

EVALUATING MACHINE LEARNING CLASSIFIERS FOR SENTIMENT ANALYSIS ON SOCIAL MEDIA DATA

Ruby Gupta

Phd Scholar, Al-Falah University, Faridabad

Dr Saoud Sarwar

Head - Deptt of Comp. Sc. Engg. , Al-Falah University, Faridabad

Abstract. Social media platforms have become integral parts of our daily lives, creating an overwhelming volume of user-generated content that reflects human emotions, thoughts, and experiences. Sentiment analysis, a powerful application of natural language processing and machine learning, has emerged as a revolutionary technology to analyze and interpret the emotional tone of this vast social media data. This noise was eliminated during preliminary processing of the data using noise elimination software. Researchers used a wide variety of methods in their studies. Multi-layer Perceptron (MLP) and Convolutional Neural Networks (CNN) were used to sort the data into categories. Based on the sentiment classification from Twitter data and consumer affairs website, these two classifications were compared to the other classed such as Support Vector Machine (SVM), Random Forest, Decision tree, Nave Bayes, etc. Results from the proposed study show that, compared to other Machine Learning Classifiers, Multi-layer Perceptron and Convolutional Neural Networks perform the best.

Keywords: Classifiers, Machine Learning, Twitter, Sentiment, Social media

1. Introduction

In an age where social media has become an integral part of our daily lives, the amount of data generated by online interactions has reached unprecedented levels. Every second, millions of users across the globe share their thoughts, feelings, and experiences on platforms like Twitter, Facebook, Instagram, and countless others. Within this vast digital landscape lays a treasure trove of valuable information waiting to be deciphered - the sentiments of people towards various topics, events, brands, and more. Sentiment analysis, a powerful application of natural language processing and machine learning, has emerged as a cutting-edge technology to unravel the emotional pulse of the digital world.

Sentiment analysis, also known as opinion mining, is the process of determining the emotional tone and polarity of a piece of text, be it positive, negative, or neutral. Its application on social media data has revolutionized how we understand public opinion, consumer behavior, and even political sentiments. In this era of big data, sentiment analysis on social media has become an

indispensable tool for businesses, marketers, researchers, and policymakers to gain actionable insights into their target audience's preferences, perceptions, and attitudes.

To comprehend the significance and impact of sentiment analysis on social media data, it is essential to delve into the underlying technology driving this fascinating field. Natural language processing (NLP), a subfield of artificial intelligence (AI), plays a pivotal role in extracting meaning from textual data. NLP algorithms are designed to handle the complexity and nuances of human language, enabling machines to interpret and comprehend written text much like humans do. From basic tokenization and part-of-speech tagging to advanced techniques like deep learning and recurrent neural networks, NLP algorithms have evolved significantly, making sentiment analysis on social media data increasingly accurate and efficient.

Social media, with its diverse and dynamic nature, presents unique challenges and opportunities for sentiment analysis. Unlike traditional surveys or feedback forms, social media data is unstructured, informal, and often riddled with abbreviations, slang, and emoticons. Moreover, the brevity of posts on platforms like Twitter, with its character limitation, demands innovative approaches to capture sentiment in a concise manner. Researchers and developers have risen to the challenge, developing robust sentiment analysis models that can navigate through these complexities and deliver meaningful results.

The implications of sentiment analysis on social media data stretch far beyond marketing and business applications. Social media has become a hotbed for political discourse, activism, and misinformation. Understanding the sentiments of the online population is crucial in monitoring public opinion and predicting social movements. Sentiment analysis has been employed to analyze political sentiments during elections, track the spread of misinformation and hate speech, and identify trends that may have societal repercussions.

However, despite its enormous potential, sentiment analysis on social media data is not without challenges and limitations. The rapidly evolving nature of language, cultural differences, and sarcasm pose hurdles to accurate sentiment classification. Furthermore, the ethical concerns of data privacy and potential biases in the training data require careful consideration to ensure that sentiment analysis is conducted responsibly and fairly.

