International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

A Survey on Sentiment Analysis of Threads Application: A Comprehensive Study

Afsheen Pothiyawala[1], Arin Malewar[2], Mrs. Ashu Nayak[3]
[1&2]BCA-6th Sem, Kalinga University, Raipur, India
[3]Assistant Professor, Kalinga University, Raipur, India
[1]affi.m810@gmail.com, [2]amalewar75@gmail.com, [3]ashunayak7791@gmail.com

***Abstract*—Sentiment analysis, also considered as opinion mining, is a crucial element in NLP (Natural Learning Processing), which distinguishes the polarity of any text, or word, or phrase. The project aims to examine and enhance sentiment analysis abilities using machine learning models: Support vector machine (svm), Naīve Bayes, Linear regression. The raw datasets are collected and preprocessed from diverse domains to ensure a comprehensive study of this paper. Text preprocessing includes the steps: manual labelling, normalization, tokenization, stop words removal, preparing the texts for sentiment analysis. The project focuses on comparison of models to identify the suitable approach out of all. This project offers insights to different techniques, its application, limitations, strengths contributing to NLP methodologies for elevation of sentimental analysis.**

Keywords — Sentiment analysis, NLP, Support vector machine, Naive Bayes, Linear regression, text preprocessing

I. Introduction

Sentimental Analysis:-

Sentimental analysis has been one of the most crucial topics among the researchers since past decade. Sentimental analysis is required for identification of sentiments behind the text sentence whether the sentence is a positive sentence or a negative sentence.

Analyzing the past few decades, the usage of internet has increased significantly which has led to people using social media, online review platforms, personal blogs, etc. for sharing their opinions and views on a particular subject or a product.

Sentimental analysis makes use of the natural language processing (NLP) techniques and categories text data into all kinds of sentimental categories (positive, negative, neutral, etc.)

This provides corporations with a better understanding of customer or public opinions and helps in analyzing trends.

It is also defined as the name given to emotional analysis by computational methods aiming at identifying, measuring, and classifying attitudes or emotions represented by any text. It therefore means that, in that respect, the main goal of sentiment analysis in general, being a part of NLP studies, is to try to determine or identify the inner sentiment—whether negative, positive, or neutral—behind spoken or written material. This approach estimates a message's complete emotional tone by evaluating linguistic aspects, including word usage, tone of voice, and context. Generally utilized for product reviews, social media posts, and customer feedback, sentiment analysis provides useful information to a company for the monitoring of brand perception, understanding consumer thoughts, and developing customer experiences. Advanced models for sentiment analysis leverage contextual embeddings through the use of transformers and deep learning to accurately pinpoint ambiguous language, sarcasm, and subtle feelings.

Natural learning processing (NLP) helps computer to analyze and understand human language via using computational operations and methods. It follows the steps as text preprocessing being the first step includes tokenization (breaking down of the texts in smaller units, words or phrases.), the next step lowercasing (converting all the texts into lowercase improving its consistency).

Research Background:-

This research investigates some machine learning model performances for classifying sentiment analyses, based on comments extracted from the Google Play Store-Threads, with a focus on sentiment analysis related to user reviews. It is important to realize that, in the competitive atmosphere of the social media platform sector, Threads is an application that continuously adopts and changes its direction to meet both changing user demands and technical development.

The classification models used in this study will be based on research by Majid et al. work, which examined the accuracy of three classification models—Naïve Bayes (92.7%),

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

Random Forest (93.7%), and Support Vector Machine (94.2%)—will serve as the foundation for the models employed in this investigation. The definition of important sentiment analysis terms, an overview of the classification models in use, and a discussion of the methods used to examine user sentiment on Threads are covered in the parts that follow. To give a thorough grasp of the opinions spoken by Threads users, the figures and data produced for this study will be shown and examined. In order to contribute to the continuing discussion regarding social media sentiment analysis, the conclusion will include a summary of the results as well as future research areas and implications.

Threads:-

The Instagram team has introduced an application known as Threads (in july 2023), which enables users to share their thoughts with other individuals to discuss problems or even share information in textual format. During its introduction in 100 nations, the Threads have generated more than 73 million program installations within one month of its circulation. By December 2023, the Threads had scaled up to host 141 million registered users with an average of about 73 million Month on Month active users globally. Given the evidence presented is convincing, considering that Threads is a new social media app that gained a large user base in a short period of time. Even so, when compared to its former competitor X, Threads users are arguably less than half of X's 500 million users by September 2023.

