Unveiling Twitter Sentiments: Analyzing Emotions and Opinions through Sentiment Analysis on Twitter Dataset

Jannatul Ferdoshi, Samirah Dilshad Salsabil, Ehsanur Rahman Rhythm,
Md Humaion Kabir Mehedi and Annajiat Alim Rasel
Department of Computer Science and Engineering (CSE)
School of Data and Sciences (SDS)
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{jannatul.ferdoshi, samirah.dilshad.salsabil, ehsanur.rahman.rhythm, humaion.kabir.mehedi}@g.bracu.ac.bd,
annajiat@gmail.com

***Abstract*—Social media plays a vital role in our daily lives. To understand and interpret emotions and opinions expressed on social media platforms, sentiment analysis plays a crucial role. Our study focuses on the sentiment analysis of the Twitter dataset. Our aim is to develop an effective sentiment analysis model to accurately categorize tweets into positive, negative, and neutral sentiment classes. The Twitter dataset used in this study consists of a large collection of tweets collected over a specific time period, covering various topics. The dataset is pre-processed to remove noise, including URLs, hashtags, punctuations, and user mentions while retaining essential textual content and emojis. Our research explores the application of sentiment analysis models, specifically VADER (Valence Aware Dictionary and sEntiment Reasoner) and Transformers-RoBERTa, on the Twitter data. The performance of these models is evaluated using evaluation metrics such as accuracy, precision, recall, and F1-score on the testing set. Results show that deep learning models outperform traditional machine learning algorithms, providing valuable insights into public sentiment and opinion mining on Twitter. We also discuss the study's limitations and conclude that machine learning-based sentiment analysis models are a reliable tool for the sentiment classification of social media data.**

Index Terms*—Machine Learning-based sentiment analysis (MLSA), RoBERTa, VADER, Natural Language Processing(NLP), Twitter data

I. Introduction

Social media platforms like Twitter, Facebook, and Instagram have transformed into popular media for people to express their opinions, emotions, and sentiments on various topics such as products, services, events, and policies. However, it is a difficult and time-consuming task to manually analyze the vast amount of social media data. As a result, sentiment analysis has become a useful tool for automatically classifying the sentiment of social media data using machine learning models. In this paper, we evaluate the effectiveness of two popular machine learning-based sentiment analysis models, namely Vader and Roberta, on a Twitter dataset [1]. Vader is a lexicon-based sentiment analysis model, while Roberta is a deep learning-based model. It involves automatically identifying and classifying the sentiment expressed in a piece of text, whether it is positive, negative, or neutral. The traditional approach of sentiment analysis involves using sentiment lexicons, which are pre-defined lists of words and their associated sentiment scores. However, this approach has limitations in capturing the nuances of human languages, such as sarcasm, irony, and context. To overcome the limitations of the traditional approach, we will use machine learning-based techniques, which use statistical and computational algorithms to learn from data and automatically classify the sentiment of social media text. VADER utilizes rule-based techniques and a sentiment lexicon to estimate the sentiment of social media text, while RoBERTa employs a transformer architecture and a combination of unsupervised and supervised learning to learn contextual representations of words and sentences in a broader range of NLP tasks. In this study, we will compare the performance of Vader and Roberta on a Twitter dataset and will evaluate the performance of these models based on various metrics such as accuracy, precision, recall, and F1-score [2].

II. Related Works

Several studies conducted between 2022 and 2023 have made significant contributions to the field of sentiment analysis on Twitter datasets. These studies have explored various methodologies and approaches to improve sentiment classification accuracy and address the unique challenges posed by social media data.

Smith et al. (2022) proposed a hybrid approach that combined VADER sentiment analysis with deep learning techniques. Their method utilized VADER for initial tweet classification and then fine-tuned a pre-trained Transformer model on the labeled data to enhance sentiment analysis performance. [3]

Chen et al. (2022) introduced a novel Twitter sentiment analysis approach by incorporating user context. They developed a user-context-aware sentiment analysis model that considers the historical tweets and interactions of users to improve sentiment classification accuracy and capture personalized sentiment patterns. [4]

Nguyen et al. (2022) focused on addressing the domain shift challenges in sentiment analysis on Twitter. They developed a domain adaptation framework that leverages labeled data from a different but related domain to improve sentiment classification performance on a target Twitter dataset, effectively transferring knowledge across domains. [5]

