



## Sarcasm Detection an Explainable AI Approach for Reddit Political Text

Rupali Amit Bagate<sup>1</sup>, Aparna Shashikant Joshi<sup>2</sup>, Anup Kadam<sup>3</sup>, Chandan Kumar Choubey<sup>4</sup>, Nilesh Sable<sup>5</sup>,  
Ajay Kumar<sup>6\*</sup>, Namrata Dogra<sup>7</sup>, Durgesh Nandan<sup>8</sup>

<sup>1</sup> Department of Information Technology, Army Institute of Technology, Pune 411015, India

<sup>2</sup> Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune 411044, India

<sup>3</sup> Department of Computer Engineering, Army Institute of Technology, Pune 411015, India

<sup>4</sup> Electronics and Telecommunication Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

<sup>5</sup> Department of Computer Science & Engineering (Artificial Intelligence), Vishwakarma Institute of Information Technology, Pune 411048, India

<sup>6</sup> Department of Mechanical Engineering, School of Engineering and Technology, JECRC University, Jaipur 303905, India

<sup>7</sup> Department of Orthodontics, Faculty of Dental Sciences, SGT University, Gurugram 122006, India

<sup>8</sup> School of Computer Science & Artificial Intelligence, SR University, Warangal 506371, India

Corresponding Author Email: [ajay.kumar30886@gmail.com](mailto:ajay.kumar30886@gmail.com)

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.120123>

### ABSTRACT

**Received:** 16 May 2024

**Revised:** 18 September 2024

**Accepted:** 24 September 2024

**Available online:** 25 January 2025

#### Keywords:

*deep learning, Explainable Artificial Intelligence, hashtag, political text, machine learning, sarcasm detection*

Sentiment analysis, often known as opinion mining, determines how people feel about any subject. Sarcasm is the expression of irony or mocking through the use of derogatory language. These phrases change the polarity of a positive feeling into a bad one or the other way around. The focus of the proposed research is to implement a domain-oriented sarcasm detection model with a XAI approach to justify classification results. This strategy emphasizes on bringing value to domain-specific text using an explainable approach. The research referred to political domain content from Reddit platform to get insights in the area. It identifies sarcastic context from textual information present on social media. As a part of sentiment analysis and sarcasm detection natural language processing (NLP) plays important role. The suggested methodology is identifying sarcasm using weighted average approach of long short term memory (LSTM) and support vector classifier (SVC). implemented system has a fresh strategy to identify such sarcastic terms from sentences or forecast such sarcastic sentences. The suggested strategy used an Explainable Artificial Intelligence (XAI) approach to identify sarcasm in a specific political domain text. As a part of XAI, counterfactual explanation is implemented to identify the sarcastic words from given text which is given with more weightage by model while training the model. The system has generated 75.75% F1 score as result of weighted average approach.

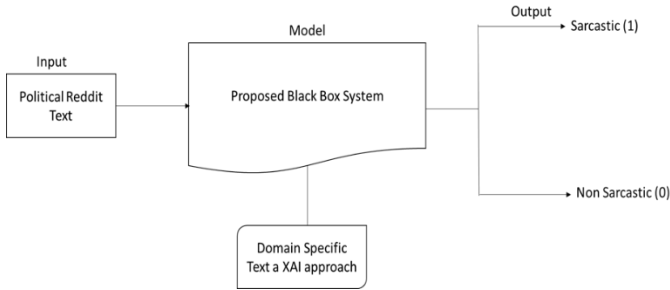
## 1. INTRODUCTION

In the social media era, a sizable amount of data is produced at a high volume and faster rate. A variety of data is gathered, consisting of text, video, audio, and combinations of these in addition to unstructured, semi-structured, and structured data. A significant problem is effectively handling various types of data, especially when verifying its accuracy. Many areas are tackling similar issues, where pre-processing ambiguous input results in a concrete judgement. Social media data is providing a tone of value for various organisations in the web 2.0 era. One social media platform where people from all backgrounds can share their unique ideas is Twitter. Their unique and divergent viewpoints generate a wealth of data that we can use to study topics like opinion mining, sentiment analysis, and sarcasm recognition on various social media sites. The way people express their thoughts has changed as a result of the

internet. The most popular forms of communication are blog posts, online forums, product review websites, and other channels. User-generated material is incredibly important to people when they're selecting whether or not to buy a product. Customers frequently look up reviews online. The user-generated content is far too extensive for the ordinary user to understand. Different emotional analysis techniques are applied to accomplish this automatically. To understand and detect sarcasm, sentiment analysis plays a very important role using natural language processing [1-4].

To increase commercial value, researchers are concentrating on domain-specific text sarcasm detection with an explainable approach. After conducting a thorough state-of-the-art analysis, researchers employed a combination of ML and DL methods with an ensemble approach to create a better sarcasm detection system. With the proposed methodology, researchers are attempting to use Explainable

Artificial Intelligence (XAI) to identify sarcasm in writings that are peculiar to a given topic. Explainable AI approach is used in many languages by many researchers [5-7] for sarcasm detection to interpret the model outcome SVC and LSTM make up the proposed architecture for sarcasm detection. The counterfactual explanation approach is used to analyze the explainability of AI models. Figure 1 shows a proposed architecture of experiment analysis.



