The Eurasia Proceedings of
Educational & Social Sciences (EPESS)
ISSN: 2587-1730

The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2023

Volume 31, Pages 189-195

ICRESS 2023: International Conference on Research in Education and Social Sciences

Decoding Emotions: Harnessing the Power of Python for Sentiment Analysis in Social Media

Mohammed Rizvi
Exelon Corporation

Abstract: Social media usage is increasing tremendously, and it has become a necessity. A person needs to be able to use social media in order to compete in this ever-developing world of technology's large number of people use social media, some for-entertainment purposes for educational purposes, Some for political, and others for economic purposes. To accommodate this tremendous amount of information that is being disseminated in social media to reflect the views of all these individuals. And all those views (information) have different sentiments echoed in them. To gain some data from the list of information, we need to analyze the feelings of the posts on social media. Sentiment analysis is a powerful tool that utilizes machine learning and natural language processing (NLP) to detect the sentiment - whether it be positive, negative, or neutral - in text. Two primary methods for conducting sentiment analysis are rule-based and automated. Convolutional neural networks (CNNs) and deep learning have been found successful in uncovering meaningful sentiments from texts, allowing for accurate classification of views expressed through written data. By breaking down each step thoughtfully with new ideas, active sentences instead of passive ones, stronger verbs for increased intensity, and synonyms to replace words that could be better used elsewhere, this makes up a successful rewrite of the original text.

Keywords: Sentimental analysis, Natural language processing, Speech detection, Jupyter

Introduction

Social media sentiment analysis seeks understanding how people feel about a product, service, or brand. It is an invaluable tool for businesses that are looking to gain insight into what consumers think of them. Manual data collection can be arduous and time-consuming due to the sheer amount of users and content available online; thus, machine learning models such as neural networks and deep learning algorithms offer efficient solutions for analyzing sentiments in large datasets. This paper will discuss a Python algorithm that provides an effective yet straightforward approach to sentiment analysis.

Related Work

Recently, automatically detecting hate speech has been widely studied by researchers. This section will review related works on traditional machine learning-based methods, deep learning-based methods, and multi-task learning-based methods of hate speech detection. Chen, (2012) proposed a variety of linguistic rules to determine whether a sentence constitutes hate speech or not. Gitari (2015) designed several sentiment features and achieved good performance in experiments. Previous studies have shown that sentiment features play an important role in hate speech detection. Deep learning-based methods have recently garnered considerable success in detecting hate speech (Krause & Grassegger, 2016). Mehdad and Tetreault (2016) extracted text's n-gram, character-level and sentiment features and used support vector machines (SVM) to detect hate speech. The semantics of hate speech contains a strong negative sentiment tendency.

The deep learning methods of predecessors often only used pre-trained models or deeper networks to obtain semantic features, ignoring the sentiment features of the target sentences and external sentiment resources, which also makes the performance of neural networks unsatisfactory in hate speech detection. Waseem (2016), stated that hate speech is very dependent on the nuance of language. Even if it is manually distinguished whether certain sentences contain hate semantics, a consensus is rare Del Vigna12 (2017) used the sentimental value of words as the main feature to measure whether a sentence constitutes hate speech. Artificial features can only reflect the shallow features of text and cannot understand content from the deep semantic features.

Our intuition is that most hate speech contains words with strong negative emotions, which are usually the most direct clues to hate speech. Meanwhile, as claimed by Davidson (2017). Sentimental analysis: It is not very easy (Kenyon Dean, 2018). Kshirsagar (2018**)** proposed a transformed word embedding model (TWEM), which had a simple structure but could achieve better performance than many complex models (Qian, 2018). A Lexicon-based sentiment analyzer was proposed to determine the polarity and measures for tweet data of a particular candidate (Nausheen & Begum, 2018).

Wang (2018) compared the performance of various neural network models in detecting hate speech and used visualization techniques to give the models better interpretability Zhang (2018) and fed input into a convolutional neural network (CNN) and a gated recurrent unit (GRU) to learn higher-level features. Rodríguez (2019) constructed a dataset of hate speech from Facebook and proposed a rich set of sentiment features, including negative sentiment words and negative sentiment symbols, to detect hate speech. Liu et al. (2019) introduced a novel formulation of a hate speech type identification problem in the setting of multi-task learning through their proposed fuzzy ensemble approach. Badjatiya (2019) found that due to the limitation of the training set, the deep learning model would have “bias” and he designed and implemented a “bias removal” strategy to detect hate speech. Multi-task learning can learn multiple related tasks and share knowledge simultaneously.

