Grenze International Journal of Engineering and Technology, June Issue

A Novel Approach to Optimizing Social Media Marketing Strategies using NLP and DL

Dr. Ch. Rajendra Babu[1], Basavaboyina Sai Keerthi[2], Madugula Amani[3] and K. Rahul Praneeth[4]
[1-4]Lakireddy Bali Reddy College of Engineering Department of Artificial Intelligence and Data Science Mylavaram, Andhra Pradesh, India
Email: chikkalarajendra@lbrce.ac.in, keerthi.basavaboyina@gmail.com, madugulaamani123@gmail.com, 21761a54b4@gmail.com

***Abstract*—Social media marketing, or SMM uses Facebook, YouTube, Twitter, Instagram, etc., rather than sales experience to sell products and services. Its real-time engagement, such as likes, comments and shares enhances the visibility & trust. Businesses can use important statistics to change their strategies and focus on target users better. Our work is a continuation, where this team has processed feedback from previous publications in social networks with the participation of influencers and further provide marketing managers more clear information about various types of reviews. This also helps them to refine their marketing strategy not just in who they choose as influencers for the next campaigns. This is achieved through the system analyzing previous interactions and trends, predicting who your main influencers are for each product has been historically helping you in future planning as to what should be pushed forward marketing wise.**

***Index Terms*— NLP, Sentiment Analysis, Deep Learning, Analysis, Prediction, VADER, LSTM, RNN.**

I. INTRODUCTION

Sentiment analysis is the study of text to figure out the emotional tone, that is whether it is positive, negative or neutral, or even the scale of emotion. Deep learning, which is a subset of machine learning, uses neural networks with layers to automatically learn complicated data patterns, which make it very useful in tasks such as sentiment analysis. Natural Language Processing (NLP) authorizes the machines to get and decode the human-made language which is the primary task for analyzing text data in social media posts. Deep learning (DL) models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are often used in NLP tasks to improve the accuracy of text and image sentiment analysis. It is NLP, which is a subset of artificial intelligence, that drives machines to understand human language, while deep learning technologies, like Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN),are great tools for analyzing sequential data, such as natural language texts. These technologies record elaborate language patterns, which is why they are so helpful in the tasks of sentiment analysis. The development of existing algorithms has a focus on text data from such platforms as YouTube and Instagram and the approach of sentiment analysis to measure public opinion and/or customer feedback. Sometimes, the models are trained only on text-based data, from which they can infer what to say in the marketing strategy, such as posts, comments, and hashtags. The groundwork of the prior approaches is an issue given that text analysis is mostly emphasized and they usually cannot deal with the multimedia

Grenze ID: 01.GIJET.11.2.568

content provided, like images and videos. This pretty much restricts sentiment analysis as well as the prediction of influencers, especially on platforms where visual content is a priority. In this endeavor, we go beyond the limitations of previous models that were primarily focused on the analysis of written speech by extending the scope to include multimedia content—text, images, and videos. Social networks are filled with all kinds of content, hence, sentence analysis that only consists of word data can superficially be missing the real knowledge that pictures and videos convey. The fusion of Natural Language Processing (NLP) and deep learning models such as LSTM and RNN with multilevel processing allows both textual and visual aspects of social posts to be taken into account when identifying the sentiment of users and the influence of influencers. NLP methods enable us to get the best scenes from the text comments. Moreover, we are training neural networks in order to learn on multimedia content (e.g., images and videos) and to recognize patterns and emotions that are not directly represented in the text. If we use data from both sources, our product can forecast which influencers have the highest chance to engage with a brand's target audience because of their involvement in multimedia content. Such an extensive solution enables marketing managers to make adverts in a way that resonates well with both the visual and written preferences of their audience also helping them to come up with modules for campaigns better. The fusion of text, images, and video data provides a more in-depth, nuanced understanding of influencer effectiveness, as a consequence, the predictions are more accurate and thus the product promotions with differently structured contents can be better tuned.