**Figure 1.** Proposed methodology

Input to the model is political text extracted from reddit platform. Here scope of this experiment is to work on plain text for sarcasm detection which is very tricky and complicated as sarcastic text looks positive in nature but is it actually negative in context. To identify and classify test data in proper class is very challenging in natural language processing. Therefore, research has targeted this problem statement.

To target this text is pre-processed as a part on natural language processing. Black box of model is a combination of different models like SVC and LSTM with some hyper parameter tuning. While training a model for domain specific text system tried to identify the combination of words which are responsible for text as sarcastic. System is trying to solve binary classification problem where 0 means predication is non-sarcastic and 1 stands for sarcastic outcome. Here future scope of the research is to work on multi modal sarcasm detection to make it more realistic which can be applied to real world.

Further article is divided into different section. Section 2 is about literature survey of sarcasm detection. Section 3 is explaining methodology of research work. Section 4 explains results of experimental work. Last Section concludes the experimental work along with discussion.

## 2. LITERATURE REVIEW

This section describes what kind of work was carried out in past and currently being done in the field of sarcasm detection. Kalaivani and Thenmozhi [7] used traditional approach of machine learning along with deep learning approach recurrent neural network (RNN)+LSTM. Author did sarcasm detection on twitter dataset and reddit dataset and compared the F1 score. Research has come up with a 0.72% and 0.67% F1 score for twitter and reddit dataset respectively.

Kumar and Garg [8] have focused on context in sentence for sentiment analysis. The research experimented two different datasets semEval2015 task11 twitter dataset and reddit dataset with 20k post. The research is done with three models. First is combination of TF-IDF with three classifiers random forest (RF), multinomial naïve Bayes and gradient boosting. Further next model is a combination of pragmatic and semantic features where results are observed using different classifiers such as decision classifier, SVM, RF, K-nearest neighbour and multi-layer perceptron. Last model is combination of Glove, LSTM and BiLSTM. As a result of classification best F1 score produced as 86.32% and 82.91% for twitter and reddit dataset respectively.

Sharma et al. [9] used hybrid auto encoder-based model for sarcasm detection, and took three datasets for analysis purpose namely headline dataset, self-annotated reddit corpus (SARC) and Twitter dataset. The accuracy received was 92.80%, 83.92% and 90.8%, respectively. They also did sarcasm detection on text along with images., used BERT, LSTM and universal sentence encoder for sentence embedding and processing considering sentence context rather than word.

Poria et al. [10] have extracted features using pre trained convolution neural network (CNN) responsible for sarcastic sentence as a positive one. During training phase, the model sentiment shift is observed and taken as input for identification of sarcasm detection. They proposed this sentiment shift as feature key point for sarcasm detection, worked on three various datasets for sarcasm detection. These three datasets are imbalanced dataset, balanced and test dataset. Every dataset is trained on combination of same set of classifiers. Combination of CNN+SVM generates respectively F1 score as 97.71%, 94.80% and 93.30% by combining baseline features, sentiment, emotions and personality features all together. Table 1 shows a summary of literature survey performed for proposed research work.

**Table 1.** Summary of literature survey

Authors	Dataset	Methodology	Results
Kalaivani and Thenmozhi [7]	Twitter and Reddit	RNN + LSTM	0.72% F1 Score 0.67% F1 Score
Kumar and Garg [8]	semEval2015 twitter	TFIDF, RF, NB, gradient boosting decision classifier, SVM, RF, KNN, MLP	86.32% F1 Score
	task11 twitter		82.91% F1 Score
Sharma et al. [9]	Reddit Dataset	BERT, LSTM and universal sentence encoder	92.80% Accuracy
	Headline		83.92% Accuracy
	SARC		90.8% Accuracy
Poria et al. [10]	Twitter	CNN + SVM	97.71% F1 Score
	Balanced		94.80% F1 Score
	Imbalanced Test		93.30% F1 Score

Research work shows many applications of sentiment analysis, such as sentiment analysis locates and extracts illogical information from text using text mining and natural language processing. Wankhade et al. [11] and Rodrigues et al.

[12] did sentiment analysis which has variety of application such as spam detection in tweets along with its emotion classification. Ali et al. [13] did a sentiment analysis of election Tweets of United States. As observed from state of art

authors have used twitter and reddit dataset for research work. Therefore, author decided to work on reddit political text for justification of said work. Along with dataset selection very few have worked on without #sarcasm as features in text as their work. As observed from literature survey drawback of existing work is lack of better pre-processing techniques and lack of explanation system for predicted outcome therefore proposed system has come up with novel work.