2. Applications of Sentiment Analysis

Sentiment analysis, a powerful branch of natural language processing (NLP), has a wide range of applications that span various industries and fields. This technology, which involves the automated identification and classification of sentiments in text data, has transformed the way businesses, researchers, and organizations extract insights from textual information. Some key applications of sentiment analysis include:

- **Social Media Monitoring and Brand Management:** Sentiment analysis is extensively used to monitor social media platforms, tracking mentions, comments, and discussions about brands, products, or services. This enables companies to assess public sentiment, manage their online reputation, and engage with customers effectively.
- **Customer Feedback and Reviews:** Businesses leverage sentiment analysis to analyze customer reviews, feedback, and survey responses. By categorizing sentiments as positive, negative, or neutral, companies can gain insights into customer satisfaction, identify areas for improvement, and make informed decisions to enhance their offerings.
- **Market Research and Consumer Insights:** Sentiment analysis aids market researchers in understanding consumer preferences, opinions, and trends. Analyzing online conversations and reviews provides valuable insights into market sentiment, enabling businesses to tailor their products, services, and marketing strategies accordingly.
- **Financial Analysis and Trading:** In the financial sector, sentiment analysis is employed to analyze news articles, social media posts, and financial reports. This data helps investors and traders gauge market sentiment, predict stock price movements, and make informed decisions in trading and investment.
- **Political Analysis and Public Opinion:** Sentiment analysis is used in political campaigns and public opinion research to assess sentiments towards political candidates, parties, and policy issues. By analyzing social media discussions and news coverage, analysts can gauge public sentiment and tailor political strategies.
- **Healthcare and Patient Experience:** Healthcare providers utilize sentiment analysis to analyze patient feedback, reviews, and surveys. This helps them understand patient experiences, identify areas for improvement in healthcare services, and enhance patient satisfaction.
- **Academic Research and Social Sciences:** Researchers apply sentiment analysis to study emotions, attitudes, and opinions expressed in textual data. This has applications in fields like psychology, sociology, linguistics, and literature, providing insights into human behavior and cultural trends.
- **Customer Support and Service:** Sentiment analysis is integrated into customer support systems to categorize and prioritize customer inquiries based on sentiment. This ensures that urgent or negative issues are addressed promptly, leading to improved customer satisfaction.
- **Product Development and Innovation:** Businesses use sentiment analysis to gain insights into customer needs, preferences, and pain points. This information guides product

development efforts, helping companies create offerings that align with customer sentiments and demands.

- **Tourism and Hospitality:** Sentiment analysis is employed in the tourism and hospitality industry to analyze online reviews and feedback. Hotels, restaurants, and attractions can understand guest experiences, identify strengths and weaknesses, and enhance overall customer satisfaction.
- **Content Creation and Marketing Campaigns:** Sentiment analysis informs content creation and marketing strategies by identifying emotional tones that resonate with the target audience. This ensures that content is engaging and aligns with consumer sentiments.
- **Crisis Management and Public Relations:** Organizations use sentiment analysis during crises to monitor public sentiment and assess the effectiveness of communication strategies. By understanding public reactions, companies can respond promptly and effectively to manage crises.
- **Voice of the Employee Analysis:** Sentiment analysis can be applied internally to analyze employee feedback, understand workplace sentiment, and identify opportunities for improving employee satisfaction and engagement.
- **Legal and Compliance:** Sentiment analysis is used in legal contexts to monitor public sentiment and assess public reactions to legal cases, regulations, and policy changes.
- **Educational Insights:** Sentiment analysis can provide educators with insights into student sentiments and attitudes, helping tailor teaching approaches and identifying areas where additional support is needed.

3. Review of Literature

Alsabah, Ruaa (2020) The scale and function of social media sites are expanding rapidly. Sentiment analysis is a component of social media platforms that allows users to get understanding of a user's mental state based on their online writings. Sentiment analysis is attracting a lot of attention since it's a potentially lucrative field with many tailored service applications. Because of Twitter's widespread use, people from all over the world and speaking all kinds of various languages may interact with one another there. This means it has the potential to help in decision-making by providing context for a wide variety of data. In this study, we focus on Arabic "Tweets" and comments sent on Twitter during the 2018 World Cup and investigate the sentiment orientation of the platform's textual features and emoji-based components. The importance of evaluating messages with or without emojis is also measured in this study. Extracted from hundreds of tweets, the data is used to discover the sentiment analysis findings for words and

emojis independently. The results suggest that emojis help to reinforce the texts' emotional orientation, and that both the texts and the emojis are necessary to convey the whole meaning.