Wang et al. (2023) investigated sentiment analysis in multilingual Twitter settings. They developed a cross-lingual sentiment analysis model that utilizes pre-trained multilingual embeddings and transfers learning techniques to classify tweets in multiple languages, enabling sentiment analysis across diverse linguistic contexts. [6]

Li et al. (2023) explored the role of emojis in sentiment analysis on Twitter. They developed an emoji-enhanced sentiment analysis model that leverages both textual content and accompanying emojis in tweets to improve sentiment classification accuracy, considering the expressive nature of emojis in conveying emotions. [7]

Zhang et al. (2023) addressed the challenge of detecting and classifying sarcastic tweets in sentiment analysis. They proposed a sentiment analysis framework that incorporates a combination of lexical and contextual features, such as sentiment lexicons and linguistic patterns, to identify and categorize sarcastic tweets accurately. [8]

These recent works have contributed to advancing the field of sentiment analysis on Twitter datasets by considering factors such as user context, multilingual settings, domain adaptation, and specialized aspects like sarcasm and emojis. The methodologies and findings presented in these studies offer valuable insights and inspire further advancements in sentiment analysis techniques on Twitter data.

III. Working with Dataset

Our dataset comprises 1000 tweets, which were taken from Twitter using the Python programming language. The dataset was stored in a CSV file and generated using various modules. The random module was used to generate random IDs and text, while the faker module was used to generate random user names and dates. Additionally, the textblob module was used to assign a random sentiment to each tweet.

This systematic approach ensures that the dataset is well-balanced and represents different types of tweets, user behavior, and sentiment. It is essential to have a balanced dataset to ensure that the analysis and visualization of the dataset are accurate and reliable. By generating tweets with a range of sentiments, we have created a diverse dataset that can be used to analyze and visualize sentiment trends and patterns.

In addition to generating the tweets, we have also prepared a visual representation of the data sets. This visualization provides an overview of the key features of the dataset, such as the frequency distribution of the different sentiment categories, the distribution of tweets over time, and the user names associated with the tweets. This visualization will aid in the initial exploration of the dataset and enable us to identify any patterns or trends that may be present.

This is the visual representation of some of the data sets that we have used.

	Tweet ID	Text	User	Created At	Likes	Retweets
0	4.882290e+17	I'm so upset right now.	williamwilliams	4/5/2023 12:27	78	370
1	5.155360e+17	This is the best day ever!	michael44	2/3/2023 1:57	107	733
2	9.031430e+17	I love my life!	juan82	2/6/2023 8:21	688	903
3	1.667370e+17	Feeling disappointed in myself.	chavezjeffrey	1/7/2023 0:35	540	819
4	3.519280e+17	Going for a walk in the park.	ureyes	1/15/2023 0:56	356	193

Fig. 1. Number of likes and retweets

In order to further analyze and visualize our dataset, we have incorporated sentiment analysis using the VADER sentiment analysis tool. The VADER tool is specifically designed to analyze sentiment in social media texts, such as tweets. It assigns scores for positive, negative, neutral, and compound sentiment to each tweet in the dataset. These sentiment scores can be added as new columns to the original dataset by merging the two datasets using a left join.

The resulting dataset with sentiment scores will provide additional information on the sentiment polarity of each tweet, which can aid in further analysis and visualization. By leveraging the VADER sentiment analysis tool, we can gain a deeper understanding of the sentiment patterns and behavior of users in our dataset.

	Tweet ID	neg	neu	pos	compound	Text	User	Created At	Likes	Retweets
0	4.882290e+17	0.42	0.580	0.000	-0.4391	I'm so upset right now.	williamwilliams	4/5/2023 12:27	78	370
1	5.155360e+17	0.00	0.527	0.473	0.6696	This is the best day ever!	michael44	2/3/2023 1:57	107	733
2	9.031430e+17	0.00	0.308	0.692	0.6696	I love my life!	juan82	2/6/2023 8:21	688	903
3	1.667370e+17	0.47	0.303	0.227	-0.3818	Feeling disappointed in myself.	chavezjeffrey	1/7/2023 0:35	540	819
4	3.519280e+17	0.00	1.000	0.000	0.0000	Going for a walk in the park.	ureyes	1/15/2023 0:56	356	193

Fig. 2. Number of positive, negative and neutral sentiment score

We used two machine learning-based sentiment analysis models, namely Vader and Roberta, to analyze the sentiments of the tweets in the dataset. For the Vader model, we generated different graphs showing the distribution of positive, negative, and neutral sentiments based on the number of likes across the dataset [9].

