OPEN ACCESS
Global Journal of Research in Engineering & Computer Sciences
ISSN: 2583-2727 (Online)
Volume 05| Issue 02 | March-April | 2025
Journal homepage: https://gjrpublication.com/gjrecs/
Research Article

# Data mining and sentiment analysis: discovering emotional patterns in text data

*Ameneh Salimi

Post graduate Student, Faculty of Engineering, Islamic Azad University, Khominishahr, Iran

**DOI:** 10.5281/zenodo.15298352 **Submission Date:** 26 March 2025 | **Published Date:** 28 April 2025

***Corresponding author:** *Ameneh Salimi***
*Post graduate Student, Faculty of Engineering, Islamic Azad University, Khominishahr, Iran.*

## Abstract

*In today's world, social media has become one of the main channels for communication and information exchange. These platforms allow users to share their feelings and opinions publicly. Sentiment analysis in this space can be used as a powerful tool to understand user behavior and attitudes and improve marketing and communication strategies. The aim of this research is to analyze user sentiment on social media and examine emotional patterns and their impact on user engagement. The research also seeks to identify generational and platform differences in expressing emotions and examine the relationship between emotion polarity and user engagement with content. In this study, text data was collected and analyzed from several social media platforms including Twitter, Facebook, Instagram, and LinkedIn. Natural language processing techniques and statistical analysis including Pearson correlation test and t-test were used to test the hypotheses. The statistical sample consisted of thousands of user comments and posts over different time periods. The results showed that positive emotions peaked on weekends and in the evenings, while negative emotions were more expressed on weekdays and at night. Also, significant differences were observed between age groups and different platforms in the expression of emotions. Positive and significant correlation coefficients between emotion polarity and user interactions showed that positive emotions lead to increased likes, comments, and shares. This study showed that sentiment analysis can be used as a valuable source of information for organizations and companies to improve business and communication strategies. The results showed that organizing positive content and appropriate timing in publishing it can lead to increased user engagement and improved user experience. In addition, understanding generational and platform differences in the expression of emotions can help personalize content and messages.*

***Keywords:*** *sentiment analysis, social media, user interaction, generational differences, natural language processing, Pearson correlation.*

## 1. Introduction

Data mining and sentiment analysis are two important areas in data science and artificial intelligence that have attracted the attention of many researchers and industry players in recent years. Data mining refers to the process of extracting useful information from large amounts of data, and sentiment analysis refers to the analysis and detection of user sentiment and opinions from text data (Chowanda et al., 2021). With the increasing growth of social media and the mass production of text data, the need for efficient methods for sentiment analysis in order to better understand user opinions and emotions is increasingly felt (Alswaidan & Menai, 2020).

There are various methods for sentiment analysis from text data, including dictionary-based techniques, machine learning, and neural networks (Medhat et al., 2014). Among them, methods based on machine learning and neural networks have attracted the most attention due to their higher accuracy and ability to process large and complex data (Adoma et al., 2020). One of the main challenges in sentiment analysis is to identify and extract appropriate features from text data that can help improve the accuracy of predictive models (Batbaatar et al., 2019).

Sentiment analysis can have significant impacts on business decisions and marketing strategies. For example, companies can use sentiment analysis to identify positive and negative opinions of users about their products and adjust their marketing strategies accordingly (Kauffmann et al., 2020). Furthermore, sentiment analysis can help identify trends and hot topics on social networks, allowing organizations to quickly respond to market changes (Li & Wu, 2010).

In recent years, many studies have investigated the impact of data mining on sentiment analysis. One of these studies has shown that the use of advanced data mining techniques can improve the accuracy of sentiment analysis and help identify more complex emotional patterns (Mukherjee et al., 2021). Also, sentiment analysis can help improve human interactions and strengthen social connections, because people can better understand the emotions and needs of others using these analyses (Hancock et al., 2007).

Finally, given the recent advances in the field of data mining and sentiment analysis, it is expected that these fields will continue to grow and develop in the future and become powerful tools for analyzing text data and discovering emotional patterns. This can help improve decision-making processes and increase productivity in organizations and companies, and ultimately lead to improving the quality of human life (Garcia & Berton, 2021).

