



SENTIMENT ANALYSIS USING NLP AND ML

Prof. J. I. Nandalwar, Prof. D. B. Parase, Prof. P. S. Pandhare, Prof. S. M. Ishwargond

Assistant Professor, A, Assistant Professor, Assistant Professor, Assistant Professor
Computer Science and Engineering,
Shree Siddheshwar Women's College of Engineering, Solapur, Maharashtra, India

Abstract : In today's digital world we all know how sentiment analysis is important and their work of website building and creating any company profile in any social media and digital platform. It helps to tackle customer inquiries digitally and they're a way to determine whether to purchase any particular company product and maintain the review platform good to protect the face of the company. In this research paper, we discuss how NLP and ML work together to function the sentiment analysis. How NLP use certain tools and how they codify the human language and their way of transferring the data to meaningful conclusion and how MI uses python for sentiment analysis.

IndexTerms - sentiment analysis, NLP, MI, nltk, GPT.

INTRODUCTION

Behind in a situation, TO detect the view or emotion sentiment analytics inspect and discover the feeling and aim behind a bit of work or articulation or any manner of reporting. Humans speak different languages on this earth; the language is an essential tool to express our views and feelings. Whatever humans say by language sentiment is associated with that. The sentiment can be constructive or pessimistic or impartial as well [1]. Like take an example, a restaurant business company put on sale various food items like milkshakes, burgers, sandwiches, pizza, etc. They also design a website to reach customers, right now where the customer could order any type of food according to their choice, and they can give reviews and suggestions any time, to upgrade the value of food or may they like the taste food or they hate the taste [2]. By using these examples we can have 3 types of reviews given by customers. First, the constructive one signifies that the food is liked by customers. Second, the review going to be a rejection and hence company should focus on developing a strategy to improve the food. The third one going to be like the customer neither show any reaction, we can consider it as a neutral statement. By examining all the reviews the company should focus on the increase the quality and value of the food or promoting different strategies of brand awareness to increase their annual sales. But there are thousand and millions of reviews of their food and over time it's impossible to scan each customer's review and come to any decision making [4]. To solve this problem that time sentiment analytics comes to play, which analyzes huge amounts of reviews and helps to make the decision for future progress by using real-life evidence better than based on a tiny sample of data, and that time natural processing and machine learning help to analyze it [5].

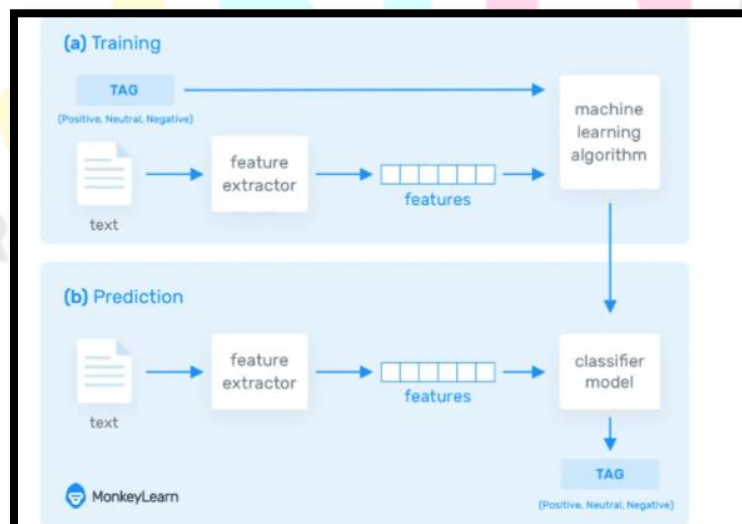


Figure 1: How sentiment analysis work

(Source: 3)

II.OBJECTIVE

1. To the importance of natural language processing.
2. To analyze the data through NPL.
3. To the importance of machine learning.

4. To do sentiment analysis by using python by using machine learning.
5. To the best mode of NPL is best for sentiment analytics.
6. To the best mode of ML algorithm is best for sentiment analytics.

.METHODOLOGY

social media handles and services to any digital marketing for any present or upcoming company [7]. Eventually, it helps to research market size and exponential growth, do a competitive analysis of other's competitor businessmen in the market, and create solutions for good industry-related problems and future growth. It has some disadvantages like sarcasm and irony negotiation types and ambiguity of words.