In recent years, there have been some achievements in hate speech detection. Ousidhoum (2019) presented a new multilingual multi-aspect hate speech analysis dataset. Its Challenging due to the inherent complexity of the natural language constructs. Most of the existing works revolve either around rules. Kapil and Ekbal (2020) and propose a deep multi-task learning (MTL) framework to leverage useful information from multiple related classification tasks in order to extend the performance of hate speech detection. Tekiroglu (2020) constructed a large-scale dataset based on hate speech and its responses and used the pre-trained language model, GPT-2, to detect hate speech. Obviously, deep learning models can extract the latent semantic features of text, which can provide the most direct clues for detecting hate speech. In this work, we approach suicide counselling aiding from a more practical aspect. Instead of creating a chatbot to replace human counsellors, our goal is to propose a model for suicidal ideation detection (Ji, 2021) for cross-modality interactive learning. The study of ordinal suicidal ideation detection Sawhney (2021) designed an architecture including Bi-LSTM layers, temporal attention layer, and ordinal regression layer, to comprehensively analyze the posts from the past. Then, they adapted C-SSRS to make an assessment of suicidality.

Verma (2022) addressed the application of sentiment analysis to build a smart society based on public services. Xiao (2022) exploit the self-attention mechanism combined with densely connected graph convolutional networks to learn inter-modality dynamics. Latifian (2022) was conducted to delve into the bipolar patients' family experiences of the outcomes of encountering stigma. Sun (2022) verify the correlation between social network characteristics, Weibo textual sentiment characteristics, and port forwarding. Various perspectives are taken into consideration in order to determine the impact of social media on the public, police and government with the help of face recognition technology. Sushith, (2022) and Tania (2022) focus on thinking aloud or screaming inside: an exploratory study of sentiment around work.

Sentimental juxtaposition revealed through the labelled data set was supported by the n-gram analysis as well. Shu (2022) attempts to identify consumers' opinions on the health impact of online game products through non-structured text and large-size social media comments. The research of Joloudari (2022) and Upadhyaya (2022) conclusively demonstrate the superiority of BERT models over other deep learning architectures in sentiment analysis. The former study leverages such models to identify denier statements on Twitter, accurately classifying tweets into either a believer or denier stance towards climate change. A sudden change in the stock movement due to COVID-19 appearance causes some problems for investors. From this point Sharaf (2022) propose an efficient system that applies sentiment analysis of COVID-19 news and articles to extract the final impact of COVID-19 on the financial stock market. Lokala (2023) analyze the substance use posts on social media with opioids being sold through crypto market listings.

Material and Methods

For building the algorithm we will be using python programming language. The program will be coded on Jupyter notebook due to the reason Jupyter Notebook allows users to compile all aspects of a data project in one place making it easier to show the entire process of a project to your intended audience. The first step is encoding a list of emotions with their enormous lists of examples into the algorithm. The second step is giving the model a text post to be analyzed. The algorithm will have functions that will eliminate white spaces, punctuation marks, pronouns, conjunctions as such words will not have any impact on the overall emotion depicted on the text. The overall all text will then be converted to a lowercase letter for a consistent and smoother working of the algorithm then the text will be split into a single word and all of the words will be stored inside an efficient data structure.

The Algorithm will loop through each line and then checks if the words inside the emotion.txt file are also present inside the collection of words from the previously cleaned file, if a word from a specific emotion is found it will increment a counting variable and after going through all the files it will then display a graph based on the found information. The sample emotions given to the algorithm consists of happy, Angry, sad, delusional, satire and as such. Please refer Table 3.1 and 3.2.

Table 3.1. List of sample positive emotions used by the Algorithm

Joy	Gratitude	Love
Satisfaction	Serenity	Interest
Amusement	Affection	Attraction
Affection	Excitement	Happy
Ecstasy	Love	Thrilled

Table 3.2. List of sample Negative emotions used by the algorithm

Anger	Fear	Guilt
Shame	Anxiety	Loneliness
Sadness	Disgust	Depression
Agony	Agitation	Happy
Sad	Rage	Melancholy

And after analyzing the texts the output gave the percentage of each of the emotions available and finally generalize the true intention of the post. Please refer to the image 3.3.

