Sentiment Analysis using Machine Learning Algorithms

Adarsh Umadi, Abrar Kadri, Jyotsna Dekhale, Hamza Shaikh

Dr. D.Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India

ARTICLEINFO

ABSTRACT

Sentiment analysis has grown in importance in both the scientific and commercial spheres as a result of its enormous potential to completely transform a wide range of sectors. Many companies have responded to this increasing significance by incorporating sentiment analysis and customer perception as essential elements of their overall strategies. But the automated analysis of social network posts, where the rich tapestry of human emotions and expressions is sewn, is one of the most fascinating uses of sentiment analysis. This chapter sets out on a revolutionary adventure to bridge the gap between the sophisticated emotional fabric of social networks and the state-of-the-art sentiment-based approaches and technology that have the potential to completely change this rapidly developing field. The main result of our investigation is a thorough study, a complex tapestry in and of itself, that presents the most engaging and cutting-edge methods for the comparison and classification of communications in the ever-changing world of social media platforms. This survey includes a thorough explanation of cutting-edge methods and technologies that have the potential to revolutionize sentiment research and shed light on the complexities of human emotion in the digital era.

Keywords : Big Data, Data Cleaning, Social-Media, Regression, Machine Learning, Supervised Learning, Text Analysis, Classification

I. INTRODUCTION

The area of sentiment analysis, which has its roots in information extraction and natural language processing, has seen a tremendous historical transformation. It continues to influence our comprehension of and interactions with the immense amount of textual data available on the internet, both in its historical context and in its modern applications. We explore the background, inspiration, and fundamental findings that have propelled the development of sentiment analysis in this introduction, with an emphasis on comments and emojis in particular.

1.1 History

The field of sentiment analysis originated with early linguists and philosophers who attempted to decipher the feelings and opinions expressed in written texts.

But the true change came with the introduction of the internet and the explosion of user-generated content on many platforms. The amount of textual material has increased dramatically as a result of the widespread use of social media, blogs, and internet forums. Sentiment analysis underwent a sea change from its early phases to the current ones as a result of this. Thanks to the development of new techniques, sentiment analysis has greatly improved in terms of accuracy and scalability over time. These techniques have improved the field's capabilities by moving from earlier rule-based approaches to more advanced machine learning algorithms.

1.2 Motivation

Sentiment analysis is in high demand due to the internet's unrelenting growth and people's ease of sharing thoughts and emotions online. Businesses, academia, and governments are among the stakeholders who have been forced to fully embrace sentiment analysis for a variety of reasons. In the business sector, making educated decisions requires having a thorough awareness of client emotion. Businesses use sentiment analysis to track how consumers feel about their brands, evaluate customer feedback, and quickly resolve issues raised by customers. In the sociological and political sciences, sentiment analysis has also proven to be useful in examining public opinions toward important problems and leaders. Sophisticated techniques, resources, and software tools devoted to sentiment analysis have been developed as a result of the need to capitalize on the abundance of digital expression available online.

1.3 Basic Observation

Providing significant insights into the attitudes and emotions portrayed in text is the main goal of sentiment analysis. It acknowledges that textual data is a rich reservoir of human sentiment that spans a wide range of emotions, from contentment and disappointment to rage and delight. Sentiment analysis explores the complex realm of human emotions and perceptions, going beyond the simple gathering of words. The core of this area is comprehending people's ideas and feelings, particularly in the context of their interactions on digital platforms. This sentiment's analysis and quantification provides important new information on the psychology of online communities. Emojis have given this project a layer of depth and nuance that makes it much more exciting and difficult because of their ability to portray emotions efficiently

II. Literature Survey

Paper Title	Year	Authors	Methods	Dataset	Abstract
Sentiment Analysis of Comments In Social Media	2020	Abdulrahman Alrumai, Ruaa Alsabah, Hiba J Aleqabie, Ahmed Yaseen Mjhool, Ali Al-Sabbagh, and James Baldwin	Chi SquareRandom Forest Classifier	Personalized dataset created	Emojis support the text's sentiment orientation, according to a sentiment analysis of Arabic comments on Twitter during the 2018 World Cup, but neither text nor emojis by themselves are reliable indicators of sentiment.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

65

Nonverbal Communication with emojis in Social media: Dissociating Hedonic intensity From frequency	2022	Lu Li, X. T. Wang	Emojis· Hedonic expressions· Public sentiment· Social-media Nonverbal communication	Personalized dataset created	Emojis, a usual type of nonverbal communication in social media, can predict swiftly how the general public will feel about a social event. Emoji usage studies in the past revealed that positive emojis are used with greater frequency than negative ones.
Analysis of Word Vector Representation Techniques with Machine-Learning Classifiers for Sentiment Analysis of Public Facebook Page's Comments in Myanmar Text	2020	Hay Mar Su Aung; Win Pa	High-dimensional word vectors, sentiments classification Machine-Learning classifiers, word vector representation analysis, logistic regression	Personalized dataset created	This essay compares three alternative machine learning approaches for sentiment analysis in the context of Myanmar. The core task of sentiment analysis (SA) is to extract and pinpoint the arbitrary data, or social sentiment, present in the original text.
Prediction of Rating from comments based on information retrieval and sentiment analysis	2016	Eissa M. Alshari; Azreen Azman; Norwati Mustapha	Multiclass Decision Jungle Multiclass Decision Forest Multiclass Neural Network Multiclass Logistic Regression	Information retrieval Internetretail data processing	Online shopping websites regularly experience an increase in users. These websites frequently offer consumers the option to comment on and rate the goods that are sold on the websites.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

