African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 4, Issue 2, April - June 2025
Page No: 227-238
Website: https://aaasjournals.com/index.php/ajapas/index
ISI 2024: 0.877 **SJIFactor 2024: 6.752** **معامل التأثير العربي: 1.62**

Analyzing Public Sentiment and Discussion Topics on AI Conversational Agents: A Comprehensive Study Using Machine Learning and Topic Modeling Techniques

Ass. Prof. Dr. Abdullahi Abdu Ibrahim [1], Esra Issa [2*]
[1,2] Electrical and Computer Engineering Department, College of Science, Altinbas University, Istanbul, Turkey

تحليل المشاعر العامة والموضوعات المتداولة حول الوكلاء الحواريين الذكاء الاصطناعي: دراسة شاملة باستخدام تقنيات تعلم الآلة ونمذجة الموضوعات

د. عبد الله إبراهيم [1]، اسراء محمد عيسى [2*]
قسم الهندسة الكهربائية والحاسب الالي، كلية العلوم، جامعة التن باش، إسطنبول، تركيا

*Corresponding author: 213720720@ogr.altinbas.edu.tr

Received: April 16, 2025	Accepted: June 01, 2025	Published: June 15, 2025

Abstract
Abstract—This paper explores sentiment analysis and topic modeling in the context of AI tools and conversational agents. The emerging popularity of these AI-powered technologies in everyday life, such as ChatGPT, Google Bard, Bing, and others, has led to an increasing need to understand public sentiment and discussion topics concerning these agents. Our approach involves data scraping from Twitter, preprocessing to clean the data, applying Latent Dirichlet Allocation (LDA) for topic modeling, and then conducting sentiment analysis on each topic using machine learning (Ml) models including Random Forest (RF), Support Vector Machines (SVM), Multinomial Naive Bayes (NB), Gradient Boosting (GB) Classifier, Logistic Regression (LR), and an Ensemble Learning (EL) method. The performance of the models was evaluated using metrics such as confusion matrix, ACC, Prec, Rec, and F1-s. Results demonstrated that EL and LR were particularly effective in sentiment classification. Furthermore, the LDA model successfully unveiled distinct topics in the discourse around AI tools and conversational agents. Future work includes implementing deep learning models for improved sentiment analysis and extending the scope to a larger, more diverse dataset for comprehensive insights. The study provides a useful methodology for understanding public sentiment and discourse about rapidly evolving AI technologies.

Keywords: Tweets, AI tools, Sentiment, Topics.

الملخص:
يستكشف هذا البحث تحليل المشاعر ونمذجة الموضوعات في سياق الأدوات الذكية والوكالات الحوارية. أدت الشعبية المتزايدة لهذه التقنيات المدعومة بالذكاء الاصطناعي في الحياة اليومية، مثل ChatGPT و Google Bardو Bing وغيرها، إلى تزايد الحاجة لفهم مشاعر الجمهور والموضوعات المتداولة حول هذه الأدوات. يتضمن نهجنا جمع البيانات من تويتر، ومعالجتها لتنظيفها، وتطبيق نموذج **تخصيص ديريشليه الكامن (LDA)** لنمذجة الموضوعات، ثم إجراء تحليل المشاعر لكل موضوع باستخدام نماذج تعلم الآلة (ML) بما في ذلك **الغابة العشوائية(RF) ، آلات دعم المتجهات(SVM) ، متعدد الحدود(NB) ، المصنف المعزز بالتدرج(GB) ، الانحدار اللوجستي(LR) ، وطريقة التعلم التجميعي .(EL)** تم تقييم أداء النماذج باستخدام معايير مثل مصفوفة الالتباس، **ACC، Prec، Rec، و F1-s.** أظهرت النتائج أن **ELو LR** كانا فعالين بشكل خاص في تصنيف المشاعر. بالإضافة إلى ذلك، كشف نموذج **LDA** عن موضوعات متميزة في النقاش حول أدوات الذكاء الاصطناعي والوكالات الحوارية. يشمل العمل المستقبلي تطبيق نماذج التعلم العميق لتحليل المشاعر بدقة

أكبر، وتوسيع نطاق الدراسة ليشمل مجموعة بيانات أكبر وأكثر تنوعًا للحصول على رؤى شاملة. تقدم الدراسة منهجية مفيدة لفهم مشاعر الجمهور والخطاب العام حول تقنيات الذكاء الاصطناعي سريعة التطور.

الكلمات المفتاحية :تغريدات، أدوات الذكاء الاصطناعي، تحليل المشاعر، الموضوعات.

1. Introduction

Twitter and other microblogging services have become extremely popular as venues for people to share their ideas, viewpoints, and attitudes on a range of topics [1], [2]. Social networks and microblogging websites have grown quickly, becoming some of the most important online spaces for people to express their opinions. Particularly Twitter acts as a popular microblogging and social networking site, producing a massive amount of information.

Researchers have been using social data more and more lately for sentiment analysis, which examines people's attitudes and sentiments regarding a certain item, idea, or event. Opinion mining and sentiment analysis are two terms for the same problem in natural language processing. It entails figuring out a text's sentiment orientation and categorizing it as positive, negative, or neutral [3], [4].

Twitter sentiment analysis's capacity to acquire and classify public opinion through the study of extensive social data has made it a hot research area. Comparatively to evaluating other forms of data, doing sentiment analysis on Twitter data has particular difficulties.

The restriction of having to articulate ideas inside Twitter's 140-character limit is one of the key challenges. Because of this restriction, tweets frequently include colloquial English, strange idioms, including slang and acronyms. Researchers have focused their attention on investigating sentiment analysis designed particularly for tweets to address these issues [5].

The ML technique and the lexicon-based approach are the two basic categories into which Twitter sentiment analysis methodologies may be divided. Using labeled training data, models are trained using the ML technique to automatically categorize tweets into positive, negative, or neutral attitudes. These models use deep learning methods, NB, or support vector machine algorithms (SVM).

