on various social media platforms. The paper focuses on the three research question. Based on the following method the paper is processed. Firstly the emotion detection dictionary is created both in Chinese and English, then the sentimental analysis are performed based on SentiWordNet resource and uses graph model to find the relationship between the posts shared. And, finally sentimental influence of the post is identified using their receptors and their sentiment of the influencers. Prakruthi v et al., [5] paper focuses on the analyzing sentiments of tweets that are useful in deciding people's option as positive, negative or neutral. Tweets from twitter are directly taken from the twitter API to construct a sentimental classification. The outcome analysis is remarked using visualization methods. Wanxiang Che et al., [6] author addressed a framework called Sent_Comp (sentiment sentence compression) step earlier to aspect based sentimental analysis. The Sent_Comp will remove unwanted data from sentiment analysis and also converts the complicated sentences into easier parses. This step is especially proposed to improve the performance of the aspect based sentimental analysis.

Vanaja, Satuluri, and Meena Belwal., [7] proposed a work based on analyzing the E-Commerce portals that generates huge data every day depends on customer reviews. Amazon customer review data are used and aspect feature for each review data were found along with part of speech tagging, by applying classification algorithm positive, negative and neutral of each review is determined. Schouten, Kim, and Flavius Frasincar., [8] surveyed a paper focuses on aspect based sentimental analysis. The fine grained methods are useful for various applications. Present solutions are divided depends on either aspect detection or sentiment analysis or both. The breakdown based type of algorithm is provided. Aspect level sentiment analysis is considered as one of the promising future direction is determined [16]-[17].

The above survey are based on the sentimental analysis about sentence level and also some papers are based on aspect based sentimental analysis which helps to analysis the sentence in more deeper manner based on aspect features. From this survey, it is appropriate to undergo the methods of aspect based sentimental analysis to get better performance is been determined.

3. Proposed Methodologies

The methodology consists of three processes mentioned in figure 1. Initially, process of data preparation and cleaning is done. Secondly, analyses of the content posted in the social media group and its replies/comments. Last part is to identify association between content posted and its reply based on association rules.

