

SARCASM DETECTION USING NLP

GROUP - 22

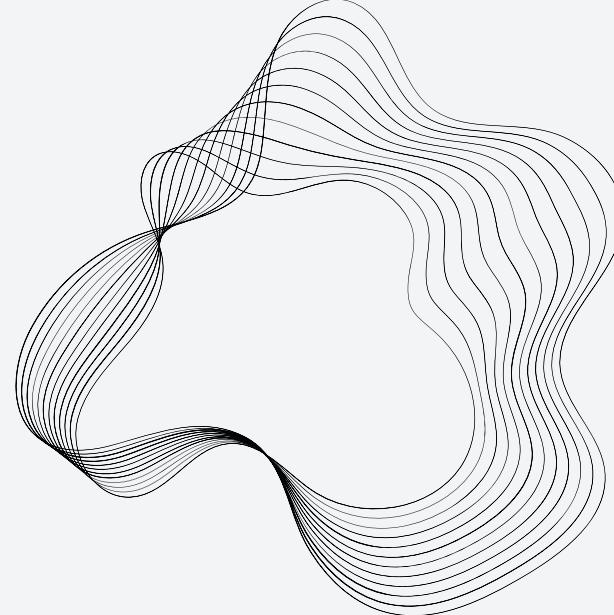
**MENTOR - PROF. MANJUSHREE
CHAKRABORTY**

**GROUP MEMBERS - UTSA GHOSH
MAYUKH PAUL
MANASH ROY
ISHA NATH**

PROBLEM STATEMENT

- Sarcasm, a form of verbal irony, often contrasts literal meaning with intent, making it difficult for NLP systems to accurately interpret sentiment.
- Detecting sarcasm is vital for accurate sentiment analysis and improving interactions in social media and customer feedback systems.
- Leveraging advanced NLP, the approach captures contextual and linguistic complexities to detect sarcasm with precision.

CHALLENGES IN SARCASM DETECTION



01

Contextual Understanding: Sarcasm relies on context for accurate detection.

02

Ambiguity: Sarcasm and non-sarcasm share surface meaning, causing ambiguity.

03

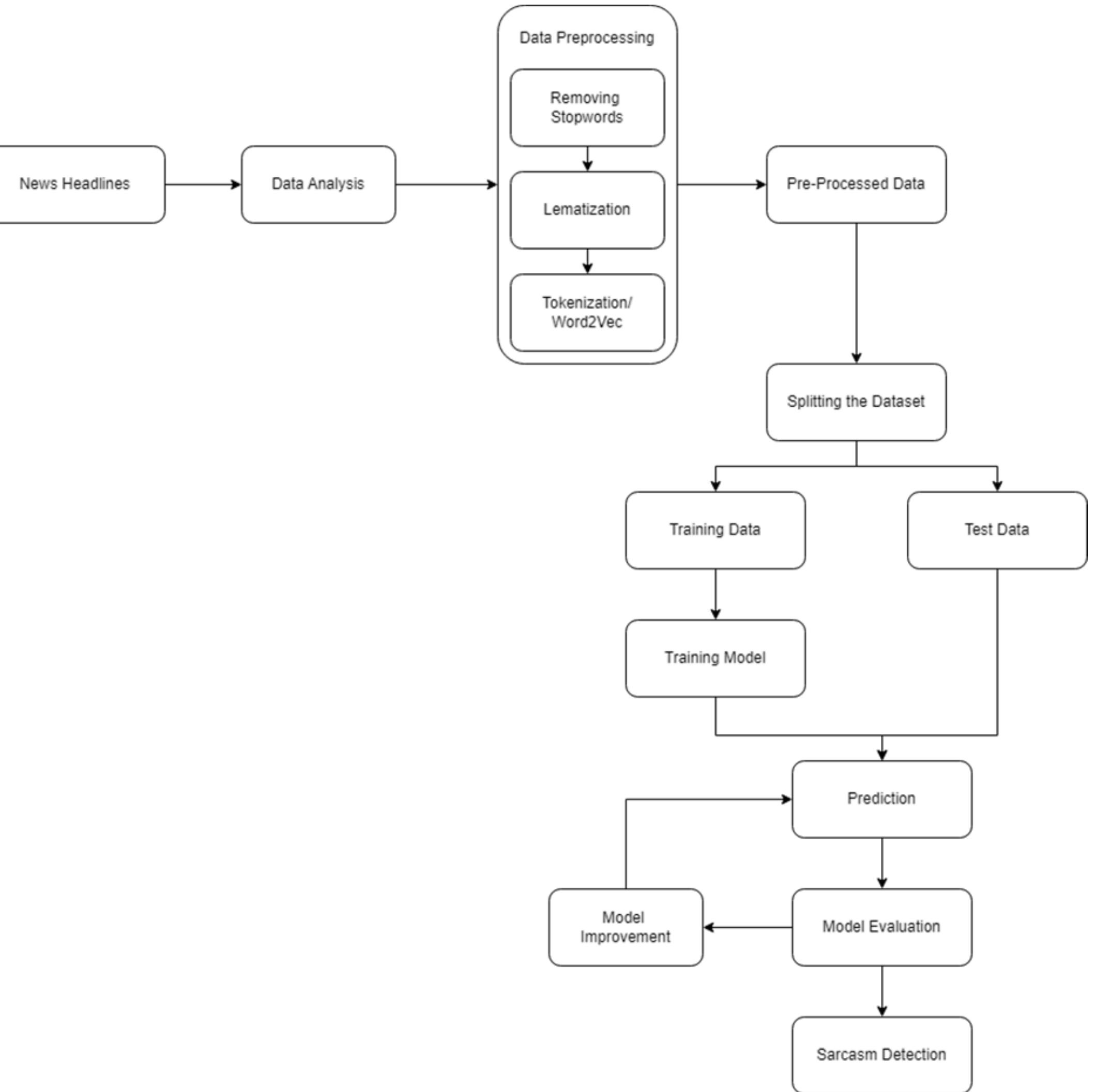
Lack of Non-verbal Cues: Text-based sarcasm detection lacks vocal and non-verbal cues.

