Vol. 7, Issue 2, pp. 22-28 | DECEMBER 2022
DOI: https://doi.org/10.53840/myjict7-2-58

Toward Hate Speech Analysis: Sentiment Analysis for Food Brands Review on Twitter

[i*]Vijay A/L Sandara Segaran, [i]Mohd Aminuddin Hamzah, & [i]Juhaida Abu Bakar, [ii]Sonia A/P Puventheran,

[i]Data Science Research Lab, School of Computing, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia.

[ii]NXP Semiconductor Malaysia, No. 2 Jalan SS 8/2 FIZ, Sungai Way Selangor, 47300 Petaling Jaya, Selangor, Malaysia

*(Corresponding author) e-mail: juhaida.ab@uum.edu.my

Article history:
Submission date: 12 Oct 2022
Received in revised form: 15 Nov2022
Acceptance date: 29 December 2022
Available online: 31 December 2022

Keywords:
Sentiment analysis, hate speech, food brand review, machine learning classifier.

Funding:
This research was supported by Ministry of Higher Education (MoHE) of Malaysia through Fundamental Research Grant Scheme for Research Acculturation of Early Career Researchers (RACER/1/2019/ICT02/UUM/1).

Competing interest:
The author(s) have declared that no competing interests exist.

Cite as:
Segaran, V. S., Hamzah, M. A., Bakar, J. A., & Puventheran, S. (2022). Toward Hate Speech Analysis: Sentiment Analysis for Food Brands Review on Twitter. Malaysian Journal of Information and Communication Technology (MyJICT), 22-28. DOI: https://doi.org/10.53840/myjict7-2-58

ABSTRACT

Social media has been a real-world sensor to observe the pulse of society. Although it provides unique communication opportunities, it also brings along vital challenges. One of them is hate speech, which attacks a single individual or targeted groups. Previous researchers claim that among social networks, the Twitter platform is mostly used to spread hate speech. However, data on a larger scale makes it hard to capture and understand the nature of hate speech on Twitter within specific food brands. In this study, sentiment analysis techniques were used to filter hate speech on Twitter and on three popular food brands in Malaysia. This study was conducted in five phases, namely raw data collection, pre-processing, sentiment analysis, visualization, and performance evaluation. This corpus consists of 28,898 data samples based on user tweet searches. A Twitter API was created and SQLite was used to store all the sample data. VADER sentiment analysis is used to classify tweet sentiment into positive, negative, and neutral. In the visualization phase, these three food brands are visualized using a histogram to gain sentiment analysis insights. Then, three machine learning methods were implemented to predict the best model for sentiment analysis. The Decision Tree classifier outperforms the average accuracy in Support Vector Machine and Logistic Regression with 99.99% for the training data set. This study provides insights to assist humans in making decisions. With the growth of opinions expressed in multimedia on social media, such as spoken feedback on Twitter, sentiment analysis has the potential to become a more news aggregation and low-cost endeavour.

Introduction

Currently, with the exponential growth of global exchanges on online social networks, there has also been an exponential growth of hate-filled events that take advantage of this technology. There are many platforms, such as online forums, blogs, and comment sections on review sites, containing various financial, educational, religious, and political problems (Gitari et al., 2015). On the popular platform on social media such as Twitter, hateful tweets target individuals, such as products, and target a particular group, such as an organization or the more specific food brand itself.

Sentiment analysis is particularly beneficial for social media tracking because it enables one to develop a better understanding of the broader public opinion around such issues. Sentiment analysis has become a discipline of study for locating opinionated data on the Web and categorizing it based on its polarity (Nezhad & Deihimi, 2021). Knowing the sentiment behind everything from forum posts and news stories enables us to develop more effective tactics and plans.

Social media has been a real-world sensor to observe the pulse of society. It is easy to communicate on social media by creating a free account. Although it provides unique communication opportunities, it also brings along vital challenges (Silva, 2016). One of them is hate speech, which attacks a single individual or targeted groups. According to Pereira-Kohatsu (2019), the massive and unfiltered feed of messages posted on social media is a phenomenon that nowadays raises social alarms. Many countries recognize hate speech as a serious issue in society. Mondal (2017) claims that social networks, especially Twitter, have been used to spread hate speech. The data is on a larger scale and hard to capture and understand the nature of hate speech on Twitter within McDonald’s, KFC, and Domino’s Pizza food brands. Therefore, in this study, hate speech between three food brands was classified, and the best model to predict the sentiment class was identified. In this study, sentiment analysis techniques will be used to filter hate speech on the Twitter social media platform on the popular food brands, which are KFC, McDonald's, and Domino's. Chapter 2 discusses related work, Chapter 3 on methodology, Chapter 4 on results and findings, and Chapter 5 on conclusions.

