PAPER • OPEN ACCESS

Twitter Sentiment Analysis: A Case Study for Apparel Brands

To cite this article: Abdur Rasool *et al* 2019 *J. Phys.: Conf. Ser.* **1176** 022015

IOP Conf. Series: Journal of Physics: Conf. Series **1176** (2019) 022015 doi:10.1088/1742-6596/1176/2/022015

Twitter Sentiment Analysis: A Case Study for Apparel Brands

Abdur Rasool[1,*], Ran Tao[1], Kamyab Marjan[1], Tayyab Naveed[2]

[1]School of Computer Science & Technology, Donghua University, Shanghai, China.
[2]College of Textile Engineering, Donghua University, Shanghai 201620, China.

*Corresponding author e-mail: 317089@mail.dhu.edu.cn

Abstract. Social media especially Twitter is providing a space for expression and opinions, where users discuss various events, services, and brands. Entrepreneurs are in continuous need of the feedback about their services to improve the quality and quantity. However, due to the bulk amount of data, it's difficult to detect the consumer's opinions. This article deliberates the problems about the Twitter data for the sentiment analysis. Furthermore, it implements the text mining and document-based sentiment on the preprocessed Twitter data through the machine learning techniques, Naïve Bayes and lexicon dictionary. Our case study is to find the public opinion about the top two apparel international brands and compare the positive and negative attitude of common users about each brand. We found that positive reviews of Adidas are more than the Nike while there is the slight difference in negative reviews. It founds that people want to discuss the other brands as comparisons when they are talking about just one brand.

1. Introduction

In the current century, the world is a global village due to the internet availability. According to [1] 54.4% of the world population was using the internet until Dec 2017. Now a day's social media (Facebook, YouTube, Twitter, and, content sharing sites, etc.) has efficiently and effectively shared useful information. In the learning resources, it is statistically [2] verified that 71% of the internet has been used through social media by the consumers. Studies [3] show that more than half customers prefer to read the other's comments about that products before purchasing. Thus public opinions are the best source of feedback for business stakeholders about their products and services which enables them to redesign the quality factor and disclose the opportunity of a new business [4]. The social network like Twitter and Facebook provided the important marketing, selling, branding, and promotional chances to the brands [5].

In the last decades, a hot research gets popular among the researchers to get the useful information from all these social media especially from Twitter. It's a mathematically procedural study of people' thoughts and opinions which can be positive or negative about any product or event through the natural language processing namely as 'sentiment analysis'. Sentiment analysis is correlated with text mining or data mining. The basic purpose of sentiment analysis is to assure the polarity of natural language by performing supervised and/or unsupervised classification. Recently available sentiment analysis techniques are useful for [6] political predictions, marketing strategy, e-commerce, and brand reputation management.

In one study, Sharma et al. have performed an apparel brand study, in which he found the trust level using regression tool for the Facebook data [5] but this study implements the sentiment for

Twitter apparel data. Another study was acquainted with apparel brands based on web text data in order to extract the customer emotion by implementing the sentiment techniques [1].

The purpose of our research is to apply the sentiment classification methodology to the Twitter apparel datasets. It will illustrate the relationship between the consumers and the apparel enterprise. In our work study, preprocessing step has been taken to achieve the better analysis results. The Bernoulli Naïve Bayes algorithm has been used with the lexicon dictionaries, e.g. VEDER dictionary. Furthermore, this study is a document-based approach and extracts the polarity from the tweets. This comparative study will assist the new researcher to analyze the social media for the sentiment detection.

The main contribution of this paper:

- To conduct a comparative analysis of two online international apparel brands.
- To find the answer, which apparel brand is most popular among the online viewers at the base of sentiment analysis?

The rest of the paper is: section-2 describe the related work, section-3 explain the case study which will discuss the preprocessing of data and will implement the sentiment techniques at our proposed case study and disclose the experimental results. Finally, section-4 clarify conclusion and future work.

2. SECTION 2: Related Work

The power of social media marketing is influencing the consumer and companies as well by spreading useful information and exchange of positive or negative values. Companies are learning the customer views and discussion to support their own mission and performance goals. Now, a direct link among consumers and companies about satisfaction and loyalty level has been found [7]. Twitter is an important source of information and data analysis for numerous sentiment analysis application by using machine learning techniques recently (Asghar et al. 2017) [8].

Shukri and Yaghi et al (2015) implement the sentiment and text mining techniques to analyze the unstructured data of Twitter to find the polarity and emotion classification about the automotive industry (such as Mercedes, Audi and BMB) and concluded that positive polarity of BMB (72%) is more than other brands [9]. However, this work will implement for the apparel brands comparison.

