International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

ANALYZING SARCASM IN TWITTER DATA USING MACHINE LEARNING ALGORITHUM

MS. NAMRATA GOKUL WAGH[1], PROF. HARISH BARAPATRE[2]

[1] Research Scholar, SES, Group of Institute, Bhivpuri Road Karjat (MS), w.namrata3@gmail.com, University of Mumbai.

[2] Professor, Yadavrao Tasgaonkar Institute, Bhivpuri Road Karjat (MS), harishkbarapatre@gmail.com, University of Mumbai.

Abstract: - It seems like We're providing information about sarcasm, sentiment analysis, and the evaluation of various machine learning algorithms for sentiment analysis. We text touches on how sarcasm is a form of expression where people say the opposite of what they mean, often to convey criticism. We also mention that sarcasm can be difficult to detect due to its obscurity and that sentiment analysis can be improved by analyzing and interpreting sarcasm in sentences. We go on to discuss Twitter as a platform where people express their thoughts and emotions in real-time, highlighting its significant volume of information and the potential challenges related to accuracy and trustworthiness since anyone can post anything. We mention the idea of Twitter battles where individuals attack each other. Lastly, we mention the evaluation of several machine learning algorithms (Random Forest, Naive Bayes, Bayesian Network, Support Vector Machine, K-Nearest Neighbour, Multi-Layer Perceptron) and a deep learning algorithm (Recurrent Neural Network) for sentiment analysis. According to We are evaluation, Random Forest performs the best in terms of prediction accuracy and error rate.

Keywords: K-nearest neighbours (KNN), Random Forest (RF), Naive Bayes (NB), Application programming interface (API), Online social networks (OSNs).

I. INTRODUCTION

We have highlighted some key points about the role of social networking websites in sharing ideas, viewpoints, and information, as well as the challenges related to interpreting sentiments and opinions accurately. Social media platforms indeed serve as powerful tools for exchanging thoughts and connecting with a global audience. Facebook and Twitter, as you mentioned, have massive user bases, making them influential platforms for communication. One crucial aspect of social media is the ability for users to express their thoughts on various topics, including emotions, activities, people, and products. This user-generated content often includes reviews, comments, and opinions about different aspects of life, including consumer products and experiences. People often turn to these platforms to read others' reviews and opinions before making purchasing decisions. Similarly, companies and organizations leverage these platforms to gather insights into consumer reactions to their offerings, helping them refine their products or services based on user feedback. However, there are challenges associated with interpreting opinions and sentiments accurately from the vast amount of content posted on social media. One notable challenge is the presence of sarcasm or irony.

Sarcasm can be difficult to detect algorithmically due to its reliance on context, tone, and cultural nuances. This challenge is compounded by the sheer volume of data generated on social media platforms daily. While natural language processing and sentiment analysis algorithms have made significant advancements in recent years, accurately identifying sarcasm remains a complex task. Sarcasm often relies on subtleties that can be challenging for algorithms to capture, and even human readers might struggle to accurately identify sarcasm in short texts like tweets or product reviews, as you pointed out. Researchers and developers are continually working on improving sentiment analysis models to better understand the nuances of language, including sarcasm and irony. Some approaches involve using context, user history, and linguistic patterns to enhance the accuracy of sentiment analysis. Additionally, as AI and machine learning technologies progress, there's hope that these challenges can be addressed more effectively in the future. In summary, social networking websites have become a significant platform for expressing opinions and sharing viewpoints, but challenges like accurately interpreting sarcasm in user-generated content still exist and are being actively researched and addressed by the technology community. Example, might seem like saying you enjoy when people don't notice you, but the use of the hashtag #sarcasm indicates that you actually don't like being ignored. Detecting sarcasm in online posts is challenging because it involves understanding the difference between the literal meaning of the words and the intended ironic or opposite meaning. This difficulty is amplified by the brevity and informal language often used on social media platforms like Twitter. Various natural language processing (NLP) techniques and machine learning algorithms are used to develop computer programs that can automatically detect sarcasm in text. These programs analyze patterns, context, and linguistic cues to identify instances of sarcasm. For example, they might look for exaggerated statements, contradictions, unexpected word choices, or the presence of hashtags like #sarcasm. Despite advancements in technology, detecting sarcasm accurately remains a complex task. Sarcasm is highly dependent on context and cultural nuances, which can be difficult for algorithms to grasp. Researchers continue to refine these algorithms by training them on large datasets of sarcastic and non-sarcastic text, allowing them to learn the subtleties of language and sentiment.