Related Works

In today's world, where humans are experiencing data exhaustion, which may not necessarily equate to stronger or deeper insights, it could have amassed mountains of data. Yet it is also difficult for mere humans to interpret it without making a mistake or prejudice. Oftentimes, humans need insights to assist in making decisions. Companies had to conduct surveys or build focus groups before using opinion analysis, which was much slower and more costly. Sentiment analysis has the potential to become a more popular news aggregation and low-cost undertaking. With the development of opinions shared in multimedia on social media, such as spoken feedback on Twitter (Soleymani et al., 2017).

One study created a Python app named Polarized opinions sentiment analyzer (POSA), which uses an ego social network analytics technique to mine and characterize an individual's conduct (Udanor et al., 2019). POSA employs a Python N-Gram dictionary of local context-based words that may be construed as hate terms. It then used the Twitter API to stream tweets about politics, race, religion, social justice, bigotry, and other topics from common and trending Nigerian Twitter handles and filtered the tweets against a custom dictionary using unsupervised classification of the texts as positive or negative sentiments. Tables, pie maps, and word clouds are used to visualize the results. A similar implementation was carried out using R-Studio codes, and the results were compared using an at-test to see whether there was a substantial difference between the outcomes. Both qualitative and quantitative analysis methodologies can be used. Quantitative in terms of being able to classify the findings as either negative or positive from the computation of text to vector, and qualitative in terms of data description. The results of two sets of POSA and R tests are as follows: in the first, the POSA program discovered that the Twitter handles examined contained between 33 and 55% hate material, while the R results indicated hate content ranged from 38 to 62%. Using t-test for both positive and negative ratings for both POSA and R-studio, the results reveal P-values of 0.389 and 0.289, respectively, on a value of 0.05, meaning that the results from POSA and R are not significantly different. The authors conclude that the percentage of hate content classified by

POSA is 40%, while the percentage of hate content classified by R is 51%, based on the results of the second experiment, which was conducted on 11 local handles with 1,207 tweets. The POSA's hate speech recognition accuracy is 87 percent, while free speech accuracy is 86 percent. R predicts hate speech with 65% accuracy and free speech with 74% accuracy. According to the findings, neither Twitter nor Facebook have an advanced hate speech detection scheme, and no benchmark has been established to determine the amount of hate speech that can be tolerated in a document. Humans, rather than machines, oversee the tracking.

The other study, Salminen et al. (2020) is working on a cross-platform online hate classification system. The model works well for identifying hateful remarks through various social networking channels, employs sophisticated linguistic features such as Bidirectional Encoder Representations from Transformers (BERT) (see "BERT"), and is freely accessible for researchers and practitioners to use and grow. Although the authors do not claim to have created a universal classifier that solves all problems in online hate detection, the findings show that this line of research has potential for the wider OHR group and can be further developed. The findings suggest that it is possible to train classifiers that can detect hateful remarks across various social media sites with good outcomes and a low rate of false positives and negatives. The fact that the models perform well when trained with BERT features backs up recent research that shows bidirectional neural networks can generate useful feature representations for online hate detection. When switching from simplified to more complex functions, the results reveal a linear pattern in the success of the classifiers, with BERT providing the better results. The author collected 197,566 comments from four different platforms: YouTube, Reddit, Wikipedia, and Twitter, with 80% of the comments being non-hateful and the other 20% being hateful. The author then tries out different classification algorithms and feature representations (Logistic Regression, Naive Bayes, Support Vector Machines, XGBoost, and Neural Networks) (Bag-of-Words, TF-IDF, Word2Vec, BERT, and their combinations). Although all the models outperform the keyword-based baseline classifier, XGBoost with all features has the best results (F1 = 0.92).

Another study by Mattila et al. (2018) examined how various variables affected the mood in response to a tweet posted on Twitter for promotional purposes by corporations in the fast-food industry in North America. The features have been considered, such as the time of posting, the length and sentiment of a message, as well as the presence of media other than text in the tweet. Sentiment was derived from regular samples of responses to advertisement tweets. It was collected between March 27th and April 28th and plotted against the factors mentioned. The findings show that the advertisement tweet's sentiment, as well as the time of publishing, had the greatest influence on the reaction, but no conclusive conclusions can be made about their consequences.