Sharma and Alavi et al. (2017) did a comparative study of 5 local apparel brands by proposing 3 hypothesis, based on the volume of posts, number of reaction on that posts, and number of comments. It decided [5] that 3 brands Craftsvilla, indiaRush, and Stalkbuylove maintained the above-mentioned hypothesis by positive coefficient while other 2 brands have negative beta coefficients for comments and number of posts.

In [1] the author proposes a study of Chinese based apparel brands by testing the web text, NPL, domain dictionary, sentiment analyses and visual techniques to extract the subjective customer emotions. It observed that T-shirts and coats have the highest emotions than the pants and, sales and emotions are directly proportions to each other as well as price and sales are directly influencing.

Gautam et al. (2014) engaged the probabilistic classifier, Naïve Bayes (NB) approaches which is supervised learning. Supervised learning is based on the labeled dataset which provided the model during the process. These labeled datasets are trained to achieve rational outcomes when encountered decision- making [10]. Jitendra et al (2018) [11] used both machine learning method, supervised and unsupervised approaches on different datasets obtained from Twitter. It achieved 80.68% accuracy on tweets for an unsupervised approach. For supervised, they combined unigram, bigram, and part-of-speech as features and achieved an accuracy of 67%.

According to Go et al. (2009) used emoticons as positive and negative in the test of tweet dataset [12]. It compared the accuracy of three machine learning algorithms including Naïve Bayes (NB), maximum entropy (MaxE) and support vector machines (SVM). And found that Naive Bayes's accuracy is better than other algorithms. Parikh and Movassate (2009) test the classification of the sentiment of two algorithms, Naïve Bayes Bigram and at Maximum Entropy model. It provides that the Naïve Bayes classifiers worked much better than the Maximum Entropy model [13].

3. SECTION 3: Case Study

3.1. Data Preprocessing:

Neuromas techniques are used to retrieve the data from the Twitter. For our research, we used a streaming API, Tweepy by executing a Python-based script. We retrieved 99,850 tweets by using the apparel brand's name as a keyword including “Nike” and “Adidas”. We fetched this data in irregular dates of 2018. Table 1 is representing the total number of tweets for each apparel brands. Sentiment analysis over the Twitter data is a very difficult and challenging task due to tweet's character limitation. Such kind of issues impact on the lexicon search. We resolved it the pre-processing step to make machine learning algorithms perform better.

Table 1. The total number of tweets for each brand.

Brand's Names	No. of Tweets
Nike	54788
Adidas	45062
Total Tweets	*99850*

As our dataset have 99,850 tweets but after removal of the Non-English tweets, the remaining dataset contains 30,895 tweets. Then, after removing the duplicate tweets (retweets); the total remaining tweets are 17006. We reform some replicated words like “I looovvvveee adidas apparel. This is my number one choice!” We reform ‘looovvveee’ word by eliminating extra o's, v's, and e's. Then we apply Porter Stemmer Algorithm to get the root word to reduce the derivational noise. For instance; governance and governing are stemmed into government. Twitter hashtag (#) often consist of some useful words about any event, product or discussion. We replace the hashtags with the same word without the hash. Twitter URL's perform as noise in sentiment detection so we subtract all URL's from our datasets to get the better accuracy of sentiment analysis.

3.2. *Implementation of Sentiment Analysis:*

An algorithm, Naïve Bayes (NB) has been used for the classification of emotions and polarity at each document based on the sentiment analysis. The NB algorithm is a probabilistic model which use Bayes' Theorem to solve the classification problems by assuming the data attributes as an independent. Mathematically, Bayes theorem is as:

$$P(c|z) = \frac{p(z|c) \times p(c)}{p(z)}$$

Where C is a class label, z is the attributes set, while $P(c)$ and $P(x|c)$ is the prior probability of the class. Fig.1 is the visual presentation of our sentiment process for Twitter data. NB classifier trained by processed data set which is annotated by three classes: Positive, negative, and neutral tweets. Machine learning natural language processing (NLP) implement to distribute the sentiment scores. To get the better engagement between the tweet words and lexicon words, the NB polarity classifier is used. This testing process is used to evaluate the accuracy of our model. Lastly, we validate the model for the extraction of the polarity % for three classes: positive, neutral, and negative to evaluate the results.