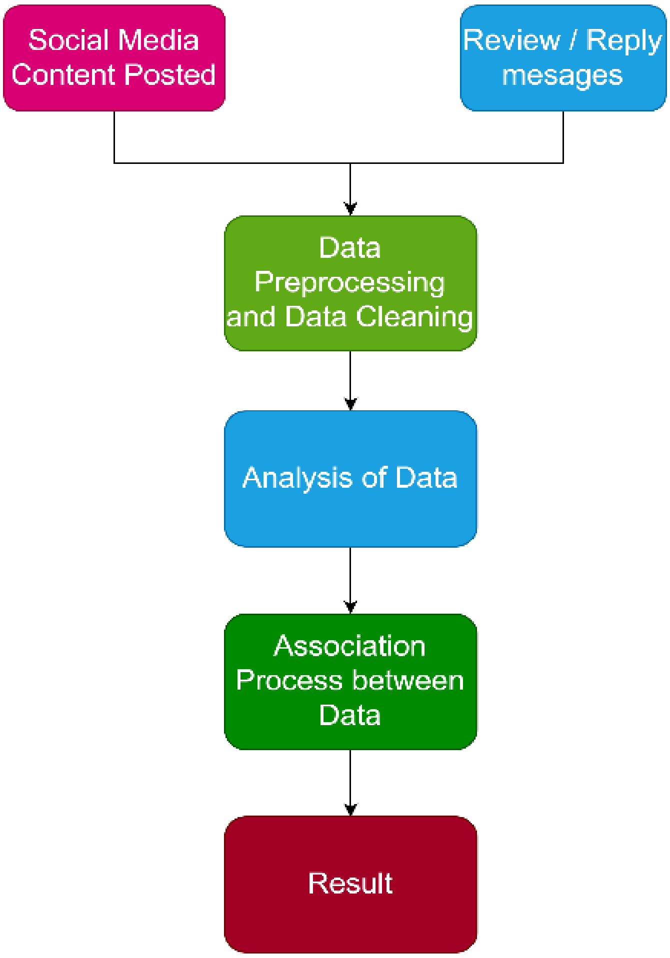

Figure 1; Overall Methodology

3.1. Data Preparation and Data Cleaning

In this phase the collected social media content posted are pre-processed. Each word in the data is tokenized. Tokenization is the act of breaking the text into phrases, words, symbols, date, time and other elements. [12] It plays a major role in removing the unnecessary words in text, like removing special symbol associated with data. [13] Cleaning of social media content is done by removing additions of text like URL's, numbers and special characters which shortens the size for comparison.

Table 1. Data Preprocessing and Data Cleaning

Post	Input	Output
1	@westdalesschools Hi, Here is a free classroom app. Could you give it a try? https://t.co/5VwXNOQgbF	Hi, Here is a free classroom app. Could you give it a try
2	Let us celebrate our #EagleStrong heroes! We want to honor and celebrate our Grand Rapids Christian Schools alumni and parents who are working in response to the COVID-19 crisis. If you or someone you know is helping to combat this pandemic – let us know! ow.ly/qLKK50zjtgF	Let us celebrate our heroes! We want to honor and celebrate our Grand Rapids Christian Schools alumni and parents who are working in response to the COVID-19 crisis. If you or someone you know is helping to combat this pandemic – let us know.

The above table 1 represents the data preprocessing and data cleaning of the social media post. In both post 1 and 2 the words next to '#', '@' symbols and the hyperlink content are removed. Then symbols like 'exclamatory' and 'question marks' are been removed by the rules of data preprocessing and cleaning. So that the analysis of the sentence would be easy while POS tagging. After this process the data has been exported into the framework for the analysis process.

3.2. Analysis of Social Media Content Post and its Replies

Aspect-based sentiment analysis performs more in-depth sentiment analysis of data. It categorizes the data by aspect and analysis the sentiment to each one of the texts [14]. It is used to analysis massive quantity of data in detail. In our proposed work aspect based sentimental analysis is used to analyze the content posted and replies or commands based on the posts.

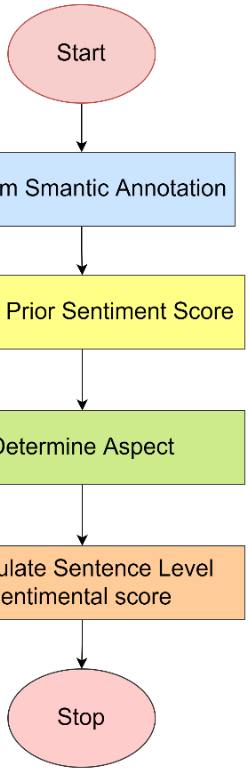

Figure 2; Work Flow of Social Media Content Analysis

Figure 2 gives an overview of the process involved in analysis of the social media content and its review. Firstly, for the review sentence and content posted sentences are initialized. Then the semantic annotation is performed and the sentences are broken into independent sentences. The prior score value is assigned for the individual sentence. Later, the aspect of the sentence is determined and the finally the sentence level score value is determined.

3.2.1. Semantic Annotation

The objective of semantic annotation is to tag the text with semantic categories, such as subject, software, community, action, entertainment based on the post represented in table 1. Data about each sentence related to each category are collected and stored in specific feature list represented in table 2. From this dataset, sentences or paragraphs covering each review aspect are read and important feature terms specific to each aspect are extracted. Annotation is performed by tagging the terms that matched the ones in the aspect and specific feature lists. For the matching, firstly, terms in sentences are tagged with part-of-speech (POS) [11], and they are compared with the entries with the same POS tags. For instance, the entry 'application' in the list matches any form of the feature in the list e.g. 'app' or 'apps'. [14] Similarly, the noun entry 'parent' matches either singular or plural noun 'parents', but not the verb term ‘parent’. The longest matching method is applied, and each entry in the feature lists is given a unique annotation.