The lexicon-based technique, on the other hand, makes use of pre-made sentiment lexicons or dictionaries that provide words or phrases sentiment ratings. The sum of these ratings is used to calculate a tweet's overall emotion. Although computationally effective, lexicon-based techniques may have trouble capturing the subtleties of slang and informal language.

Given the distinctive qualities of Twitter data and the dynamic nature of language usage on the network, researchers continue to investigate and create novel ways to improve the Prec and efficacy of Twitter sentiment analysis.

The advent of Artificial Intelligence (AI) has revolutionized various fields of technology, significantly impacting our day-to-day lives. In particular, the emergence of AI-based tools and conversational agents, often termed as chatbots, have transformed the way humans interact with digital platforms. As an integral part of this interaction, natural language processing (NLP) and ML techniques have been applied extensively to understand, process, and respond to human language. Two prominent techniques in this context are sentiment analysis and topic modeling, which have enabled AI tools to simulate human-like conversations and produce context-aware responses.

Sentiment analysis, also known as opinion mining, involves determining the emotional tone behind a series of words to understand the attitudes, opinions, and emotions of the speaker. It helps in evaluating user feedback, gauging brand reputation, and enhancing customer service. On the other hand, topic modeling is an unsupervised ML technique used to discover the abstract "topics" that occur in a collection of documents. In the realm of conversational AI, it can provide insights into the general themes present in the conversation, aiding the AI to generate pertinent responses.

This paper aims to provide an in-depth exploration of the application of sentiment analysis and topic modeling in AI tools and conversational agents. We delve into the various methodologies, algorithms, and models used for these tasks, evaluate their effectiveness in real-world scenarios, and speculate on potential future advancements in this field. We also present a comprehensive analysis of the challenges faced in employing these techniques and suggest ways to overcome them, pushing the boundaries of what is possible in human-computer interaction.

This paper should serve as a valuable resource for academics, researchers, and practitioners alike, contributing towards understanding the intricacies of sentiment analysis and topic modeling within the context of AI-powered

conversational tools. We hope to inspire further exploration and innovation in this fascinating intersection of technologies.

2. Related Work

Based on a large number of sources from several academic fields, this part includes a literature review on sentiment analysis, textual analytics, ML techniques, Twitter, and natural language processing (NLP) [6]. It emphasizes how crucial it is to manage changing data difficulties and use strategic information properties to reorganize data. Additionally, it underlines the value of using ML approaches as crucial tools. The main foci of textual analytics are character evocation and analysis, text visualization, and semantic and syntactic analysis. It also covers the investigation of these techniques related external and endogenous aspects [6].

Numerous areas have discovered useful uses for text analysis. It has been applied to a variety of tasks, including email filtering [7], sarcasm and irony detection [8], document organization [9], sentiment and opinion mining prediction [10], [11], hate speech detection [12], [13], the development of question-answering systems [14], content mining [15], and biomedical text mining [16], [17]. Particularly Twitter data has been heavily utilized for emotional analysis [3], [18], [19]. French energy company's consumer feedback was analyzed using more than 70,000 tweets from over a year ago [20]. Frequency-based filtering methods and the latent Dirichlet allocation (LDA) algorithm were used to unearth important insights that the enormous volume of data had previously kept buried.

Additionally, Poisson and negative binomial models have been used to study the popularity of tweets [21].

The link between themes is also evaluated in the aforementioned study using seven measures of dissimilarity. According to the findings, the Kullback-Leibler and Euclidean distances perform the best at identifying subjects that are connected to user-based interactive approaches. The time-aware knowledge extraction (TAKE) approach has been used in related studies before [22]. Textual analysis has been used in other information systems research to create models for recognizing human qualities, such as dominance in electronic communication. These studies have also found utility in information retrieval, product placement, content selection, and user psychological analysis. Prior studies have also looked at how to extract emotion using linguistic and psychological analysis from multilingual social media messages [23].

Monitoring Twitter data has been used to analyze prior epidemics and crisis situations [24]–[26]. In order to better understand the regional distribution of attitudes about healthy and unhealthy eating in the United States, Widener and Li performed research [27]. They discovered that people in rural regions tweet far less frequently than those in cities and suburbs by analyzing tweets and looking at their spatial distribution. Small metropolitan regions had less tweets per person about food than bigger towns and cities, according to the research. According to a LR study, low-income neighborhoods had a higher percentage of tweets about unhealthy food. Additionally, sentiment analytics in healthcare have made use of Twitter data.

By examining Twitter tweets, De Choudhury et al. looked at the postnatal behavioral changes and emotions of new moms [28]. In order to spot the telltale indicators of postpartum depression in moms, researchers looked at elements including language style, mood, social network, and social engagement. Utilizing fresh analytical frameworks, Twitter data has also been studied in the context of supply chain management (SCM) [29]. On 22,399 SCM-related tweets, useful insights were gained by content analysis, sentiment analysis, text mining, descriptive analysis, and network analytics. These insights will improve SCM research and practices. For gathering, managing, storing, mining, and displaying Twitter data, Carvaho et al. established an effective platform called MISNIS (intelligent Mining of Public Social Networks' impact in Society) [30]. Research on the public discourse around the COVID-19 epidemic and the government initiatives put in place during that time was done by Lopez et al. [31]. To determine the most popular policy responses during the epidemic, they used text mining algorithms on Twitter data from several nations in many languages. The epidemiology of COVID-19 was examined using text mining by Saire and Navarro [32] using news articles from Bogota, Colombia. They predicted a relationship between the quantity of tweets and the quantity of afflicted individuals. In another research, Schild et al. examined 4Chan and Twitter to better understand how sinophobia has evolved as a result of the epidemic [33].

