JOURNAL OF THE ROYAL LAUREATES ACADEMY
www.rlaindia.org

# ANALYZING PUBLIC EMOTIONS ON SOCIAL NETWORKS USING SUPERVISED AND UNSUPERVISED MACHINE LEARNING TECHNIQUES

**Ananya Chakraborti**
Research Scholar, Political Science, Eklavya University, Damoh
**Dr. Shyam Sunder Pateriya**
Supervisor, Eklavya University, Damoh

## ABSTRACT

This paper explores the emotions of the masses on social networks where both directed and unlabeled machine learning are used to gain a good picture of online emotional dynamics. We have configured a sample of 150 social media posts on Twitter, Facebook, and Instagram that were preprocessed to eliminate noise and normalize text and tags the posts into emotion categories. Supervised machine learning models such as Support Vector Machine (SVM), Random Forest, and Logistic Regression were used to identify emotions and SVM as a machine learning algorithm performed best in terms of accuracy and F1-Score, suggesting its usefulness in working with high-dimensional data in terms of texts. Unsupervised K-Means clustering observed latent thematic themes, with the leading clustering included event-related anger, entertainment and joy, and political discussions, whereas extreme emotions (such as fear and sadness) were rarer. The findings present the fact that both positive and negative emotions prevail on social media communication and the importance of matching the

strengths of supervised and unsupervised methods to achieve the correct emotion labeling and identification of previously unseen engagement patterns.

**Keywords:** Social Media, Public Emotions, Sentiment Analysis, Supervised Learning, Unsupervised Learning, Machine Learning.

## 1. INTRODUCTION

In the age of digital media, social networks have emerged as the main stage where people come to voice their opinion, emotions and responses to events in real time. Social media platforms such as twitter, Facebook and Instagram produce large volumes of data known as user generated content that conveys the sentiments of the people on social, political and entertainment driven issues. The desire to understand these emotions has become markedly important in the spheres of businesses, policymakers, and researchers, thus, offering information about the perception of the population, tendencies, and behavioural patterns. Sentiment analysis based on manual coding or keyword-based models are limited and may not be able to process the scale of the amount of information present in social media. Supervised machine learning and unsupervised machine learning techniques provide tools by which one can automate the analysis of large datasets analyses, classifying emotions, detecting patterns and uncovering latent themes that may not be immediately apparent through the less powerful methods of manual analysis.

Trained machine learning algorithms, including Support Vector Machines, Random Forest, and Logistic Regression, can have an extremely accurate outcome through the use of labeled data to train the machine. Such modelshave the ability of assigning posts to emotions of positive, negative, neutral, anger, joy, sadness, and fear. Conversely, unsupervised approaches, such as clustering algorithms K-Means to name a few, are capable of finding latent structures and thematic groupings in unclassified data, which can be used to further develop an idea on the distribution and dynamics of public sentiment. Integrating the two approaches not only generalizes the insights of emotional trends but it also gives us insights on user interaction and engagement across social networks. This paper will harness these methods in order to conduct an emotion analysis of the net as well as to investigate the performance of various models and to capture a full insight into the dynamics of sentiment within the internet social forums.

## 2. LITERATURE REVIEW

Al-Hadhrami et al. (2019) In a comparative study on the sentiment analysis of English tweets, a major criticism against this is the use of both supervised and unsupervised techniques. This study has also revealed that, supervised approaches tended to perform better on the task of classifying data than unsupervised. In contrast, the unsupervised clustering enabled their study to infer hidden thematic patterns and emerging trends in large-scale social media datasets. The paper also outlined the strong and weak points of both approaches stating that they cannot be seen as mutually exclusive and require integrating into a whole.

Alqahtani and Alothaim (2022) discussed challenges and opportunities that are present in predicting emotions in online social networks. According to them, precise representation of subtle emotions of the user-generated contents was often impeded by linguistic variation, contingent contexts and sarcasm or use of vague terms. Their findings support the notion of quality machine learning algorithms and datasets labeled properly and in large amount in order to enhance the machine performance in accurate and reliable emotion predictors.