### 3. METHODOLOGY

This methodology is used to design and develop a domain-oriented sarcasm detection model that justifies classification result using XAI approach. A model that works on domain specific text for adding more value to a specific domain text with its explainable approach. Therefore, research chose political domain text from reddit to work and gain more insights in the specific area.

Many researchers did sarcasm detection with presence of hash tags in tweets. It makes system training on dataset very easy, however during testing and validation phase, system becomes vulnerable for sarcasm prediction of tweets where hint of explicit sarcasm or hint of hash tags are absent. Therefore, research methodology has decided to perform sarcasm detection on tweets and text where explicit clues are absent. Additionally, research is aiming a for domain-oriented sarcasm detection for better research contribution in the field of sarcasm detection. Proposed work is divided into three phases as shown in Figure 2. Workings of models are divided into different stages such as pre-processing, training testing and ensembling the result of different classifiers.

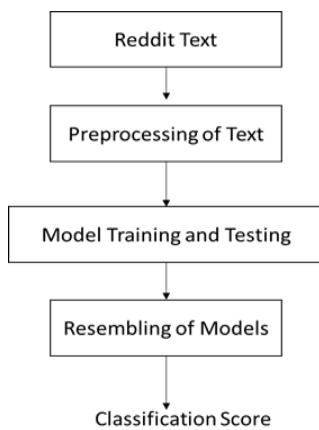


Figure 2. Overview of proposed methodology

The main aim of pre-processing is to remove noise from tweets/text and make it ready for word embedding. Noise removal is one of the biggest and very important stage of natural language processing. Until text is noise free, model can't process it further for better precision and accuracy. Training and testing phase includes various stages like word embedding, vectorization, creating input vector as per requirement of models etc. After word Embedding stage once vector is ready for training it is fed to different machine learning model and deep learning model for further processing. Different models are extracting features from given tweets/text such as sentiments and emotions. Once features are extracted tweets are trained for sarcasm detection and classification. In testing phase tweets/text is input to different models for

classification. Further best of working models is inputted to ensemble learning. Stacking technique is used to enhance the accuracy, F1 score and many other parameters of proposed model.

Proposed methodology is trying to generate a value from social media text to help society in better way. Working on generalized data won't help business in certain direction. For e.g. if someone wants to predicts the result of stock market, one has to collect the data specific to the domain. Such as stock market news related to stock movements, NSE and BSE data will help to identify which stock will rise and which stock will fall. Similarly, in opinion mining or sentiment analysis being a customer-oriented industry wants to cater the requirement of client needs to work on specific data to create a value out of it. As shown in Figure 3, the proposed methodology is working on two major parts. First part where the concept sarcasm detection is implemented. Second part where the concept of XAI is applied on outcome of classifier.

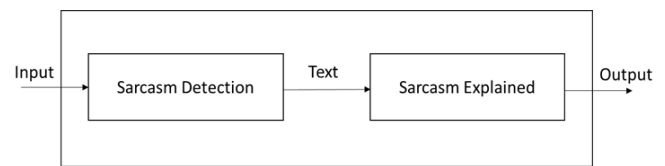


Figure 3. Proposed system outline

#### 3.1 Sarcasm detection

To work upon on such methodology, author has retrieved text from reddit platform and made dataset for sarcasm detection system on domain specific task. Proposed system predicts the political parties' trends along with XAI approach. In proposed work author referred Self-Annotated Reddit Corpus (SARC) dataset from the study conducted by Khodak et al. [14] where 533M text extracted for research work. Out of which 1.34M comments were sarcastic. For research purpose selected shape of political filtered dataset is (39496, 11). For research work new dataset is constructed from selected dataset. Newly constructed dataset consists of 3 columns namely label, comment and parent comment. After observing dataset, found that 40% comments are non-sarcastic and 60% comments are sarcastic resulting balanced dataset. Poria et al. [10] worked on balanced dataset and proved balanced dataset produces better F1 score as compare to unbalanced dataset.

The proposed methodology black box consists of pre-processing, feature extraction, model training and classification. Pre-processing is combinations of various steps such as noise removal, hash tag sarcasm removal, stop word removal, url dropping, conversion to lowercase, tokenization, lemmatization etc. In first stage, dropping of unnecessary data such as urls, which are not relevant to opinion mining is done. Further irrelevant characters are removed to reduce the noise. Further string words are lowercased for uniformity to improved machine learning ability. As a part of natural language, processing tokens of every string is very important. After tokenization, noise removal removes unnecessary words such as comma, full stop, punctuation marks etc. This step makes text more understandable and relevant for further processing. After removal of stop words, system proceed for lemmatization. Lemmatization is very important stage in pre-processing, which establishes a root meaning of words. Finding root of a given word has two methods named