In this study, this work will apply analysis and text mining to explore users' patterns and issues aligned with each brand theme tweeted.

Methodology

This study was conducted in five phases, namely raw data collection, pre-processing, sentiment analysis, visualization, and performance evaluation. The first phase is raw data collection. The raw dataset from three fast food companies, which are Kentucky Fried Chicken (KFC), McDonald's (McD), and Domino's Pizza, were collected from Twitter. The data is collected by creating an API account, which is Twitter REST APIs (Twitter Search API). The dataset collected is based on the search keyword. The target search keyword is mainly related to the tweets that users or customers mention about these three food brand companies. The software that is used for collecting data is KNIME. This raw dataset will then be stored in a database called SQLITE. This process took 2 weeks from the end of March to early April 2021.

The second phase consists of pre-processing. In this phase, the Knowledge Discovery in Databases (KDD) technique is used when the focus is more on meaningful extraction. This research looks at how users react to these three companies through emotion or sentiment analysis. R language is used for cleaning raw data and Python is used for sentiment analysis. In the cleaning data process, three terms have been removed; missing values, "blob" data, and Boolean types, which include Unicode. Column data and time are also divided. Then, clean data was saved into a new csv file.

In the third phase, the sentiment analysis process, the Python programming language is used. The Natural Language Toolkit (NLTK) is used to process and analyze text. It helps to determine the ratio of positive, negative, and neutral in terms of score using a sentiment analysis tool (Valence Aware Dictionary and sEntiment Reasoner), also known as VADER. Such as, a score of less than 0 is negative sentiment, 0 is neutral, and more than 0 is positive polarity are learned. A polarity dictionary with words and sentences annotated by its semantic orientation is required for a lexicon-based method (Mahmood et al., 2020).

The fourth phase consists of information visualization. Jupyter Notebook, Spyder, and KNIME are used to visualize the result of the sentiment analysis. Figure 1 shows the process of reading a file and displaying the top 5 data using the head () function.

Figure 1: Reading a file and displaying the top 5 data

```
In [1]: import pandas as pd
        df = pd.read_csv(r'C:/Users/sonia/Documents/project soc ana/dominosRaw.30.csv')
        df.head()
```

Out[1]:

	Tweet	Tweet.ID	Time	User	Date	score	sentiment	sentiment_N
0	@dominos_india @dominos \r\n#Dominos https://t...	1.39E+18	21:49:46	anujagarwala3	4/28/2021	0.0000	neutral	1
1	Dog Helps Himself to @dominos ' Pizzas Left on...	1.39E+18	21:44:02	itsstevenhudson	4/28/2021	0.3818	neutral	1
2	@Let__It_be___ @dominos @pizzahut Ye mere fav ...	1.39E+18	21:40:08	uniquesadu	4/28/2021	0.4588	neutral	1
3	@23XIRacing I'm still waiting for the @dominos...	1.39E+18	21:36:48	Scott_Derek	4/28/2021	0.0000	neutral	1
4	ok today I'm having @dominos\r\nI'm sold thank...	1.39E+18	21:34:09	tteateetl	4/28/2021	0.6249	positive	2

In the fifth phase, performance evaluation of the result is calculated. Figure 2, 3 and 4 shows the results for sentiment analysis for three food brand reviews using a histogram chart. Positive, neutral, and negative sentiment values are represented on the x-axis. In this chart, a total of 13,000 instances are collected for Domino's Pizza. The highest sentiment for Domino's Pizza shows a neutral sentiment, followed by negative and positive sentiment.

The KFC food brand has the highest neutral sentiment. Almost 3,000 instances show that the customer's tweets are neutral about KFC, followed by the second highest, which is more negative than positive. For McDonald's, the data also shows that, based on 10,000 instances, 6,000 instances show that customer tweets are neutral. Meanwhile, we can see that both the negative and the positive were nearly at the same level as shown in Figure 2, 3 and 4.