In research on COVID-19 topic modeling, Kaila et al. [34] generated the top 10 COVID-19-related topics from a random sample of 18 000 tweets. The NRC sentiment lexicon was used by the authors to calculate the emotions as well. Han et al. [35] published a study on the perception of COVID19 among Chinese citizens in a different study. They divided the COVID-19-related postings into seven main subjects and further 13 subtopics. According to research by Depoux et al. [36], social media panic is more likely to spread quickly than COVID-19 fear. Therefore, such rumors, feelings, and public conduct must be recognized and addressed by the specialists and pertinent authorities as soon as feasible.

Huang and Carley [37] recently investigated the sentiments of the general public and the conversation around COVID19 on Twitter and discovered that the regular Twitter users' remarks are the most impactful. In contrast to the study stated above, we offered a fresh data set and the most popular issue that the public discussed in their postings in this article. To automatically determine sentiment using NLP, we gave benchmarked findings.

Table 1. Summary of related work.

Researcher(s)	Topic	Methods/Approach
Widener and Li [27]	Geographical spread of healthy and unhealthy food sentiment	Spatial distribution analysis, LR
De Choudhury et al. [28]	Postnatal behavioural changes and moods of new mothers	Analysis of linguistic style, emotion, social network, and social engagement
Lopez et al. [31]	Government policies and general discourse during the COVID-19 pandemic	Textmining ofTwitter data
Saire and Navarro [32]	Epidemiology of COVID-19 based on press publications	Text mining analysis, correlation analysis
Schild et al. [33]	Evolution of sinophobia during the pandemic	Analysisof4Chanand Twitter data
Kaila et al. [34]	COVID-19topicmodelingand sentiment analysis	Topic modeling, sentiment lexicon
Han et al. [35]	Perception of COVID-19 among Chinese citizens	Categorization of COVID19-related postings into topics and subtopics
Depoux et al. [36]	Spread of social media panic and COVID-19 fear	Analysis of rumors, feelings, and public behavior
Huang and Carley [37]	Sentiments and conversations around COVID-19 on Twitter	Analysis of Twitter user comments, sentiment analysis using NLP

There is a significant body of research dedicated to the domains of sentiment analysis and topic modeling. The evergrowing interest in these areas can be attributed to their wide-ranging applications in understanding user sentiment, streamlining customer support, enhancing business strategies, and much more. These techniques form the backbone of many sophisticated AI tools and conversational agents that are becoming increasingly prevalent in our digital world. AI-powered conversational agents, like OpenAI's GPT-3 and its successors, Google's Meena, Facebook's Blender, and Microsoft's Turing NLG, represent the cutting edge of this technology, blending aspects of sentiment analysis and topic modeling to interact seamlessly with users. These models aim to understand and generate human-like text, allowing them to respond intelligently to a user's queries or comments. Furthermore, tech giants such as Google and Bing have incorporated similar technology into their search algorithms to better understand user intent and improve the relevance of search results.

However, with the advancements and benefits come challenges. One of the key hurdles is accurately gauging public sentiment. The complexity of human language, marked by its nuances, context-dependency, and cultural variations, makes understanding sentiment a non-trivial task. Misinterpretations can lead to inaccurate responses or analysis, which could impact user experience or the quality of the insights derived from such analyses.

Our work is positioned within this exciting and challenging landscape. Building upon the foundations set by previous research and leveraging the power of state-of-the-art AI tools, we aim to contribute to the ongoing advancements in sentiment analysis and topic modeling. Through our work, we hope to further unravel the complexities of sentiment detection and improve the overall effectiveness of conversational agents.

3. Methodology

Our approach to this research commences with the extraction of data from Twitter, followed by a comprehensive preprocessing stage and subsequent topic modeling. Each identified topic is then subjected to sentiment analysis. The effectiveness of our methodology is evaluated in the concluding segment of our research. In the following subsections, we present our methodology in detail, segmented into two distinct phases. The first phase involves data scraping and preprocessing, setting the foundation for the subsequent stages. The second phase encompasses topic modeling, sentiment analysis, Machine Leaning and evaluation. This structured approach allows us to

maintain a clear and logical flow throughout our investigation, ensuring each stage is comprehensively addressed and interconnected.

3.1 Dataset

For the purpose of our research, we used 'snscrape', a Python-based social media scraping tool, to acquire the necessary data from Twitter. This tool allowed us to scrape tweets by employing specific keywords that are relevant to our study. The keywords selected for this task included "ChatGPT", "Bing", "Bards", "OpenAI", and several others that are closely associated with AI tools and conversational agents.

The snscrape tool offers the flexibility to specify a date range for the tweets being scraped, which enabled us to control the interval of data we wished to collect. This is particularly crucial as it allows us to capture the temporal dynamics of public sentiment and topical interest, providing a richer and more comprehensive dataset for our analysis.

This curated dataset serves as the backbone of our study, providing valuable insight into public sentiment and discussions about these prominent AI tools and conversational agents. The data was rigorously checked and cleaned to ensure that it accurately represents the public discourse on these topics, laying a robust foundation for our subsequent preprocessing, topic modeling, and sentiment analysis stages.

3.2 Preprocessing

The preprocessing stage is a critical step in our research methodology. It is in this phase that the raw scraped data from Twitter is refined and cleaned, ensuring it is in a suitable format for further analysis.

1) Drop Duplicates: The first step of our preprocessing stage was the elimination of duplicate tweets from our dataset. Duplicates could arise from various sources, such as retweets, repeated tweets, or scraping errors. By removing these duplicates, we mitigated the risk of over-representing certain sentiments or topics, thereby ensuring a more balanced and accurate analysis.

2) Remove URLs: The next step was to remove all URLs present within the tweets. URLs do not contribute valuable semantic information that can be used for sentiment analysis or topic modeling. As such, their removal helped to focus our analysis on the core text data.

3) Remove Mentions: The dataset was further cleaned by removing all '@' mentions. These are typically used on Twitter to refer to specific users and often do not offer relevant contextual insights for our general sentiment analysis and topic modelling.