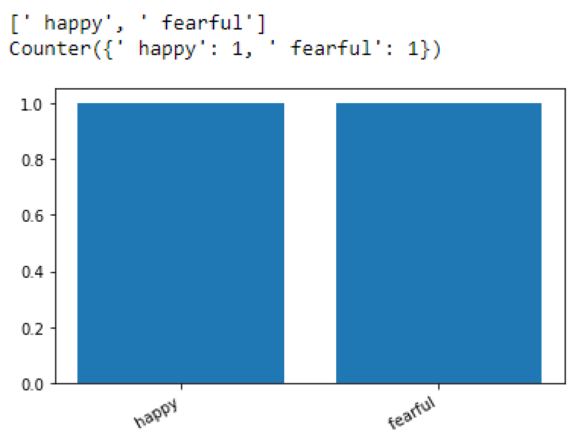

Figure 1. Output graph of the Algorithm after analyzing an emotion-based text

Results

Datasets and Evaluation Metrics

After programming, the input was fed to the algorithm and the following outputs were found. The datasets given to the model are carefully selected text posts that include almost all emotions available and they are supposed to reflect the view and opinions of majority of extreme social media users. Please refer to the table 4.1.

Table4.1. Example of the questionnaire used to gather the datasets from the public

Sample text	Positive emotion	Neutral emotion	Negative emotion
1	….	…..	…..
2	….	…..	…..
.	.	.	.
.	.	.	.

A survey will be conducted on 100 random individuals who are substituted from different walk of life and they will be requested to rate the sample texts from a range of 1 to 10 based on the feelings they felt when they read the texts, the gathered information is then tabulated and catalogued. Please refer to the table 4.2.

Table 4.2. List of example of sentences the Algorithm is will be able to analyze

Sample text 1	I am thrilled that my new job starts Monday!
Sample text 2	You are stupid and delusional, Get out of my face.
Sample text 3	I am not quite sure I will let you know if it is possible or not.
Sample text 4	Covid vaccination is deadly and unorthodox as it can cause…
Sample text 5	If voting changed anything, they would make it illegal
Sample text 6	She trembled with fear when she find out the door was not locked all night.

The evaluation metric is conducted by comparing the algorithm output with the survey conducted earlier, then the two outputs are compared and the output was as follows. Based on the comparison of the two outputs it is concluded that the model is 95 % effective. Since the error margin is less than 5 % it can be said the AI model works as intended and it is successful.

Discussion

After properly following the methodology, the output of the model was found to be highly accurate with a 95% success rate. This algorithm could accurately detect hate speech and also identify what type of emotions it evoked in text-based social media posts. When this innovation was implemented on a local website customer service page, they observed an improvement in their services as well as better customer satisfaction due to its low complexity and high efficiency. The advantages of this model are manifold - from being easy to implement and maintain, compared to complex models that require substantial time for training; upscaling emotion lists so that more feelings can be detected easily; incorporating image or voice based posts etc., there is much potential for further development. On the other hand, one major limitation lies in its inability to work beyond text-based social media posts which restricts its application range significantly. Considering these aspects of this model, future work must focus on integrating both image and voice based postings along with providing greater accuracy by increasing emotion list size thereby improving prediction results even further.

Conclusion

The research centers on Natural Language Processing and Sentiment Analysis using Python. This algorithm is simple, fast, and efficient in gauging emotions from a given text. It can be integrated into any program or webpage as a web-based application with ease. However, there are some drawbacks to this method such as the difficulty of recognizing true context when negative sentiment is expressed through backhanded compliments -

resulting in an overestimation of positive feedback. Additionally, this algorithm cannot identify sarcasm, negation, grammar mistakes or irony; making it ideal for analyzing data gathered from social media platforms but not suitable for others forms of communication like voice recordings and image files. To improve upon these limitations further exploration should be done on refining the ability to detect nuances while processing complex language structures including different tones and dialects used by people all over the world – ultimately leading to more accurate results when performing sentiment analysis tasks using Python algorithms.