04

Cultural and Linguistic Variations: Cultural and linguistic differences challenge sarcasm detection.



PROJECT WORKFLOW



DATASET

```
df.head()
```

	is_sarcastic	headline
0	1	thirtysomething scientists unveil doomsday clo...
1	0	dem rep. totally nails why congress is falling...
2	0	eat your veggies: 9 deliciously different recipes
3	1	inclement weather prevents liar from getting t...
4	1	mother comes pretty close to using word 'strea...

- Dataset: Labeled headlines: sarcastic (1) or non-sarcastic (0).
- Content: Includes URL, text, and label.
- Sources: The Onion, Huffington Post.
- Purpose: Trains NLP models for sarcasm detection.

DATA PREPROCESSING

Tokenization splits text into smaller units for NLP analysis.

Lemmatization reduces words to their root form for consistency.

Stopwords are common words removed for cleaner analysis.

Punctuation is removed to simplify text for analysis.

TOKENIZATION

LEMMATIZATION

STOPWORDS

PUCTUATION

```
lem = WordNetLemmatizer()
stop_words = set(stopwords.words("english"))
punctuations = string.punctuation
```

VECTORIZATION

Word2Vec

- Word2Vec converts words into vector representations based on context.
- It uses CBOW and Skip-Gram models to learn word associations.
- It captures semantic relationships, placing similar words closer in vector space.

MODEL BUILDING

RECURRENT NEURAL NETWORK (RNN)

- RNNs process sequential data with feedback loops.
- Used in speech recognition and NLP.
- Captures temporal dependencies for sarcasm detection.

LONG SHORT-TERM MEMORY (LSTM)

- LSTM overcomes the vanishing gradient problem in RNNs.
- It stores long-term dependencies using memory cells.
- Used in speech recognition, language modeling, and time series prediction.

BIDIRECTIONAL MODEL (BI-LSTM)

- Bi-LSTM processes data forward and backward.
- Captures context from past and future inputs.
- Enhances sarcasm detection by improving conversation understanding.

MODEL PERFORMANCE

- Models were trained for sarcasm detection using RNN, LSTM, and Bi-LSTM, with accuracy as the performance metric.