Table 2. Aspects based Feature List

Aspects	Aspect based feature list
Subject	Classroom, school, Alumni, Parents, COVID-19, Pandemic.
Software	Application, Install, Free trial, Uninstall, Upload, Download, Upgrade.
Community	Hindu, Muslim, Christian
Action	Try, Give, Want, Honor, Work, help, Response
Entertainment	Celebrate, Dance, Song, Music, Act

3.2.2. Prior Sentiment Scoring

Once the semantic annotation is completed the sentence is shattered into individual words. Those words are considered for the prior sentiment scoring method [15]. It is used to calculate the sentiment score of the particular words. The prior sentiment scores of the words range between –1 (negative) and +1 (positive), with 0 being neutral. The prior sentimental score of the content posted and reviews are stored for further uses of analysis.

3.2.3. Determining Aspects

The next process after prior sentiment score is to determine the aspects for each word in the sentence. In this process the semantic tags used from the semantic annotation process. If a semantic tag representing a specific aspect is found in a sentence, [9][10] then it is categorized into the specific aspect. If no semantic tag is found in a sentence and there is a previous or a following word in the sentence, then it is categorized into the same target aspect of the following word. Otherwise, it is categorized into the overall review aspect. After determining the aspects, sentence level scoring is processed.

3.2.4. Sentence-level Sentiment Score

A sentence can have one word about one aspect, multiple words about the same aspect, or multiple words about various aspects. Therefore, after calculating the prior sentiment scores and determining their review aspects in a sentence, [15] the sentiment score for each review aspect is calculated by grouping together the same aspect words and taking the average score. Then, the sentence-level sentiment score is calculated by averaging the sentiment scores for all aspects based on scores. The above four process is used for analysis of content posted and its review. Once the analysis part is ended, the content posted and its review sentences are evaluated using the association rule to determine their dependency.

3.2.5. Association between Social Media Content Post and Replies

In this process the above formula is applied to determine the relationship between the content posted and its replies (r1, r2, r3,…,rn). Where the score of content posted and its each reply score is taken separately for evaluation. Based on the formula given above the scores are evaluated and the total score is determined for finding association between the sentences. Here we can assume that if the total score is above standard value 2 then the sentences are considered to be associated else not associated.

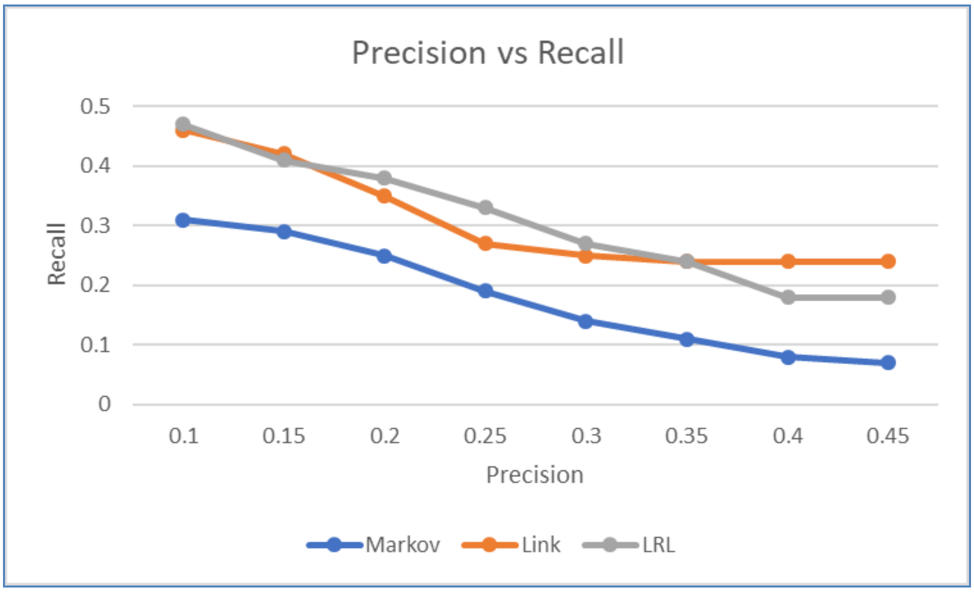

Figure 3; Accuracy Analysis

Let us consider the score of replies = 1, -1, 1,…,n and Content score = 3

For r1, Total score = 70 % (3) + 30 % (1) = 2.4 evaluated based on the formula. So, based on the total score reply one is associated with the content is determined and that is tabulated in below section of Table 3.