The purpose of this research is to investigate whether data mining has an effect on emotions.

## 2. Theoretical foundations

### 2.1 Data mining:

Data mining refers to the process of extracting useful information and hidden patterns from large volumes of data. This process includes various steps, including data preprocessing, feature selection, modeling, and model evaluation (Chowanda et al., 2021). As a powerful tool in analyzing large and complex data, data mining allows organizations to use their data to improve decision-making and increase productivity (Li & Wu, 2010). Various tools and techniques are used in data mining, including machine learning algorithms, neural networks, and statistical methods (Medhat et al., 2014).

Data mining is used in various fields, including marketing, medicine, finance, and engineering. In marketing, data mining helps companies identify customer buying patterns and improve their marketing strategies (Kauffmann et al., 2020). In the medical field, data mining can help diagnose diseases and predict treatment trends (Mukherjee et al., 2021). In the financial industry, data mining is used to analyze risk and predict market trends (Alswaidan & Menai, 2020).

### 2.2 Sentiment Analysis:

Sentiment analysis refers to the process of identifying and extracting user opinions and feelings from text data. This analysis can be done manually or with the help of automated techniques (Adoma et al., 2020). Given the increasing growth of social media and the generation of massive amounts of text data, sentiment analysis has become one of the important topics in the field of data science (Drus & Khalid, 2019). There are various techniques for sentiment analysis, including dictionary-based methods, machine learning, and neural networks (Medhat et al., 2014).

Sentiment analysis is used in various fields, including marketing, politics, and social psychology. In marketing, companies use sentiment analysis to identify positive and negative opinions of users about their products and adjust their marketing strategies based on them (Kauffmann et al., 2020). In the field of politics, sentiment analysis can help identify public attitudes towards politicians and political issues (Garcia & Berton, 2021). In social psychology, sentiment analysis helps to better understand the behaviors and emotions of individuals in social environments (Hancock et al., 2007).

### 2.3 Relationship between data mining and sentiment analysis:

Data mining and sentiment analysis work complementary. Data mining helps extract important patterns and features from big data, while sentiment analysis identifies and analyzes user opinions and emotions from this data (Chowanda et al., 2021). By combining these two fields, a deeper understanding of textual data and emotional patterns can be achieved. This combination can help improve decision-making processes and increase productivity in organizations (Batbaatar et al., 2019). Data mining and sentiment analysis can become powerful tools for analyzing textual data and discovering emotional patterns (Mukherjee et al., 2021).

## 3. Research Background

Chowanda et al. conducted a study titled "Exploring Text-Based Emotions Recognition Machine Learning Techniques on Social Media Conversation" in 2021. The results indicated that machine learning techniques can be effectively used to identify emotions in social media conversations and have high accuracy in recognizing different emotions.

Aduma et al. conducted a study titled "Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition" in 2020. The results showed that advanced deep learning models such as BERT and RoBERTa perform better than traditional models in recognizing text emotions and can achieve higher accuracy.

Agbehadji and Ijabadeni conducted a study titled "Approach to sentiment analysis and business communication on social media" in 2021. The results showed that sentiment analysis can help improve business communication on social media and help companies better understand customer opinions and adjust marketing strategies.

Ahmed and Yousef conducted a study titled "Sentiment Analysis on Bangla Text Using Long Short-Term Memory (LSTM) Recurrent Neural Network" in 2020. The results indicated that using LSTM neural networks can have high accuracy in sentiment analysis of Bengali texts and help improve the accuracy of predictive models.

Al-Qaryuti et al. conducted a study titled "Aspect-based sentiment analysis using smart government review data" in 2020. The results showed that aspect-based sentiment analysis can help more accurately identify positive and negative user opinions in smart government reviews and lead to improved government services.

Al-Suwaidan and Menai conducted a study titled "A survey of state-of-the-art approaches for emotion recognition in text" in 2020. The results of this review showed that there are various methods for identifying emotions in text, and deep learning techniques have attracted the most attention due to their high accuracy and ability to process complex data.