In the context of online promotion or messaging, detecting sarcasm is important because it can help businesses and individuals better understand the sentiment behind social media posts. It ensures that automated tools for sentiment analysis provide more accurate insights into how people are actually feeling, which in turn can guide marketing strategies, customer engagement, and public relations efforts for example It seems like you're discussing the concept of irony in tweets and how sentiment can be expressed in a counterintuitive or sarcastic manner. Irony often involves using words to convey a meaning that's opposite to their literal interpretation. In the case you've described, love is being used in a positive manner within a generally negative context, creating irony. This type of sentiment can be challenging to detect accurately, and it requires analyzing a substantial amount of text data, such as reviews, comments, or feedback messages, to understand the context and sentiment. Sarcastic tweets are a subset of ironic expressions where positive words are used to convey a cynical or pessimistic perspective. Detecting such sentiments involves using algorithms to classify the tweets correctly. These algorithms could be based on natural language processing (NLP) techniques that take into account the choice of words, sentence structure, and overall context to determine the sentiment's true intention. To accurately predict the orientation of such ironic or sarcastic tweets, a combination of machine learning, NLP, and sentiment analysis can be applied. By training models on a diverse dataset of tweets and their associated sentiments, the algorithms can learn to recognize patterns and differentiate between sincere and ironic expressions. Keep in mind that understanding irony and sarcasm often involves a nuanced grasp of cultural and contextual factors, which can be challenging for automated systems to fully capture.

Deep learning is a form of machine learning aims to solve perceptual problems such as speech and image recognition. A deep learning model is designed to analyse data with a logic structure similar to how a human 's brain work. To achieve this, deep learning uses a layered structure of algorithms called an Artificial Neural Network (ANN). Deep ANNs contain multiple hidden layers to recognise patterns and structure in large datasets. Each layer learns a concept from the data from subsequent layers. The core difference between machine learning and deep learning lies on the feature engineering. In machine learning, feature engineering can be done manually while deep learning does not depend on prior data processing and automatically extracts features. Twitter is one of the most popular social communication tools, where millions of users participate and discuss everything including their mood, news and events around them through a simple interface that enables the post of messages, photos and videos, you find in the trend many of the topics that may contain spam messages. Due to the popularity of Twitter, it becomes an attractive platform for spammers to spread spam. It has become a severe issue on Twitter.

II. RELATED WORK

In [2] Yi Tay, Tuan, and Cheung that focuses on detecting and classifying hateful speech using content from self-identifying hateful communities on Twitter. According to your description, the researchers developed an approach that utilizes a Naive Bayes classifier for this purpose. The results of their experiments indicate that the Naive Bayes classifier outperformed existing methods in terms of hate speech detection algorithms. Specifically, the precision, recall, and accuracy values achieved were 70.69% and 85.25%, respectively. Sarcasm is a form of communication where words are used to convey meanings that are opposite or different from their literal interpretations. This can be employed to convey humour, wit, offense, or other intentions. Sarcasm can sometimes be embedded within messages to convey information that the recipient is expected to understand, even though the literal words may seem contradictory. Sarcasm, like jokes or analyses, can serve various purposes and can be challenging for individuals to recognize, especially in written text. Detecting sarcasm is important for understanding the sentiment expressed by a user in platforms like Twitter and Facebook, which are often used for opinion sharing. The ability to detect sarcasm can enhance sentiment analysis and provide a deeper understanding of users' emotions and intentions. As a result of the potential benefits of detecting sarcasm for sentiment analysis and opinion mining, there has been a growing interest in automated sarcasm identification as a research topic.