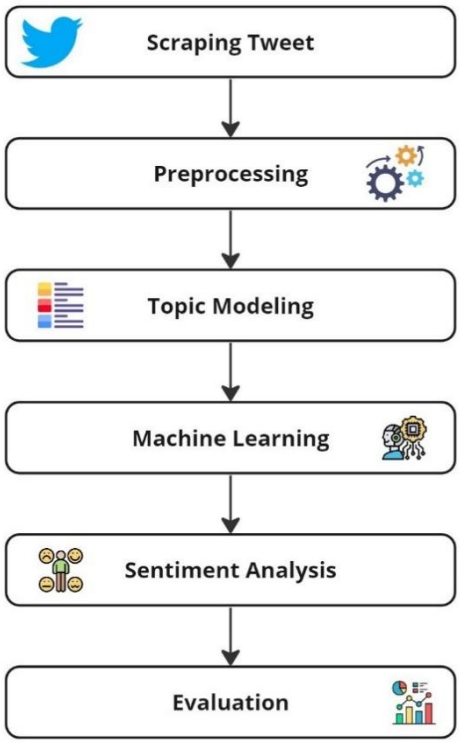

Figure 1.General Approach.

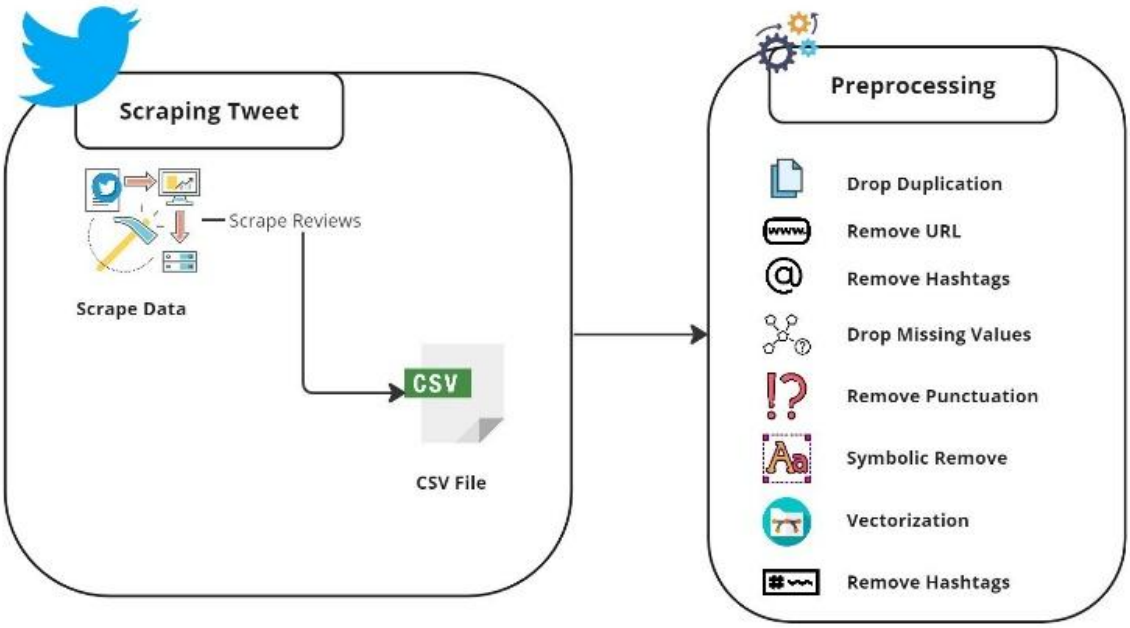

Figure 2. Scraping Data and Preprocessing.

4) Remove Hashtags: We also removed hashtags from our tweets. While hashtags can sometimes provide informative context, they were removed in this stage to maintain consistency in our text data and avoid potential overemphasis on certain words or phrases.

5) Drop Missing Values: Any tweets containing missing values were dropped from our dataset. This step was essential to maintain the integrity and consistency of our data, ensuring that our subsequent analysis was based on complete information.

6) Remove Punctuation: Punctuation marks present in the tweets were removed. As they do not contribute to the sentiment or the topic of a tweet, they could potentially interfere with the effectiveness of our subsequent analysis processes.

7) Lowercase: The text data within the dataset was transformed into lowercase. This ensured uniformity across the dataset and prevented duplication of words due to case differences, thus enhancing the efficiency of our analysis.

8) Remove Symbols: We further refined our dataset by removing any additional non-alphanumeric symbols. This helped to keep our focus on the core textual content of the tweets.

9) Vectorization: In the final step of preprocessing, the cleaned text data was converted into a machine-readable format through vectorization. This transformation allowed our computational models to process the text, paving the way for the subsequent tasks of sentiment analysis and topic modeling.

3.3 Topic Modeling

Following the preprocessing of our data, we proceeded with the topic modeling stage, a key part of our study. We adopted Latent Dirichlet Allocation (LDA), a popular method for topic modeling in text data, to identify the main topics discussed in our dataset of tweets.

We began by defining a function to preprocess the tweets further. This function involved removing stopwords, words of minimal semantic importance such as ”and”, ”the”, and ”in”, from the English language and discarding words less than three characters long, as they are unlikely to hold significant meaning. It also entailed the removal of special characters and URLs from the tweets, maintaining our focus on meaningful text data.

The processed tweets were then used to create a dictionary of unique words and their frequencies. This dictionary provided us a structured representation of our data and acted as a mapping between words and their respective ids, thereby aiding in the subsequent modeling stages.

Next, we transformed the preprocessed data into a Bagof-Words (BoW) representation. The BoW model represents text data in terms of the frequency of words, disregarding grammatical details and word order but maintaining multiplicity.