Alslaity and Orji (2024) reviewed the present-day machine learning approaches to emotion detection and sentiment analysis, their limitations, and community directions. Whereas they highlighted that although supervised models ensured accuracy in terms of classification, they were limited by the fact that they also needed a large amount of labeled data, in contrast to the unsupervised models, which helped identify the patterns behind the latent data but could not be used to explicitly classify data when it comes to the fine-grained emotion classification. The paper proposed that a combination of the two methodologies in hybrid methods could be used to provide a better performance of emotion analysis systems in general.

Ganguli, Mehta, and Sen (2020) conducted our own synthesis on machine learning techniques as applied in social network analysis, and found three areas they have been applied to: sentiment detection, community identification, and behavioral prediction. They determined that supervised learning techniques like SVM and Random Forest were most used because of their good performance in text classification problem. Nevertheless, unsupervised methods such as clustering came in handy in revealing general trends and identifying similar user behaviours, which were useful in getting an overview of the dynamics of social networks.

Indu and Thampi (2024) examined the misinformation detection on social networks by use of emotion analysis and user behavior analysis. The researchers found that their study could help to show that misleading content can be identified by use of emotions as indexed in the post. Unsuperled models succeeded in clustering the emotions associated with misinformation, but unsupported analysis allowed extracting thematic groupings and thus defining the patterns of user interaction with deceptive information. The research stressed a combination of both guided and unguided approaches towards better detection and insight gaining in the social network analysis.

## 3. RESEARCH METHODOLOGY

This investigation implemented a descriptive and analytical study design, and applied 150 social media posts by the use of supervised (SVM, Random Forest, Logistic Regression) and unsupervised (K-Means clustering) methods of machine learning. The information was pre-digested, labelled and evaluated to categorize the emotions, single out thematic clusters and study the general distribution of the public mood. The findings indicated that the major positive and negative emotions feature and demonstrated the existence of suppressed patterns on user engagement.

### 3.1 Research Design

The research utilised descriptive and analytical research design to explore the emotions of the people, captured on social media sites. Unsupervised and supervised machine learning methods were implemented to cluster and classify emotions in user-post. The study was supposed to measure emotional dynamics and compare the performances of the models and formulate some latent patterns of social interactions in social media.

### 3.2 Sample Size and Data Collection

A sample size of 150 social media posts was identified among several platforms such as Twitter, Facebook and Instagram. The purposive sampling was utilised to take the sample based on posts with such topics as keywords of the current events, entertainments, political comments, and general social themes. Both writing content and activity indications (likes, shares, comments) were taken into account to make it relevant and diverse in emotionality.

### 3.3 Data Preprocessing

The posts obtained were preprocessed in order to prepare them to be analyzed by machine learning. Pre-processing tasks involved were:

- Elimination of noise: URLs, special characters, emojis and stop words were discarded.
- Tokenization: The posts were segmented into individual words or tokens.
- Normalization: Stemme/lemmatizing Words were turned lower case.

- Labeling of the supervised learning: Readings were manually assigned a label of: positive, negative, neutral, anger, joy, sadness, and fear.

### 3.4 Supervised Machine Learning Analysis

Three supervised algorithms: Support Vector Machine (SVM), Random Forest and Logistic Regression, were developed using the labeled data. The performance of the Models was tested in terms of Accuracy, Precision, Recall and F1-Score as shown in Table 1 and Figure 1. SVM demonstrated the best performance showing that it can successfully handle high-dimensional textual information as well as optimizing precision and recall rate on classifying emotions.