Characteristics of Threads :-

1. Threads was created as a direct rival to X/Twitter, a microblogging site for sharing brief articles and amusing anecdotes.
2. Even though Meta Threads and X/Twitter are similar, Meta Threads has a number of distinctive features that have drawn in over 137 million members so far:
3. Up to 500 characters can be posted in text (compared to 280 characters on X/Twitter).
4. Up to five-minute videos.
5. The ability to follow accounts you already follow on Instagram.
6. An algorithm-driven feed that displays content from followed accounts and suggested content rather than the most recent content first.
7. The ability to "mute" accounts instead of banning or unfollowing them.

Importance of This Research:-

Investigating social media threads is essential because of the extensive, real-time data they offer. Public opinion, new trends, and social issues are all reflected in the threads. This information is useful for a number of reasons:

1. Business: Thread sentiment analysis provides real-time input on goods and services, supporting advertising tactics.
2. Politics: By assessing public opinion on politicians and policies, thread analysis aids in the development of campaign tactics.
3. Crisis Management: By identifying early warning indicators of crises, Threads facilitates prompt responses.
4. Public health: Monitoring Threads keeps tabs on the spread of illnesses and how the general public responds to health campaigns.

All things considered, Thread analysis aids in data-driven decision-making in a variety of fields, improving our comprehension of social trends and human behavior.

II. LITERATURE REVIEW

Numerous academic studies have been carried out to investigate sentiment analysis in social media situations, primarily using techniques like Naïve Bayes and SVMs, or support vector machines. Positive sentiment is more common than negative sentiment, according to these studies. Majid et al., for example, used a Natural Language Processing (NLP) method in a study that concentrated on sentiment analysis within the Threads platform. Despite not naming the classification model they employed, their analysis of 1,000 reviews yielded an average accuracy of 76.92%, with precision and recall rates of 80% and 74%, respectively.

The field of sentiment analysis, sometimes referred to as sentiment mining, has grown in importance within machine learning and natural language processing (NLP). According to **Error! Reference source not found.** , it entails the automatic extraction and categorization of concepts found in the literature. Opinion polls are becoming more important than ever due to the rise in user-generated material on social media platforms, online review sites, and discussion forums (Bollen et al., 2011). The evolution of emotion measurement techniques, sentiment analysis's difficulties, and the function of machine learning algorithms in emotion categorization are all covered in this review of the literature.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

Sentiment analysis, Prediction of trends by Threads:-

In ML and NLP, Sentiment analysis is becoming more and more necessary day by day. Bollen, Mao & Zeng (2011), stated Opinion polls are becoming more and more necessary as user-generated material on social media platforms, online review sites, and discussion forums grows. Concepts from the literature are automatically extracted and categorized. Earlier, the predictions were based on longer texts, which includes product reviews, movies reviews (Pang & Lee, 2008).

Public opinion and political trends, sentiment analysis in threads:-

Opinions are differentiated between positive comments, negative comments, and neutral too. Sentiment analysis predicts mood of the public opinion on trends on app. **Error! Reference source not found.** stated that for theoretical models, the existence of noisy data, many languages spoken by various populations and geographical areas, and various classes within the data set present significant challenges. Despite such challenges, emotional analytical intelligence have concurring opportunities in several fields of business, tech, politics, medicines, well being **Error! Reference source not found.**.

Comparisons with Twitter/ X :

Twitter founded in 2006, rebranded as X, by Elon Musk in 2022 faced a backlash for sometimes as Instagram's Threads was introduced as its competitor. Sudden launching of threads, piqued interest in public and the growth of Threads up scales in millions in 100 nations worldwide, still the population cannot be compared to twitter/X, as it can be considered threads have half of the population as Twitter/ X.

III. METHODOLOGY

In this process, the Data is processed to determine whether the review posted are negative or positive or neutral. This section includes Data mining and Data preprocessing of the datasets.

Data Mining:

This section involves collection of data sets, and data mining of the collected data for the proposed research. The dataset proposed is downloaded from Kaggle.