Iglesias, Carlos & Moreno, Antonio (2019) The use of sentiment analysis to glean intelligence from social media has rapidly expanded in recent years. The area has matured to the point that it may be used in many other sectors, including marketing, healthcare, banking, and even politics. Recent technical developments, like as deep learning methods, have helped overcome some of the problems that have plagued the field in the past due to a lack of lexical resources. Several methods that further this field are discussed in this Special Issue. There are two main types of papers that have been submitted: contributions to technology and applications.

Drus, Zulfadzli & Khalid, Haliyana (2019) This study reports on an overview of sentiment analysis in social media, including its history, current state of development, and potential future uses. Raw data in the form of text, videos, photographs, and audio files shared by users abounds on social media. Sentiment analysis may be used to extract useful insights from the data. The databases ACM, Emerald Insight, IEEE Xplore, Science Direct, and Scopus were used to conduct a comprehensive literature assessment of papers published between 2014 and 2019. Out of the 77 articles that were subjected to the initial and thorough paper screening, 24 were ultimately accepted for review. The publications were evaluated in light of the study's objectives. The findings demonstrate that the majority of the publications utilized the opinion-lexicon approach to analyzing text sentiment in social media, with the majority of the data being gathered from Twitter, the most popular microblogging site.

Abdullah, Nur Atiqah Sia et al., (2017) Methods for analyzing public opinion and social media data are discussed in this article. Classifier, social media platform, algorithm, data review, and polarity categories are all used in the comparison. In this research, we apply a systematic literature review to delve into the methods of sentiment analysis employed in the categorization of social media data. Some examples of these methods are lexicon-based methods, hybridization techniques, and supervised machine learning. Data from online sources including Twitter, Email, Youtube, and Websites are included in the examined literature. We examine and contrast all of the methods by classifier, social media platform, data review, and polarity categories. Most studies identify social media data using a mixed strategy, according to the comparison. Hybrid methods employ algorithms that draw from both unsupervised and lexicon-based machine learning. These studies rely heavily on Twitter data for their social media analysis. Classifying political tweets uses a lexicon of dictionary-based terms and a support vector machine. Malay social media posts have also been the subject of literary works. Blog and Facebook data have been used in prior studies. Then, lexicon-based and support vector machine-based methods are used for the analysis of sentiment. Only the states of being happy, sad, or emotionless make up the polarity classes. The optimum hybridization strategy for classifying the mood of Malay political tweets is a combination of lexical dictionary-based methods and support vector machine techniques, it has been concluded.

Jayasanka, Sachira et al., (2013) Over the last decade, scholars have paid a great deal of attention to sentiment analysis, or the automatic extraction of expressions of positive or negative opinions from text. As new technologies appear, so does the number of people using the internet; these people frequently share their ideas through online review sites, social networks, and personal blogs. They have both good and negative opinions on a wide variety of individuals, groups, institutions, locations, and concepts. Natural language processing, machine learning, and other methods for analyzing massive amounts of text make it feasible to begin the process of collecting sentiments from social media. In this paper, we examine the problems that can arise during the sentiment extraction process, as well as the methods that have been developed to solve them. Our own method analyzes Twitter social media sentiments, providing insights that can be applied to product profiling, trend analysis, and forecasting. Positive findings indicate that the method may be refined to meet the requirements of the corporate world by analyzing public opinion on social media.

Neri, Federico et al., (2012) The Internet provides a massive forum for people to voice and discuss their views, which has far-reaching ramifications for marketing and communication. Consumers' tastes are being molded by the attitudes and actions they observe on social media. By keeping tabs on what people are saying about your brand or product on social media, you may get a fair idea of how devoted your consumers are. The social media platforms are the obvious next step for advertising. Facebook is now the leader in digital advertising, with Twitter coming in a close second. In this work, we present the findings of a Sentiment Analysis research that compared the reactions of viewers of the Italian public broadcaster Rai to those of the newer, more innovative private firm La7. This research draws parallels between the Osservatorio di Pavia's analysis of political communication in the mass media and the findings of the study. The Osservatorio di Pavia is an Italian research center specializing in media analysis at both the theoretical and empirical levels. Correlating the examination of Social Media, and Facebook in particular, with quantifiable data, available in the public domain, this research also considers Auditel's data on broadcast audience.

