**SARCASM DETECTION USING CONTEXTUAL EMBEDDING**



**A Synopsis**

submitted for the degree of

**Bachelor of Technology**

**in**

**Computer Science / Information Technology**

Submitted by:

**[GroupID: CSIT6]**

**Aahna Sethi [2116091]**

**Ananya Singh Gautam [2116104]**

**Akanksha Sharma [2116728]**

**Sakshi [2116814]**

Under the Supervision of:

**Dr. Nisheeth Joshi**

Associate Professor

Department of Computer Science

Banasthali Vidyapith

Rajasthan.

**DEPARTMENT OF COMPUTER SCIENCE**

**BANASTHALI VIDYAPITH**

**Rajasthan, India**

**2023-24**

**TABLE OF CONTENTS**

[**1.** **INTRODUCTION** 3](#_Toc148698170)

[**2.** **LITERATURE REVIEW** 4](#_Toc148698171)

[**3.** **RESEARCH GAPS** 15](#_Toc148698172)

[**4.** **OBJECTIVES** 16](#_Toc148698173)

[**5.** **PROPOSED METHODOLOGY** 16](#_Toc148698174)

[**i.)** **Corpus Creation:** 16](#_Toc148698175)

[**ii.)** **Extraction of Linguistic and Cognitive factors:** 17](#_Toc148698176)

[**iii.)** **Deep Learning Model:** 18](#_Toc148698177)

[**iv.)** **Evaluating The Model:** 18](#_Toc148698178)

[**6.** **Tools and Techniques** 19](#_Toc148698179)

[**7.** **Work Plan** 20](#_Toc148698180)

[**8.** **Expected Outcomes** 20](#_Toc148698181)

[**REFERENCES** 21](#_Toc148698182)

# **INTRODUCTION**

As per Merriam Webster[[1]](#footnote-1), sarcasm is said to be "a mode of satirical wit depending on its effect on bitter, caustic, and often ironic language that is usually directed against an individual". According to Dews and Winner (1995), it has the supernatural ability to mask the speaker's animosity while boosting the effect of ridicule or comedy on the audience.

Today's social media platforms are rife with sarcasm, and automatic identification of it is crucial for activities like customer support, opinion mining, online harassment detection, and other ones that demand understanding of people's true emotion. Self-deprecating, brooding, deadpan, polite, obnoxious, manic, and rage are a few examples of the various sorts. Sarcasm detection is difficult due to its dependency on context, linguistic complexity, and different forms across cultures. The process is made more difficult by the use of numerous communication channels, such as text and images, because automated systems must understand both verbal and nonverbal clues. Accurate identification is further complicated by the subjective nature of sarcasm and its range of tones.

Irony and sarcasm are examples of figurative language that include expressing one thing while actually meaning something another. Their aims and tonality are the main points of distinction. Sarcasm is a form of irony that uses a mocking or critical tone to ridicule or convey contempt, while irony, in general, refers to a disconnect between expectations and reality, often used for humor or highlighting paradoxes. Effective communication requires an awareness of sarcasm because it enables others to comprehend the intended meaning behind nuanced remarks, improving interpersonal relationships and preventing misunderstandings.

Accurate sarcasm detection is becoming more and more important in the age of social media, which is widely used, and pervasive digital communication, where sarcasm is common. Tools for sarcasm detection are essential in many fields, such as sentiment analysis, customer service, and online content control. They aid in identifying the real intention behind text-based interactions, enabling companies to precisely assess customer happiness and sentiment.

Additionally, in the context of online platforms, efficient sarcasm detection is essential for avoiding misunderstandings, reducing the dissemination of false information, or containing inflammatory content, and fostering a more positive and courteous online environment.

# **LITERATURE REVIEW**

Riloff et al. (2013) developed Sarcasm as Contrast between a Positive Sentiment and Negative Situation. They worked on the contrast between the situation and the sentiment of the author. For identifying contrast between emotion and sarcasm they developed a sarcasm recognizer. They presented an algorithm that accordingly learns positive sentiment phrases and negative situation phrases from sarcastic tweets. And showed that the algorithm yielded better recall for sarcasm recognition. Their study had limitations like the phrase they used was limited to the same type structure and context. And they wanted to learn more on how to identify stereotypical negative activity because it is essential to have that world knowledge to identify sarcasm. If we can identify the intensity of the negative situation then also it is useful to differentiate between positive and negative contrast

Maynard et al. (2013) presented an integrated method for social media analysis that integrates textual and multimedia data. The study suggested a rule-based textual method enhanced by multimedia analysis in light of the difficulties brought on by the dynamic nature of social web material, including noisy text and complicated multimedia interactions. The research emphasised the significance of structured preservation based on semantic categories along with emphasis on helping archivists choose pertinent content for communal memory preservation. The article, which offers promising insights into fields including sentiment analysis, sarcasm detection, and discourse analysis, is notable for introducing unique text and multimedia mining algorithms customised to the nuances of social media. The modular design of the study paved the way for future developments in comprehending and examining the complex world of social media interactions.