	Unnamed: 0	source	review_id	user_name	review_title	review_description	rating	thumbs_up	review_date	de
0	0	Google Play	7cd90e5b-4829-43b9-9fb4-c8c6d1e339c1	Eddie Clark Jr.	NaN	Good	5	0.0	2023-08-07 19:14:36	Na
1	1	Google Play	6deb8265-2bac-4524-bcb6-f90829fa4e69	Rasa RT	NaN	Weak copy of Twitter	1	0.0	2023-08-07 19:07:04	Na
2	2	Google Play	91ef61ce-0f05-4f3b-b3d3-5d19cd408ab8	SITI NUR HAFIZA BINTI AZIZ	NaN	i wish threads have a save button for images a...	3	0.0	2023-08-07 18:57:07	Na
3	3	Google Play	b7721b78-6b77-4f8c-a1d3-a854af4c1f0f	Asap Khalifah	NaN	Love it	5	0.0	2023-08-07 18:37:16	Na
4	4	Google Play	c89ef522-c94c-4171-878f-1d672dce7f11	Syed Hussein	NaN	Very god	5	0.0	2023-08-07 18:14:15	Na
...	...	...	...	...	...	...	...	...	...	...
36938	1995	App Store	d0503900-7c8b-4cf3-8f16-1a628f25d99b	lleila888	Hybrid of IG and Twitter with 0 UVP	Threads have mediocre UX with 0 unique value p...	2	NaN	2023-07-06 00:37:57	Na
36939	1996	App Store	d55529a3-42b9-4a55-a17c-b49f30d7b419	MaxW239	Outstanding	Twitter (Instagram's Version)	5	NaN	2023-07-06 00:00:39	Na
36940	1997	App Store	8818ddd0-1ce4-4d82-b0df-b43a6c68ff81	Anne Marie C	Let the battle begin!	😂😂😂	5	NaN	2023-07-05 23:16:15	Na
36941	1998	App Store	81561238-d8c1-467d-b9dc-be25d64a1aa7	alexcookiedough92	No search bar??	How do you expect a social media app to succee...	1	NaN	2023-08-06 12:31:54	Na

Fig 1-Snippet of the data collection in CSV format

Data PreProcessing:-

Data or text preprocessing is the initial step which is done to understand the sentiment of comments from the users. The comments rated as 1, 2 & 3 are automatically considered as negative by the Google Inc., whereas the comments with rating 4 & 5 are considered as positive. Data Preprocessing involves various steps, followed as:

- Removal of all the unwanted variables such as emojis, url's, hashtag, referral codes or any other targets.
- Manually converting the chosen as positive and negative. Classifying of rating 1, 2 & 3 as negative whereas rating 4 & 5 as positive.
- Translating the comments into English (with Google Translate), and applying auto correct, then reverting it back to original (or source) language. This process also includes removal of common slangs for better understanding.
- Removal of all the punctuations, numbers and symbols.

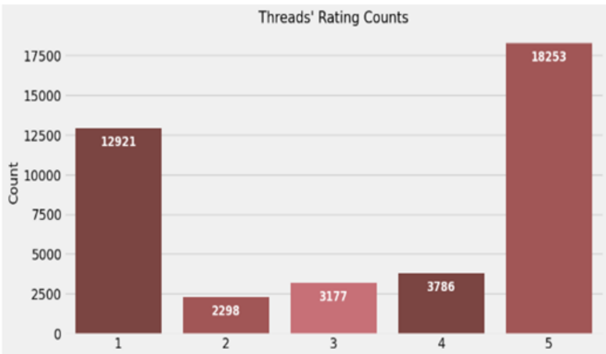

Fig 2-Snippet of Threads Rating counts all across the world

The above fig.2 shows the chart of the ratings of threads app all across the world.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

EDA:-

EDA (Exploratory Data Analysis) is a method in data analysis which helps to understand the patterns (maybe, occurring repetitively) in the data sets.

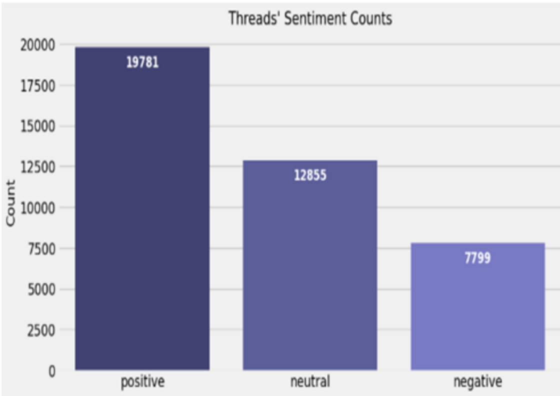

Fig 3- Snippet of Threads app's Sentiment counts

The above fig.3 shows the chart of the Threads App, depicting the sentiment counts as Positive, Negative Or Neutral, as rated by the users across the world.

Feature Extraction:-

The main method used in this study to transform text data into numerical vectors based on their frequency of occurrence in a document and throughout the dataset is the TF-IDF approach. Because it can handle massive volumes of text data, identify words and phrases within a text, and assign greater weight to unique phrases, the TF-IDF is useful for sentiment analysis. Because of its computational efficiency, it is a sensible option for handling large datasets.