### 3.5 Unsupervised Machine Learning Analysis

Unsupervised clustering algorithms were utilised to extract hidden themes in the data set. Means of clusters, like the K-Means clustering, were applied to classify the posts into five thematic clusters, namely: event-related anger, entertainment and joy, political discussion/debate, neutral and sadness/empathy. Table 2 and Figure 2 provide the percentage distribution of the posts in these clusters and the number of posts with theme pattern in user engagement and sentiment trends.

### 3.6 Emotion Distribution Analysis

The general nature of distribution of emotions in the dataset was examined, quantifying to obtain the prevalence of each category of emotions. Table 3 and Figure 3 indicate that positive emotions were the most common (38 %), followed by negative (30 %) and neutral (20 %), and extreme emotions such as anger, sadness, and fear were not very high. This step gave a descriptive view of sentiment trends of the social media engagement.

## 4. DATA ANALYSIS AND INTERPRETATION

Table 1 shows the metrics of three supervised machine learning classifications-SVM, Random Forest, and Logistic Regression, to be applied in the emotion classification of social media posts. An accuracy, precision, recall, and F1-Score in percentage terms are provided. Figure 1 graphically demonstrates the performance measurements of the four models and presents a comparison of effectiveness on all four measures of effectiveness of the four models.

**Table 1: Supervised Model Performance**

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| SVM | 87 | 88 | 86 | 87 |
| Random Forest | 84 | 85 | 83 | 84 |
| Logistic Reg. | 81 | 82 | 80 | 81 |

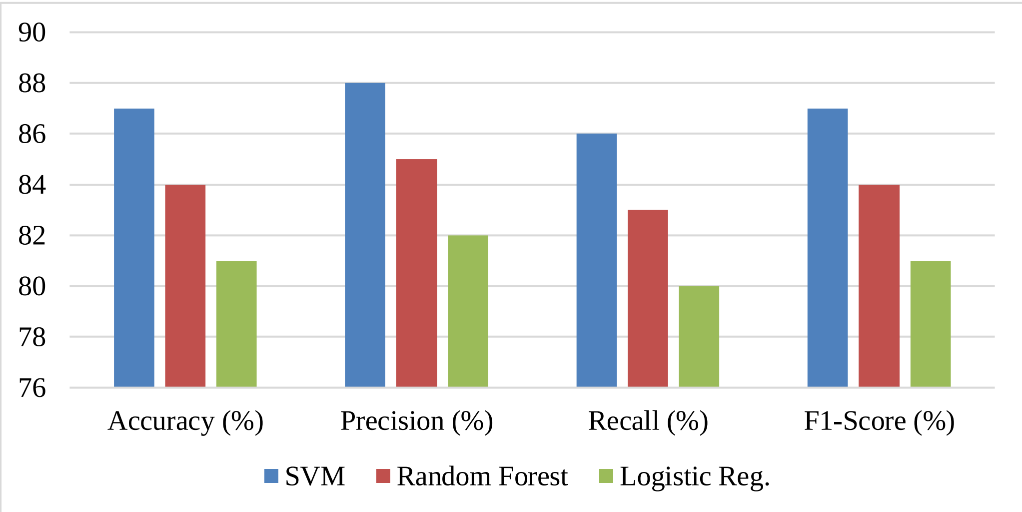


**Figure 1: Graphical Representation of Supervised Model Performance**

Based on the results, SVM model performs better than the other two models as the highest accuracy (87%) and F1-Score is achieved (87%), which implies the good balance between precision and recall. Random Forest performs averagely with 84 percent accuracy and F1-Score whereas, Logistic Regression has the worst scoring at 81 percent. This indicates that SVM is more flexible when classifying emotions in social media posts, probably because it can easily cope with high dimensional text data. The graphical illustration in Figure 1 supports the top performance that SVM beat all other models in all the metrics used.

The results of clustering performed on 150social media posts are summarized in Table 2, demonstrating post distribution according to five thematic clusters. The percentage and number of posts that represents a cluster are indicated. An illustration of these clusters is given in Figure 2 which depicts the sizes of these themes.