4. Results and Discussion

This section gives the details about experimental process of total number of reply messages received for the particular post. Followed by the tabulation and a graph that describe reply messages associated with content posted based on the score values.

4.1. Dataset and Experimental Setup

For the time being, there is no dataset for the purpose of detecting social media sentiments. As a result of the explosive development in user-generated content production and sharing on platforms like Instagram, the number of people creating and sharing their own material has skyrocketed. There are genuine people who are happy to talk about their favorite brands and outfits, making them a reliable and current resource for fashion sentiment analysis. We provide the community with a large-scale sentiment detection dataset linked to fashion.

Over 12,000 social media posts are included in this database, each with an accompanying image and text. To build the dataset, we use a set of pre-defined hashtags to capture similar Instagram posts. Those postings with no hashtags or with popular hashtags that don't always convey strong feelings are also included in our database. Following the collection of postings, the data is cleaned in the manner listed below.

A pre-trained object identification algorithm is used to remove postings that lack any human face or body parts. If a post has fewer than five tokens or emojis exclusively, we delete it. In the event that there are two posts with the exact same image, one of them will be deleted. Only the top five hashtags are retained when a post includes more than five hashtags. We use the approaches to preprocess the noisy texts. Manual annotations and classification of all postings into positive, negative, or neutral sentiment are the last steps in getting a true sense of community sentiment.

A team of three annotators has been engaged to assess the emotional polarity of all of the entries on the site. For every post we make sure that both the image and the text contribute to the sentiment label. Annotators are asked to debate and re-annotate postings whose sentiment findings differ following the initial round of annotation. Advertising and promotions are categorized as neutral since they don't convey any kind of emotional content. The average number of words in a post in our sample is 25.5.

4.2. Experimental Outcomes

The Table 3 is designed to decide whether the reply messages are associated with the content posted or not based on the score values. Here the 'Post' column is filled by social media content posts, followed by the 'Replies' part. Table 3 shows only first seven reply messages of the post. Then the 'Content posted score' and 'Reply message score' determined using sentence level sentimental analysis are noted in the following columns, proceeded with the 'Final score' and 'Result'.

Table 3. Determine Reply Messages Associations with Score Values

Post	Replies	Content Posted Score	Reply Message Score	Final Score	Result
Hi, Here is a free classroom app. Could you give it a try	R1	3.00	1.00	2.40	Associated
	R2	3.00	-1.00	1.80	Not Associated
	R3	3.00	1.00	2.40	Associated
	R4	3.00	0.96	2.38	Associated
	R5	3.00	0.74	2.32	Associated
	R6	3.00	0.00	2.10	Associated
	R7	3.00	-1.00	1.80	Not Associated

From the above table it is determined that if the ‘Final score’ values are greater than ‘2’ then the reply messages are considered to be associated (i.e. the messages are related to the content posted) else if the scores are lesser than ‘2’ then the reply messages are considered to be not associated (i.e. the messages are not related to the content posted) which is denoted using a graph in Figure 4 as shown below. In this graph the straight line represents the ‘Standard score’ of ‘2’ which is the fixed value. Where, the upper part of line represents the associated messages and the lower line below represents the not associated messages.