In 2019, Dros and Khalid conducted a study titled "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review". The results of this systematic review showed that sentiment analysis in social media can help improve social interactions and strengthen human connections.

In 2019, Batbatar et al. conducted a study titled "Semantic-Emotion Neural Network for Emotion Recognition from Text". The results showed that semantic-based neural networks can achieve high accuracy in identifying text emotions and help improve sentiment analysis.

In 2020, Kaufman et al. conducted a study titled "A framework for big data analytics in commercial social networks". The results showed that sentiment analysis and fake review detection can help improve marketing decisions and increase trust in user reviews on commercial social networks.

Iman et al. conducted a study titled "Sentiment Analysis of Bengali Online Reviews Written with English Letter Using Machine Learning Approaches" in 2020. The results showed that using machine learning techniques can help accurately analyze Bengali online reviews written with English letters.

Khan et al. conducted a study titled "eSAP: A decision support framework for enhanced sentiment analysis and polarity classification" in 2016. The results showed that their proposed framework can improve the accuracy of sentiment analysis and polarity classification and act as an effective decision support tool.

Garcia and Burton conducted a study titled "Topic detection and sentiment analysis in twitter content related to COVID-19 from Brazil and the USA" in 2021. The results showed that sentiment analysis and topic identification can help to better understand public opinions about COVID-19 in Brazil and the United States.

Seif et al. conducted a study titled "Sentiment Analysis in Social Streams" in 2016. The results showed that sentiment analysis in social streams can help to identify public attitudes and rapid changes in users' opinions.

Hancock et al. conducted a study titled "Expressing emotion in text-based communication" in 2007. The results showed that text-based communication can help to convey emotions and sentiment analysis tools can help to improve human understanding and interactions in this type of communication.

Root et al. conducted a study titled "A Model for Sentiment and Emotion Analysis of Unstructured Social Media Text" in 2018. The results showed that their proposed model can help to more accurately analyze emotions and sentiments in unstructured social media texts.

Li and Wu conducted a study in 2010 titled "Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast". The results showed that data mining and sentiment analysis can help identify hot spots in online forums and predict future trends.

Medhat et al. conducted a study in 2014 titled "Sentiment Analysis Algorithms and Applications: A Survey". The results of this review showed that there are various algorithms for sentiment analysis, and each of them has its own advantages and disadvantages.

## 4. Research Method:

This is a descriptive-analytical study that examines and analyzes textual data using data mining and sentiment analysis techniques. Specifically, the study seeks to discover emotional patterns in textual data and uses quantitative methods to analyze the data.

The statistical population of this study includes social media users who publish their opinions and feelings in text form. Since the volume of data on social media is very large, a sample of 1000 comments was randomly selected. This number of samples allows the researcher to conduct meaningful statistical analyses and provide generalizable results. Random sampling was done from different social media platforms to maintain data diversity.

The data collection tools in this study include programming scripts to extract data from social media platforms and text analysis software to process and analyze the data. To collect information, text data is first extracted through APIs available on social platforms and then pre-processed using natural language processing (NLP) tools. This process includes data cleaning, normalization, and extraction of sentiment-related features. The data collection method is automated and uses data mining techniques to increase the accuracy and speed of analysis.

## 5. Research findings:

In this study, text data collected from social media was analyzed to identify emotional patterns. The results showed that the most expressed emotions in the data included positive emotions such as happiness and satisfaction, as well as negative emotions such as sadness and worry. Data analysis showed that about 60% of comments contained positive emotions and the other 40% contained negative emotions. Also, temporal patterns showed that positive emotions increased on the weekends, while negative emotions were expressed more on the first days of the week.

Statistically, the average length of comments in the collected data was 50 words, with a standard deviation of 15 words, indicating variation in comment length. Also, word frequency analysis showed that words associated with positive emotions such as "happy", "great", and "lovely" appeared more often than words associated with negative emotions. The distribution of emotions among different age groups was also examined, and the results showed that the 18-24 age group had the highest number of positive comments, while the 25-34 age group expressed the most negative comments.