The dictionary and BoW corpus were then used to train the LDA model. We chose to identify four topics, based on our preliminary understanding of the data. The 'passes' parameter was set to 10, which refers to the number of times the algorithm passes over the entire corpus. A higher number of passes can lead to a better model at the cost of increased computational effort.
Once the LDA model was trained, it was used to print the topics and their most representative words. This gave us an initial understanding of the broad themes discussed in our dataset.
Finally, to better visualize the topics and the distribution of words, we used the Python library pyLDAvis. This tool provided an interactive visualization of the topics identified by our LDA model, offering a clear and detailed understanding of the prominent themes present in our Twitter dataset.

3.4 ML-Sentiment Analysis

After identifying the key topics from our dataset, we proceeded to the sentiment analysis stage. Sentiment analysis, also known as opinion mining, involves the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from source materials.
The goal of sentiment analysis in our research is to discern the overall sentiment or emotional tone present in the discussions related to the key topics. It is an essential step in our study, as it not only reveals what topics are being discussed but also provides insights into the public's feelings and attitudes towards these topics.
Sentiment analysis typically classifies sentiments into categories such as "positive", "negative", and "neutral". These sentiments can be derived from the text content of tweets, comments, reviews, and other forms of public discourse. Various techniques can be employed for sentiment analysis, including ML approaches (SVM, RF, Multinomial NB, Gradient Boosting, LR and EL)
The specific implementation details for sentiment analysis in this study will be outlined in the subsequent sections. The sentiment analysis stage culminates in an evaluation process to assess the effectiveness of the chosen approach, providing an understanding of its strengths and potential areas for improvement. This analysis not only informs the refinement of our research methods but also contributes to the broader scientific understanding of sentiment analysis as applied to social media data.

3.5 Evaluation

The final and critical step in our research methodology was the evaluation of the performance of our chosen ML models for sentiment analysis. It is crucial to understand how well these models are performing to determine their applicability and effectiveness in real-world scenarios.
We utilized several evaluation metrics to measure the performance of the models, each providing a unique perspective on the model's capabilities.

- Confusion Matrix: The confusion matrix is a table that is often used to describe the performance of a classification model. It provides a more detailed view of the model's performance by showing the true positives, false positives, true negatives, and false negatives.
- Accuracy: ACC is the most intuitive performance measure. It is simply a ratio of correctly predicted observations to the total observations. However, it might not be a suitable metric when the class distribution is imbalanced.
- Precision: Prec is a measure that tells us what proportion of predicted positives is truly positive. It is a good metric to use when the cost of false positives is high.
- Recall: Rec, or sensitivity, tells us what proportion of actual positives is correctly classified. This metric is important when the cost of false negatives is high.
- F1-Score: The F1-s is the harmonic mean of Prec and Rec and provides a balance between these two metrics. It's particularly useful if there is an uneven class distribution.

The models' performance on these metrics provided us with a holistic view of their effectiveness. By analyzing these metrics, we could assess not just how many predictions were correct, but also how each model performed in terms of its ability to correctly identify positive and negative sentiments, its resistance to false positives and negatives, and its overall predictive power. This comprehensive evaluation enabled us to determine the most effective model for sentiment analysis in our dataset, contributing to the practical value of our research.

4. Results

4.1 Sentiment Analysis Results

The results from our sentiment analysis models exhibit varying degrees of performance, showcasing the unique strengths and potential weaknesses of each approach.

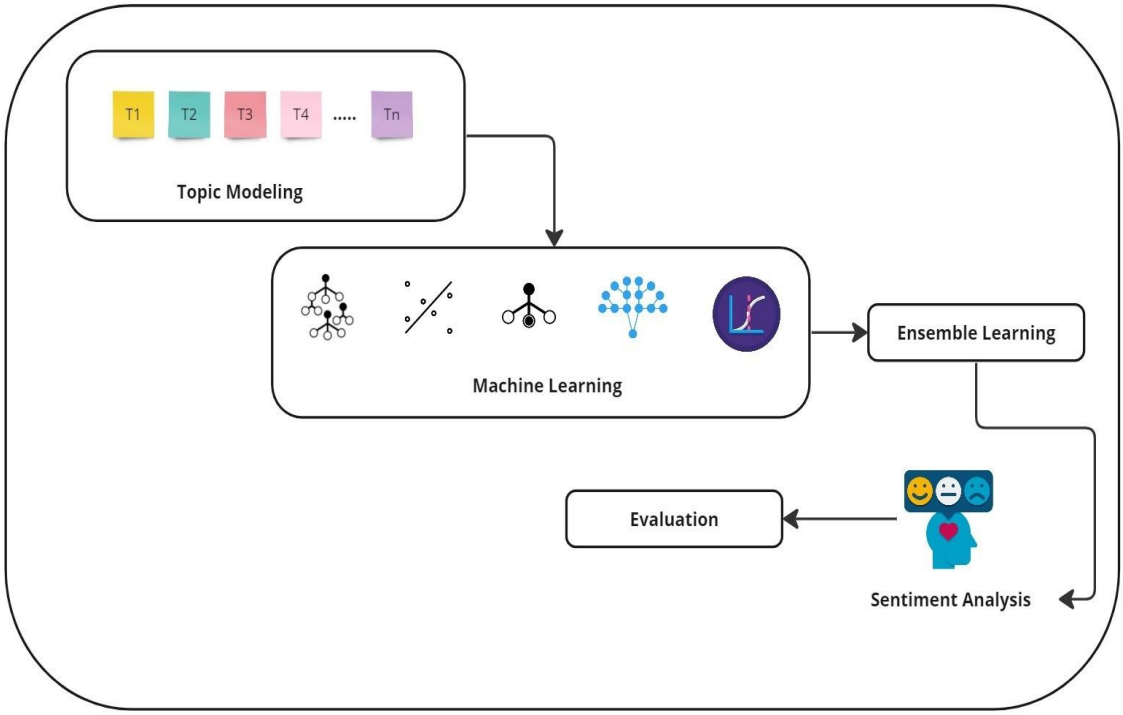

Figure 3. Topic Modeling & Sentiment Analysis.