IV. IMPORTANCE OF NATURAL LEARNING PROCESSES.

At the current stage, humans speak rightly about 6500 languages, and they also have another sub and sign language. It really not possible to analyze all the languages and give us normal data. Language is not just a tool of communication, many sentiment and identity and also many cultures attach to it and become the carrier of all of this to the next generation [8]. But for machines, it's very challenging to analyze all of this sentiment and give us data. For that, there is called natural processing language. The NPL sub file is sentiment analysis, it helps identify and takes out the important point from the data.



Figure .2 analysis of human language

(Source: 5)

Natural language processing: what is it?

The field of computer science known as "natural language processing" (NLP) is more particularly the field of "artificial intelligence" (AI) that is concerned with providing computers the capacity to comprehend written and spoken words in a manner similar to that of humans. NLP blends statistical, machine learning, and deep learning models with computational linguistics—rule-based modelling of human language. With the use of these technologies, computers are now able to process human language in the form of text or audio data and fully "understand" what is being said or written, including the speaker's or writer's intentions and sentiment.

NLP powers computer programmes that translate text between languages, comply with spoken requests, and quickly summarise vast amounts of text—even in real time. You've probably used natural language processing (NLP) in the form of voice-activated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and other consumer conveniences. However, NLP is also taking on a bigger part in enterprise solutions that assist reorganise workplaces, boost worker output, and make complex but vital company procedures simpler.

Tasks involving NLP

The uncertainties in human language make it extremely challenging to create software that correctly ascertains the intended meaning of text or voice input. Here are just a few examples of the anomalies in human language: homophones, sarcasm, idioms, metaphors, unusual syntax and usage, and different sentence structures. The process of accurately translating voice data into text is known as speech recognition, commonly referred to as speech-to-text. Any programme that responds to voice commands or questions must use speech recognition. The way individuals speak—quickly, slurring words together, with varied emphasis and intonation, in various dialects, and frequently using improper grammar—makes speech recognition particularly difficult.

The act of identifying a word's part of speech based on its use and context is known as part of speech tagging, also in order to help the computer understand the text and speech data it is absorbing, several NLP activities deconstruct human text and voice data. Few of these jobs are listed below:

- Word sense disambiguation is the act of choosing a word's meaning from among its possible meanings using semantic analysis to discover which word makes the most sense in the context at hand. Word sense disambiguation, for instance, clarifies the difference between the meanings of the verbs "make" and "make the grade" (achieve) and "make a bet." (place).
- Words or phrases are recognised as useful entities using named entity recognition, or NEM. NEM identifies "Kentucky" as a place or "Fred" as the name of a guy.noun as grammatical tagging. 'Make' is classified as a part of speech.
- The task of determining whether and when two words refer to the same item is known as co-reference resolution. The most typical example is figuring out who or what a certain pronoun refers to (e.g., "she" = "Mary"), but it can also require figuring out a metaphor or idiom that is used in the text (e.g., when "bear" refers to a big, hairy person rather than an animal).
- Sentiment analysis looks for intangible elements in text, such as attitudes, feelings, sarcasm, bewilderment, and mistrust.
- Natural language generation is the process of converting structured data into human language; it is frequently referred to as the opposite of voice recognition or speech-to-text.For a more in-depth examination of how these concepts work.

NLP techniques and tools

Natural Language Toolkit and Python (NLTK)

For tackling particular NLP tasks, a variety of tools and libraries are available in the Python programming language. The Natural Language Toolkit (NLTK), an open source collection of libraries, tools, and educational resources for developing NLP programs, contains several of them.. The NLTK offers libraries for many of the above-mentioned NLP tasks as well as libraries for subtasks including sentence parsing, word segmentation, stemming and lemmatization (techniques for removing all but the most essential letters from words), and tokenization. (for breaking phrases, sentences, paragraphs and passages into tokens that help the computer better understand the text). Additionally, it contains libraries for developing skills like semantic reasoning, the capacity to draw logical conclusions from facts extracted from text.

Deep learning, machine learning, and statistical NLP

The original NLP applications were hand-coded, rules-based systems that were capable of carrying out certain NLP tasks, but they were unable to easily scale to handle an apparently never-ending stream of exceptions or the growing amounts of text and speech data. Enter statistical natural language processing (NLP), which combines computer algorithms with deep learning and machine learning models to automatically extract, classify, and label parts of text and speech input before estimating the statistical likelihood of each potential interpretation. Currently, deep learning models and learning methods based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) allow NLP systems to "learn" as they go along and extract ever-more-accurate meaning from enormous amounts of unlabeled, unstructured, raw text and voice data.