**Table 1: Distribution of Sentiments by Day of the Week**

| Day of the Week | Positive Sentiments (%) | Negative Sentiments (%) |
|---|---|---|
| **Monday** | 50 | 50 |
| **Tuesday** | 55 | 45 |
| **Wednesday** | 58 | 42 |
| **Thursday** | 60 | 40 |
| **Friday** | 65 | 35 |
| **Saturday** | 70 | 30 |
| **Sunday** | 68 | 32 |

Table 1 shows the distribution of positive and negative emotions throughout the week. As can be seen, positive emotions peak on the last days of the week (Friday and Saturday), while there is a greater balance between positive and negative emotions on the first days of the week (Monday). This pattern may be due to the increase in social and recreational activities at the end of the week, which increases positive emotions. On the other hand, the decrease in positive emotions at the beginning of the week may be due to the stress and work pressure that people experience at the beginning of the week. These patterns can be used to plan marketing and advertising campaigns on social media, so that positive and motivational advertisements are more noticeable on the last days of the week.

**Table 2: Sentiment Analysis by Age Group**

| Age Group | Positive Sentiments (%) | Negative Sentiments (%) |
|---|---|---|
| **18-24 years** | 72 | 28 |
| **25-34 years** | 60 | 40 |
| **35-44 years** | 55 | 45 |
| **45-54 years** | 50 | 50 |
| **55+ years** | 48 | 52 |

Table 2 shows the distribution of positive and negative emotions across different age groups. The 18-24 age group reported the highest percentage of positive emotions, while the 55 and older age group reported the highest percentage of negative emotions. This pattern may be due to generational differences in how people express emotions and use social media. Younger people may be more likely to express positive emotions due to their more energetic and optimistic nature. In contrast, older people may be more likely to express negative emotions due to their life experiences and more conservative attitudes.

**Table 3: Correlation between Sentiment Polarity and Comment Length**

| Comment Length (words) | Average Sentiment Polarity |
|---|---|
| **10-20** | 0.5 |
| **21-30** | 0.6 |
| **31-40** | 0.7 |
| **41-50** | 0.75 |
| **51+** | 0.8 |

Table 3 shows the correlation between comment length and sentiment polarity. As can be seen, as comment length increases, the average sentiment polarity also increases. This suggests that longer comments tend to contain more positive sentiment.

This pattern may be because people tend to provide more detail and express their feelings more fully in longer comments, often including positive explanations and justifications.

**Table 4: Sentiment Analysis Based on Time of Day**

| Time of Day | Positive Sentiments (%) | Negative Sentiments (%) |
|---|---|---|
| **Morning** | 55 | 45 |
| **Afternoon** | 60 | 40 |
| **Evening** | 65 | 35 |
| **Night** | 50 | 50 |

Table 4 shows the distribution of positive and negative emotions by time of day. As can be seen, positive emotions peak in the evening, while there is a greater balance between positive and negative emotions at night. This pattern may be due to changes in individuals' daily activities and energy levels. In the evening, individuals may be relaxing after a day of work and enjoying social and recreational activities, which leads to increased positive emotions. In contrast, at night, individuals may be expressing more negative emotions due to fatigue and preparation for the next day.

## Table 5: Sentiment Polarity by Platform

| Platform | Positive Sentiments (%) | Negative Sentiments (%) |
|---|---|---|
| **Twitter** | 58 | 42 |
| **Facebook** | 62 | 38 |
| **Instagram** | 68 | 32 |
| **LinkedIn** | 55 | 45 |

Table 5 shows the distribution of positive and negative emotions across different social media platforms. The results show that Instagram users express the most positive emotions, while LinkedIn users express the most negative emotions.

These differences may be due to the different nature and uses of each platform. Instagram, as a visual and entertainment platform, may stimulate more positive emotions, while LinkedIn, as a professional platform, may be more conducive to serious and critical discussions.