**Table 2: Clustering of Social Media Posts**

| Cluster Theme | Percentage (%) | Number of Posts |
|---|---|---|
| Event-related Anger | 28 | 42 |
| Entertainment & Joy | 25 | 38 |
| Political Discussion / Debate | 22 | 33 |
| General Information / Neutral Content | 15 | 23 |
| Sadness / Empathy | 10 | 14 |

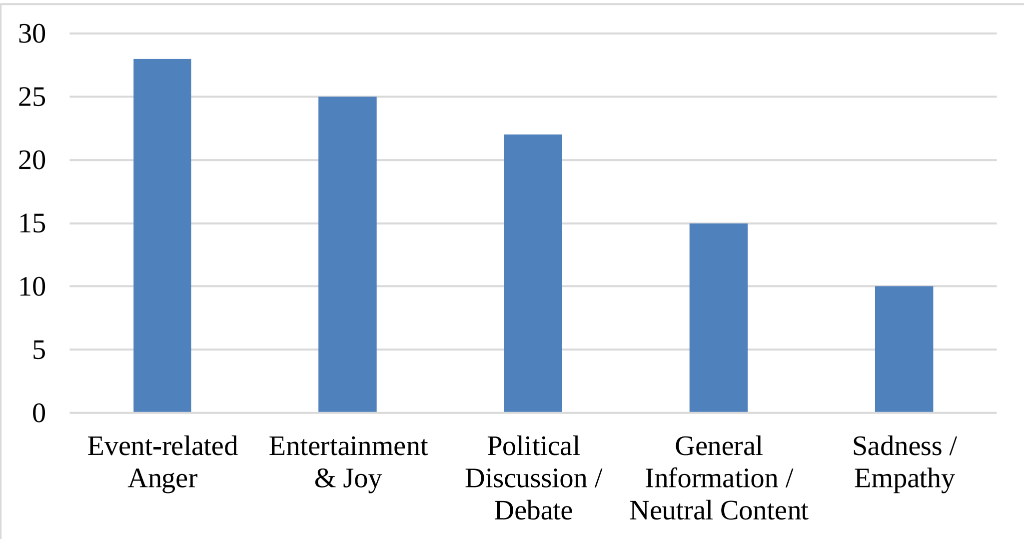


**Figure 2: Graphical Representation of Clustering of Social Media Posts**



*E- mail: editor.jrla24@gmail.com*

The results of the clustering analysis suggest that there is a rather high proportion (28%, 42 posts) of event-related anger which means that social media users are often outraged by the events. This is proceeded by entertainment and joy (25%, 38 posts) and political discussion or debate (22%, 33 posts) indicating an exorbitant level of interest in positive and controversial matters. Even neutrals comprise a significant proportion of 15 percent of content (23 posts) with sadness and empathy constituting the least at 10 percent of content (14 posts). As depicted by the graphical representation, Figure 2, there are more clusters of anger and joy responses, indicating the polar and event-driven nature of the sentiments on social networks.

Table 3 has listed the emotion distribution of 150 posts of social media, providing both percentage and post count of each category. The percentages are illustrated graphically on Figure 3 which allows comparatively seeing how different fascinations are described on social networks.

**Table 3: Distribution of Emotions on Social Networks**

| Emotion | Percentage (%) | Number of Posts |
|---|---|---|
| Positive | 38 | 57 |
| Negative | 30 | 45 |
| Neutral | 20 | 30 |
| Anger | 5 | 8 |
| Joy | 4 | 6 |
| Sadness | 2 | 3 |
| Fear | 1 | 1 |