We are comparation with Accuracy and Methods [1]: In the study, two methods were employed to classify tweets based on the polarity of user sentiment towards specific topics. The first method used the Support Vector Machine (SVM) algorithm and achieved an accuracy of 74.59%. The second method utilized a voting classifier, which achieved the highest accuracy of 83.53%. This indicates that the second method, with the voting classifier, was more successful in classifying tweets based on sentiment. [2][3]

Content and Features: The focus of the study was on classifying tweets based on the content and polarity of user sentiment. Different features and techniques were explored to achieve this classification. Some of these features include the presence of diagrams, word frequency, and non-text features like emoticons. These features can contribute to understanding the sentiment expressed in the tweets.

Emotion Context: The author of the study (possibly referenced as [3]) introduced a framework that involves learning to classify words and understanding the emotional context of these words. This approach likely helps in better capturing the emotional nuances present in the text, which is important for accurate sentiment analysis. It seems like the study aimed to improve sentiment analysis on social media platforms like Twitter by using different classification methods and considering various features, including both text and non-text features, to determine the sentiment polarity of tweets. It seems like you're discussing the challenges and potential approaches to utilizing Machine Intelligence for detecting sarcasm in Twitter and other semi-structured knowledge formats. You've mentioned that the absence of a static form for sarcasm in the data stream makes it difficult to predict. Instead of relying on simple heuristics like pattern matching or context, you're suggesting a more comprehensive approach. In this paper, His refer to a study by [11] that demonstrates how different digital technologies can be used to address societal issues and barriers to free speech. This involves classification schemes for descriptions and tweet classifications, using a hyperbolic feature set. They are also proposing a potential analysis that involves resolving semantic uncertainty using a radical Recurrent Neural Network paradigm. This paradigm would utilize the functionality and metadata generated by the current model. To achieve this, you suggest considering Bidirectional LSTMs for context identification and the VADER library for performing a comprehensive emotion analysis. It seems like your work is aiming to develop a sophisticated system that can detect sarcasm in a more nuanced and accurate manner, even in a semi-structured format like Twitter. By leveraging advanced techniques like Recurrent Neural Networks, Bidirectional LSTMs, and emotion analysis, you're looking to enhance the understanding of context and emotional cues to better identify instances of sarcasm. This is indeed a challenging yet promising endeavour in the realm of natural language processing and sentiment analysis

It appears that their implementation is focused on real-time detection, analysis, and mitigation of shaming tweets. Shaming tweets can be harmful, and your project aims to address this issue using advanced techniques. However, they noted a potential limitation. If the language to be detected is provided in multiple languages or is not within the dataset provided to his project, it results in an error that can impact the detection of tweets. This highlights the challenge of dealing with multilingual and diverse content that goes beyond the scope of the dataset of his project is trained on. To address this limitation, we might consider incorporating a broader range of languages in our training data or implementing language detection mechanisms. This way, your system could better adapt to different languages and increase its effectiveness in detecting shaming tweets across various linguistic contexts. Additionally, we might explore techniques like transfer learning or using pre-trained language models that support multiple languages to enhance the language coverage of your system.