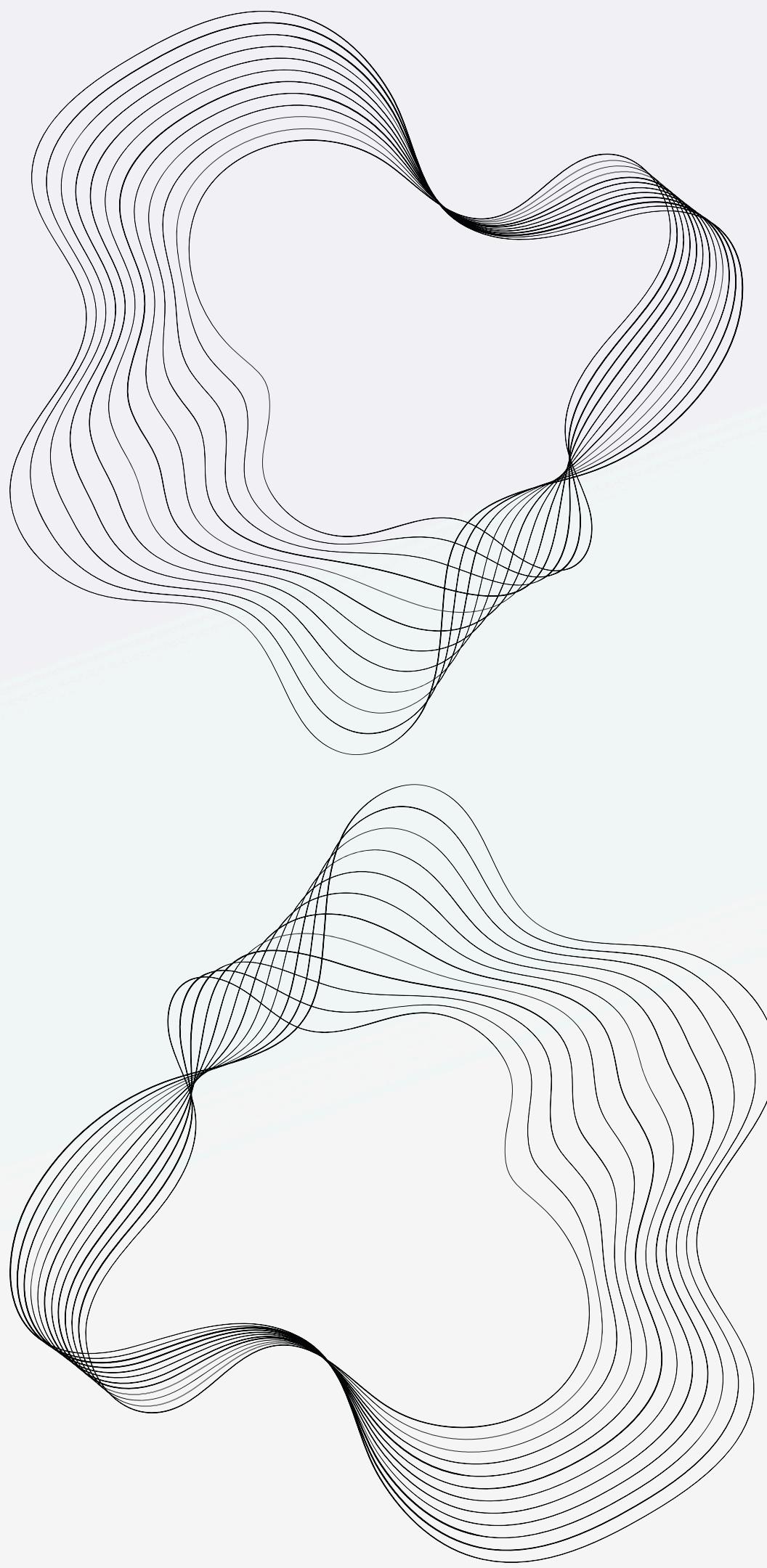
- The accuracy of the RNN model was **85.09%** on the test set.
- The LSTM model achieved an accuracy of **86.25%**.
- The **Bidirectional LSTM** model gave the highest accuracy at **86.30%**.

ACCURACY



FUTURE SCOPES

- Use transformers, BERT, and LLMs for improved sarcasm detection.
- Include full article content or external context to enhance detection.
- Expand the dataset for multi-language sarcasm detection.



CONCLUSION

- The Sarcasm Headlines Dataset supports sarcasm detection in NLP.
- It addresses challenges like ambiguity and cultural context.
- Advancements in contextual and multilingual analysis enhance model applications.

REFERENCES

1. Das, Dipto. "A multimodal approach to sarcasm detection on social media." (2019).
2. Sentamilselvan, K., P. Suresh, G. K. Kamalam, S. Mahendran, and D. Aneri. "Detection on sarcasm using machine learning classifiers and rule based approach." In *IOP Conference Series: Materials Science and Engineering*, vol. 1055, no. 1, p. 012105. IOP Publishing, 2021.
3. Bouazizi, Mondher, and Tomoaki Otsuki Ohtsuki. "A pattern-based approach for sarcasm detection on twitter." *IEEE Access* 4 (2016): 5477-5488.
4. Reyes, Antonio, and Paolo Rosso. "Mining subjective knowledge from customer reviews: A specific case of irony detection." In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis (WASSA 2.011)*, pp. 118-124. 2011.
5. Deshmukh, Pooja, and Sarika Solanke. "Sarcasm Detection and User Behaviour Analysis." *Int. J. Recent Innov. Trends Comput. Commun* 6, no. 6: 6. Shubhodip Saha, Jainath Yadav and Prabhat Ranjan (2017) *Proposed Approach for Sarcasm Detection in Twitter*. ISSN: 0974-5645.
6. Shubhodip Saha, Jainath Yadav, and Prabhat Ranjan. "Proposed approach for sarcasm detection in twitter." *Indian Journal of Science and Technology* 10, no. 25 (2017): 1-8.
7. Sarsam, Samer Muthana, Hosam Al-Samarraie, Ahmed Ibrahim Alzahrani, and Bianca Wright. "Sarcasm detection using machine learning algorithms in Twitter: A systematic review." *International Journal of Market Research* 62, no. 5 (2020): 578-598.
8. Šandor, Daniel, and Marina Bagić Babac. "Sarcasm detection in online comments using machine learning." *Information Discovery and Delivery* ahead-of-print (2023).
9. Sundararajan, Karthik, and Anandhakumar Palanisamy. "Probabilistic model based context augmented deep learning approach for sarcasm detection in social media." *International Journal of Advanced Science and Technology* 29, no. 6 (2020): 8461-8479.
10. Safrin, Raheesa, K. R. Sharmila, TS Shri Subangi, and E. A. Vimal. "Sentiment analysis on online product review." *International Research Journal of Engineering and Technology (IRJET)* 4, no. 04 (2017).