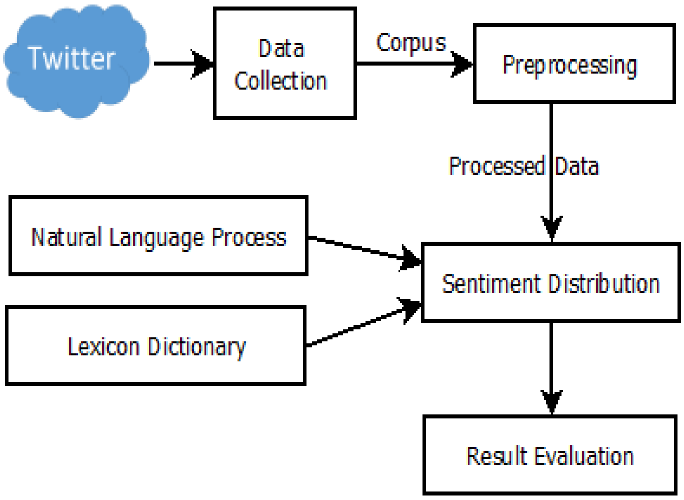

Figure 1. Sentiment Process Diagram on Twitter Data

4. Results

After the implementation of sentiment classification, we got the values of sentiment distribution of Nike and Adidas' Figure 2 display the Nike sentiment distribution which shows the positive, negative and neutral views 24.5, 11.9%, and 63.6% respectively. Similarly, figure 3 is the distribution of Adidas sentiment with positive (27.2%), negative (11.7%) and neutral (61.1%). It was found that positive reviews of Adidas are more than the Nike. While the neutral values record the satisfaction level among the online Twitter' users for both brands which is more than 60% of total reviews.

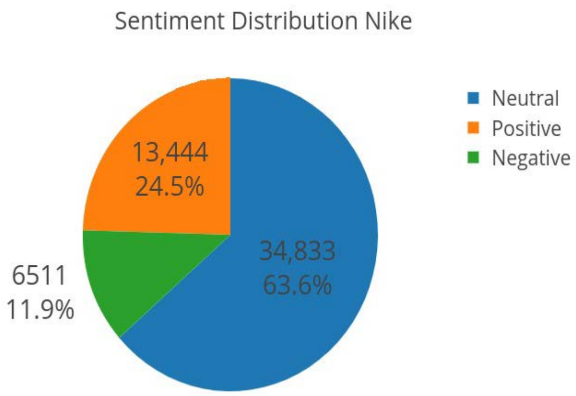

Figure 2. Sentiment distribution of Nike

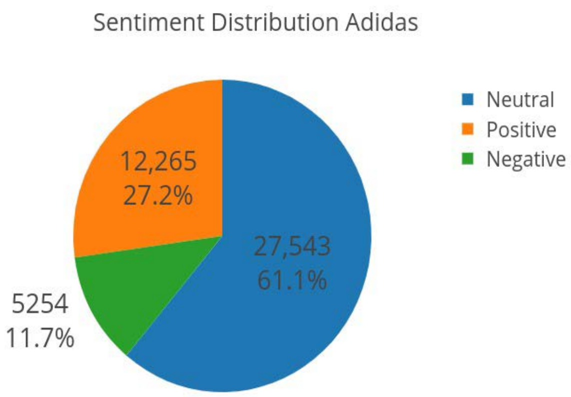

Figure 3. Sentiment distribution of Adidas

However, the negative opinions are more valuable than the positive for the shareholders and investors of both brands. These stakeholders could evaluate the negative reviews to detect the common users' requirements from the Twitter. For example in the word cloud of Adidas keywords (as shown in figure 4) has some negative words e.g. low. Similarly, in the word cloud of Nike (as shown in figure 5), containing the word e.g. los, indicate the bad comment. It was observed that people discuss the other brands as well when they want to talk about any single brand product. It's concluded, online consumers always try to compare the same product of different brands before making a purchase decision. For instance, in both word clouds of Adidas and Nike, people are discussing about third apparel brand, GUCCI.

Figure 4. Word Cloud of Adidas

Figure 5. Word cloud of Nike

Figure 6 has shown as a comparison of both brands for sentiment distributions about the online Twitter users. It is indicated the sentiment of each brand, Adidas and Nike, and total distributions of sentiments for our apparel dataset. Positive reviews were found greater than the negative and the peak of neutral reviews in both brands is a representation of most popular brands on the online users. As compared to the positive amount of comments, the neutral comments of Nike are better than the Adidas. While the more negative views of Nike is alarming for its market value as compare to Adidas.