III. Proposed System

It's absolutely mind-blowing how data mining techniques struggle to unravel the intricate web of sarcasm within conversations. I mean, who wouldn't be impressed by their valiant attempts to decipher the hidden nuances of tone, pitch, and intensity? And let's not forget the eloquent dance of eyes and hands that come into play, truly the epitome of straightforward communication. And printed sarcasm? Oh, it's a sheer delight to witness the delicate dance of irony across official letters, emails, and the highbrow realm of product reviews. It's almost like a game of hide-and-seek where the meaning slyly disguises itself in a cloak of formality, leaving us mere mortals to grapple with its enigmatic essence. Truly, sarcasm is a realm that data mining can only dream of conquering. In the proposed work that involves different categories to achieve a desired result, which seems to be related to sarcasm detection. We've mentioned four categories: data collection, data preprocessing, feature creation, and sarcasm detection. Additionally, we've referred to a table for feature creation. "Data Collection" phase of we proposed work, where we are emphasizing the importance of collecting valid and unbiased data for your analysis. We've mentioned that We're using a tool called "twint," which is a Python library used for scraping tweets from Twitter without using the official Twitter API. This tool allows you to gather data related to the keyword "sarcasm," and you've collected a total of 10,000 tweets for further processing. "Data Preprocessing" phase of your work, which involves transforming the collected data into a more suitable format for analysis. This phase typically includes several steps to clean and prepare the data for further processing. Let's break down the steps you've mentioned:

Hashtag Removal: You start by identifying hashtags in the input data (presumably the tweets you've collected) and then removing these hashtags from the data. This step aims to eliminate the hashtags that might not directly contribute to the analysis of sarcasm.
Field Selection: In this step, you choose specific fields or attributes from the collected data that are relevant to your analysis. This could involve selecting specific parts of the tweets, such as the text content, timestamps, user information, etc., that are necessary for your sarcasm detection task.
Data Cleaning and Noise Removal: This step involves cleaning the data by removing any irrelevant or noisy information. Noise could include typographical errors, irrelevant characters, or data points that don't add value to your analysis.
Tokenization: Tokenization is the process of splitting text into individual words or tokens. This step is crucial for further analysis because it breaks down the text into units that can be processed more easily.
Stemming: Stemming is a technique in natural language processing that involves reducing words to their root forms. This helps in treating similar words with different forms as the same word, which can improve the accuracy of analysis.