V. ANALYZE THE DATA THROUGH NPL

In this process,nltk (natural language toolkit) helps it by providing sets of archives for natural process language. Then the stop words provide a set of words that don't give any meaning to any sentences, then the WordNet lemmatize converts every human language word into the meaning full word according to the machine with keeping context behind it. Basically, it helps AI (artificial intelligence) to proceed and recognize human language that helps repetitive tasks to perform monotonously [9]. Many processes like checking spells, translation of machine, and classification of ticket summary. It also has five phases, such as analysis of discourse, analysis of pragmatics, analysis of syntactic, analysis of lexical, and analysis of semantics.

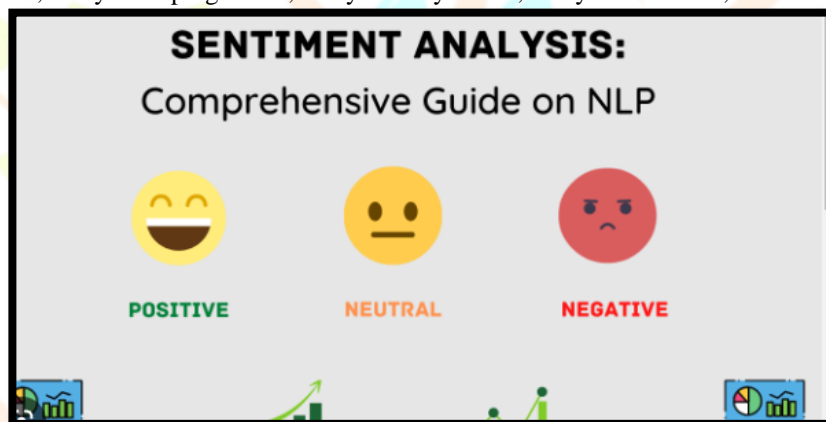


Figure 3: NLP guide

(Source: 2)

VI. IMPORTANCE OF MACHINE LEARNING

To analyze the polarity from negative to positive we need machine learning. By using it machines can determine the emotions of any sentence without the input of humans, basically, computers or AI can learn new tasks with no systematic program and perform them in a perfect way [10]. It can also be trained and it can be understood like sarcasm, definitions, context, and words that are misplaced. So, it's a much needed program for AI.

VII. SENTIMENT ANALYSIS IN PYTHON BY USING MACHINE LEARNING

Using MI sentiment analysis can define different architectures by using lstm rnn for new tasks. Bidirectional rnn of the multilayer is being used and that help to create rnn layers. Its major advantage is having more context in a directional in a single frame, such as a model in a flowing forward, when the model estimates the net sentence, it's going to be because of due reliance on the basis of earlier knowledge [11]. In a directional network, a machine has knowledge of the opposite directions of two networks and how to input catching flows of two opposite directions, Input parameters numbers, a layer of hidden, and the dimension of output can be defined by using the rate throughout and boolean of bidirectional. As a result, we get weights of embedding through pre-trained and duplicate the same model. In any case, the machine does not require the embedding of learning and also be the straight focal point directly to the job and can also learn those embedding related to sentiments. Some models regarding the parameter optimize the fast convergence by using Adam optimizer along with a long model of the loss function of logistics. By loping the epochs through the numbers and each epoch iterations by the number it can measure the batch size that already is there [12]. It can pass the model of text, the prediction, the calculation of every iteration loss, and also calculate the loss of propagate backward. The evaluating function of the function of training is the only major change and by that, we can evaluate the back ward propagate loss of torch and model. It's not gradient signifying while evaluating the descent way [13]. By the helpers of epoch helper functioning machine can calculate each epoch times and takes to its run completing and it's a way of print it.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Figure 4: naive Bayern formula

(Source: 8)

Mode	BEST
The mode for NLP	GPT is the best in case anybody uses NLP in the AI [14]. It helps the language of a pre-trained model that can help various tasks to be fine-tuned
The best mode of MI algorithm	There are several, such as vector machines of support, neural networks of multilayer perception, naive Bayes, and trees of decisions.