II. Related Work

This analysis used both machine learning and Data envelopment analysis for the analysis of the twitter data and its classification with efficiency ranking upon the chosen products and typologies of messaging that should be impactful to the messaging of furniture retail chosen as Stroes[1]. Online marketing trends allow access to global platforms on the same platform; many persons rely on social media, especially because they can freely access various applications. In the research focused developed a model for marketing purpose who required trend and influential profiles for their product lunch taking twitter dataset applies correlation from influential profiles[2]. Social media affects society to a great extent, and in teenagers most social media activities take place through mobile apps. Access is therefore made easy 22% of the youth use platforms 10 times a day. RSA works to mitigate the negative effects of social media while using its positive potential for social change. Marketing and boosting sales at low costs Social media enhances communication but raises misinformation and privacy issues [3]. This research took Facebook data, comments for extracting English and Malay words and found emotion based on the text that was extracted and clustered into emotions. This emotions are positive, negative and emotionless. In this paper a lexicon-based approach to analyse sentiments in Facebook comments.it is helping stockholders monitor discussions and enhance their services.[4]. This research investigates how integrating search engine marketing (SEA) with unsupervised machine learning brings insights into consumer purchase behavior influenced by social networks in this research taking survey with response is 98.1%.and also used hierarchical and Kmeans unsupervised algorithms to cluster formation.[5]This paper investigates social media data analytics through machine learning, focusing on the Waikato Environment for Knowledge Analysis (WEKA). The results indicate that WEKA has surpassed all the other algorithms in precision, recall, and F-measure with J48 and PART reaching about 96.72% accuracy. This study analyzes the visually incongruent content of brand- related posts on Instagram and develops a computer vision and machine learning model to detect image, text, and metadata. This end therefore highlights the efficiency of deep learning in detecting mismatches and shows that these techniques are superior to others for research purposes in advertising. [7]. This study uses the Systematic Literature Review for the analysis of the use of social media in e- commerce through different industries and countries with a lexicon-based approach as the most commonly used in big data analysis. This research would therefore shed light on how companies employ the use of social media to strengthen their customers' trust as well as their engagement.

III. Methodology

A. Dataset Information

This data is derived from Kaggle. This contains a pretty good amount of reviews of products available in various categories, clothing, sports, devices, etc. It's really very wide as a source for the collection of customer feedback with respect to social media marketing. Because the manager in marketing is concerned about the customer's sentiments that serve as a significant index for adjustment of areas within such services, this set becomes the ultimate source for establishing probable improvements. Once such areas are realized, this does an effective

adjustment regarding the selection of the influencer concerning how the promotion must match up with differences in preference of customers regarding product categories. By the attributes of review text and review ratings, we classify each review into the positive or negative sentiment group by utilizing a randomly generated rating scale of 1 to 5. Now we present the steps taken in processing and analyzing the review data.
Rating: It stores the rating given by the customer in a range of I to V.
Verified Reviews: The primary textual content of the review, where customers express their viewpoints and encounters regarding the product.

	review_text	review_rating
0	Works as it claims. Could see the difference f...	5.0
1	It does what it claims . Best thing is it smoo...	5.0
2	I have been using this product for months now....	4.0
3	i have an oily skin, while this whip acts as a...	3.0
4	It's not that good. Please refresh try for oth...	2.0

Fig.1. Dataset for the sports shoes.

Using the VADER [2](Valence Aware Dictionary and sentiment Reasoner) tool in natural language processing (NLP),[8] we have changed our dataset. By taking the values of the Sentiment scores we made the changes in our dataset. VADER, or Valence Aware Dictionary and Sentiment Reasoner, is a useful tool in the domain of natural language processing (NLP) that excels in sentiment analysis.Each text segment receives a polarity score from VADER, which indicates whether the sentiment is positive, negative, or neutral. In Movies category dataset we used Vader to map review rating using the text. The polarity score provided by VADER[2] ranges from -1 to +1, where: • Negative polarity in the text represents the scores closer to -1, indicating a negative sentiment. • Positive polarity in the text represents the scores closer to 1, indicating a positive sentiment. • Neutral Polarity in the text represents the scores closer to 0 indicating a neutral sentiment. By using the NLTK library[8], specifically the VADER sentiment analysis tool to analyze a dataset[1] of text reviews. Subsequently, a function is defined to convert sentiment scores into ratings. This function maps sentiment scores to a scale of 1 to 5(ratings). The loop iterates over each review.

TABLE I: TABLE DESCRIBING ABOUT THE VARIATIONS IN THE RATING ACCORDING TO THE SENTIMENT SCORES

Sentiment score	change in rating
>=0.5	5
>=0.3	4
>=0.1	3
>= -0.1	2
Other than those	1

And we as-well changed the values of feedback using the sentiment scores to get better accuracy for our model. For this a function is defined to convert sentiment scores into binary feedback, categorizing them as either positive or negative based on whether the sentiment score is greater than or equal to 0. The loop iterates over each review in the dataset, calculates its sentiment score using the sentiment analyzer, and then converts this score into binary feedback using the defined function.