Appendice

```
import string
from collections import Counter
import matplotlib.pyplot as plt
text=open('read',encoding='utf-8').read()
lower_case = text.lower()
cleaned_text =lower_case.translate(str.maketrans(",",string.punctuation))
tokenized_words=cleaned_text.split()
##print(tokenized_words)
stop_words = ["i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "your", "yours", "yourself",
"yourselves", "he", "him", "his", "himself", "she", "her", "hers", "herself", "it", "its", "itself", "they", "them",
"their", "theirs", "themselves", "what", "which", "who", "whom", "this", "that", "these", "those", "am", "is",
"are", "was", "were", "be", "been", "being", "have", "has", "had", "having", "do",  "does", "did", "doing", "a",
"an", "the", "and", "but", "if", "or", "because", "as", "until", "while", "of", "at", "by", "for", "with", "about",
"against", "between", "into", "through", "during", "before", "after", "above", "below", "to", "from", "up",
"down", "in", "out", "on", "off", "over", "under", "again", "further", "then", "once", "here", "there", "when",
"where", "why", "how", "all", "any", "both", "each", "few", "more", "most", "other", "some", "such", "no",
"nor", "not", "only", "own", "same", "so", "than", "too", "very", "s", "t", "can", "will", "just", "don", "should",
"now"]
final_words=[]
for word in tokenized_words:
   if word not in stop_words:
       final_words.append(word)
emotion_list = []
with open('emotions', 'r') as file: for line in file:
clear_line = line.replace("\n", ").replace(",", ").replace(""", ").strip()
     word, emotion = clear_line.split(':')
       if word in final_words:
          emotion_list.append(emotion)
print(emotion_list)
w = Counter(emotion_list)
print(w)
# Plotting the emotions on the graph
fig, ax1 = plt.subplots()
ax1.bar(w.keys(), w.values())
fig.autofmt_xdate()
plt.savefig('graph.png')
plt.show()
```

Scientific Ethics Declaration

The author declares that the scientific ethical and legal responsibility of this article published in EPESS journal belongs to the author.

Acknowledgements

*This article was presented as an oral presentation at the International Conference on Research in Education and Social Sciences (www.icress.net) held in Budapest/Hungary on July 06-09, 2023.

*I would like to express my appreciation for the time and effort that the reviewers and editorial team will invest in evaluating my submission.

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012). Detecting offensive language in social media to protect adolescent online safety. *International Conference on Privacy, Security, Risk and Trust (PASSAT).*

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media, 11*.

Del Vigna, F., Cimino, A., Dell'Orletta, F., Petrocchi, M., & Tesconi, M. (2017). Hate me, hate me not: Hate speech detection on Facebook. *Proceedings of the First Italian Conference on Cybersecurity (ITASEC17),* Venice, Italy.

Dinan, E., Zhang, S., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too?. *Proceedings of the 56 th Annual Meeting of the Assocation for Computational Linguistics, 1,* 2204-2213, Mellbourne Australia.

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.

Glaz, A. L., Haralambous, D. H., Kim Dufor, P. L., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of Medical Internet Research, 23(5).*

Hoque Tania, M., Hossain, M R., Jahanara, N., Andreev, I., & Clifton, D. A. (2022)."Thinking aloud or screaming inside: Exploratory study of sentiment around work." *JMIR Form Research, 6*(2).

Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang. Z. (2021). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*.

Joloudari, J. H., Hussain, S., Nematollahi, M., Bert, A., Bagheri, R., Fazl, F., Alizadehsani, R., Reza Lashgari, R., & Talukder, A. (2022). Bert- deep CNN: State of the art for sentiment analysis of Covid-19 tweets. *Social Network Analysis and Mining, 13*(1), 14.

Kapil, P., & Ekbal, A. (2020). Leveraging multi-domain, heterogeneous data using deep multitask learning for hate speech detection. *Proceedings of the 17th International Conference on Natural Language Processing (ICON*).

Kenyon Dean, K., Ahmed, E., Fujimoto, S., Georges-Filteau, J., Glasz, C., Kaur, B., Lalande, A., Bhanderi, S., Belfer, R., Kanagasabai, N., Sarrazingendron, R., Verma, R., & Ruths, D. (2018). Sentiment analysis: It's complicated!. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*

Krause, T., & Grassegger, H. (2016). Facebook's secret rules of deletion. *Proceedings of the Conference Suddeutsche Zeitung*.