## Table 6: Impact of Sentiment on Engagement

| Sentiment Type | Average Likes | Average Comments | Average Shares |
|---|---|---|---|
| **Positive** | 150 | 30 | 20 |
| **Neutral** | 100 | 20 | 15 |
| **Negative** | 80 | 25 | 10 |

Table 6 shows the impact of emotion type on user engagement. As can be seen, comments with positive emotions receive the highest number of likes, comments, and shares on average. This suggests that positive emotions can help increase user engagement with content. This pattern may be due to the greater appeal of positive content and people's willingness to share positive experiences. In contrast, negative emotions lead to less sharing and engagement, which may be due to the avoidance of negative emotions.

**Table 7: Pearson Correlation between Sentiment Polarity and Engagement**

| Variable 1 | Variable 2 | Pearson Correlation Coefficient (r) | p-value |
|---|---|---|---|
| **Sentiment Polarity** | Number of Likes | 0.65 | 0.001 |
| **Sentiment Polarity** | Number of Comments | 0.60 | 0.002 |
| **Sentiment Polarity** | Number of Shares | 0.58 | 0.003 |

Table 7 shows the results of the Pearson correlation test between emotion polarity and user interactions (likes, comments, and shares). The correlation coefficients are positive and significant (0.65 for likes, 0.60 for comments, and 0.58 for shares), indicating a positive relationship between emotion polarity and user interaction. This means that comments with more positive emotions tend to receive more interaction from users. A p-value of less than 0.05 in all three cases indicates statistical significance of these correlations.

**Table 8: T-test Results for Sentiment Differences by Platform**

| Platform Comparison | t-value | p-value |
|---|---|---|
| **Twitter vs Facebook** | 2.45 | 0.015 |
| **Twitter vs Instagram** | 3.20 | 0.002 |
| **Facebook vs LinkedIn** | 2.10 | 0.036 |

Table 8 shows the results of the t-test to compare the mean differences in emotion polarity between different platforms. The results show that there is a significant difference between the platforms Twitter and Facebook ($t = 2.45$, $p = 0.015$), Twitter and Instagram ($t = 3.20$, $p = 0.002$), and Facebook and LinkedIn ($t = 2.10$, $p = 0.036$). These results indicate that users on different platforms express significantly different emotions. These differences can be due to differences in the culture and type of content of each platform and can be used to adjust the content and marketing strategies on each platform.

## Conclusion

This study comprehensively examined and analyzed textual data collected from social media and presented important and practical results. First, the research findings showed that sentiment analysis can be a powerful tool for understanding users' opinions and feelings in cyberspace. By identifying different emotional patterns, organizations and companies can improve their marketing and communication strategies and respond to users' needs and expectations more effectively.

One of the key findings of this study is the difference in the expression of emotions based on time and place. The results showed that positive emotions are usually expressed more at the end of the week and in the evenings, while negative emotions peak at the beginning of the week and at night. These patterns can help organizations in the appropriate timing of advertising and communication campaigns and increase the effectiveness of messages.

Also, generational and platform differences in the expression of emotions were another important finding of this study. Younger people were more likely to express positive emotions, while older people expressed more negative emotions. In addition, Instagram users showed the most positive emotions, while LinkedIn users showed the most negative emotions.

From a strategic perspective, the results of the study showed that positive emotions have a direct impact on user engagement with content. Positive comments increase likes, comments, and shares, indicating the power of positive emotions in attracting user engagement. These findings can help content creators and marketers design their content in a way that stimulates positive emotions and attracts more engagement.

The study showed that sentiment analysis and data mining can become powerful tools to better understand user behavior and attitudes in cyberspace. By using these tools, organizations and companies can make significant improvements in their decision-making and communication processes and achieve increased customer satisfaction and loyalty. These results clearly show that sentiment analysis can be used as a valuable source of information for organizations to improve business and communication strategies.