TABLE II: TABLE DESCRIBING ABOUT THE VARIATIONS IN THE FEEDBACK ACCORDING TO THE SENTIMENT SCORES

Sentiment score	feedback
>=0	Positive (1)
<0	Negative(0)

Before analysis, the dataset underwent many preprocessing [14] steps measures to guarantee data performance and quality and consistency. These steps involved:

B. Data Preparation

We then included two new columns; review_len: captures the length of each review text. review_word_count: calculates the number of words in each review. We cleaned the review text by removing special characters like punctuation using regular expressions, focusing on meaningful words for analysis.

C. What is NLP6?

Natural Language Processing (NLP) is a subarea of computer science and AI that enables computers to process, analyze, and generate human language. Through understanding the structure and context of text, NLP derives insights and aids in the development of intelligent systems for natural communication.

a)Word Tokenization:- Tokenizing words splits text into words or tokens, allowing the analysis of words in NLP. Calculating sentiment scores, then applying a function to convert them into ratings and then updating to get new ratings.

b)Stemming:- it is a text normalization technique that reduces words to their root forms by removing prefixes and suffixes.

c)Lemmatization:- Lemmatization converts words to their dictionary forms, known as lemmas, by taking into consideration the context and linguistic analysis, unlike stemming, which often results in a valid word, thus helping to standardize.

d)Stop Word Analysis:- Stop words are everyday words that come up too often in any language. Often, stop words bring very little, if any, semantic value. Their identification and exclusion from a given text would be needed for further processing. Hereby, through removal of these stop words-"the," "and," and "is"-NLP models might pay more attention to meaningfully used words and might have a more accurate analysis on tasks like sentiment and text classification.

D. Data Exploration

When analyzing product reviews, several important pre-analysis and analysis steps are performed to remove the found content from the data. This includes cleaning the data to remove text and stop words from the review so that the dataset can be optimized, especially removing words like “and” or “o” that do not have a specific opinion or information about the product.

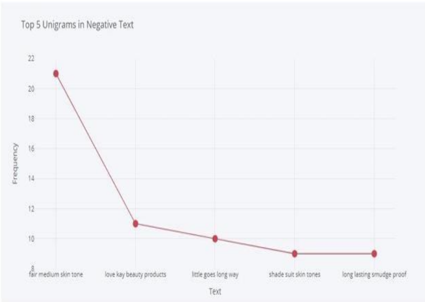

Fig.2.Top 5 negative unigrams in the reviews

These metrics can be a good indicator of whether a longer review is associated with a particular opinion or rating. See. We use the “CountVectorizer” to extract the most common ones from each group of sentiments. From here, we see that each group has a specific language that provides a better understanding of behavior and experience. We used Cufflinks to draw two graphs showing the top row of positive and negative reviews, and we could see the content that customers submitted for each opinion. This step provides insight into consumer sentiment and valuable information that can guide product development, marketing strategies.

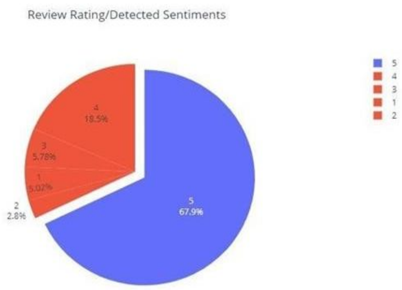

Fig.3. Review rating vs Sentiment

Word Cloud:- We generate a word cloud visualization for reviews with negative sentiment in a dataset. What it does first is to use the function df.query() that allows us to filter the dataset, so we are filtering our dataset where sentiment is equal to neg. After that, it merges the cleaned reviews into one string negative text. Finally, the WordCloud library is used to generate a word cloud image (i.e. words sized by their number of incidences in the negative reviews). We plot the generated word cloud using Matplotlib without any axes for clarity. This lets us know which words are present the most when it comes to negative feedback.

Fig.4. Positive word cloud of a beauty product

Then we created a word cloud to visualize the most common words in positive reviews from a dataset. It extracts reviews marked with positive sentiment and combines them into a single text. A word cloud is then generated, where the size of each word reflects its frequency in the positive reviews. This provides a clear and engaging way to identify key themes or frequently mentioned terms in the feedback.