66

Retrieving youtube video by sentiment analysis on user comment	2017	Hanif Bhuiyan; Jinat Ara;	K-nearest neighbor, Support vector machine, Decision Tree, Navier buyers	Personalized dataset created	One of the most wide-ranging online video information sources, Youtube regularly and swiftly uploads new videos. It is one of the most widely recognized social media platforms.

III. Problem Statement

In the age of pervasive online communication, the imperative to develop highly effective machine learning models for sentiment analysis has become paramount. These models need to be able to correctly identify the emotions hidden in user comments and emoticons on review and social networking sites. This urgency arises from the issue of managing the growing amount of unstructured data created in the digital realm, which cuts across industrial borders. Sophisticated sentiment analysis techniques are necessary for both businesses and government organizations to obtain accurate insights into public opinion, developing trends, and consumer sentiments. This project's main goal is to deal with the numerous difficulties involved in managing enormous, disorganized data sources.By doing this, it aims to provide decision-makers from a variety of sectors with the information and understanding needed to take well-informed, data-driven decisions. Essentially, the goal of this project is to close the gap that exists between the ever-growing landscape of online communication and the vital requirement for precise sentiment analysis. This will improve the capacity to traverse and utilize the abundance of information present in these digital spaces.

IV. Proposed Solution

This research article offers a thorough and multifaceted answer to the complex problems associated with sentiment analysis in the context of social media. This strategy includes a number of essential elements, each of which adds to a comprehensive solution:

1. Data Collection:
The first step of the process involves gathering data from different social media platforms, where users contribute a large amount of textual data as well as, occasionally, emojis. The first phase entails gathering this raw data in a methodical manner to make sure it is a representative sample of the target domain.

2. Data Preprocessing:
Extensive preprocessing is carried out after the data is collected. Emoji extraction, text normalization, and handling missing values are some of the responsibilities that fall under this stage. Since social media raw data might be noisy and unstructured, these procedures are essential to guaranteeing data consistency and quality.

3. Feature Engineering:
This crucial stage involves extracting pertinent features from the preprocessed data. These capabilities could be text-based as well as, most famously, emoji-based. To capture the richness of emotional expression in text, methods such as sentiment scores associated with emojis, word embeddings, and TF-IDF (Term Frequency-Inverse Document Frequency) are used.

4. Machine Learning Models:
The research paper emphasizes the utilization of machine learning models for sentiment analysis. Choosing the right models, training them on the prepared dataset, and thoroughly assessing their

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

67

performance are all steps in the process. Ensuring accurate sentiment classification and grading is the goal in order to make the models dependable for practical use.

5. Emojis in Real-World Applications:

This research is unique in that it examines emojis' function in sentiment analysis in great detail. Emojis are emotive graphic expressions that are becoming more and more common in digital communication. It is important to comprehend how they affect sentiment, and the research sheds light on this aspect. Furthermore, the study highlights the adaptability and usefulness of the created sentiment analysis tools by proposing real-world uses for sentiment analysis, such as business intelligence and website reviews.

6. Continuous Improvement:

The research supports continuous improvement because it acknowledges that sentiment analysis and natural language processing are always changing fields. In order to improve the models and procedures in light of new trends and advancements in the industry, feedback loops are used. Maintaining the relevance and efficacy of the established tools requires keeping up with technological improvements in sentiment analysis.

7. Enhanced Decision-Making:

Enhancing decision-making and informed action across a wide range of disciplines is the ultimate goal of this approach. The methodology offers businesses, governments, and other stakeholders significant insights into consumer attitudes, trends, and public opinion through accurate sentiment classification and efficient data processing. This gives them the ability to make data-driven decisions that are in sync with how online communication is evolving constantly.