stemming and lemmatization. After comparing both processes, lemmatization is accurate as compare to stemming. As ‘change’ is actual required word which is being generated after the process of lemmatization rather than ‘chang’ after stemming process. Therefore, lemmatization preserves the actual meaning of words. After lemmatization, pre-processing removes words having length less than 2 characters. As a final step, all tokens are combined back into a string. Once string is ready after so many steps and filters, next stage is the feature extraction. Feature extraction is a process of identifying different parameters, which helps in enhancing the result of classification. Different features such as emotions (Happy, Angry, Surprise, Sad and Fear) along with sentiments (Negative, Positive, Neutral and Compound) are extracted. For every emotion and sentiment some score is generated by model as a part of weight calculations. After feature extraction, different models are trained on features and comments for sarcasm detection. In proposed methodology SVC classifier and LSTM are used for sarcasm detection work. SVC is a valuable baseline model for classification. Support vector classification belongs to the broader family of support vector machines (SVMs). It aims to locate the hyperplane that best splits the classes in a high-dimensional space. For sarcasm identification, SVC will categorise text into sarcastic and non-sarcastic classes. LSTM networks are a powerful approach for detecting sarcasm due to its ability to identify long-term dependencies in textual data by handling vanishing gradient problems occurred in RNN. However, the effectiveness of an LSTM-based sarcasm detector relies greatly on the quality and diversity of the training data.

For feature extraction process, text2emotion and Vader sentiment library is used to extract the sentiment from text. Various types of sentiments in the form of emotions are extracted using these libraries such as happy, sad, angry, fearful, surprised etc. along with sentiment polarity. These explicit extractions supplying as an input to SVC classifiers helps the training process in accurate way. In parallel with SVC, LSTM neural network also trained.

LSTMs are commonly used for language translation, language creation, and text categorization. They have also been employed in applications like as voice recognition and time series forecasting. Many researches [15-17] used long short term memory, recurrent neural network and convolution neural network for sarcasm detection.

LSTM is capable of feature extraction automatically. In Proposed Methodology LSTM is infused with many layers such as embedding layer, dropout layer, convolution 1D, maxpooling layer, LSTM layer and dense layer. Every layer has specific task and contribution in training the model. While training data Stochastic Gradient Descent (SGD) optimizer is used for better optimization while training a model. During training, network weights are updated using SGD to minimize the error. The size of the steps, which is defined by the one of the hyper parameters as learning rate, is a crucial Gradient Descent (GD) parameter. The algorithm will need to go through many iterations in order to converge, which will take a lot of time. If the learning rate is too high, we risk jumping the ideal value. After many trials model set learning rate as 0.01.

In short LSTMs prefer to perform well due to their capability to perceive and remember context, which is critical for effectively identifying sarcasm in text. However, SVMs can still be useful, especially when combined with the efficient feature engineering provided by LSTMs. After training both

classifiers, weighted average approach is been applied on outcome of SVC and LSTM to improve the result. Figure 4 shows working of proposed methodology.

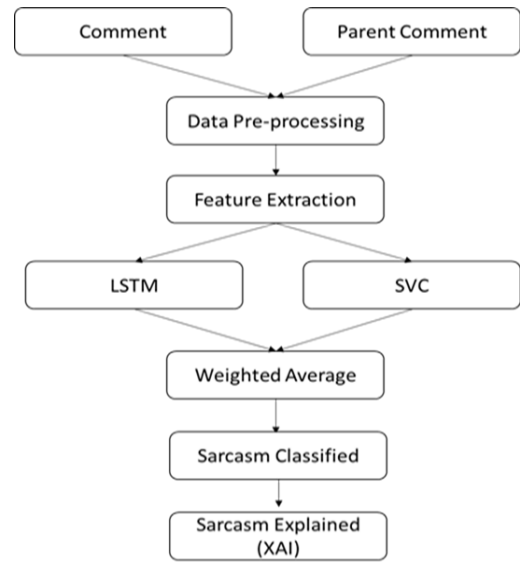


Figure 4. Sarcasm detection workflow

### 3.2 Explainable AI

XAI attempts to make AI models easier to understand and interpret, by providing explanations for their predictions and decisions. This enables users to understand how models works, and gain insight into why the AI made certain decisions. Explainable AI is critical for detecting sarcasm, particularly in the political sphere, where openness, accountability, and justice are important. By offering unambiguous facts about decision-making processes, XAI improves confidence, accuracy of models, bias detection and mitigation, regulatory compliance, and understanding of context. This results in more trustworthy and ethical AI systems that can efficiently negotiate the nuances of sarcasm in political debate. Figure 5 shows concept of XAI used for sarcasm detection. XAI is used to improve decision-making, reduce bias, and increase trust in AI systems.



Figure 5. XAI using counterfactual explanation

Many researchers referred LIME, SHAP, IntGrad and attention-based methods for XAI. Bodria et al. [18] used and compared attention-based mechanism for explainability of sarcasm detection model with LIME and IntGrad. Authors have used the concept of heat map visualization to describe the black box of their model. Authors did comparison on test dataset, where author found as compare to LIME and IntGrad, attention mechanism is working well. By referring the state of art, research found counterfactual explanation is better method for XAI as it gives more attention on permutations and combinations of words from a given sentence. In proposed methodology system has passed some sentences to counterfactual model for interpretability. Counterfactual explanation model is trying to explain the sentence by highlighting the words which are responsible for sentence to be sarcastic or non-sarcastic.