Ptacek et al. (2014) conducted a study on sarcasm detection in both English and Czech tweets using a machine learning approach. They assessed two classifiers with diverse feature combinations on datasets from both languages. The study addressed the complexities of Czech morphology by exploring various preprocessing techniques. Notably, their results demonstrated the superiority of a language-independent approach over an adapted state-of-the-art method in English (F-measure 0.947). The research primarily emphasised the rich morphology and syntax of both languages, especially Czech, and incorporated language-independent methods and preprocessing recommendations from Habernal et al. (2013). Overall, their work focused on document-level sarcasm detection through supervised machine learning and contributed by extensively evaluating classifiers, feature sets, and preprocessing techniques for both Czech and English datasets.

Nagwanshi and Madhavan (2014) presented an innovative approach for detecting sarcasm in text, focusing on key features such as semantic content and sentiment reversal. They supplemented their findings with human evaluation results and aimed to enhance their approach by integrating both linguistic and cognitive aspects of text processing. Recognizing the complexity of automatically processing sarcastic statements, they introduced a novel technique that relies on semantics and the length of sentiment progressions. Their research demonstrated the feasibility of distinguishing between sarcastic and negatively toned statements with an accuracy of up to 75%. Although the set of features proposed may not represent the optimal differentiator, the framework they developed holds promise for incorporating additional semantic and cognitive elements into sarcasm detection.

Liu et al. (2014) delved into the nuances of sarcasm within both English and Chinese languages. Detecting sarcasm holds substantial significance for various Natural Language Processing (NLP) applications, such as sentiment analysis, opinion mining, and advertising. They introduced an innovative ensemble learning approach known as the Multi-strategy Ensemble Language Approach (MSELA), to address the challenge of class imbalance. They evaluated their model on datasets in both English and Chinese, their experiments demonstrated superior performance compared to existing sarcasm detection methods and imbalanced classification techniques. Sarcasm, classically defined as the deliberate use of words to convey a meaning contrary to their literal interpretation, served as their focal point. They further introduced an explicit sarcasm feature set, aiming to capture distinctive characteristics of sarcasm in both languages, encompassing low-level and high-level properties. Through a series of experiments, they compared their model with existing approaches, with the ultimate goal of automating sarcasm annotation and uncovering new features, enabling the creation of an enhanced model capable of detecting sarcasm across a variety of textual contexts.

Rajadesingan et al. (2015) developed a novel approach to detect sarcasm on Twitter by incorporating the behavioural traits of users expressing sarcasm. It utilised users' past tweets and drew from behavioural and psychological theories to construct a behavioural modelling framework. Unlike prior studies focusing solely on linguistic cues, this approach combined content analysis with user behaviour analysis. The paper formulated the sarcasm detection problem as discerning whether an unlabeled tweet from a user, accompanied by that user's past tweets, is sarcastic. The proposed SCUBA framework (Sarcasm Classification Using Behavioral Modeling Approach) evaluates the likelihood of a user being sarcastic and employs supervised learning algorithms with constructed features to detect sarcastic tweets. Through experiments, the authors demonstrate SCUBA's effectiveness in identifying sarcastic tweets. SCUBA stood out for considering psychological and behavioural aspects of sarcasm and leveraging historical information, even with limited data, enhancing sarcasm detection efficiency. These findings made SCUBA suitable for real-time applications with computational constraints.

Kunneman et al. (2015) discovered that the inclusion of hashtags effectively reduced the necessity for employing other linguistic indicators such as exclamations and intensifiers to convey sarcasm. They proposed that these explicit markers, like hashtags, function as digital counterparts to non-verbal expressions commonly used in face-to-face interactions to convey sarcasm. Their research involved the development and testing of a system designed to identify sarcastic tweets within a sample of Dutch tweets posted on a single day. Their findings indicated that distinguishing between sarcastic and literally intended tweets proved to be quite challenging. They observed that most tweets carried a genuinely positive message and identified four categories of markers associated with sarcasm: intensified and non-intensified evaluative words, exclamations, and non-sarcastic hashtags. It was noted that intensified evaluative words and exclamations induced hyperbole, yet they occurred less frequently in sarcastic tweets compared to non-intensified evaluative words. While their study shed light on the utilisation of sarcasm in Twitter messages and successfully trained a machine learning classifier for automated detection, they emphasised their focus on hyperbole as a primary linguistic marker used for sarcasm. Notably, they found that typical hyperbole inducers, including exclamations and intensifiers, ranked among the most predictive features of their classifier.