In simple terms, the approach offered in this research study provides a flexible and methodical framework for sentiment analysis, tackling the particular difficulties brought about by social media and online communication. In an increasingly digital world, it helps to make decisions that are more informed and more successful by bridging the gap between unstructured data and actionable insights.
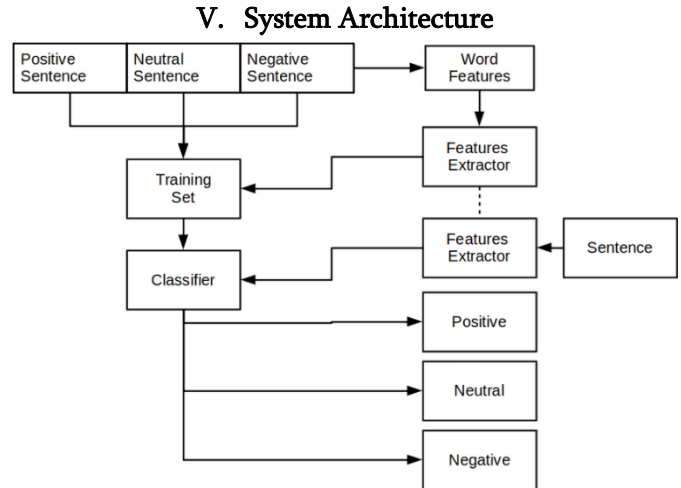
V. System Architecture

Fig. 1. System architecture for the system

A sequence of actions is taken during the categorization process in order to ascertain the emotion of a particular text. The process starts with the construction of a training set, which consists of a wide variety of texts classified into three mood categories: neutral, negative, and positive. Text stemming is used to reduce words to their root form for more consistent analysis in both the training set and the input text. In order to improve the emphasis on content that is relevant, common stop words—which do not convey any major sentiment information—are also eliminated. Next, the input text's polarity is ascertained by classifying it as positive, negative, or neutral in accordance with the patterns and terms found during the training stage.Iteratively applying this categorization technique to every phrase in the text results in a cumulative output that, when paired with lexical information, helps determine the overall sentiment intensity of the text. Through the use of this methodical technique, sentiment analysis is able to accurately determine the emotional content of textual material, which opens up a variety of applications such

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

68

as sentiment scoring and the interpretation of customer and public opinion.

VI. Conclusion

In conclusion, this study has increased our understanding of the urgent need for sophisticated sentiment analysis methods, particularly in the setting of social media and online communication. A new era of information interchange has been brought about by the digital age, when individuals' thoughts and feelings are frequently hidden under enormous amounts of textual data. Our study has illuminated the possibility of using data cleansing techniques and machine learning algorithms to glean insightful information from this torrent of data. We have demonstrated the efficacy of feature engineering and thorough data pretreatment in enhancing sentiment grading and classification.Our results show how sentiment analysis can be applied in a variety of contexts and industries. e-commerce and government intelligence are only two of the many real-world uses for our study. Sentiment analysis is a powerful tool that government organizations can use to monitor changes in public opinion, assess popular sentiment, and effectively address new challenges. In the interim, companies might utilize sentiment analysis to acquire more profound understanding of customer sentiments, assess product reviews, and adjust their approaches accordingly. This research has implications for the political, marketing, and market research domains, among others, as better decision-making can result from a knowledge of sentiment.Although we acknowledge the inherent limitations of our study, including the difficulties in managing noisy data and the always changing context of online communication, our work lays the groundwork for further research. The dynamic digital landscape offers a range of prospects and obstacles, necessitating ongoing enhancement and flexibility. As our research has shown, feedback loops are crucial for improving the accuracy and resilience of sentiment analysis

models.Our work adds to the continuing discussion on enhancing automated tools for comprehending human emotions and views in the larger context of machine learning and sentiment analysis applications. We think that the future of sentiment analysis will be greatly influenced by our research, which will also assist people and organizations grasp the nuances of human sentiment in the digital era and make more educated decisions. In a quickly changing digital context, our work opens the path for more efficient communication, more focused plans, and ultimately better informed actions by bridging the gap between unstructured web data and actionable insights.

VII. REFERENCES

[1]. Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP; 2005

[2]. Lambov Dinko, Pais Sebastiao, Dias Gael. Merged agreement algorithms for domain independent sentiment analysis. In:Presented at the Pacific Association

[3]. Barbosa, L., and Feng, J., Robust sentiment detection on twitter from biased and noisy data, In 23rd International Conference on Computational Linguistics: Posters, 36–44, 2010

[4]. C. Whitelaw, N. Garg, and S. Argamon, Using appraisal groups for sentiment analysis, In Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM), pp. 625–631, ACM, 2005

[5]. Hirst, G. & St-Onge, D., 'Lexical chains as representation of context for the detection and correction malapropisms, 1997

[6]. Rupesh Mahajan, "An Analytical Evaluation of Various Approaches for Load Optimization in Distributed System". International Journal of Intelligent Systems and Applications in Engineering, August 2023, Volume 12 Issue 1, 526–548.

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

69

[7]. Rupesh Mahajan, "Novel Load Balancing Optimization Algorithm to Improve Quality-of-Service in Cloud Environment", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 11, Issue 2, March 2023.

[8]. Rupesh Mahajan, "A Survey on Intelligent and Effective Intrusion Detection System Using Machine Learning Algorithm", IJERT, Volume 9, Issue 1, 2020.

Cite this article as :

International Journal of Scientific Research in Science, Engineering and Technology | www.ijsrset.com | Vol 10 | Issue 6

70