Kshirsagar, R., Cukuvac, T., McKeown, K., McGregor, S. (2018). Predictive embeddings for hate speech detection on Twitter." *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*, 26-32, Brusells, Belgium.

Latifian, M., Raheb, G., Abdi, K., & Rosa Alikhani, R. (2022). The bipolar patient's family experiences of the outcomes of encountering stigma in Tehran: A qualitative study. *International Journal of Social Psychiatry, 69*(2).

Liu, N. F., Schwartz, R., & Smith, N. A. (2019). Inoculation by fine-tuning: A method for analyzing challenge datasets. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1.*

Lokala, U., Phukan, O. C., Dastidar, T. G., Lamy, F., Daniulaityte, R., & Sheth, A. (2023). Can we detect substance use disorder?": Knowledge and time aware classification on social media from darkweb. *ARXIV-CS.LG.*

Mehdad, Y., &Tetreault, J. (2016). Do characters abuse more than words?. *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.

Nausheen, F., & Begum, S. H. (2018). Sentiment analysis to predict election results using Python. *Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018).* IEEE Explore Compliant.

Ousidhoum, N., Lin, Z., Hongming Z., Song, Y., & Yeung, D. Y. (2019). Multilingual and multi-aspect hate speech analysis. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing - 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).*

Parikh, P., Abburi, H., Badjatiya, P., Krishnan, R., Chhaya, N., Gupta, M., & Varma, V. (2019). Multi-label categorization of accounts of sexism using a neural framework. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing - 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP),* 1642-1652.

Qian, J., ElSherief, M., Belding, E., & Yang Wang, W. (2018). Leveraging intra-user and inter-user representation learning for automated hate speech detection. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2*.

Sawhney, R., Joshi, H., Gandhi, S., & Shah, R. R.(2021) Towards ordinal suicide ideation detection on social media. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining,* 22-30.

Sharaf, M., Hemdan, E. E. D., El-Sayed, A., & El-Bahnasawy, N. A. (2023). An efficient hybrid stock trend prediction system during Covid-19 pandemic based on stacked-lstm and news sentiment analysis." *Multimedia Tools and Applications, 82,* 23945-23977.

Shu, T., Wang, Z., Jia, H., Zhao, W., Zhou, W., & Peng, T. (2022). Consumers opinions towards public health effects of online games: an empirical study based on social media comments in China. *International Journal of Environmental Research and Public Health, 19*(19).

Sun, K., Han Wang, H., & Zhang, J. (2022). The impact factors of social media users forwarding behavior of Covid-19 vaccine topic: Based on empirical analysis of Chinese Weibo users". *Frontiers in Public Health, 10*

Sushith, M. (2022). Semantic feature extraction and deep convolutional neural network-based face sentimental analysis". *Journal of Innovative Image Processing,4(*3).

Tekiroglu, S. S., Yi Ling, C., & Guerini, M. (2020). Generating counter narratives against online hate speech: Data and strategies. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.*

Upadhyaya, A., Fisichella, M., & Nejdl, W. (2022). A multi- task model for sentiment aided stance detection of climate change tweets. *Athmospheric Sciences.*

Verma, A., Pruksachatkun, Y., Kai Wei, C., Aram G., Jwala D., & Yang Trista, C. (2022). *Proceedings of the 2nd Workshop on Trustworthy Natural Language Processing (TrustNLP 2022)*. Seattle, USA: Association for Computational Linguistics.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. (2018).Glue: A multi-task benchmark and analysis platform for natural language understanding. *Proceedings of the 2018 EMNLP Workshop Blackbox NLP: Analyzing and Interpreting Neural Networks for NLP*.

Xiao, L., Wu, X., Wu, W., Yang, J., & He, L. (2022). Multi-channel attentive graph convolutional network with sentiment fusion for multimodal sentiment analysis."ICASSP.

Author Information

Mohammed Rizvi
Exelon Corporation
Chicago, USA
Contact e-mail: *mrizvi0310@gmail.com*

To cite this article:

Rizvi, M. (2023). Decoding emotions: Harnessing the power of Python for sentiment analysis in social media. *The Eurasia Proceedings of Educational & Social Sciences (EPESS), 31,* 189-195.