Figure 2: Domino's Pizza sentiment analysis

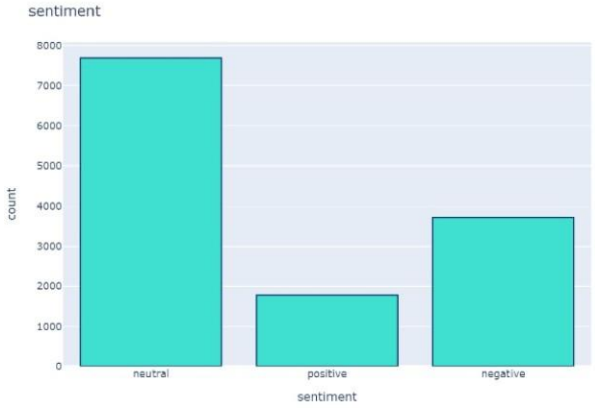


Figure 3: KFC sentiment analysis

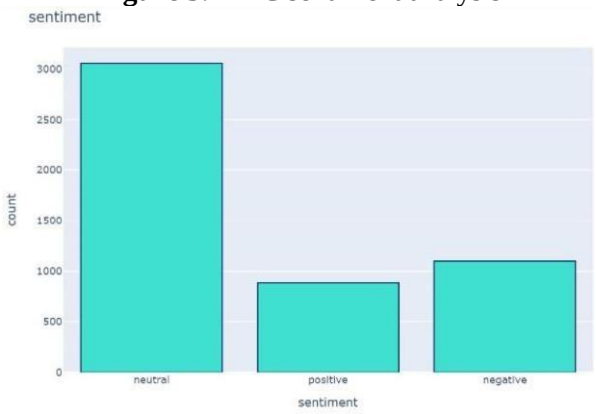
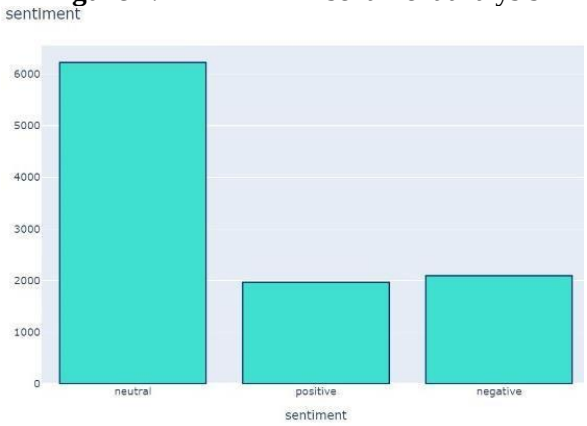


Figure 4: McDonald's sentiment analysis



Results and Findings

Three algorithms were used to predict the sentiment of the dataset, whether the sentiment of the dataset is neutral, negative, or positive. To make sure the model we use is the right model for what it needs to be, we use the score test model. The algorithms that we use are the Logistic Regression algorithm, Decision Tree, and Support Vector Machine (SVM). For the logistic regression and SVM, we used the Jupyter notebook to build a model of prediction. For the decision tree, we used the KNIME software to build the prediction model. Jupyter is used for regression and SVM because the algorithm is simpler to understand and implement than KNIME.

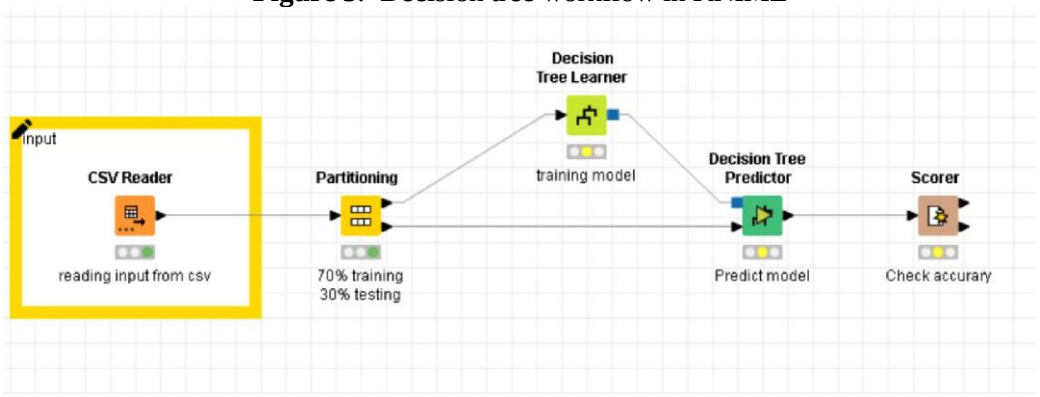
When the dependent variable is binary, logistic regression is the proper regression strategy to use. The logistic regression, like all regression studies, is a predictive analysis. We use the logistic to predict whether the data is neutral, negative, or positive. Using the python code in the Jupyter notebook, all three datasets, which are Domino's pizza, KFC, and McDonald's, showed the same accuracy. The accuracy was 73% accurate for the negative sentiment, 92% accurate for the neutral sentiment, and 58% accurate for the positive sentiment.