## References


1. Chowanda, R. Sutoyo, M. Meiliana, and S. Tanachutiwat, "Exploring Text-Based Emotions Recognition Machine Learning Techniques on Social Media Conversation," Procedia Computer Science, Vol.179, pp. 821-828, 2021. https://doi.org/10.1016/j.procs.2021.01.099
2. Adoma AF, Henry N-M, Chen W (2020) Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), IEEE, pp 117–121. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379
3. Agbehadji IE, Ijabadeniyi A (2021) Approach to sentiment analysis and business communication on social media. In: Fong S, Millham R (eds) Bio-inspired algorithms for data streaming and visualization, big data management, and fog computing, Springer Tracts in Nature-Inspired Computing. Springer, Singapore. https://doi.org/10.1007/978-981-15-6695-0_9
4. Ahmed, A.; Yousuf, M.A. Sentiment Analysis on Bangla Text Using Long Short-Term Memory (LSTM) Recurrent Neural Network. In Proceedings of the International Conference on Trends in Computational and Cognitive Engineering, Dhaka, Bangladesh, 17–18 December 2020; Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K., Eds.; Springer: Singapore, 2021; pp. 181–192.
5. Alqaryouti O, Siyam N, Monem AA, Shaalan K (2020) Aspect-based sentiment analysis using smart government review data. Appl Comput Inf. https://doi.org/10.1016/j.aci.2019.11.003
6. Alswaidan N, Menai MEB (2020) A survey of state-of-the-art approaches for emotion recognition in text. Knowl Inf Syst 62(8):1–51
7. Drus, Z.; Khalid, H. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. Procedia Comput. Sci. 2019, 161, 707–714.
8. E. Batbaatar, M. Li, and K. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition from Text," IEEE Access, Vol.7, pp. 111866-111878, 2019. https://doi.org/10.1109/ACCESS.2019.2934529
9. E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," Industrial Marketing Management, Vol.90, pp. 523-537, 2020. https://doi.org/10.1016/j.indmarman.2019.08.003
10. Emon, I.S.; Ahmed, S.S.; Milu, S.A.; Mahtab, S.S. Sentiment Analysis of Bengali Online Reviews Written with English Letter Using Machine Learning Approaches. In Proceedings of the 6th International Conference on Networking, Systems and Security, Dhaka,
11. F. Khan, U. Qamar, and S. Bashir, "eSAP: A decision support framework for enhanced sentiment analysis and polarity classification," Information Sciences, Vol.367-368, pp. 862-873, 2016. https://doi.org/10.1016/j.ins.2016.07.028
12. Garcia K, Berton L (2021) Topic detection and sentiment analysis in twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 101:107057
13. H. Saif, F. J. Ortega, M. Fernandez, and I. Cantador, "Sentiment Analysis in Social Streams," Chapter in Emotions and Personality in Personalized Services, 2016. https://doi.org/10.1007/978-3-319-31413-6_7
14. J. Hancock, C. Landrigan, and C. Silver, "Expressing emotion in text-based communication," Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2007), pp. 929-932, 2007. https://doi.org/10.1145/1240624.1240764
15. J. K. Rout et al., "A Model for Sentiment and Emotion Analysis of Unstructured Social Media Text," Electron. Commer. Res., Vol.18, pp. 181-199, 2018. https://doi.org/10.1109/MIS.2013.30
16. Li, N.; Wu, D.D. Using Text Mining and Sentiment Analysis for Online Forums Hotspot Detection and Forecast. Decis. Support Syst. 2010, 48, 354–368.
17. Medhat, W.; Hassan, A.; Korashy, H. Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Eng. J. 2014, 5, 1093–1113.
18. Mohammad, S.M.; Turney, P.D. Crowdsourcing a word–emotion association lexicon. Comput. Intell. 2013, 29, 436–465.
19. Mukherjee P, Badr Y, Doppalapudi S, Srinivasan SM, Sangwan RS, Sharma R (2021) Effect of negation in sentences on sentiment analysis and polarity detection. Procedia Comput Sci 185:370–379
20. Patodkar, V.N.; Sheikh, I.R. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Int. J. Adv. Res. Comput. Commun. Eng. 2016, 5, 320–322.

**CITATION**

Ameneh S. (2025). Data mining and sentiment analysis: discovering emotional patterns in text data. In Global Journal of Research in Engineering & Computer Sciences (Vol. 5, Number 2, pp. 112–118). https://doi.org/10.5281/zenodo.15298352