4. Research Methodology

4.1 Data Extraction

The datasets are presented as comma-separated values; one contains tweets extracted via the Twitter API in real time; the other is a Google Sheets template called TAGS that automates the retrieval of relevant tweets from Twitter; and the third contains tweets from globally famous persons. The second piece of data is composed of reviews of Ola rides that were scraped in real time from a consumer affairs website with the delicious soup python module.

4.2 Preprocessing

Unfiltered Evaluation data that has been "scratched" from the internet is typically quite loud. This is because social media use is so informal. The dataset underwent extensive preparation in order to standardize it and reduce its overall size.

4.3 Feature Representation

One of the biggest obstacles in sentiment analysis is the ambiguity of words. A term can have a favorable or negative connotation depending on the context. Sentiment polarity can't be determined without first extracting features.

Multi-layer Perceptron (MLP) and Convolutional Neural Networks (CNN) were used to categorize the data. Based on the sentiment classification from Twitter data and consumer affairs website, the results from these two classifications were compared to those from other classifications, such as Support Vector Machine (SVM), Random Forest, Decision tree, Nave Bayes, etc.

5. Data Analysis and Interpretation

The goal of any sentiment analysis is to ascertain the general public's feeling toward a certain topic or piece of writing. In sentiment analysis, the simplest task is to label a sentence or text as positive, negative, or neutral, as shown in Table 1.

Table 1: Sentiment and emotion Analysis

Datasets	Emotions	Percentage
Ola Ride Review	Negative	40.0%
	Positive	33.0%
	Neutral	27.0%
Famous Personality	Negative	30.0%
	Positive	42.0%
	Neutral	28.0%

A number of different classifiers were used in the tests. Models were validated using 10% of the training datasets to check for over fitting, unless otherwise stated. Table 2 summarizes the results of using a sparse vector representation of multiple reviews and tweets using Naive Bayes, Maximum Entropy, Decision Tree, Random Forest, SVM, and Multi-layer Perceptron (MLP). Better results were achieved by implementing Convolutional Neural Networks (CNN) trained on the dense vector representation of each review.

Table 2: Comparison of various classifiers

Algorithms	Ola Ride Review	Famous Person
Naïve Bayes	83.77%	74.20%
MaxEnt	87.38%	80.05%
Logistic Regression	93.43%	91.38%
Decision Tree	81.50%	79.43%
Random Forest	84.42%	78.22%
SVM	83.73%	84.35%
MLP	96.69%	95.89%
CNN	96.86%	97.05%

As can be seen graphically in Figure 1, Neural Networks techniques like Multi-layer Perceptron (MLP) and Convolutional Neural Network (CNN), as described above, provide superior accuracy in relation to the provided social media data.