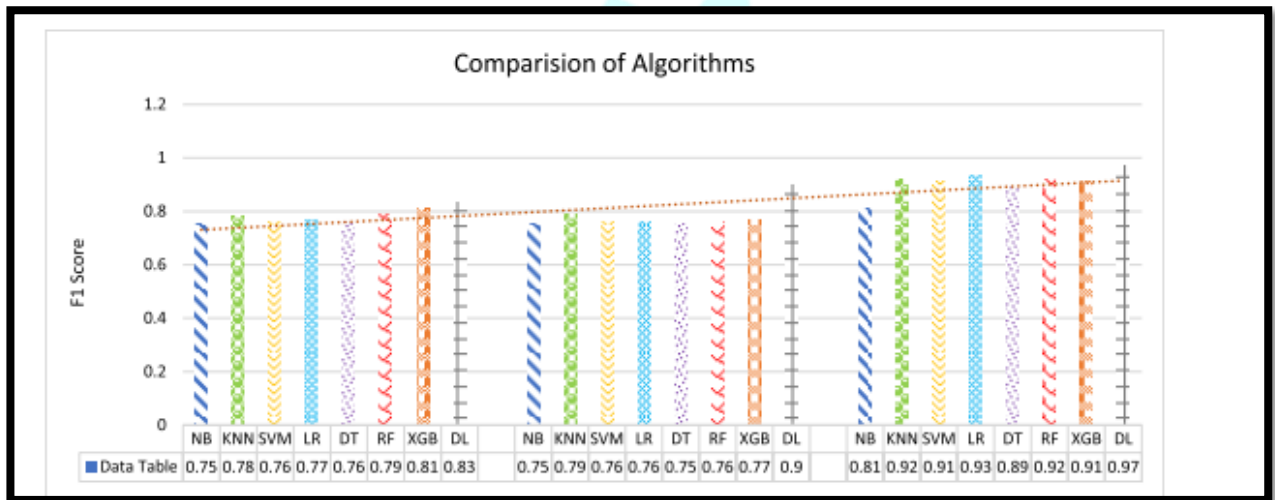


Figure.5 Algorithm uses in MI

(Source: 9)

VIII. PROBLEM STATEMENT

Due to the shortage of time, we could not extract primary sources of data, so we heavily rely on secondary sources. To conduct this research we get help from research articles conducted by notable researchers and also take help from their research data and their recent public general article [15]. Other than this, data is also extracted from various print media article or their online plat form. To extract the secondary source data get help from google scholar and online available interviews with researchers and software developers in this same field. Primary sources are the spot interview, known to investigate journalism, and research work is done by method given by researchers, so, not check the parameters, the surety of that data can vary from person to person due to changes in parameters.

CONCLUSION

Here this research paper discussed how sentiment analysis is needed for today's generation to improve business and how they need NLP to understand the human language and simplified for the machine. Along with that how MI helps in sentiment analysis. Which mode of analysis is the best available NLP and MI in the market and the algorithm behind it and the way it concludes that data? And give the solution to any real-life problem.

REFERENCES

1. Tusar, M. T. H. K., & Islam, M. T. (2021, September). A comparative study of sentiment analysis using NLP and different machine learning techniques on US airline Twitter data. In *2021 International Conference on Electronics, Communications and Information Technology (ICECIT)* (pp. 1-4). IEEE. Retrieved from: <https://arxiv.org/pdf/2110.00859> [Retrieved on:29/03/2023]
2. Su, G. M., Zhang, C., Wang, H., Chen, M., Kankanhalli, M., & Lienhart, R. (2019). Pushing the boundary of multimedia big data: an overview of ieemipr. *IEEE MultiMedia*, 26(2), 87-91. Retrieved from: <https://opus.bibliothek.uni-augsburg.de/opus4/files/58083/58083.pdf> [Retrieved on:29/03/2023]
3. Lucky, E. A. E., Sany, M. M. H., Keya, M., Khushbu, S. A., & Noori, S. R. H. (2021, July). An attention on sentiment analysis of child abusive public comments towards bangla text and ml. In *2021 12th international conference on computing communication and networking technologies (ICCCNT)* (pp. 1-6). IEEE. Retrieved from: https://www.researchgate.net/profile/Mumenunnessa-Keya/publication/355900946_An_Attention_on_Sentiment_Analysis_of_Child_Abusive_Public_Comments_Towards_Bangla_Text_and_ML/links/61dac854d4500608169be864/An-Attention-on-Sentiment-Analysis-of-Child-Abusive-Public-Comments-Towards-Bangla-Text-and-ML.pdf [Retrieved on:29/03/2023]