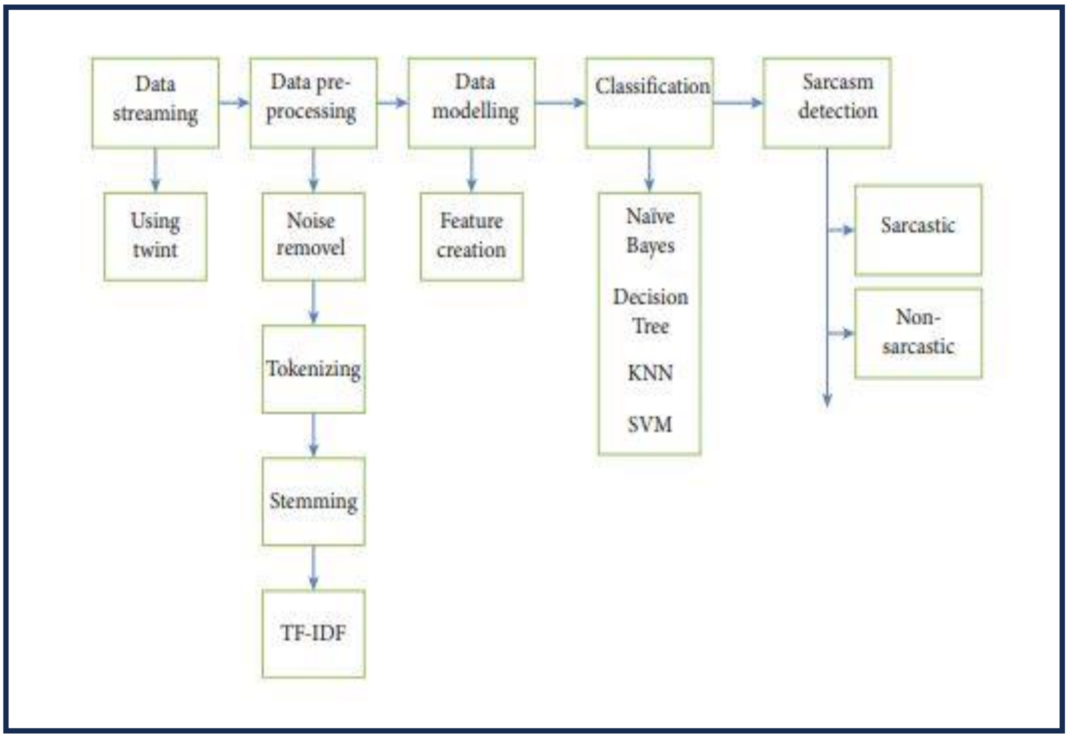

Figure No-1: Proposed Block Diagram

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

	Target	0	1	2	3	4	5	6	7	8	...	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
0	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
8568	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8569	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8570	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8571	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8572	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

8573 rows × 2001 columns

Figure N0-2 Data set ID description.

New text can be found while we searching tweets using data set as shown in figure-3 the table, and a new data frame is made with just the column recognized as “tweets”. Figures 3 describe the words stated above, where the only concerned data field, i.e., tweets are considered, and a new data frame is made which will commence the further processing. Further Table 3 and Table 4 are summarizing the results and accuracy level obtained through different classifiers. 8174 entries; 0 to 8573. Data columns (total 1 column).

	Target	new_text
0	1	leader squandered money abandoned parthic hosp...
1	1	rt million u resident trillion wealthy economy...
2	1	baloch student pushed away education system we...
3	1	rt already happening rule r shame farmersprote...
4	1	inspite clear word calling dint recieve call pm
...	...	...
8568	0	made rafa fact newcastle premierleague
8569	0	nothing satisfying stuffing strawberry cheesec...
8570	0	playstation suck fact via
8571	0	half working office dedicating talking shit pe...
8572	0	littlemarco fyi none candidate could win presi...

8573 rows × 2 columns

Data Cleaning is the process of the passage emphasizes that the primary goal of data cleaning is to remove irrelevant or uninformative data from the dataset. This process helps in making the dataset more suitable and useful for the desired study or analysis. The passage mentions that the dataset contains special symbols, which might be irrelevant or disruptive to the analysis. Therefore, it's important to remove these special symbols from the data. Special symbols could include punctuation marks, mathematical symbols, or any character that is not relevant to the analysis. The passage indicates that the "re" library, which is a built-in library in Python, is commonly used for data cleaning tasks that involve pattern matching and text manipulation using regular expressions. Regular expressions are powerful tools for identifying and manipulating specific patterns within text. Unfortunately, the passage does not provide the example of data cleaning as mentioned. However, data cleaning examples might include tasks like removing punctuation, converting text to lowercase, removing extra spaces, and handling missing values.

Noise Removal in Text Analysis: Noise refers to any irrelevant or disruptive elements in the text data that can hinder the analysis process. In the context of text analysis, removing noise is essential to ensure that the data is as relevant and informative as possible for the intended study or analysis. Importance of Noise Removal in Text Classification is the passage emphasizes the importance of noise removal, especially in text classification tasks. For accurate classification, it's crucial to have clean and focused data that highlights the significant features for classification. Maximizing Output through Data Processing is the passage suggests that data processing techniques are applied to text data to achieve maximum output. These techniques are designed to enhance the quality and relevance of the data, making it more conducive to successful analysis. Noise Removal Techniques is the passage mentions various techniques used for noise removal in text data: Removing Characters and Digits is a Characters and digits that don't contribute to the context or meaning of the text are removed. Removing URL is nothing but URLs or web links are often removed, as they might not be relevant to the analysis and can introduce noise. Removing Stop Words is Stop words are common words like "the," "and," "is," etc., that are often removed because they occur frequently in text but typically don't carry significant meaning. Removing Punctuation: Punctuation marks are removed to focus on the actual words and their arrangement. Removing Pieces of Text: Portions of text that are irrelevant to the analysis are removed to ensure the dataset is streamlined and meaningful Cleaned Data for Next Phase: After noise removal, the cleaned data is ready for the next phase of analysis. This phase might involve tasks like feature extraction, text classification, sentiment analysis, and more. In summary, noise removal is a fundamental step in text analysis, particularly in text classification tasks. By removing irrelevant elements such as characters, digits, URLs, stop words, punctuation, and unnecessary text portions, the data becomes more focused and suitable for the subsequent phases of analysis. Noise removal contributes to the accuracy and meaningfulness of the analysis outcomes.