Sentiment Analysis Classification:-

In machine learning, classifier training is an essential stage that includes training models on the available data and evaluating their performance. For classification tasks, Support Vector Machine (SVM), Linear Regression and Naïve Bayes classifiers are often utilized techniques.

First, the data must be preprocessed, which requires cleaning, converting, and getting it ready for training. After preprocessing, the SVM model classifier is trained using the processed data.

The SVM classifier discovers the best decision boundary between the various classes in the data during the training phase. By optimizing the margin between the support vectors—the data points nearest to the decision border—this decision boundary is established. On unseen data, the SVM classifier may produce precise predictions by determining the ideal decision boundary.

As soon as, the SVM classifier model is trained, the performance is evaluated using Performance Metrices. The metrics involves the calculation of Precision, Accuracy, Recall, F1 Score, and Overall Accuracy.

Similarly, Naïve Bayes model classifier is trained with the process data to determine the probabilistic classification algorithm on the texts, or reviews, to find the sentiment behind (whether positive, negative or neutral).

After the data is processed in Naïve Bayes classifier, it predicts the frequency of the words in a text with each sentiment class. Following the prediction of frequency of words, calculates the probability of each sentiment class on the basis of words present in the texts, and correspondence of the words in data preprocessing at the time of training data.

Evaluation:-

The section includes the study that utilizes the Confusion Matrix method to accurately measure each classification technique's capacity to correctly identify cases that belong in different categories with the aim to thoroughly assess its performance. The categories are named as TP (True Positive), TN (True Negative), FP(False Positive), FN(False Negative).

A. Accuracy:-

The percentage of all classifications—whether positive or negative—that were accurate is known as accuracy. It is calculated as:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

B. **Precision:-**

The precision of a model is the percentage of all positive classifications that are truly positive.

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

C. Recall:-

Recall is a metric that quantifies the frequency with which it properly detects positive occurrences, or true positives, out of all the actual positive samples in the dataset.

$$\textbf{Recall} = \frac{TP}{TP + FN}$$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

IV. RESULT

In this research study, the threads app data was collected, cleaned, modeled, and prepared for analysis as part of an initial step. Modeling the text by fixing grammar and deleting icons, as well as eliminating noise like URLs, hashtags, and reviews written in languages other than English, prepares the data for extraction.

Use TF-IDF to extract features and transform data into number vectors suitable for machine learning models, creating a solid foundation for emotional analysis.

Fig 4- Confusion Matrix

Fig 5-Word Cloud of reviews of Threads App

Use TF-IDF to extract features and transform data into number vectors suitable for machine learning models, creating a solid foundation for emotional analysis.

Classifier models such as SVM (Support Vector Machine) and Naïve Bayes Model were used in the occurring research study. As a result, both the model were found effective. Following the study, Measurements such as Accuracy, Precision, Recall are considered to have strong capabilities. The classifier's performance is further optimized and its efficacy in prediction theory is increased through hyper parameter tuning. The results demonstrate how well these categories may categorize the concepts presented in reviews, enabling decision-makers to make rational decisions on various kinds of topics.

Important details regarding the distribution of emotions in the data are revealed by information collected via data analysis (EDA). By exposing common themes and content in expressing various points of view (reviews), a keyword analysis of each position adds more context. This insights helps to deeper understanding of the sentiment of the social media such as Threads and highlights the importance of the case study of sentiment analysis.

All things considered, the study's findings offer concepts and perceptions on analytical reasoning, paving the way for data-driven choices and implementation across a number of fields.

V. CONCLUSION

In summary, sentiment analysis utilizing machine learning and natural language processing techniques shows how well these approaches can categories and forecast sentiments in Threads data. It provides a strong basis for the classifiers' training and assessment. The study emphasized how crucial it is to perform data preprocessing, such as feature extraction and cleaning, to guarantee the accuracy and applicability of the input data. The application of Naīve Bayes, and Support Vector Machine, classifiers demonstrated how well they handled sentiment analysis tasks. The capacity of these classifiers to produce accurate predictions was demonstrated by a thorough evaluation of each one utilizing performance metrics like accuracy, precision, recall.Further optimizing the classifiers' performance through hyper parameter tuning highlights how crucial it is to fine-tune machine learning models for better outcomes.The results of this study highlight the considerable promise of sentiment analysis across a range of applications, such as business, politics, crisis management, and public health. Through the examination of social media data, organizations can acquire important insights into public sentiment, emerging trends, and societal challenges, thus facilitating data-informed decision-making in various sectors.