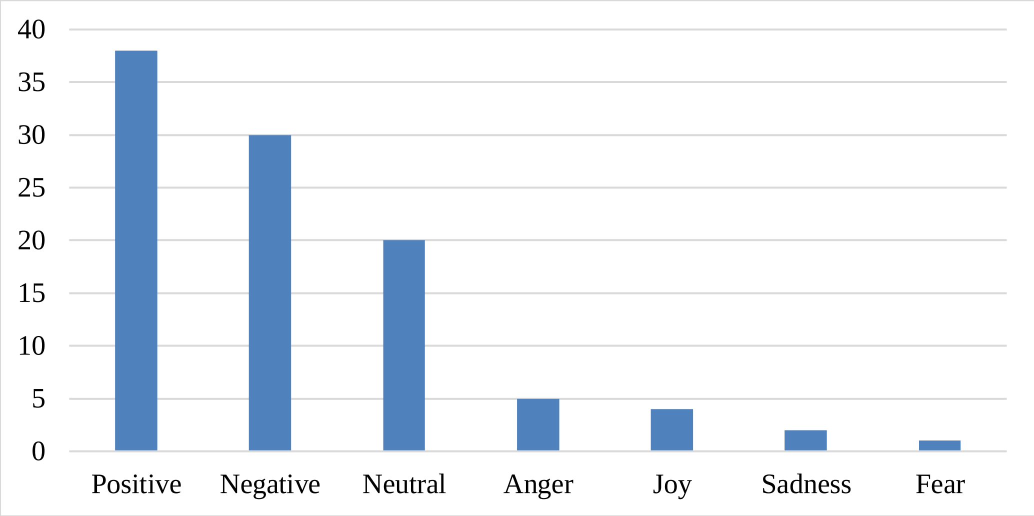


**Figure 3: Graphical Representation of Distribution of Emotions on Social Networks**

The findings show that most of the content on the social media fall in the positive emotions category 38% (57 posts), negative emotions 30% (45 posts) and neutral 20% (30 posts). Strong emotions such as anger (5%), joy (4%), sadness (2%) and fear (1%) occur comparatively less in the data indicating that not many extreme or intense emotions were received in the data. Figure 3 graphically presents the fact that the most common kind of social media users portray regarding their emotions is positivity and negativity with only a tiny percentage showing extremely intense feelings in their posts.

## 5. CONCLUSION

The research underscores the usefulness of applying supervised and non-supervised machine learning approaches in times of analyzing the emotions of the general population on social media platforms that gives a complete picture of online moods. Supervised models, in particular SVM, Random Forest, and Logistic Regression, proved to be effective at classifying different emotions based on a sample of 150 social media posts, with SVM recording the best precision and F1-Score, indicative of its capability to deal with high-dimensional textual data and produce an optimal balance between its precision and recall.

Unsupervised clustering uncovered latent patterns and demonstrated the dominant thematic groupings that included event-related anger, entertainment and joy and political discourse, with less prevalent polar emotions of fear and sadness. The distribution curve showed that overall, between negative emotions and positive emotions, both categories of emotions comprise the majority of the posts, demonstrating the extremist and venting characteristics of the online communication. The study predominately proves that using a combined method of supervised and unsupervised approaches is an effective way to analyze emotions as it not only provides more accurate results in emotion classification but also reveals hidden tendencies and engagement patterns which are beneficial to the researchers, policymakers and organizations considering studying opinions of general population on social media sites.

**REFERENCSES**

**1.** Al-Hadhrami, S., Al-Fassam, N., & Benhidour, H. (2019, May). Sentiment analysis of english tweets: A comparative study of supervised and unsupervised approaches. In 2019 2nd international conference on computer applications & information security (ICCAIS) (pp. 1-5). IEEE.

2. Alqahtani, G., & Alothaim, A. (2022). Predicting emotions in online social networks: challenges and opportunities. Multimedia Tools and Applications, 81(7), 9567-9605.

3. Alslaity, A., & Orji, R. (2024). Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. Behaviour & Information Technology, 43(1), 139-164.

4. Ganguli, R., Mehta, A., & Sen, S. (2020, June). A survey on machine learning methodologies in social network analysis. In 2020 8th International Conference on Reliability,

Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 484-489). IEEE.