Bamman and Smith (2015) emphasised sarcasm's dependence on shared knowledge between speakers and audiences, highlighting its contextual nature. They demonstrated that incorporating non-linguistic context data from Twitter, including author and audience properties, improved sarcasm detection accuracy over linguistic features alone. Sarcasm's effectiveness hinged on the principle of inferability, as per Kreuz (1996), where speakers used sarcasm when confident it will be understood, typically more among acquaintances than strangers. The author-audience relationship is pivotal, but it becomes intricate on social media, where the audience is unknown or vague. Users often add #sarcasm when unfamiliar with their audience, signalling intent to those who may not discern sarcasm otherwise. Their model used Tweet, Author, Audience, and Response features, where Author features provided the greatest accuracy improvement, reaffirming the role of author-audience interaction in sarcasm recognition. The #sarcasm hashtag served as a communication tool rather than a natural indicator and aided the audience in interpreting sarcasm. Detecting sarcasm without explicit markers required alternative approaches.

Zhang et al. (2016) explored the application of a neural network for detecting sarcasm in tweets. They conducted a comparison between continuous automatic features and discrete manual features in their analysis. Their approach involved employing a bi-directional Gated Recurrent Neural Network (GRNN) to capture both syntactic and semantic information in tweets, along with a pooling neural network to automatically extract contextual features from past tweets. To assess the performance of the neural network, they compared it with a traditional discrete model. They first created a baseline discrete model, incorporating commonly used features from previous literature, such as characteristics of the target tweet content and historical tweets from the same author. They used a GRNN to model tweet content and employed a gated pooling function for feature extraction. To represent significant words from contextual tweets, they utilised pooling for direct feature extraction. Their deep neural network model for tweet sarcasm detection was innovative in that it eliminated the need for manual feature engineering and external resources like POS taggers and sentiment lexicons. Instead, it leveraged distributed embedding inputs and recurrent neural networks to induce semantic features. The neural network model outperformed the state-of-the-art discrete model, demonstrating improved results.

Bouazizi and Ohtsuki (2016) developed a pattern-based approach for sarcasm detection on Twitter. Sarcasm is an ironic form of criticism or teasing, which is even difficult for humans to decipher. They used the idea of sentiment analysis to group the opinions of internet users on a specific topic. They highlighted the necessity of pattern–based features to distribute tweets as sarcastic and non–sarcastic. Sarcasm detection’s necessity and effectiveness depend on accurately identifying sarcasm in sarcastic statements, as per Maynard and Greenwood (2014). Their approach took into account the different types of sarcasm and analysed the tweets regardless of their temporal context. They uncovered the key purposes of sarcasm in social networks and proposed an efficient method to detect tweets. They utilised part-of-speech tags to extract patterns that define the degree of sarcasm in tweets. Their goal was to enhance sentiment analysis accuracy by using this information.

Amir et al. (2016) introduced a novel deep neural network that effectively recognizes and leverages user embeddings in conjunction with lexical cues to identify sarcasm. Their method did not rely on intricate feature engineering; rather, it solely utilised data from users' prior posts. They observed that earlier computational techniques often relied on surface-level patterns in the frequency of specific words, whereas they recognized that the interpretation of a given statement as literal or sarcastic depends heavily on the speaker. To address this, they proposed a fresh approach to sarcasm detection that eliminated the need for extensive manual feature engineering. Their method centred around a neural model that harnessed both content embeddings and context embeddings. They incorporated a vector representation of lexical cues using a convolutional layer. Subsequently, their model acquired user embeddings. This combined technique resulted in the creation of CUE-CNN; a deep neural network designed to broadly analyse sarcastic comments on social media.

Khodak et al. (2017) introduced the Self-Annotated Reddit Corpus (SARC), a groundbreaking dataset for sarcasm research housing 1.3 million self-labelled sarcastic statements annotated with "/s". Diverging from previous datasets, SARC offers a rich tapestry of data, embedding both self-annotations and contextual intricacies like user profiles, topics, and conversation context. Encompassing a staggering 533 million comments, this corpus stands as a beacon for large-scale machine learning endeavours. Confronting challenges such as noise stemming from the "/s" marker, the authors meticulously validate SARC's credibility by comparing it with sources like Twitter and the Internet Argument Corpus. Through meticulous manual evaluations, subtle false positive and false negative nuances emerge. While baseline methods like Bag-of-n-Grams and Sentence Embeddings strive, human evaluators shine, underscoring the complexity of the sarcasm detection task. This paper, heralding a new era, not only presents a substantial resource but also paves the way for the development of cutting-edge sarcasm detection algorithms.