An explainable AI method to identify sarcasm not only increases the trustworthiness and reliability of AI systems, yet it also provides out a broad spectrum of applications across multiple areas. By offering greater transparency and understanding into AI decision-making processes, XAI improves the efficiency, fairness, and customer acceptance of the sarcasm detection model.

The trained political sarcasm detection model likewise improved confidence following classification using the XAI technique.

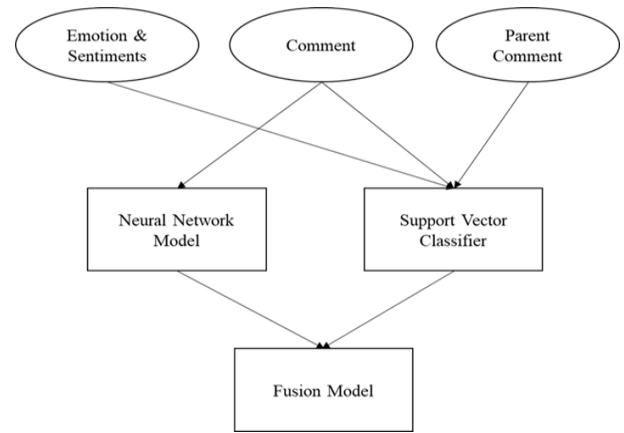
#### 4. EXPERIMENTAL RESULTS

This section describes the sarcasm detection on political domain dataset. Figure 6 describes the working methodology of domain specific sarcasm detection for a given dataset.

Said experiment is trained on comments, parents' comment, emotions and sentiments. As a research experiment, system used baseline model trained on comment and parent comment, where system used LSTM along with dense layer achieved F1 score as 73.71% and testing F1 score as 70.90%. Further LSTM model processed only comments along with Glove model. LSTM model has achieved 74.55% testing F1 score and 74.98% as training F1 score. After comparing the result of baseline model and LSTM model, LSTM is working better as compare to baseline model. Therefore, as a result of comparison research experiment has taken only LSTM in to the account for further model training and baseline model is dropped.

Along with LSTM one more model is trained on same dataset with different feature set. SVC classifier is trained on comment, parent comment and some feature combination such as emotions and sentiment. SVC generated training F1 score as 75.53% and testing F1 score as 74.57%. As a final classification result, weighted average of LSTM and SVC classifier is calculated. The final weighted average approach F1 score is 77.07% and testing F1 score as 75.75%.

Training of SVC is done on different features such as emotions, sentiments of comments along with parent comment. The features extracted along with sentiment score are playing major role while training the SVC classifier. Further, to enhance the result of a system, weighted average technique is used on both classifiers. Table 2 shows the detailed result of LSTM, SVC and weighted average model.



**Figure 6.** Working model of political domain sarcasm detection

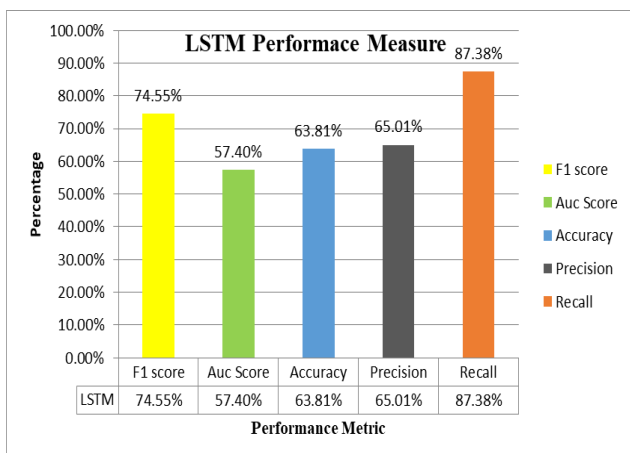
From Table 2, the fusion of LSTM and SVC is resulting into desirable F1 score as 75.75%. Figure 7 shows graphical representation of domain specific sarcasm detection of LSTM model. Figure 8 shows the SVC classifier performance analysis in graphical format.

Figure 9 shows graphical analysis of fusion model. It is an average result of LSTM and SVC classifiers.