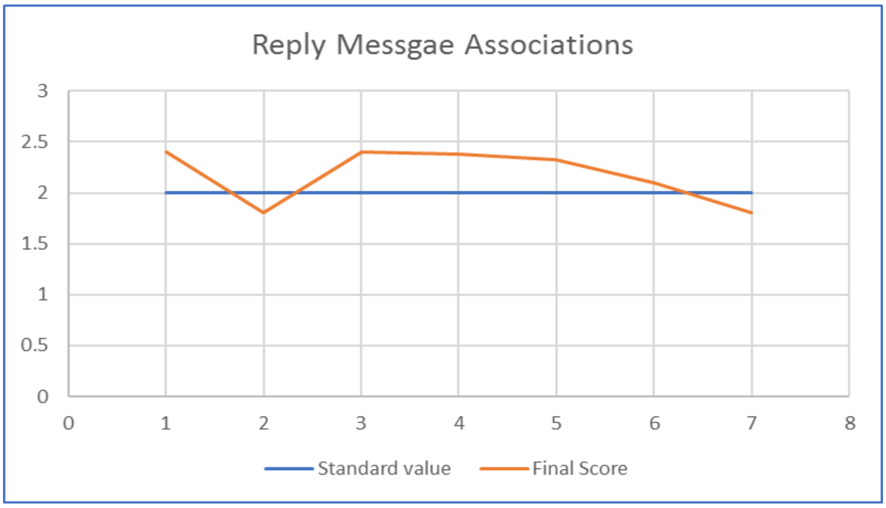

Figure 4; Reply Message Associations

In Figure 5, we have analyzed the performance of the proposed model and for social media links and LRLs, the rule counts and the support in percentage have been analyzed. In this, we infer that, while the increase in the number of rules we get better support whereas, there won't be much deviation among the usage of links or LRLs for variation in the rule counts applied with the proposed model.

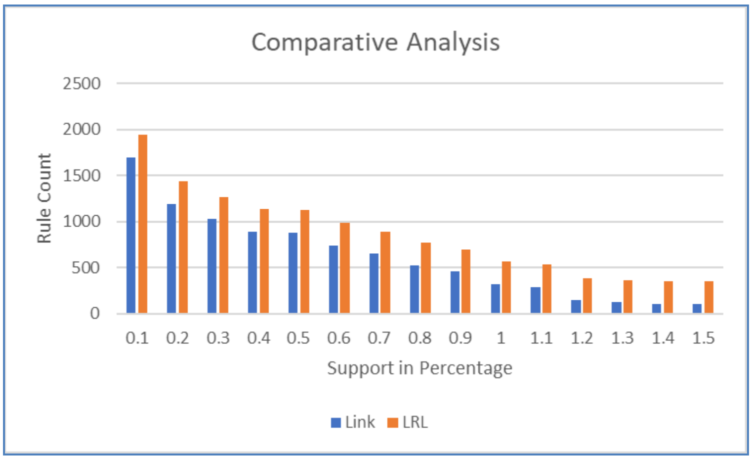

Figure 5; Comparative Analysis

The effectiveness of our approach on three types of information does not differ substantially from the baseline approaches, and it beats most other models in this area. We have the most accurate model when it comes to predicting favorable comments. Because of the framework's three components, it is possible to combine data from three separate modalities in order to arrive at a sentiment category from several points of view.