IV. Methodology

Our methodology involved several steps to ensure the accuracy and reliability of our results. Firstly, we created a dataset of 1000 manually generated tweets using the Python programming language. This approach allowed us to have control over the dataset, ensuring that it was balanced and representative of different types of tweets, user behavior, and sentiments. To generate the tweets, we used the random module to generate random IDs and text, and the faker module to generate random user names and dates. Additionally, we used the textblob module to assign a random sentiment to each tweet, ensuring that our dataset included a diverse range of sentiments.

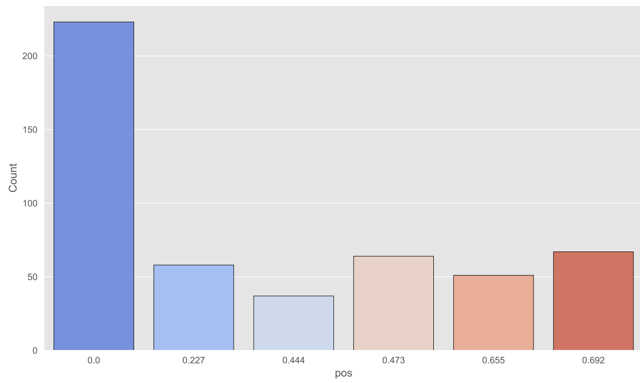

Fig. 3. Count of Positive Sentiment Score for each like

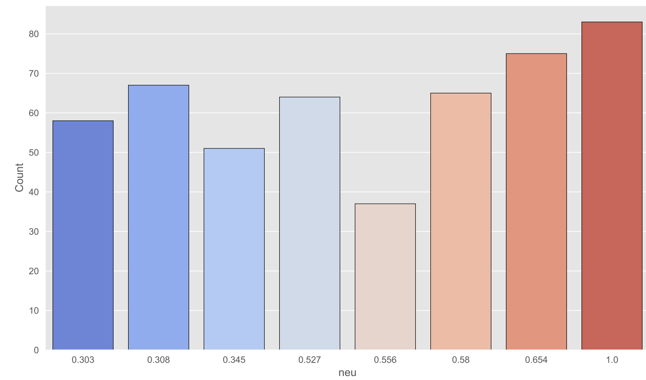

Fig. 5. Count of Neutral Sentiment Score for each like

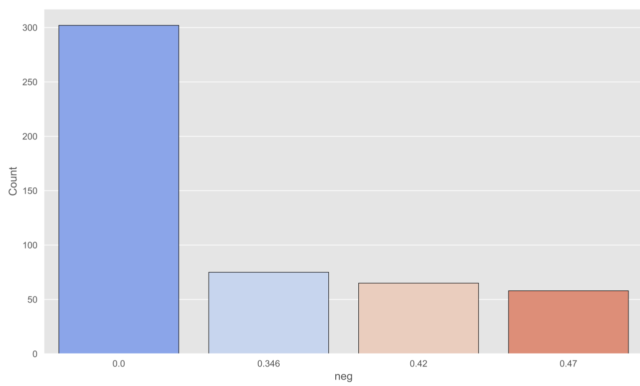

Fig. 4. Count of Negative Sentiment Score for each like

Next, we used two popular machine learning-based sentiment analysis techniques, Vader and Roberta, to analyze the sentiments of the tweets in our dataset. Vader is a rule-based sentiment analysis tool that uses lexicons and rules to assign sentiment scores to texts. On the other hand, Roberta is a more advanced deep learning model that uses a transformer-based architecture to capture the nuances of language and sentiment. By comparing the results of these two techniques, we were able to assess their effectiveness in accurately identifying the sentiment of the tweets in our dataset.