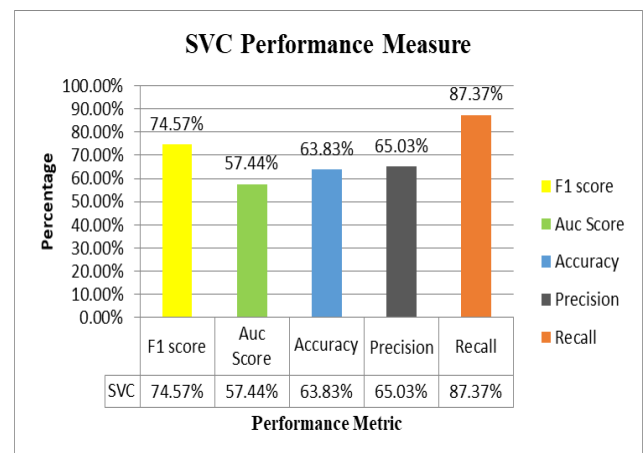
Proposed methodology is compared with existing work done by González-Ibáñez et al. [19] and Karoui et al. [20]. In comparison with mentioned work, proposed methodology is generating better F1 score for sarcasm classification and prediction.

**Table 2.** Proposed methodology result analysis

Models	Layers	Features	F1 Score %	Accuracy%	AUC Score %	Precision %	Recall %
Neural network	Glove, CNN, LSTM, Dense	Comment	74.55	63.81	57.4	65.01	87.38
Support vector classifier	SVC	Emotion & sentiment (Comment and parent comment)	74.57	63.83	57.44	65.03	87.37
Fusion result	Weighted average	Combining all feature set	75.75	63.82	57.42	65.02	87.38



**Figure 7.** LSTM performance analysis



**Figure 8.** SVC performance analysis

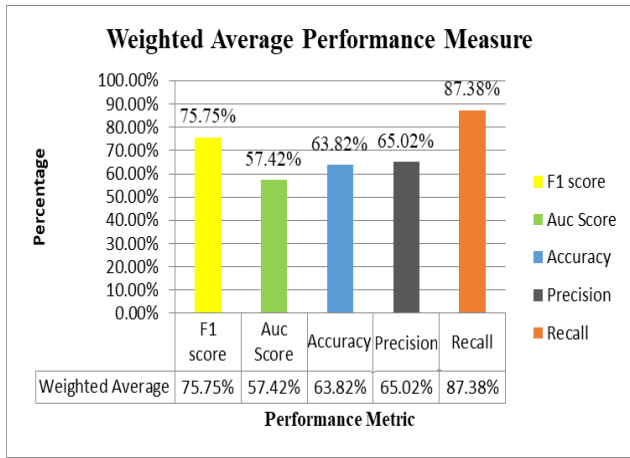


Figure 9. Performance analysis of fusion model

Table 3. Comparative analysis of sarcasm detection model with state of art

Authors	Dataset	Methodology	Language	F1 Score
Lemmens et al. [21]	Reddit	CNN, LSTM, SVM, MLP	English	66.70%
Javdan and Minaei-Bidgoli [22]	Reddit	BERT-base-cased	English	73.40%
Proposed model	Reddit	LSTM, SVC (Sentiments+emotions)	English	75.75%

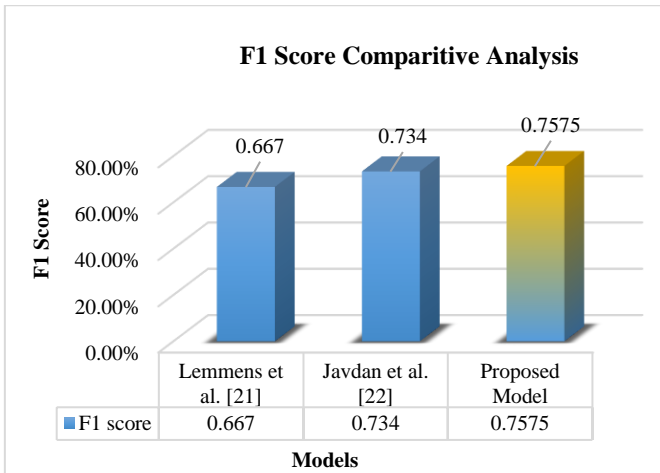


Figure 10. Graphical representation of result analysis

Research has utilized the Python multiset permutations package to implement counterfactual explanation for explainable AI. With the use of the multiset permutation library, one can run permutations and combinations of several words in order to pinpoint the necessary word that is most crucial to the sarcastic intent of the statement. As part of experiment, model have passed 10 sentences for validation.

Out of 10 sentences, 8 were predicted and explained correctly by our XAI model. Here system successfully able to find out word from text which contributes for sentence classified as sarcastic or non-sarcastic highlighted with yellow colour and red font. This highlighted word is an automatic outcome of counterfactual explanation function as a part of XAI model. Therefore, one can easily explain why the sentence is predicted as sarcastic. After passing such sentences to counterfactual model system receives an accuracy of XAI

model as 80%.

As a part of the research justification Table 3 explains the comparison of F1 score of implemented methodology with existing systems. As observed in Table 2, reddit dataset combining different methodologies such as CNN, LSTM, SVM, BERT etc. are used. Proposed model is giving F1 score as 75.75% which is comparatively very good as compared to the existing work. Figure 10 shows the graphical representation of Table 2.