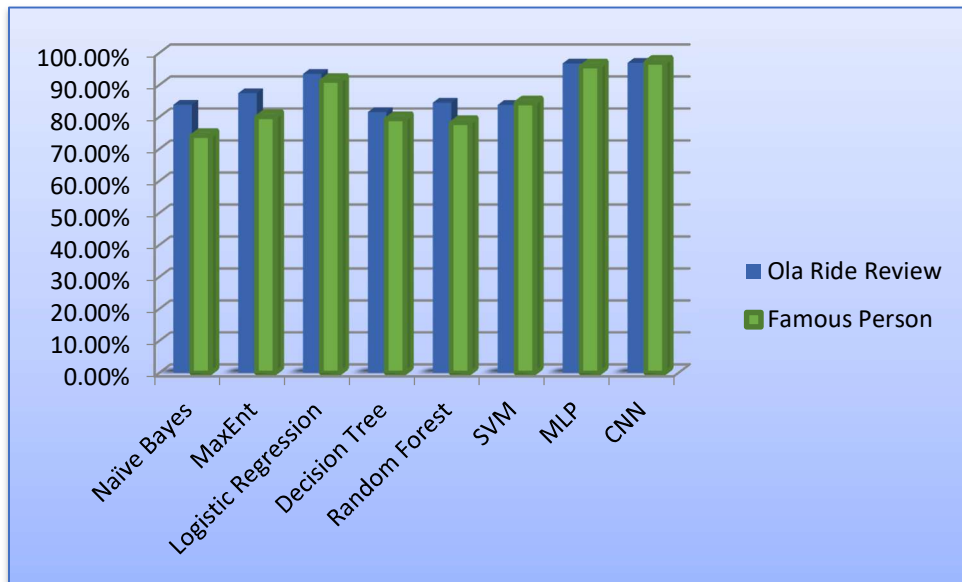


Figure 1: Performance Evaluation

6. Conclusion

Sentiment analysis on social media data has become an indispensable tool for decoding the emotional pulse of the digital world. As technology advances and our understanding of human emotions deepens, sentiment analysis will continue to play a pivotal role in shaping the way we interact with the digital landscape and each other. Harnessing the potential of sentiment analysis responsibly, while addressing challenges and ethical concerns, will pave the way for a more informed, empathetic, and understanding society in the age of social media. By utilizing sentiment analysis to its full potential, we can unlock valuable insights that will benefit businesses, governments, researchers, and ultimately society as a whole. In general, Neural Networks approaches, such as Multi-layer Perceptron (MLP) and Convolutional Neural Network (CNN), outperformed other classifiers in the experiments conducted with Twitter data and Ola ride data from the consumer affair website.

REFERENCES: -

- Alsabah, Ruaa. (2020). Sentiment analysis of comments in social media. *International Journal of Electrical and Computer Engineering*. 10. 10.11591/ijece.v10i6.pp5917-5922.
- Sharma, Dipti & Sabharwal, Munish & Goyal, Vinay & Vij, Mohit. (2020). Sentiment Analysis Techniques for Social Media Data: A Review. 10.1007/978-981-15-0029-9_7.
- Samuels, Antony & Mcgonicle, John. (2020). Sentiment Analysis on Social Media Content.
- Singh, Nikhil & Tomar, Deepak & Kumar, Arun. (2020). Sentiment analysis: a review and comparative analysis over social media. *Journal of Ambient Intelligence and Humanized Computing*. 11. 10.1007/s12652-018-0862-8.
- Iglesias, Carlos & Moreno, Antonio. (2019). Sentiment Analysis for Social Media. *Applied Sciences*. 9. 5037. 10.3390/app9235037.
- Drus, Zulfadzli & Khalid, Haliyana. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*. 161. 707-714. 10.1016/j.procs.2019.11.174.
- kumar, vinay. (2019). Sentiment Analysis Techniques for Social Media Data: A Review.
- Aqlan, Ameen & Bairam, Dr. Manjula & Naik, R Lakshman. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. 10.1007/978-981-13-6459-4_16.
- Chakraborty, Koyel & Bhattacharyya, Siddhartha & Bag, Rajib. (2019). A Survey of Sentiment Analysis from Social Media Data. *IEEE Transactions on Computational Social Systems*. PP. 10.1109/TCSS.2019.2956957.

- Abdullah, Nur Atiqah Sia & Shaari, Nurul & Rahman, A.R.A.. (2017). Review on sentiment analysis approaches for social media data. *Journal of Engineering and Applied Sciences*. 12. 462-467. 10.3923/jeasci.2017.462.467.
- Abdelrazeq, Anas & Janssen, Daniela & Tummel, Christian & Jeschke, Sabina & Richert, Anja. (2015). *Sentiment Analysis of Social Media for Evaluating Universities*.
- Wang, Yilin & Li, Baoxin. (2015). *Sentiment Analysis for Social Media Images*. 1584-1591. 10.1109/ICDMW.2015.142.
- Felciah, Lovelin & Anbuselvi, R.. (2015). *A study on sentiment analysis of social media reviews*. 1-3. 10.1109/ICIIECS.2015.7192949.
- Jayasanka, Sachira & Madhushani, Thilina & Marcus, E. & Aberathne, I. & Premaratne, Saminda. (2013). *Sentiment Analysis for Social Media*.
- Neri, Federico & Aliprandi, Carlo & Capeci, Federico & Cuadros, Montse & By, Tomas. (2012). *Sentiment Analysis on Social Media*. 10.1109/ASONAM.2012.164.