E. Predicting the Influencer

After the sentiment analysis, we begun predicting suitable influencers based on criteria such as the target country, budget, and product category. We used data from the Instagram and YouTube platforms in doing this. The two platforms give all information about influencers, including demographics of audiences and categories of content uploaded. Looking at TikTok first, the dataset was not detailed with key information, especially about the category and audience country. This gap posed a problem, since recommendations based on these criteria are necessary for businesses to identify influencers in their space and target audience. Therefore, we omitted TikTok from our consideration for the influencing campaign. Because we were working with Instagram and YouTube, our recommendations were far more precise and impactful in nature, since we could match the profiles of influencers with the business' budget, desired geographical footprint, and product category. This fine-tuned approach makes the recommendations of influencers more relevant, and more suitable for strategic goals that define targeted marketing campaigns. We then computed for each influencer the Engagement Rate (ER) after filtering the datasets from Instagram and YouTube to only include relevant columns. The Engagement Rate is the most important metric that computes the percentage of followers an influencer's content interacts with. This metric will help us identify how well the influencer connects with his or her audience and how effective his or her content is for promoting interactions.

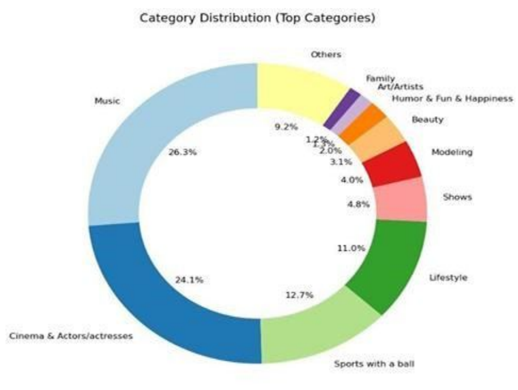

Fig.5 Frequency Distribution of Categories in the Dataset

Importance of Engagement Rate (ER) Audience Engagement: A high ER shows that the content of an influencer is well received by the audience, which means that followers actually like and engage with the posts. Content Reach: Higher engagement is likely to increase the chances of content being boosted in news feeds; therefore, a piece of content attracts more attention on the respective posts. The Engagement Rate is calculated through the following formula:

$$ER = (\text{Total Followers}/\text{Average Engagement}) \times 100\ . \quad (1)$$

This study employed two models—LSTM and RNN—to predict influencer engagement rates based on follower counts. The LSTM model achieved 77% accuracy by using 50 LSTM units, dropout for overfitting prevention, and a dense output layer for engagement rate prediction, with data preprocessed through normalization and reshaping. The RNN model, on the other hand, achieved 87% accuracy using a similar approach but with a 50-unit RNN layer and ReLU activation. Both models were trained for 100 epochs with a batch size of 32, and performance was evaluated using Mean Squared Error (MSE). The RNN showed slightly better performance, making it more effective for influencer selection in targeted marketing strategies.

F. Flask Application

This web application using Flask allows filtering and displaying of influencer data from CSVs in a tabular manner. Users can select influencers from either Instagram or YouTube and narrow results down by country, category, and follower type (mini or mega). Sorted data by engagement rate is then displayed with salient details such as name, platform, followers, and category. Data is loaded only once per session for performance, and when an invalid category is input, the app offers available choices. This provides a seamless and friendly experience for navigating influencer data.

IV. RESULTS

We investigated two distinct models as a part of our task on influencer engagement prediction: LSTM and RNN. In the case of an LSTM model, particularly designed with sequence data and long-term dependency in mind, we reach an accuracy of 77%. Such an accuracy also implies that the model can be actually capturing the trend over time, though further tuning on the dataset will be important for fine-tuning with our specific data. On the contrary, the RNN model gave an accuracy of 87%.

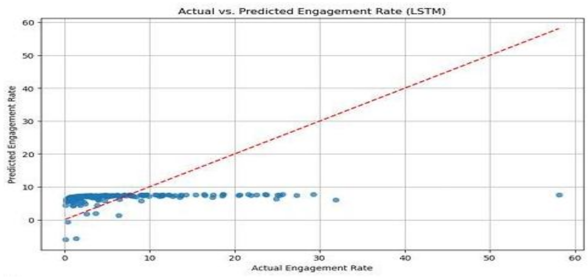

Fig.6. Actual vs. predicted rate.

This means that for this experiment, the simple architecture of RNN was more preferable for our purpose of prediction of engagement rates based on follower counts. The RNN accuracy shows it's capable of representing patterns within our data without the complexity overhead of managing long dependencies due to the fact that the follower count and the engagement rate have a more direct relation without extensive sequential dependency. This comparison shows that LSTM models are powerful for time series tasks, but sometimes more straightforward numerical relationships can result in better performance with simpler RNNs.