As a summary of Experimental research work has contributed in sarcasm detection field by focusing on domain specific dataset along with implementation XAI. A handful of work is done in sarcasm classification area with absence of #sarcasm with application of XAI.

The impact of research is contributing in many fields. Such as social media monitoring, news and media analytics and political discourse. Political discourse helps in analyzing the sentiment of people towards any political party for future campaign to gain more votes during elections.

## 5. DISCUSSION

The main objective of this research was to work in the field of sentiment analysis where rather focusing on sentiment as whole spectrum, research is focusing on sarcasm. Identifying sarcastic words from text which is flipping the polarity of positive sentence into negative or vice versa is very challenging.

Generating a value for a business is goal of any research or innovation. To accomplish the requirement of business growth by people's opinion is targeted in proposed methodology using domain-oriented dataset for sarcasm detection. Here research aimed to target two outcomes, first is sarcasm detection using a combination of machine learning and deep learning algorithms with some hyper parameter tuning and improved pre-processing techniques and second step to open a black box of sarcasm detection system for gaining an insight from working model of AI. XAI is novel part of proposed methodology which used the concept of counterfactual explanation to identify the different words from text which are responsible for class to be predicted as sarcastic. Many studies [23-25] recommended counterfactual explanation for XAI. Proposed methodology aimed to solve opinion of people about political domain. Sarcasm detection in politics has immense potential for increasing the analysis and understanding of political conversation & assessing public opinion. However, it also presents challenges in regards to accuracy, privacy, and bias, all of which must be properly addressed in order to fully realise the benefits. After training a model system has achieved good result for targeted domain.

## 6. CONCLUSIONS

Generating a value for a business is goal of any research or innovation. The objective of this proposed methodology is to achieve this goal with innovative way. To accomplish the requirement of business growth by people's opinion is targeted in methodology using domain-oriented dataset for sarcasm detection. The implemented approach aims to extract value from social media text in order to benefit society. Working with generalized data will not aid an organization in a specific direction, hence the author focused on the political domain

data to create value. To attain the goal and improve accuracy, a weighted average technique of SVC and LSTM is used, which is then compared to current models in the paper.

Here research aimed to target two outcomes, first is sarcasm detection using a combination of machine learning and deep learning algorithms and second step to open a black box of sarcasm detection system. Opening a black box means system is trying to gain an insight from working model of AI. XAI is novel part of proposed methodology which used the concept of counterfactual explanation to identify the different words from text which are responsible for class to be predicted as sarcastic. This methodology aimed to solve opinion of people about political domain. After training a model system has achieved good result for targeted domain as compared to existing system. There are some limitations of proposed research such as model is considering only text as input in dataset. Additional inputs along with text such as audio or image may help to increase the accuracy of system.

## 7. FUTURE SCOPE

Applications of Sarcasm detection and classification impacts many social aspects such as political discourse, chatbot and virtual assistant, social media monitoring etc. To analyse and improve this classification research has to consider a multiple dimension. The future scope of this work is to take efforts upon multimodal aspects. Along with text research work has to consider tonal quality and variation, image, videos and emojis as features for sarcasm detection. To work more superiorly on contextual features and semantic features different word embedding's can be used.