Stemming is another important aspect of natural language understanding, reducing the word to its stem making it viable in reducing the vocabulary and summarizing different words to their roots for input making it easier for the analysis. The main aim of this is to reduce the repetition of words by dropping the suffix of the word to arrive at the basic form of the word [16]. Utilization of stemming is done in the commenced work by reducing the word to its stem so that the vocabulary is reduced.

Sarcasm and humour are distinct human qualities that involve nuanced expressions and subtleties that can be challenging for artificial intelligence systems to comprehend and reproduce. AI is striving to become more human-like in its understanding of these aspects of communication. Challenges in Sentiment Categorization, the passage points out that there are algorithms in machine learning designed to address these challenges by categorizing text based on sentiment (which could include positive, negative, neutral, sarcastic, humorous, etc.). However, this task is not straightforward due to the inherent complexities of human language and its

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

varied expressions. Difficulties in Sentiment Categorization, the passage alludes to the fact that categorizing sentiment, especially when it involves sarcasm and humor, is particularly difficult. The specific difficulties are not elaborated upon in this excerpt, but it suggests that these challenges are complex and multifaceted. The Question of Features, the passage raises a key question regarding sentiment analysis: "What kinds of features do we use?" In the context of machine learning, "features" refer to the characteristics or attributes of the text that are used to train the algorithms. Determining which features are most relevant and effective for capturing sarcasm, humor, or other sentiments is a critical consideration.

In summary, the passage highlights the complexities associated with AI's attempt to understand and replicate human qualities such as sarcasm and humor in text. It acknowledges that while machine learning algorithms exist for sentiment analysis, categorizing sentiments involving these nuances is challenging. The passage concludes by pointing to the question of identifying appropriate features to effectively capture such nuanced sentiments. The sarcasm detection in the proposed model is classified using different supervised categorical machine classification: decision tree, Naïve Bayes, KNN, and support vector machine. Decision tree classifier poses a series of carefully crafted questions about the features that are supplied to the algorithm [8]. Naïve Bayes is a log-linear model; that is, in both cases the probability of a document belonging to a class is proportional. The model is trained and tested with a ratio of 80: 20 splits of the 8000 tweets collected. Accuracies are gained from different classifiers, and the prediction of a particular statement is done on the bases of the highest accuracy among the four classifiers.