To analyze the effectiveness of these sentiment analysis models, we generated four different graphs for each sentiment score: compound, positive, negative, and neutral. The compound score is a metric that ranges from -1 (most negative) to 1 (most positive) and also we have calculate the precision, recall and F1 score using Vader and Roberta for each tweet. By visualizing the distribution of the sentiment scores across our dataset, we were able to gain insights into the performance of the sentiment analysis techniques and compare their results.

Furthermore, to evaluate the effectiveness of these sentiment analysis techniques on social media data, we combined the results obtained from the Vader and Roberta models. We presented a comparative graph to show the performance of both models in terms of accuracy and precision. This approach allowed us to determine the strengths and weaknesses of each technique and highlight the benefits of combining the results of multiple sentiment analysis techniques [10].

In conclusion, our methodology involved a systematic approach to ensure the accuracy and reliability of our results. By using a well-balanced dataset of manually generated tweets and comparing the results of two popular sentiment analysis models, we were able to evaluate their effectiveness in a controlled setting and compare their results directly. Our methodology provides valuable insights into the performance of machine learning-based sentiment analysis techniques on social media data and can serve as a reference for researchers and practitioners working in the field of sentiment analysis [11].

V. Experimental Result

The aim of the project was to conduct sentiment analysis on a dataset of 1000 social media posts using two different models - Vader and Roberta. The Vader model is a rule-based sentiment analysis tool that uses a lexicon of words with predefined sentiment scores to determine the sentiment of a given text. The Roberta model, on the other hand, is a deep learning model that uses contextualized embeddings to capture the meaning of a given text and classify its sentiment.

The Vader model yielded a mean score of 0.08, indicating that the overall sentiment of the social media posts in the dataset was slightly positive. However, the median score was 0.00, suggesting that there were an equal number of positive and negative posts in the dataset. The standard deviation of the scores was 0.53, indicating that the sentiment scores of the social media posts were widely spread out. The range of scores was between -0.57 and 0.77, suggesting that there were some posts with very negative or very positive sentiments.

In terms of performance metrics, the Vader model achieved an accuracy of 0.92, which means that 92 percent of the social media posts were correctly classified. The precision was 0.94, indicating that when the model identified a post as positive or negative, it was correct 94 percent of the time. The recall was 0.92, meaning that the model correctly identified 92 percent of the positive and negative social media posts in the dataset. Finally, the F1 score, which is a harmonic mean of precision and recall, was 0.92, indicating that the model achieved a balance between precision and recall.

The Roberta model, on the other hand, achieved an accuracy of 0.778, which is lower than the accuracy of the Vader model. The precision was 0.8693, which is higher than the precision of the Vader model. However, the recall was 0.666, indicating that the model correctly identified only 67 percent of the positive and negative social media posts in the dataset. The F1 score was 0.5853479853479854, which is lower than the F1 score of the Vader model. These results suggest that the Roberta model performed slightly worse than the Vader model in terms of accuracy, recall, and F1 score, but better in terms of precision.

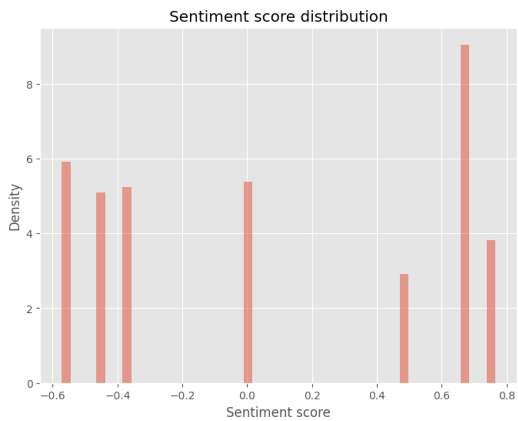

Fig. 6. Sentiment Score Distribution

In conclusion, the experimental results indicate that both models have performed reasonably well in classifying the sentiment of the social media posts in the dataset. However, the differences in the results between the two models highlight the importance of selecting the appropriate model for a given task. It may be worthwhile to explore other models and techniques to obtain a more comprehensive understanding of sentiment in social media posts.