## REFERENCES

- [1] Kumar, A., Rani, S., Rathee, S., Bhatia, S. (2023). Security and risk analysis for intelligent cloud computing: Methods, applications, and preventions. CRC Press. <https://doi.org/10.1201/9781003329947>
- [2] Kakkar, M.K., Singla, J., Garg, N., Gupta, G., Srivastava, P., Kumar, A. (2021). Class schedule generation using evolutionary algorithms. *Journal of Physics: Conference Series*, 1950(1): 012067. <https://doi.org/10.1088/1742-6596/1950/1/012067>
- [3] Bagate, R.A., Suguna, R. (2021). Sarcasm detection of tweets without# sarcasm: Data science approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(2): 993-1001.
- [4] Rajendra, P., Kumari, M., Rani, S., Dogra, N., Boadh, R., Kumar, A., Dahiya, M. (2022). Impact of artificial intelligence on civilization: Future perspectives. *Materials Today: Proceedings*, 56: 252-256. <https://doi.org/10.1016/j.matpr.2022.01.113>
- [5] Mehta, H., Passi, K. (2022). Social media hate speech detection using explainable artificial intelligence (XAI). *Algorithms*, 15(8): 291. <https://doi.org/10.3390/a15080291>
- [6] Rani, S., Tripathi, K., Kumar, A. (2023). Machine learning aided malware detection for secure and smart manufacturing: A comprehensive analysis of the state of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1-28. <https://doi.org/10.1007/s12008-023-01578-0>
- [7] Kalaivani, A., Thenmozhi, D. (2020). Sarcasm identification and detection in conversion context using BERT. In *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 72-76. <https://doi.org/10.18653/v1/2020.figlang-1.10>
- [8] Kumar, A., Garg, G. (2023). Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets. *Journal of Ambient Intelligence and Humanized Computing*, 14(5): 5327-5342. <https://doi.org/10.1007/s12652-019-01419-7>
- [9] Sharma, D.K., Singh, B., Agarwal, S., Kim, H., Sharma, R. (2022). Sarcasm detection over social media platforms using hybrid auto-encoder-based model. *Electronics*, 11(18): 2844. <https://doi.org/10.3390/electronics11182844>
- [10] Poria, S., Cambria, E., Hazarika, D., Vij, P. (2016). A deeper look into sarcastic Tweets using deep convolutional neural networks. *arXiv preprint arXiv:1610.08815*. <https://doi.org/10.48550/arXiv.1610.0881>
- [11] Wankhade, M., Rao, A.C.S., Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7): 5731-5780. <https://doi.org/10.1007/S10462-022-10144-1>
- [12] Rodrigues, A.P., Fernandes, R., Shetty, A., Lakshmana, K., Shafi, R.M. (2022). Real-time Twitter spam detection and sentiment analysis using machine learning and deep learning techniques. *Computational Intelligence and Neuroscience*, 2022(1): 5211949.
- [13] Ali, R.H., Pinto, G., Lawrie, E., Linstead, E.J. (2022). A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election. *Journal of big Data*, 9(1): 79. <https://doi.org/10.1186/s40537-022-00633-z>
- [14] Khodak, M., Saunshi, N., Vodrahalli, K. (2017). A large self-annotated corpus for sarcasm. *arXiv preprint arXiv:1704.05579*. <https://doi.org/10.48550/arXiv.1704.05579>
- [15] Salim, S.S., Ghanshyam, A.N., Ashok, D.M., Mazahir, D.B., Thakare, B.S. (2020). Deep LSTM-RNN with word embedding for sarcasm detection on Twitter. In *2020 International Conference for Emerging Technology (INCET), Belgaum, India*, pp. 1-4. <https://doi.org/10.1109/INCET49848.2020.9154162>
- [16] Kumar, A., Dabas, V., Hooda, P. (2020). Text classification algorithms for mining unstructured data: A SWOT analysis. *International Journal of Information Technology*, 12(4): 1159-1169. <https://doi.org/10.1007/s41870-017-0072-1>
- [17] Khotijah, S., Tirtawangsa, J., Suryani, A.A. (2020). Using LSTM for context based approach of sarcasm detection in twitter. In *Proceedings of the 11th International Conference on Advances in Information Technology, Bangkok, Thailand*, pp. 1-7. <https://doi.org/10.1145/3406601.3406624>
- [18] Bodria, F., Panisson, A., Perotti, A., Piaggese, S. (2020). Explainability methods for natural language processing: Applications to sentiment analysis. *CEUR Workshop Proceedings*, 2646: 100-107.
- [19] González-Ibáñez, R., Muresan, S., Wacholder, N. (2011, June). Identifying sarcasm in Twitter: A closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 581-586.
- [20] Karoui, J., Benamara, F., Moriceau, V., Patti, V., Bosco,

- C., Aussenac-Gilles, N. (2017). Exploring the impact of pragmatic phenomena on irony detection in Tweets: A multilingual corpus study. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 262-272. <https://aclanthology.org/E17-1025>.
- [21] Lemmens, J., Burtenshaw, B., Lotfi, E., Markov, I., Daelemans, W. (2020). Sarcasm detection using an ensemble approach. In Proceedings of the Second Workshop on Figurative Language Processing, pp. 264-269. <https://doi.org/10.18653/v1/2020.figlang-1.36>
- [22] Javdan, S., Minaei-Bidgoli, B. (2020). Applying transformers and aspect-based sentiment analysis approaches on sarcasm detection. In Proceedings of the Second Workshop on Figurative Language Processing, pp. 67-71. <https://doi.org/10.18653/v1/2020.figlang-1.9>
- [23] Warren, G., Smyth, B., Keane, M.T. (2022). "Better" counterfactuals, ones people can understand: psychologically-plausible case-based counterfactuals using categorical features for explainable AI (XAI). In International Conference on Case-Based Reasoning, pp. 63-78. [https://doi.org/10.1007/978-3-031-14923-8\\_5](https://doi.org/10.1007/978-3-031-14923-8_5)
- [24] Crupi, R., Castelnovo, A., Regoli, D., San Miguel Gonzalez, B. (2022). Counterfactual explanations as interventions in latent space. Data Mining and Knowledge Discovery, 1-37. <https://doi.org/10.1007/s10618-022-00889-2>
- [25] Keane, M.T., Smyth, B. (2020). Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI). In Case-Based Reasoning Research and Development: 28th International Conference, ICCBR 2020, Salamanca, Spain, pp. 163-178. [https://doi.org/10.1007/978-3-030-58342-2\\_11](https://doi.org/10.1007/978-3-030-58342-2_11)