IV Results

1. Count Analysis Of tweets

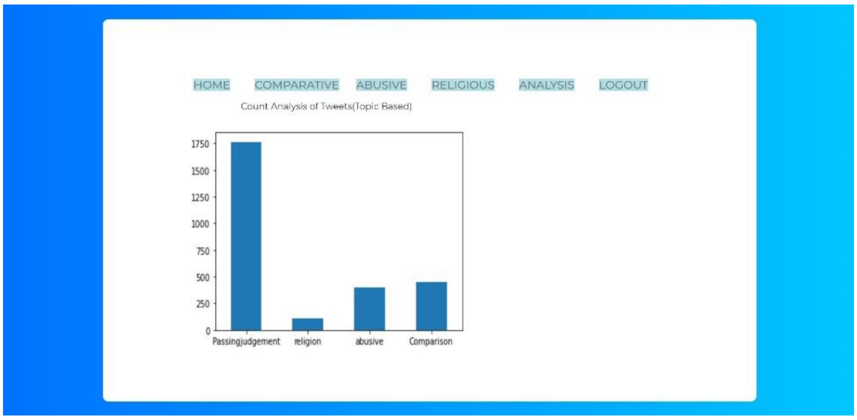

Figure N0-3 Tweets Count analysis

It's clear that we have a good understanding of the text and opinion mining area, specifically focusing on Twitter sentiment analysis using Python. Our description outlines the various steps involved in the process, including data gathering, text preprocessing, sentiment detection, classification, model training, and testing. The progress made in this field over the last decade, with models achieving efficiencies of around 85% to 90%, is impressive. We've also pointed out some important challenges and limitations in the current state of sentiment analysis. The lack of data diversity is a significant concern as it can impact the model's generalization and ability to handle various types of tweets and sentiments.

2. Tweeter Accuracy Model

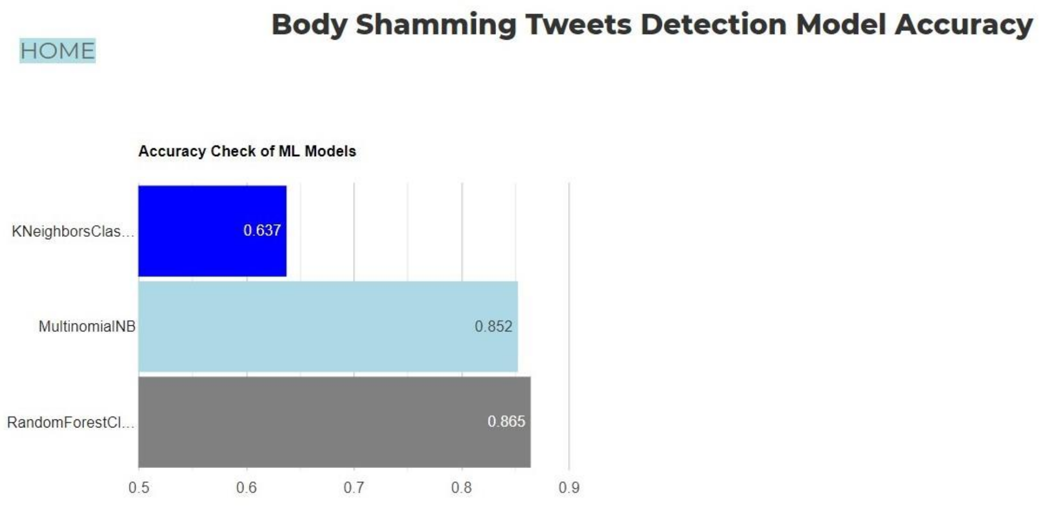

Figure N0-4 Accuracy Model

The use of slang and abbreviated forms of words further complicates the analysis and classification process. Raising the number of sentiment classes can indeed be challenging, as it requires the model to distinguish between more nuanced emotions and sentiments. This can lead to decreased performance for some sentiment analysers. The point you raised about the model's accuracy for topics other than the specific one under discussion is crucial. Models trained on one domain or topic might struggle to perform well when applied to different subjects, as the language, sentiment expressions, and context can vary significantly. Our assessment of sentiment analysis having a lot of room for growth in the future is accurate. Here are a few potential directions for improvement and growth in this field. Data Diversity and Representation is also to improve model performance, gathering diverse and representative data is crucial. This helps the model to understand and classify sentiments across different demographics, languages, and cultural contexts. Handling Slang and Abbreviations is also helps to Developing models that can accurately understand and interpret slang, colloquialisms, and abbreviated words will lead to better performance on real-world social media text. Multilingual Sentiment Analysis is Extending sentiment analysis to different languages can greatly enhance the utility of such models, allowing businesses and researchers to analyze sentiments in a global context. In conclusion, sentiment analysis has made significant strides, but challenges related to data diversity, domain adaptation, and contextual understanding still remain. As AI and natural language processing technologies continue to evolve, there's ample room for research, innovation, and improvement in the field of sentiment analysis.