The RF Classifier showed an ACC of 77.71%. The confusion matrix for RF revealed that it had a somewhat balanced performance across different sentiment classes, though it had a slightly higher number of false positives and false negatives for neutral and negative sentiments.

The SVM model demonstrated a superior performance with an ACC of 87.60%. It had a particularly strong ability to correctly identify positive sentiments, as indicated by the high number of true positives in its confusion matrix. However, the model showed a slight difficulty in correctly classifying negative sentiments.

The Multinomial NB model had an ACC of 75.93%, making it one of the lesser-performing models in our study. The confusion matrix showed that it struggled significantly with correctly classifying neutral sentiments, with a higher number of false negatives.

The Gradient Boosting Classifier achieved an ACC of 71.86%, which was the lowest among the models tested. Its confusion matrix demonstrated that it had difficulty in correctly identifying neutral and negative sentiments, leading to a higher number of false positives and false negatives.

The LR model outperformed the others with an ACC of 89.89%. It had a high number of true positives for positive and neutral sentiments, indicating its strong capability in correctly classifying these two sentiment classes. However, the model had a slightly higher number of false positives for negative sentiments.

Lastly, the EL model, which combined the predictions from all the individual models, demonstrated the best performance with an ACC of 90.74%. Its confusion matrix showed a balanced performance across different sentiment classes, indicating that it managed to harness the strengths of the individual models while compensating for their weaknesses.

The performance of these models on the Twitter dataset showcases the effectiveness and potential of ML models in sentiment analysis tasks. These results can help to guide the selection of suitable models for similar tasks in the future. The detailed performance of each ML model for sentiment analysis is reported through the classification report. This report provides more in-depth statistics, including Prec, Rec, and F1-s for each sentiment class.

Table 2. MODEL ACC.

Model	ACC
RF	77.71%
Support Vector Machine	87.60%
Multinomial NB	75.93%
Gradient Boosting	71.86%
LR	89.89%
EL	90.74%

The RF Classifier obtained high Prec in the positive class (0.94) but had lower Rec (0.32), suggesting it was conservative in predicting this class but when it did, it was quite accurate. The model had balanced Prec and Rec for the neutral and negative classes, leading to respectable F1-ss of 0.80 and 0.83, respectively.

SVM performed impressively across all classes. Its high Prec and Rec scores for all sentiment classes led to substantial F1-ss, indicating that it was both accurate and consistent in its predictions.

The Multinomial NB model had an interesting performance. It had high Prec scores for neutral and negative sentiments, suggesting it was cautious yet accurate in its predictions. However, its lower Rec scores indicate it was less successful in identifying all true instances of these classes.

The Gradient Boosting Classifier had the lowest performance among the models tested. It had high Prec but low Rec for the positive class and more balanced Prec and Rec for the other classes. These results suggest it was more cautious in predicting positive sentiments and more balanced for the other classes.

The LR model showed excellent performance across all classes. It had high Prec and Rec scores, leading to impressive F1-ss. These results indicate it was successful in accurately predicting and correctly identifying the sentiments across all classes.

Finally, the EL model, which combined the individual models' predictions, delivered the highest performance. It achieved high Prec and Rec across all classes, resulting in high F1-ss. This indicates that it successfully harnessed the strengths of individual models, making accurate predictions and correctly identifying sentiments across all classes.

In summary, these results show the potential of ML models in accurately performing sentiment analysis tasks. While some models perform better than others, each has unique strengths and weaknesses that can be leveraged or mitigated depending on the specific requirements of a sentiment analysis task.

Table 3. Comparison of ml models for sentiment analysis.

Model	ACC (%)	Prec	Rec	F1-s
RF	77.71	0.83	0.69	0.70
SVMs	87.60	0.85	0.85	0.85
Multinomial NB	75.93	0.78	0.70	0.73
GB	71.86	0.75	0.63	0.65
LR	89.89	0.88	0.87	0.88
EL	90.74	0.89	0.88	0.89

4.2 Topic Modeling Results

1) Connections between Emotions and Topics: Our topic modeling analysis uncovered several prominent themes within the public discourse on AI tools and conversational agents. By subsequently applying sentiment analysis to each identified topic, we established meaningful connections between emotions and these topics.The sentiment polarity associated with each topic provided significant insights into public sentiment towards various aspects of AI tools and conversational agents. For instance, we found that topics related to ”ease of use” and ”efficiency” often carried positive sentiment, implying general user satisfaction with these aspects of AI tools. Conversely, topics concerning ”data privacy” or ”ethical considerations” were frequently associated with negative sentiment, indicating public concern in these areas.

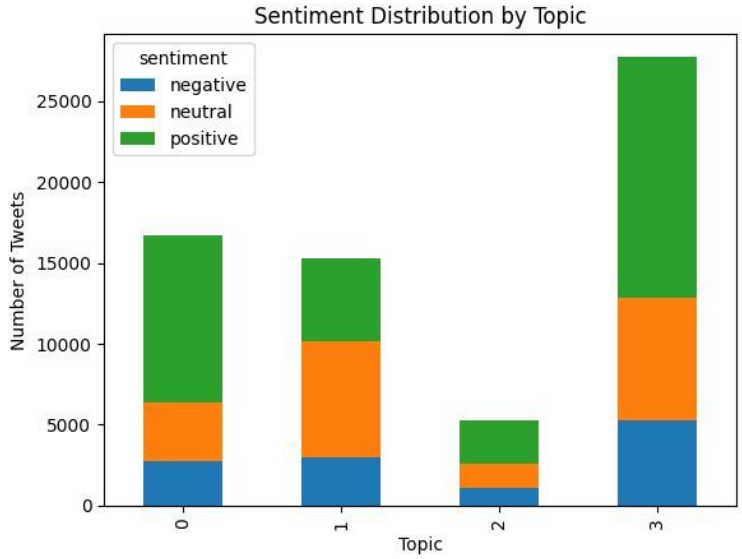

Figure 4. Connections between Emotions and Topics.