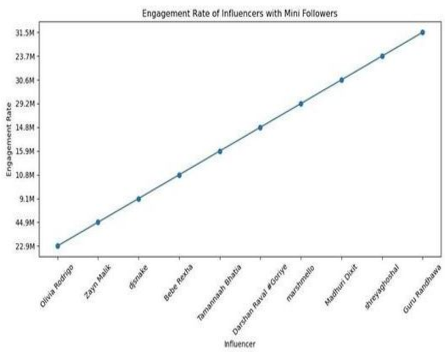

Fig.7. Itpredicted mini influencers data

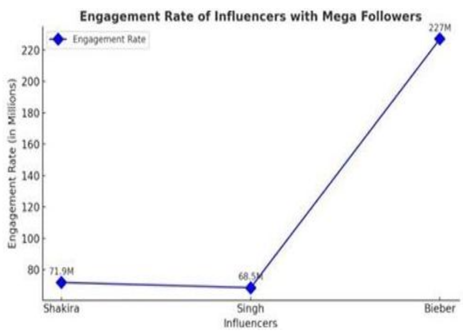

Fig.8. It predicted mega influencers data.

V. Conclusion and Future Scope

In conclusion, this work adds great value to marketing managers by providing data-driven decisions for product performance optimization. The ability to predict influencer engagement rates based on follower count has given the system a strategic choice of influencers according to the budget set for the product, its category, and the target country. The campaign will be more impactful and less expensive. In addition, the work can be improved in the future by integrating real-time data and social media sites, further developing the selection of influencers and growing digital marketing strategies.

References

[1] Elif Kongar and Olumide Adebayo "Impact of Social Media Marketing on Business Performance: A Hybrid Performance Measurement Approach Using Data Analytics and Machine Learning" IEEE Engineering Management Review (Volume: 49, Issue: 1, Firstquarter, march 2021) 11 February 2021.

[2] Somya Jain; Puneet Jaiswal; Arpit Jain; Adwitiya Sinha; Ankita Gupta "Analyzing Influential Impact of Online Trends on Social Media Channels" : 2022 8th International Conference on Signal Processing and Communication (ICSC),IEEE 12 January 2023

[3] Duaa Ali Al-Sabti; Ajay Vikram Singh; S. K. Jha "Impact of social media on society in a large and specific to teenagers" 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) ,IEEE 20-22 September 2017.

[4] Nur Azmina Mohamad Zaman and Siti Zaleha Zainal Abidin "Sentiment Analysis: Determining People's Emotions in Facebook" Proceedings of the 13th International Conference on Applied Computer and Applied Computational Science, April 2014.

[5] Pejman Ebrahimi and Marjan Basirat "Social Networks Marketing and Consumer Purchase Behavior: The Combination of SEM and Unsupervised Machine Learning Approaches" Big Data and Cognitive Computing, March 2022.https://www.mdpi.com/2504-2289/6/2/35 marketing"

[6] B. Senthil Arasu, B. Jonath Backia Seelan, N. Thamaraiselvan "A machine learning-based approach to enhancing social media Computers & Electrical 106723.https://doi.org/10.1016/j.compeleceng.2020.106723 Engineering Volume 86, September 2020,

[7] Yui Ha , Kunwoo Park , Su Jung Kim , Jungseock Joo & Meeyoung Cha, "Automatically Detecting Image–Text Mismatch on Instagram with Deep Learning" January 2021 ,50(1):1-16.https://doi.org/10.1080/00913367.2020.1843091 Conference

[8] Irma Irawati Ibrahim "Social Media Analysis in Building Customer Trust - Systematic Literature Review", IEEE, 2022 International on Information Management and Technology (ICIMTech), 21 October 2022.https://doi.org/10.1109/ICIMTech55957.2022.9915099

[9] Abu Bashar, Mohammad Wasiq, Brighton Nyagadza and Eugine Tafadzwa Maziriri "Emerging trends in social media marketing: a retrospective review using data mining and bibliometric analysis" Springer, Future Business Journal ,volume 10, Article number: 23 (2024), 15 February 2024. https://doi.org/10.1186/s43093-024-00308-6

[10] Omvir Gautam and Rahul Chander Jaitly "Impact of social media influencers on customer engagement and brand perception" International Journal of Internet Advertising 15(2):220, January 2021. Marketing and Advertising 15(2):220,January 2021. https://doi.org/10.1504/IJIMA.2021.114336