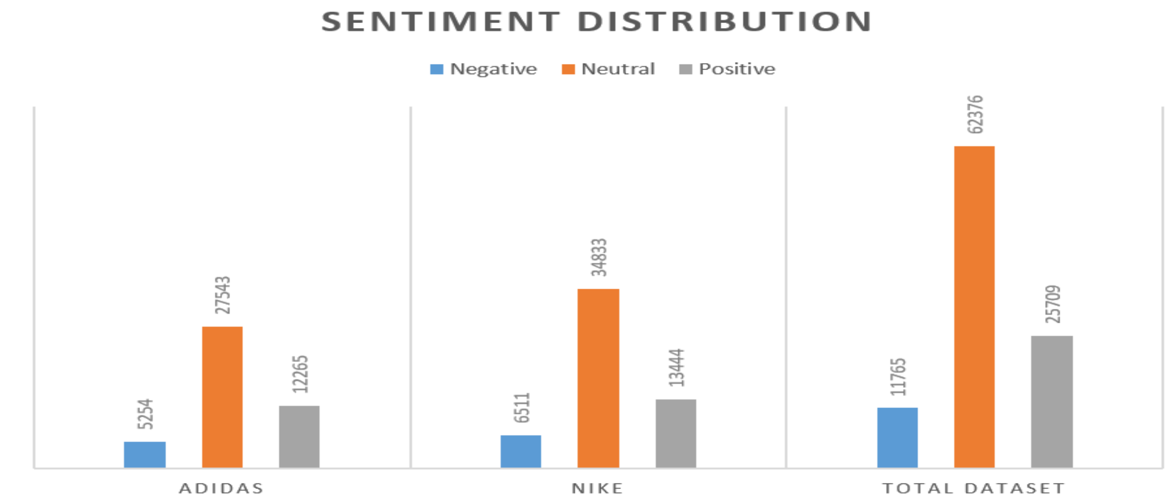

Figure 6. Sentiments distribution of each brands and total dataset.

5. Discussion

Nowadays, online shopping is replacing the regular or custom shopping. An apparel shopping is the popular one among the online users on social media. With the technologies advancement, new methods for customer satisfaction has been introduced by using the big data, text mining and sentiment analysis. Before implementing the sentiment, we apply the preprocess step at our Twitter data to make it much easier for our sentiment model. We conduct a comparative case study to analyze the sentiment for Twitter posts about Apparel brands to find which brand is popular. We concluded from our experiments that Adidas has more positive sentiment than the Nike, while the negative polarity between them is slightly different.

6. Conclusion

Sentiment analysis is one of the most shining field of text and data mining in numerous sectors. We applied the Naïve Bayes algorithm for sentiment classification. In this research, sentiment models have been applied to two international apparel brands to detect the polarity and emotions values e.g. opinions of consumers to help out the marketing and decision strategy for the apparel industry. The results disclosed that Adidas' positive polarity (27.5%) is 2.7% higher than Nike's positive polarity (24.5%). In the contras of positive polarity, the neutral polarity of Nike is more than Adidas which indicate the more customer's satisfaction level. Furthermore, we concluded that online users compare the other brands while making a decision of brand. In addition, the sentiment classification concludes the reliability of polarity classification about both brands.

In future, we will take the datasets of multiple apparel brands and implement a hybrid sentiment classification to conduct the better accuracy by constructing an apparel domain dictionary.

Acknowledgments

We are very grateful and would like to appreciate our respected professor Ran Tao[1] for his guideline and insightful feedbacks.

References

[1] Luo and Tao, "Sentiment Analysis Based on the Domain Dictionary: A Case of Analyzing Online Apparel Reviews" 2018.
[2] Statista.com, "number of social media users worldwide from 2010 to 2021 in Billions, Link: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ .
[3] DoubleClick_Inc. (2005).Understanding Buyer Search Activity as it Builds to Online Purchase. Retrieved.from.www.innovation-marketing.at/news/newsmodul/upload/429181378_searchpurchase_0502.pdf.
[4] Hu, Bhargava, Fuhrmann, "Analyzing users' sentiment towards popular consumer industries and brands on Twitter"
[5] Sharma and Alavi (2017), "Generating trust using Facebook-A study of 5 online apparel brands", Information Technology and Quantitative Management (ITQM2017)
[6] Liu, B. (2012). "Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies", 5(1), 1–167. doi:10.2200/S00416ED1V01Y201204HLT016.
[7] Kim JH, Kim M, Kandampully J. "Buying environment characteristics in the context of e-service. European Journal of Marketing". ResearchGate 2009 Sep 18;43(9/10):1188-204
[8] Asghar MZ, Khan A, Ahmad H, Kundi FM, Ismail S (2017) "A rule-based sentiment classif4ication framework for health reviews on mobile social media". Journal of Medical and Health Informatics, J Med Imaging Health Inf 7(6):1445–1453
[9] Shukri, Yaghi, "Twitter Sentiment Analysis: A Case Study in the Automotive Industry", 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). -2015.
[10] Geetika Gautam, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis". 7th International Conference on Contemporary Computing, 978-1-4799-5173-4/14/$31.00 ©2014 IEEE
[11] Jitendra Kumar, Sambit Bakshi, Karen L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," Electron Commerce Research, (2018) 18:181–199,
[12] Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. Processing 150(12):1–6.
[13] R. Parikh and M. Movassate, "Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009.