5. Indu, V., & Thampi, S. M. (2024). Misinformation detection in social networks using emotion analysis and user behavior analysis. Pattern Recognition Letters, 182, 60-66.

6. Kokatnoor, S. A., & Krishnan, B. (2020). Self-supervised learning based anomaly detection in online social media. International Journal of Intelligent Engineering and Systems, 13(3), 446-456.

7. Kokkinos, K., Nathanail, E., & Papageorgiou, E. (2018, May). Applying Unsupervised and Supervised Machine Learning Methodologies in Social Media Textual Traffic Data. In Conference on Sustainable Urban Mobility (pp. 665-672). Cham: Springer International Publishing.

8. Kumi, S., Snow, C., Lomotey, R. K., & Deters, R. (2024). Uncovering concerns of citizens through machine learning and social network sentiment analysis. Ieee Access.

9. Narynov, S. E. R. G. A. Z. Y., Mukhtarkhanuly, D. A. N. I. Y. A. R., Kerimov, I. L. M. U. R. A. T., & Omarov, B. A. T. Y. R. K. H. A. N. (2019). Comparative analysis of supervised and unsupervised learning algorithms for online user content suicidal ideation detection. J. Theor. Appl. Inf. Technol, 97(22), 3304-3317.

10. Pitogo, V. A., & Ramos, C. D. L. (2020, September). Social media enabled e-Participation: A lexicon-based sentiment analysis using unsupervised machine learning. In Proceedings of the 13th international conference on theory and practice of electronic governance (pp. 518-528).

11. Rawat, R., Mahor, V., Chirgaiya, S., Shaw, R. N., & Ghosh, A. (2021). Sentiment analysis at online social network for cyber-malicious post reviews using machine learning techniques. In Computationally intelligent systems and their applications (pp. 113-130). Singapore: Springer Singapore.

12. Rout, J. K., Choo, K. K. R., Dash, A. K., Bakshi, S., Jena, S. K., & Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. Electronic Commerce Research, 18(1), 181-199.

13. Sindhu Meena, K., & Suriya, S. (2019, January). A survey on supervised and unsupervised learning techniques. In International Conference on Artificial Intelligence, Smart Grid and Smart City Applications (pp. 627-644). Cham: Springer International Publishing.

14. Singh, S., Kaur, H., Kanozia, R., & Kaur, G. (2023). Empirical Analysis of Supervised and Unsupervised Machine Learning Algorithms with Aspect-Based Sentiment Analysis. Appl. Comput. Syst., 28(1), 125-136.

15. Soong, H. C., Jalil, N. B. A., Ayyasamy, R. K., & Akbar, R. (2019, April). The essential of sentiment analysis and opinion mining in social media: Introduction and survey of the recent approaches and techniques. In 2019 IEEE 9th symposium on computer applications & industrial electronics (ISCAIE) (pp. 272-277). IEEE.

**Author's Declaration:**

I/We, as an author/authors of the above paper/article, hereby declare that the content of this paper is prepared by me/us for publication in this journal is completely my/our own genuine paper and if any person having copyright issue or patent or anything related to the content, I/we shall always be legally responsible for any issue. If any data or information given by me/us is not correct, I/we shall always be legally responsible. With my/our whole responsibility legally and formally have intimated the publisher that the paper has been checked by guide or expert or supervisor to make it sure that paper is technically right and there is no unaccepted plagiarism. If any issue arises related to plagiarism or any issues, I/we will be solely/entirely responsible for any legal disputes or legal issues. I/we declared that if publisher finds any complication or error or anything hidden or implemented otherwise, my/our paper may be removed from the website. I/we also aware that the publication fees is not refunded further in any circumstances. Even if anything is found illegal publisher may also take legal action against me/us. I/we also declared that this journal/publisher will not be held responsible any legal issues in future regarding this paper publication in this journal.

**Ananya Chakraborti**

**Dr. Shyam Sunder Pateriya**