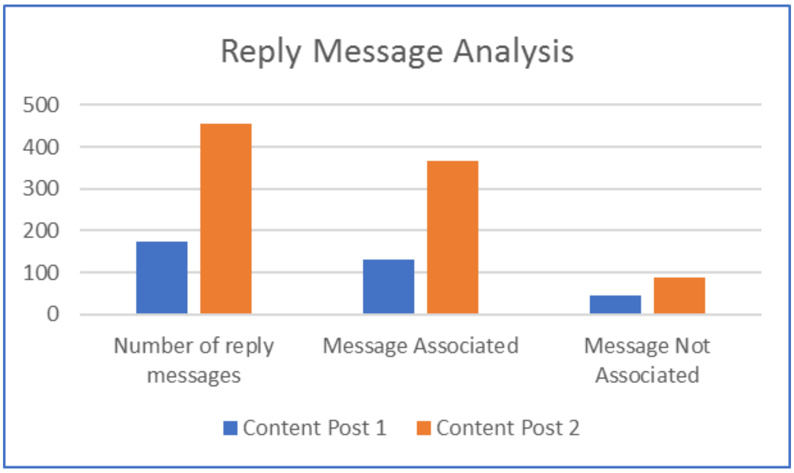

Figure 6; Reply Message Analysis

Finally, as shown in Figure 6, the results of social media content posted and its replies association are finalized and found that sentiments of each sentence is determined by the aspect based sentimental analysis with the score values.

Table 4. Aspect based Outcomes Analysis

Model	Accuracy	Precision	Recall	F1
Positive	92.54	88.24	78.54	45.65
Neutral	95.65	88.96	70.54	64.85
Negative	97.25	92.58	97.85	98.25

The aspect-based outcomes analysis has been listed on Table 4. Here all the positive, neutral and negative aspect responses from the samples we have considered for the experimentation. In Figure 7, the various metrics such as accuracy, precision, recall and F1-score are analyzed with respect to the positive, negative and neutral reviews collected from the samples.

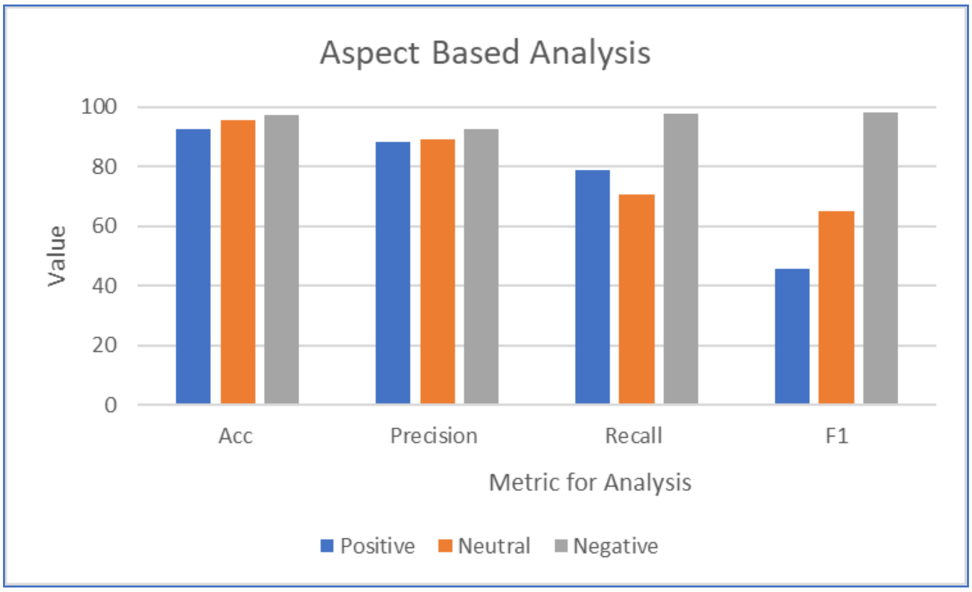

Figure 7; Aspect based Metrics Analysis

Our method can be used to find and compare group message and its reply messages association on different social media platforms. Models perform the poorest on negative postings, as indicated in the image, which may be due to an imbalance in the dataset's sample size. When using attribute-only, for example, there is a large difference in performance between various categories. Only in the negative category does it score less than 0.9, which means it is more accurate in the neutral area. In this study, it was shown that neutral-toned fashion products (mainly advertisements) might have certain commonalities in their design and style.

5. Conclusion and Future Enhancement

In the Proposed work it is clear that how sentences are analyzed with the help of aspect based sentimental analysis on their different aspects. This approach provides the sentiment score for the sentences that is useful to determine the association between content posted and its reply messages that is visualized in experimental section clearly. The experimental result part shows how the reply message are associated and analyzed based on the graphs and tabular column. Finally, the proposed work allows us to know better terms used in each type of reply messages that is associated or not associated with the content posted in social media is categorized. The future work idea of the author would be based on exploring other visualization techniques. Furthermore, to focuses the study on some hybrid methodology to increase the performance accuracy of content-based analysis.