Then, in this study, the model using a decision tree is built and the model is tested based on its accuracy. A decision tree is a diagram that shows a statistical probability or assists in determining a course of action. Due to its resemblance to the namesake plant, the chart is referred to as a decision tree, and it is depicted as an upright or horizontal figure with branches. Each "branch" of the decision tree represents a possible decision, outcome, or reaction, starting with the decision itself (called a "node"). The "leaves" are the tree's farthest branches, which indicate the final effects of a particular choice process.

The KNIME software is used to develop the model for the decision tree algorithm. Figure 5 shows the model that was built from the KNIME. First, read an input in a CSV data format. Then, a partitioning node is used to separate the dataset into training and testing datasets. The data split is 70% for training

data and 30% for testing datasets. A decision tree learner node was chosen to create a decision tree model from the training data. After the training data is completed, the decision tree predictor is used to test the remaining testing datasets. After testing the testing dataset, the Scorer node was chosen to check the accuracy of the testing dataset.

Figure 5: Decision tree workflow in KNIME



The Domino's Pizza shows an accuracy rate on the testing data of 99.98%, which is from the 3,946 instances, and only 1 instance was misclassified. Meanwhile, for KFC and McDonald's, it shows that both models get 100% accuracy on the testing dataset.

The third classifier tested is the Support Vector Machine (SVM). SVM is a supervised machine learning algorithm that can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space, with the value of each feature being the value of a particular coordinate. Then, classification by finding the hyper-plane is performed, which differentiates the two classes very well.

By using an SVM classifier, the accuracy score on the KFC dataset is 81.7%, which correctly classified 244 sentiment classes while 56 classes were incorrectly classified. Then, the accuracy score on the Domino dataset is 88.7%, which correctly classified 266 sentiment classes while 34 classes were incorrectly classified. Next, the accuracy score on McDonald's dataset is 86%, which correctly classified 258 classes while 42 classes were incorrectly classified. From the above data modelling, it shows that the SVM model performs well, quite consistently, and is able to predict the sentiment classes with almost 85% accuracy for all three datasets.

Conclusion

In a nutshell, from the data modelling process, the best model to predict the classes of sentiment was the Decision Tree. It obtains the highest average accuracy score of almost 99% among the three algorithms. As far as the accuracy of the model is concerned, the Decision Tree performs better than the SVM model, which in turn performs better than Logistic Regression.

In this paper, we suggested a tool for scraping data to identify hate speech on Twitter. Our suggested method uses tools to identify hate speech patterns automatically, and most of them combine them with sentimental and semantic functionality to label tweets as hate speech. We gathered data from Twitter using the KNIME Twitter API and compiled it in the database to make progress on this work. In future work, after data cleaning, F-score will be considered to be included to evaluate the performance of the classification sentiment analysis of hate speech.

Acknowledgment

This research was supported by Ministry of Higher Education (MoHE) of Malaysia through FundamentalResearch Grant Scheme for Research Acculturation of Early Career Researchers (RACER/1/2019/ICT02/UUM/1).

References

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(4), 215-230.

Nezhad, Z. B., & Deihimi, M. A. (2021). Sarcasm detection in Persian. *Journal of Information and Communication Technology*, *20*(1), 1-20.

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016). Analyzing the targets of hate in online social media. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 10, No. 1, pp. 687-690).

Pereira-Kohatsu, J. C., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019). Detecting and monitoring hate speech in Twitter. *Sensors*, *19*(21), 4654.

Mondal, M., Silva, L. A., & Benevenuto, F. (2017, July). A measurement study of hate speech in social media. In *Proceedings of the 28th ACM conference on hypertext and social media* (pp. 85-94).

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3-14.

Udanor, C., & Anyanwu, C. C. (2019). Combating the challenges of social media hate speech in a polarized society: A Twitter ego lexalytics approach. *Data technologies and applications*, *53*(4), 501-527.

Salminen, J., Hopf, M., Chowdhury, S. A., Jung, S. G., Almerekhi, H., & Jansen, B. J. (2020). Developing an online hate classifier for multiple social media platforms. *Human-centric Computing and Information Sciences*, *10*(1), 1-34.

Mattila, M., & Salman, H. (2018). Analysing social media marketing on twitter using sentiment analysis.

Mahmood, A. T., Kamaruddin, S. S., Naser, R. K., & Nadzir, M. M. (2020). A combination of lexicon and machine learning approaches for sentiment analysis on Facebook. *Journal of System and Management Sciences*.