4. Qaisar, S. M. (2020, October). Sentiment analysis of IMDb movie reviews using long short-term memory. In *2020 2nd International Conference on Computer and Information Sciences (ICIS)* (pp. 1-4). IEEE. Retrieved from: https://www.researchgate.net/profile/Saeed-Qaisar/publication/346511493_Sentiment_Analysis_of_IMDb_Movie_Reviews_Using_Long_Short-Term_Memory/links/626174a8bca601538b5cd022/Sentiment-Analysis-of-IMDb-Movie-Reviews-Using-Long-Short-Term-Memory.pdf [Retrieved on:29/03/2023]
5. Sufi, F. K., & Khalil, I. (2022). Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis. *IEEE Transactions on Computational Social Systems*. Retrieved from: https://www.researchgate.net/profile/Fahim-Sufi/publication/358935794_Automated_Disaster_Monitoring_from_Social_Media_Posts_using_AI_based_Location_Intelligence_and_Sentiment_Analysis/links/62df83847782323cf1788dc8/Automated-Disaster-Monitoring-from-Social-Media-Posts-using-AI-based-Location-Intelligence-and-Sentiment-Analysis.pdf [Retrieved on:29/03/2023]
6. de Arriba, A., Oriol, M., & Franch, X. (2021, September). Applying Transfer Learning to Sentiment Analysis in Social Media. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)* (pp. 342-348). IEEE. Retrieved from: <https://upcommons.upc.edu/bitstream/handle/2117/361570/2021152892.pdf?sequence=3> [Retrieved on:29/03/2023]
7. Alhujaili, R. F., & Yafooz, W. M. (2021, March). Sentiment analysis for youtube videos with user comments. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (pp. 814-820). IEEE. Retrieved from: <https://ieeexplore.ieee.org/iel7/6287639/9312710/09612169.pdf> [Retrieved on:29/03/2023]
8. Long, F., Zhou, K., & Ou, W. (2019). Sentiment analysis of text based on bidirectional LSTM with multi-head attention. *IEEE Access*, 7, 141960-141969. Retrieved from: <https://ieeexplore.ieee.org/iel7/6287639/8600701/08845615.pdf> [Retrieved on:29/03/2023]
9. Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019, November). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)* (pp. 266-270). IEEE. Retrieved from: https://www.researchgate.net/profile/Abu-Kaisar-Mohammad-Masum/publication/342221481_Comparison_of_Naive_Bayes_and_SVM_Algorithm_based_on_Sentiment_Analysis_Using_Review_Dataset/links/5ee9d951a6fdcc73be82b934/Comparison-of-Naive-Bayes-and-SVM-Algorithm-based-on-Sentiment-Analysis-Using-Review-Dataset.pdf [Retrieved on:29/03/2023]
10. Raza, M. R., Hussain, W., Tanyıldızı, E., & Varol, A. (2021, June). Sentiment analysis using deep learning in cloud. In *2021 9th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-5). IEEE. Retrieved from: https://asafvarol.com/makaleler/Sentiment_Analysis_using_Deep_Learning_in_Cloud.pdf [Retrieved on:29/03/2023]
11. Mahmud, M. S., Bonny, A. J., Saha, U., Jahan, M., Tuna, Z. F., & Al Marouf, A. (2022, March). Sentiment analysis from user-generated reviews of ride-sharing mobile applications. In *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 738-744). IEEE. Retrieved from: https://www.researchgate.net/profile/Md-Shihab-Mahmud-2/publication/359933117_Sentiment_Analysis_from_User-Generated_Reviews_of_Ride-Sharing_Mobile_Applications/links/6259bd0709c5c2adb7dd5a7/Sentiment-Analysis-from-User-Generated-Reviews-of-Ride-Sharing-Mobile-Applications.pdf [Retrieved on:29/03/2023]
12. Elbattah, M., Arnaud, É., Gignon, M., & Dequen, G. (2021). The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications. *Healthinf*, 825-832. Retrieved from: <https://pdfs.semanticscholar.org/0b09/c20584c61cadd637120495b37c09d44969a2.pdf> [Retrieved on:29/03/2023]
13. Shahbazi, Z., & Byun, Y. C. (2021). Blockchain-based event detection and trust verification using natural language processing and machine learning. *IEEE Access*, 10, 5790-5800. Retrieved from: <https://ieeexplore.ieee.org/iel7/6287639/9668973/09665772.pdf> [Retrieved on:29/03/2023]
14. Wang, X., He, J., Jin, Z., Yang, M., Wang, Y., & Qu, H. (2021). M2Lens: visualizing and explaining multimodal models for sentiment analysis. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 802-812. Retrieved from: <https://arxiv.org/pdf/2107.08264> [Retrieved on:29/03/2023]
15. Asyofi, M. H., Yang, Z., Yusuf, I. N. B., Kang, H. J., Thung, F., & Lo, D. (2021). Biasfinder: Metamorphic test generation to uncover bias for sentiment analysis systems. *IEEE Transactions on Software Engineering*, 48(12), 5087-5101. Retrieved from: <https://arxiv.org/pdf/2102.01859> [Retrieved on:29/03/2023]