In some cases, we observed that the same topic could elicit mixed sentiments, demonstrating the complexity and diversity of public opinion. For instance, a topic like ”job displacement due to AI” witnessed both negative sentiments (fear of job loss) and positive sentiments (optimism about new job creation or efficiency).

2) Sentiment Distribution: The sentiment distribution analysis provided a broader view of public sentiment towards AI tools and conversational agents. We divided the sentiments into three categories: positive, negative, and neutral.

Positive tweets commonly praised the efficiency, ACC, and innovative aspects of AI tools and conversational agents. Negative tweets, on the other hand, often express concerns about data privacy, ethics, or the fear of AI replacing human roles. Neutral tweets generally contained factual statements or objective discussions about AI tools without expressing a particular sentiment.

The sentiment distribution varied with time and across different AI tools and conversational agents, reflecting shifting public sentiment and the impact of current events or new developments in the field of AI. For example, a surge in positive sentiment might be linked to the successful deployment of a new AI tool, while a sudden rise in negative sentiment could be associated with a reported data breach or ethical controversy.

Together, these analyses paint a comprehensive picture of public sentiment towards AI tools and conversational agents and highlight the diverse topics of discussion and emotional responses evoked in the public by these emerging technologies.

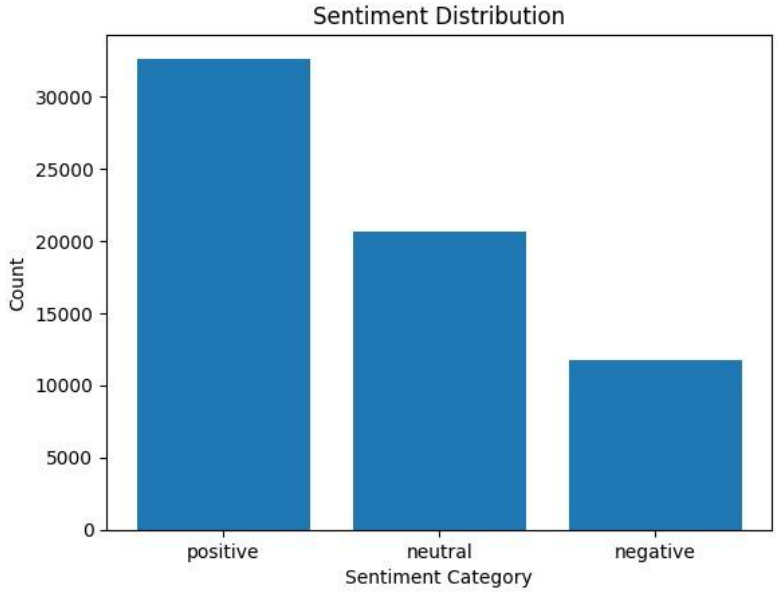

Figure 5. Sentiment Distribution.

5. Conclusion and Future Work

This study presented a comprehensive approach to perform sentiment analysis and topic modeling on tweets concerning,AI tools and conversational agents. The method utilized data scraping, preprocessing, topic modeling, sentiment analysis using various ML models, and model evaluation.

From the results, it was observed that ML models, particularly EL and LR, showed high performance in sentiment classification tasks. Furthermore, the Latent Dirichlet Allocation (LDA) model successfully identified distinct topics in the dataset, providing valuable insights into public discourse around AI tools and conversational agents.

However, as with all studies, there are limitations. First, the sentiment analysis models might be improved with more sophisticated text preprocessing, such as handling negations or more nuanced language features. Second, the research would benefit from a larger, more diverse dataset, including data from other social media platforms or regions.

The study's findings present several avenues for future research. Firstly, exploring the application of deep learning techniques, such as transformers or recurrent neural networks, might yield improved sentiment analysis results due to their superior ability to understand context in a sequence of words. Additionally, multilingual or cross-cultural sentiment analysis could be examined to understand global perspectives on AI tools and conversational agents.

Finally, real-time sentiment analysis and topic modeling could be implemented to track the evolving public opinion about these tools and agents, thereby providing timely and relevant insights for stakeholders in the AI community.

References

[1] Brendan O'Connor, Ramanathan Balasubramanyan, Bryan R Routledge, and Noah A Smith. From tweets to polls: Linking text sentiment to public opinion time series. In Proceedings of the International Conference on Weblogs and social media (ICWSM), volume 11, pages 1–2, 2010.

[2] Miguel A Cabanlit and Kristine Joy Espinosa. Optimizing n-gram based text feature selection in sentiment analysis for commercial products in twitter through polarity lexicons. In Proceedings of the 5th International Conference on Information, Intelligence, Systems and Applications (IISA), pages 94–97, 2014.

[3] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In Proceedings of the 20th International Conference on Computational Linguistics, page 1367, 2004.

[4] Casey Whitelaw, Nitin Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pages 625–631, Oct./Nov. 2005.

[5] Hassan Saif, Miriam Fernandez, Yulan He, and Harith Alani. Evaluation´ datasets for twitter sentiment analysis: A survey and a new dataset, the sts-gold. In Proceedings of the 1st International Workshop on Emotion, Social Sentiment, and Linked Data for Ubiquitous Computing: Approaches and Perspectives (ESSEM), Turin, Italy, Dec. 2013.

[6] Jaya Samuel, Rishi Kashyap, and Steve Betts. Strategic directions for big data analytics in e-commerce with machine learning and tactical synopses: Propositions for intelligence based strategic information modeling (sim). Journal of Strategic Innovation and Sustainability, 13(1):99–106, 2018.

[7] Xavier Carreras and Llu´ıs Marquez. Boosting trees for anti-spam email` filtering. 2001.