References

1. Yue, Lin, et al. "A survey of sentiment analysis in social media." Knowledge and Information Systems 60.2 (2019): 617-663.
2. Rahman, Zoha, Kumaran Suberamanian, and Hasmah Binti Zanuddin. "Social media content analysis-A study on fanpages of electronics companies." International Journal on Global Business Management & Research 5.1 (2016): 87.
3. Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data." Journal of Big Data 2.1 (2015): 1-14.
4. Sun, Beiming, and Vincent TY Ng. "Analyzing sentimental influence of posts on social networks." Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2014.
5. Prakruthi, V., D. Sindhu, and S. Anupama Kumar. "Real time sentiment analysis of Twitter posts." 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS). IEEE, 2018.
6. Che, Wanxiang, et al. "Sentence compression for aspect-based sentiment analysis." IEEE/ACM Transactions on audio, speech, and language processing 23.12 (2015): 2111-2124.
7. Vanaja, Satuluri, and Meena Belwal. "Aspect-level sentiment analysis on e-commerce data." 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2018.
8. Schouten, Kim, and Flavius Frasincar. "Survey on aspect-level sentiment analysis." IEEE Transactions on Knowledge and Data Engineering 28.3 (2015): 813-830.

9. Laskari, Naveen Kumar, and Suresh Kumar Sanampudi. "Aspect based sentiment analysis survey." IOSR Journal of Computer Engineering (IOSR-JCE) 18.2 (2016): 24-28.
10. Thellaamudhan, C., R. Suresh, and P. Raghavi. "A comprehensive Survey on aspect-based sentiment analysis." International Journal of Advanced Research in Computer Science and Software Engineering 6.4 (2016): 442-447.
11. Chauhan, T., and S. Sonawane. "The Contemplation of Explainable Artificial Intelligence Techniques: Model Interpretation Using Explainable AI". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 65-71, doi:10.17762/ijritcc.v10i4.5538.
12. Srividya, Kotagiri, and A. Mary Sowjanya. "Aspect based sentiment analysis using POS tagging and TFIDF." International Journal of Engineering and Advanced Technology 8.6 (2019): 1960-1963.
13. García, Salvador, Julián Luengo, and Francisco Herrera. Data preprocessing in data mining. Vol. 72. Cham, Switzerland: Springer International Publishing, 2015.
14. Alasadi, Suad A., and Wesam S. Bhaya. "Review of data preprocessing techniques in data mining." Journal of Engineering and Applied Sciences 12.16 (2017): 4102-4107.
15. Benamara, Farah, et al. "Sentiment analysis: Adjectives and adverbs are better than 3adjectives alone." ICWSM 7 (2007): 203-206.
16. Liu, Jingjing, and Stephanie Seneff. "Review sentiment scoring via a parse-and-paraphrase paradigm." Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. 2009.
17. K. Dasaradharami Reddy, S. Mohanraju, Dr.A. Jebaraj Ratnakumar, Dr.S. Balakrishnan, "Querying and Searching of Friendship Selection in the Social IoT, Jour of Adv Research in Dynamical & Control Systems. Vol.10, 11-Special issue, 2018, pp. 910- 914.
18. Dipon Kumar Ghosh , Prithwika Banik , Dr. S. Balakrishnan (2018), "Review-Guppy: A Decision-Making Engine for Ecommerce Products Based on Sentiments of Consumer Reviews", International Journal of Pure and Applied Mathematics, Volume 119, No. 12, 2018, pp.1135-1141.
19. Ananthakrishnan, B., V. . Padmaja, S. . Nayagi, and V. . M. "Deep Neural Network Based Anomaly Detection for Real Time Video Surveillance". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 54-64, doi:10.17762/ijritcc.v10i4.5534.