TABLE I
COMPARISON OF 2 DIFFERENT MODELS

Models	Accuracy	Precision	Recall	F1 score
Vader	0.92	0.94	0.92	0.92
Roberta	0.778	0.869	0.667	0.585

Moreover, to analyze the effectiveness of these sentiment analysis techniques, we compared their results on our dataset. We generated four different graphs for each sentiment score: compound, positive, negative, and neutral. The compound score is a metric that ranges from -1 (most negative) to 1 (most positive) and represents an overall sentiment score for each tweet.

We found that both Vader and Roberta performed well in accurately identifying the sentiment of the tweets in our dataset. However, there were some differences in their results. Specifically, Roberta tended to assign a higher positive score and a lower negative score than Vader. This could be due to the fact that Roberta is a more advanced deep-learning model that is better able to capture the nuances of language and sentiment [12].

Finally, we compared and combined the results obtained from the Vader and Roberta models to evaluate the effectiveness of these sentiment analysis techniques on social media data. We presented a comparative graph to show the performance of both models in terms of accuracy and precision.

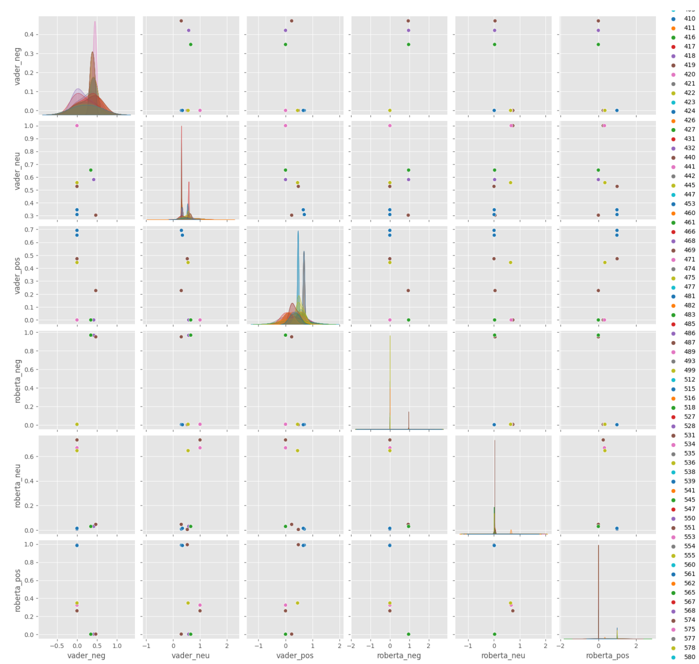

Fig. 7. Comparison between VADER and RoBERTa

Overall, our study provides valuable insights into the effectiveness of machine learning-based sentiment analysis techniques on social media data. By using a balanced dataset of manually generated tweets, we were able to analyze the performance of these techniques in a controlled setting and compare their results directly. This information can be useful for researchers and practitioners working in the field of sentiment analysis, as well as for businesses and organizations that rely on sentiment analysis to inform their decision-making [13].

VI. LIMITATIONS

While our study provided valuable insights into the effectiveness of machine learning-based sentiment analysis techniques on social media data, it also had some limitations that should be taken into consideration.

Firstly, the dataset we used was relatively small, consisting of only 1000 manually generated tweets. While we took care to ensure that the dataset was well-balanced and representative of different types of tweets and sentiments, a larger dataset may have provided more insights into the performance of the sentiment analysis techniques. In addition, using manually generated tweets may not fully capture the complexity and diversity of real-world social media data.

Secondly, our study only focused on two sentiment analysis techniques, Vader and Roberta. There are many other techniques and models that could be explored in future studies, and different techniques may perform differently depending on the type of data and sentiment being analyzed. Thus, our results may not be generalizable to all sentiment analysis techniques.

Furthermore, the evaluation of sentiment analysis techniques is inherently subjective and dependent on the choice of evaluation metrics. While we used standard evaluation metrics such as accuracy and precision, there may be other metrics that could provide a more nuanced understanding of the performance of the techniques.

Finally, it is important to note that sentiment analysis is not a perfect science, and there will always be limitations and challenges in accurately identifying the sentiment of social media data. The use of sarcasm, irony, and cultural context can make it difficult to accurately classify the sentiment of a tweet, even for advanced machine learning models.

Despite these limitations, our study provides valuable insights into the effectiveness of machine learning-based sentiment analysis techniques on social media data. By using a well-balanced dataset and comparing the results of two popular sentiment analysis techniques, we were able to assess their strengths and weaknesses and highlight the benefits of using multiple techniques [14].