VI. FUTURE SCOPE

In order to better comprehend the subtitles of text messages, including slang, abbreviations, and emoticons, future emotional analysis study could explore the preparatory level. In Machine Learning, investigating approaches for dealing with non-English languages would enhance the relevance of emotional models in various languages.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 14, Issue 4, April 2025) *DOI: 10.54380/IJRDET0425_01*

The idea of analytics could be greatly expanded by extending it beyond text to other formats present in Threads (such as pictures, videos, or user metadata).

By taking into account both textual and visual material, multimodal analysis techniques can offer a more thorough comprehension of emotional thinking and produce more precise thought classifications. Furthermore, using multimodal data can provide a fresh approach to investigating the connections between various patterns and feelings, leading to a better comprehension of user behaviour and social networking sites like Threads.

One intriguing area for further research is creating a theory of instantaneous conversational systems that can handle Threads as they are published. By identifying new trends, shifts in public sentiment, or urgent situations, instant analysis enables stakeholders to respond and make well-informed decisions. This process's implementation calls for effective and efficient algorithms that can handle massive amounts of data in a quick, complex, and user-friendly manner, opening up possibilities for the development of real-time communication.

REFERENCES

[1] Instagram. (2023). Introducing threads: A new way to share with text.

[2] Statista. (2024). Number of threads app downloads worldwide from July 2023 to March 2024.

[3] Jay, G. (2023). Threads (By Instagram) usage statistics for 2024.

[3] Drapkin, A. (2023). As advertisers flee X, threads quietly makes a comeback.

[4] Sentiment Analysis on User Reviews of Threads Applications in Indonesia(2024), Vol. 57, No. 4, August, 2024, pp. 1165-1171

[5] Caliskan, A., Bryson, J. J. & Narayanan, A., 2017. Semantics derived automatically from language corpora contain human-like biases. *Science,* 356(6334), pp. 183-186.

[6] Hamid, , . N. & Sjarif, . N., n.d. Handwritten Recognition Using SVM, KNN and Neural Network. p. 2017/02/01.

[7] Cambria, E., Schuller, B., Xia, Y. & Havasi, C., 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems,* 28(2), pp. 15-21.

[8] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval , 2* (1-2), 1-135.

[9] Huang, Y. H., Singh, V. K. & Ma, Y., 2019. Multimodal sentiment analysis using hierarchical fusion with context modeling. *IEEE Transactions on Multimedia,* 21(7), pp. 1785-1798.

[10] Harmandini, K. K., 2024/03/31. Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment. *Sinkron,* Volume 8, pp. 929-937.

[11] Hossin, . M. & M.N, . S., 2015/03/31. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process,* Volume 5, pp. 01-11.

[12] Jia, W. M. J. S., 2022/06/01. Feature dimensionality reduction: a review. *Complex & Intelligent Systems,* 8(3), pp. 2663-2693.

[13] Wankhade, M., rao, A. & Kulkarni, C., 2022. A survey on sentiment analysis methods, applications,and challenges.. *Artif Intell Rev ,* 55(7), p. 5731–5780.

[14] Wonjae , L. & Kangwon , S., 2022. Downsampling for Binary Classification with a Highly Imbalanced Dataset Using Active Learning. *Big Data Research,* Volume 28.

[15] Zhou, X., Zhu, M., Li, M. & Zhao, X., 2020. A review of sentiment analysis research in Chinese language. *Artificial Intelligence Review,* 53(1), pp. 457-486.

[16] Li, J. Y. K. M. A. A. S. K. N. I. D., 2024/01/10. Advances in Sentiment Analysis - Techniques, Applications, and Challenges.

[17] Liu, B., 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies,* 5(1), pp. 1-167.

[18] Madyatmadja, E.D., Shinta, Susanti, D., Anggreani, F., Sembiring, D.J.M. (2022). Sentiment analysis on user reviews of mutual fund applications. Journal of Computer Science, 18(10): 885-895.

[18] Liu, C., Sheng, Y., Wei, Z., Yang, Y. (2018). Research of text classification based on improved TF-IDF Algorithm.

Kadhim, A. (2018). An evaluation of preprocessing techniques for text classification.

[19] Guia, M., Silva, R., Bernardino, J. (2019). Comparison of Naïve Bayes, support vector machine, decision trees and random forest on sentiment analysis.

[20] Powers, D.M.W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation.

[21] Scikit. (n.d.). Supervised Learning, Support Vector Machines.

[22] Social Media Sentiment Analysis: A Comprehensive Analysis(2024),DOI:10.13140/RG.2.2.31094.37441