REFERENCES

[1] Da Silva, N.F., Hruschka, E.R., & Hruschka Jr, E.R. (2014). "Tweet sentiment analysis with classifier ensembles". Decision Support Systems, 66, 170-179.

[2] Das, D., & Clark, A.J. (2018). "Sarcasm detection on Facebook: A supervised learning approach". In Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct, (pp. 1-5).

[3] Fersini, E., Pozzi, F.A., & Messina, E. (2015, October). Detecting irony and sarcasm in microblogs: The role of expressive signals and ensemble classifiers. In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-8). IEEE.

[4] Vinodhini, G., & Chandrasekaran, R.M. (2013). Effect of feature reduction in sentiment analysis of online reviews. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(6), 2165-2172.

[5] Dash, B., Mishra, D., Rath, A., & Acharya, M. (2010). A hybridized K-means clustering approach for high dimensional dataset. International Journal of Engineering, Science and Technology, 2(2), 59-66.

[6] Jotheeswaran, J., & Kotees waran, S. (2016). Feature selection using random forest method for sentiment analysis. Indian Journal of Science and Technology, 9(3), 1- 7.

[7] Bharti, S.K., Vachha, B., Pradhan, R.K., Babu, K.S., & Jena, S.K. (2016). Sarcastic sentiment detection in tweets streamed in real time: a big data approach. Digital Communications and Networks, 2(3), 108- 121.

[8] Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis. Procedia computer science, 132, 937-946.

[9] Vinodhini, G., & Chandrasekaran, R.M. (2014). Sentiment classification using principal component analysis based neural network model. In International Conference on Information Communication and Embedded Systems (ICICES2014) (pp. 1- 6). IEEE.

[10] Garg, A., & Duhan, N. (2020). Sarcasm Detection on Twitter Data Using Support Vector Machine. ICTACT Journal of soft computing, 10(4), 2165-2170.

[11] Ashwitha, A., Shruthi, G., Shruthi, H.R., Upadhyaya, M., Ray, A.P., & Manjunath, T.C. (2021). Sarcasm detection in natural language processing. Materials Today: Proceedings, 37, 3324-3331.

[12] Suhaimin, M.S.M., Hijazi, M.H.A., Alfred, R., & Coenen, F. (2017). Natural language processing-based features for sarcasm detection: An investigation using bilingual social media texts. In 2017 8th International conference on information technology (ICIT) (pp. 703-709). IEEE.

[13] I lavarasan, E. (2020, March). A Survey on Sarcasm detection and challenges. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1234-1240). IEEE.

[14] Kumar, A., Narapareddy, V.T., Srikanth, V.A., Malapati, A., & Neti, L.B.M. (2020). Sarcasm detection using multi-head attention based bidirectional LSTM. Ieee Access, 8, 6388-6397.

[15]. Bouazizi, M., & Oht suki, T.O. (2016). A pattern-based approach for sarcasm detection on twitter. IEEE Access, 4, 5477- 5488. [16]. Jurek, A., Bi, Y., Wu, S., & Nugent, C. (2014). A survey of commonly used ensemble-based classification techniques. The Knowledge Engineering Review, 29(5), 551.

[16] Liyuan Liu, Jennifer Lewis Priestley, Yi Yun Zhou, Herman E. Ray and Meng Han, A2Text-Net: A Novel Deep Neural Network for Sarcasm Detection‖, IEEE First International Conference on Cognitive Machine Intelligence (Cog MI),2019.

[17] Swati Jain, Suraj Prakash Narayan, Rupesh Kumar Dewang, Utkarsh Bhartiya, Nalini Meena and Varun Kumar, "A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter", IEEE,2019.

[18] Rahul Gupta, Jitendra Kumar, Harsh Agrawal and Kunal, "A Statistical Approach for Sarcasm Detection Using Twitter Data", IEEE International Conference on Intelligent Computing and Control Systems (ICICCS),2020