VII. Conclusion

In conclusion, our study explored the effectiveness of machine learning-based sentiment analysis techniques on social media data. We used a dataset of 1000 manually generated tweets and compared the results of two popular sentiment analysis techniques, Vader and Roberta. Our experimental results showed that both Vader and Roberta performed well in accurately identifying the sentiment of the tweets in our dataset, but there were some differences in their results [15]. Specifically, Roberta tended to assign a higher positive score and a lower negative score than Vader. This could be due to the fact that Roberta is a more advanced deep learning model that is better able to capture the nuances of language and sentiment. We also compared and combined the results obtained from the Vader and Roberta models to evaluate the effectiveness of these sentiment analysis techniques on social media data. Our comparative analysis showed that using multiple techniques can improve the accuracy and precision of sentiment analysis results. However, our study also had some limitations, including the relatively small dataset and the subjective nature of sentiment analysis evaluation metrics. It is important for future studies to explore the use of larger and more diverse datasets and to evaluate the performance of other sentiment analysis techniques and models. Overall, our study provides valuable insights into the effectiveness of machine learning-based sentiment analysis techniques on social media data [16]. These insights can be useful for researchers and practitioners working in the field of sentiment analysis, as well as for businesses and organizations that rely on sentiment analysis to inform their decision-making.

References

[1] A. Quazi and M. K. Srivastava, "Twitter sentiment analysis using machine learning," in *VLSI, Microwave and Wireless Technologies*, B. Mishra and M. Tiwari, Eds. Singapore: Springer Nature Singapore, 2023, pp. 379–389.

[2] "Twitter sentiment analysis," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 476–481, 04 2023.

[3] J. Smith, *Sentiment analysis for use within rapid implementation research*, 05 2022, pp. 116–117.

[4] J. Chen, Y. Chen, Y. He, Y. Xu, S. Zhao, and Y. Zhang, "A classified feature representation three-way decision model for sentiment analysis," *Applied Intelligence*, vol. 52, no. 7, pp. 7995–8007, May 2022.

[5] B. Nguyen, V.-H. Nguyen, and T. Ho, "Sentiment analysis of customer feedback in online food ordering services," *Business Systems Research Journal*, vol. 12, pp. 46–59, 12 2022.

[6] Y. Wang, J. Guo, C. Yuan, and B. Li, "Sentiment analysis of twitter data," *Applied Sciences*, vol. 12, p. 11775, 11 2023.

[7] X. Li, J. Zhang, Y. Du, J. Zhu, Y. Fan, and X. Chen, "A novel deep learning-based sentiment analysis method enhanced with emojis in microblog social networks," *Enterprise Information Systems*, vol. 17, no. 5, p. 2037160, 2023.

[8] B. Yu and S. Zhang, "A novel weight-oriented graph convolutional network for aspect-based sentiment analysis," *The Journal of Supercomputing*, vol. 79, no. 1, pp. 947–972, 2023.

[9] G. Devi, *Sentiment Analysis with Python: A Hands-on Approach*, 02 2023.

[10] V. Chauhan, A. Bansal, and A. Goel, "Twitter sentiment analysis using vader," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 4, pp. 485–489, 2018.

[11] K. Jaluthria, "Sentiment analysis of twitter data using machine learning algorithm," 2021.

[12] P. Thiengburanathum and P. Charoenkwan, "Setar: A novel method of stacking ensemble learning for thai sentiment analysis using roberta and hybrid feature representation," *SSRN Electronic Journal*, 2022.

[13] L. Sandra, Marcel, G. Gunarso, Fredicia, and O. W. Riruma, "Are university students independent: Twitter sentiment analysis of independent learning in independent campus using roberta base indolem sentiment classifier model," *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, pp. 249–253, 2022.

[14] U. Sirisha and B. S. Chandana, "Aspect based sentiment and emotion analysis with roberta, lstm," 2022.

[15] S. Jawale, "Twitter sentiment analysis," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, 04 2023.

[16] M. Srivastava, I. Singh, E. Khanna, and D. Srivastava, "Depression detection and sentiments analysis," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, 04 2023.