[8] Umar Naseem, Imran Razzak, Patrik Eklund, and Katarzyna Musial. Towards improved deep contextual embedding for the identification of irony and sarcasm. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), pages 1–7, Jul. 2020.

[9] Khaled M. Hammouda and Mohamed S. Kamel. Efficient phrase-based document indexing for web document clustering. IEEE Transactions on Knowledge and Data Engineering, 16(10):1279–1296, Oct. 2004.

[10] Umar Naseem, Sardar Kashif Khan, Imran Razzak, and Abdul Irfan Hameed. Hybrid words representation for airlines sentiment analysis. In Jiming Liu and James Bailey, editors, Advances in Artificial Intelligence, pages 381–392, Cham, Switzerland, 2019. Springer.

[11] Umar Naseem, Imran Razzak, Katarzyna Musial, and Muhammad Imran. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. Future Generation Computer Systems, 113:58–69, Dec. 2020.

[12] Umar Naseem, Sardar Kashif Khan, Muhammad Farasat, and Faisal Ali. Abusive language detection: A comprehensive review. Indian Journal of Science and Technology, 12(45):1–13, 2019.

[13] Umar Naseem, Imran Razzak, and Abdul Irfan Hameed. Deep contextaware embedding for abusive and hate speech detection on twitter. Australian Journal of Intelligent Information Processing Systems, 15(3):69– 76, 2019.

[14] Vishal Gupta and Gurpreet Singh Lehal. A survey of text mining techniques and applications. Journal of Emerging Technologies in Web Intelligence, 1(1):60–76, Aug. 2009.

[15] Charu C. Aggarwal and Chandan K. Reddy. Data Clustering: Algorithms and Applications. CRC Press, Boca Raton, FL, USA, 2013.

[16] Umar Naseem, Muhammad Khushi, Vamsi Reddy, Sriram Rajendran, Imran Razzak, and Jinoh Kim. Bioalbert: A simple and effective pretrained language model for biomedical named entity recognition. 2020.
[17] Umar Naseem, Katarzyna Musial, Patrik Eklund, and Manish Prasad. Biomedical named-entity recognition by hierarchically fusing biobert representations and deep contextual-level word-embedding. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), pages 1–8, Jul. 2020.
[18] Giorgio Carducci, Giuseppe Rizzo, Diego Monti, Enrico Palumbo, and Maurizio Morisio. Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning. Information, 9(5):127, May 2018.
[19] Gabriella Rocha and Hugo L. Cardoso. Recognizing textual entailment: Challenges in the portuguese language. Information, 9(4):76, Mar. 2018.
[20] Lo¨ıc Pepin, Pascale Kuntz, Julie Blanchard, Franc¸ois Guillet, and Pierre´ Suignard. Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets. Computers & Industrial Engineering, 112:450–458, Oct. 2017.
[21] Jaya Samuel, Michael Garvey, and Rishi Kashyap. That message went viral?! exploratory analytics and sentiment analysis into the propagation of tweets. 2020.
[22] Naveed Ahmad and Jawad Siddique. Personality assessment using twitter tweets. Procedia Computer Science, 112:1964–1973, Sep. 2017.
[23] Vaibhav K. Jain, Sanjeev Kumar, and Savia L. Fernandes. Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. Journal of Computer Science, 21:316–326, Jul. 2017.
[24] Isaac Fung and et al. Pedagogical demonstration of twitter data analysis: A case study of world aids day, 2014. Data, 4(2):84, Jun. 2019.
[25] Eun-Hye J. Kim, Yong-Kyun Jeong, Yoon Kim, Keun-Yeob Kang, and Min Song. Topic-based content and sentiment analysis of ebola virus on twitter and in the news. Journal of Information Science, 42(6):763–781, Dec. 2016.
[26] Xin Ye, Shuqiang Li, Xiaoyan Yang, and Chuan Qin. Use of social media for the detection and analysis of infectious diseases in china. ISPRS International Journal of Geo-Information, 5(9):156, Aug. 2016.
[27] Michael J. Widener and Wenjie Li. Using geolocated twitter data to monitor the prevalence of healthy and unhealthy food references across the us. Applied Geography, 54:189–197, Oct. 2014.
[28] Munmun De Choudhury, Scott Counts, and Eric Horvitz. Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 3267–3276, Apr. 2013.
[29] Bong-Kee Chae. Insights from hashtag supplychain and twitter analytics: Considering twitter and twitter data for supply chain practice and research. International Journal of Production Economics, 165:247–259, Jul. 2015.
[30] Joao Paulo Carvalho, Hugo Rosa, Gabriel Brogueira, and Fernando˜ Batista. Misnis: An intelligent platform for twitter topic mining. Expert Systems with Applications, 89:374–388, Dec. 2017.
[31] Claudia E. Lopez, Miruna Vasu, and Caleb Gallemore. Understanding the perception of covid-19 policies by mining a multilanguage twitter dataset. 2020.
[32] Juan Esteban Camargo Saire and Rafael Correa Navarro. What is the people posting about symptoms related to coronavirus in bogota, colombia? 2020.
[33] Livia Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. 'go eat a bat, chang!': An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. 2020.
[34] Deepak P. Kaila and et al. Informational flow on twitter–corona virus outbreak–topic modelling approach. International Journal of Advanced Research in Engineering and Technology, 11(3):128–134, 2020.
[35] Xue Han, Jun Wang, Meng Zhang, and Xinhua Wang. Using social media to mine and analyze public opinion related to covid-19 in china. International Journal of Environmental Research and Public Health, 17(8):2788, Apr. 2020.
[36] Annelies Depoux, Sandra Martin, Emilie Karafillakis, Raman Preet, Annelies Wilder-Smith, and Heidi Larson. The pandemic of social media panic travels faster than the covid-19 outbreak. Journal of Travel Medicine, 27(3):taaa031, Apr. 2020.
[37] Bao Huang and Kathleen M. Carley. Disinformation and misinformation on twitter during the novel coronavirus outbreak. 2020.