Joshi et al. (2017) presented "Automatic Sarcasm Detection: A Survey" presents a comprehensive overview of computational approaches to detect sarcasm in text. It highlights the complexities of this problem due to the varied ways sarcasm is expressed. The paper is significant for being the first compilation of previous research in this field, identifying three key milestones in sarcasm detection: semi-supervised pattern extraction, hashtag-based supervision, and the incorporation of context beyond the target text.The paper also categorises the approaches used in sarcasm detection, such as rule-based methods that rely on predefined rules, and statistical approaches that use features like sentiment changes. It emphasises the importance of understanding the relationship between sentiment and sarcasm and addresses the issue of data skew in sarcasm-labelled datasets.

Chaudhari and Chandankhede (2017) explored diverse methodologies, ranging from rule-based techniques leveraging hashtags and specific indicators, to statistical approaches integrating lexical, pragmatic, and pattern-based features. Distributional methods, rooted in the Distributional Hypothesis, employ semantic contexts for nuanced understanding. Machine learning classifiers like SVM and deep learning techniques, especially multimodal fusion of textual and visual data, have emerged as powerful tools. Rule-based methods offer precision, statistical approaches blend features for accuracy, and distributional methods provide semantic depth. Machine learning classifiers discern patterns effectively, while deep learning, through multimodal fusion, captures intricate nuances. As research advances, the integration of diverse datasets, innovative feature sets, and hybrid models combining rule-based, statistical, and deep learning approaches are poised to shape the future of sarcasm detection, expanding its applications in sentiment analysis and social media analysis

Bharti et al. (2017) developed Sarcasm Analysis on Twitter Data Using Machine Learning Approaches. Sarcasm detection is one of the challenging task in the field of NLP .They proposed four approaches namely parsing-based lexical generation algorithm, likes and dislikes contradiction, tweet contradicting universal facts, and tweet contradicting temporary facts. After that they also deployed four machine learning classifiers namely support vector machine, Naive Bayes, maximum entropy, and decision tree. Classifiers were trained using extracted features Their work achieved considerable accuracy. Highest accuracy was achieved by PBLGA approach Naive Bayes performed worst out of all the other remaining approaches.

Misra and Arora (2018) examined sarcasm recognition in natural language processing and highlighted the shortcomings of earlier research that used Twitter-based datasets. The authors created a superior dataset that combines news headlines from a mock news website and a legitimate news website in order to overcome these difficulties. They put forth a novel hybrid neural network design that, in terms of classification accuracy, outperforms current models. The research contained an attention module to clarify the model's decision-making process, highlighting the value of interpretability. The authors also make recommendations for future research, which will help to increase sarcasm detection in NLP. These recommendations include an ablation study, transfer learning on the Semeval dataset, and the incorporation of common knowledge.

Kolchinski and Potts (2018) discussed the complex relationship between author characteristics and sarcasm recognition in textual data. In order to capture author-specific nuances in sarcasm expression, the study offers two approaches, Bayesian modelling and dense embeddings, combined within a bidirectional RNN architecture. The research highlighted the effectiveness of these approaches in capturing distinct author trends using the Self-Annotated Reddit Corpus (SARC), demonstrating the usefulness of the Bayesian approach in smaller forums and the dense embedding technique in bigger, more diversified datasets. The study emphasises how crucial it is to take into account author-specific traits while performing language comprehension tasks, and it recommends additional investigation of computational tools to improve our knowledge of user behaviour and contextual language analysis.

Hazarika et al. (2018) illustrated the state-of-the-art in automated sarcasm recognition in online social media conversations. The paper makes the case that sarcasm frequently depends on contextual assumptions and background information, offering a considerable difficulty for detection, despite the fact that existing research has primarily concentrated on lexical, syntactic, and semantic analysis. CASCADE fills this gap by combining context-driven modelling of discourse from discussion threads with user embeddings that include stylometric and personality traits. In comparison to other models, CASCADE outperforms them by incorporating both content and contextual data, especially when it comes to accurately capturing the subtle nature of sarcasm in online forums. The critical role that context and user behaviour play in strengthening sarcasm detection, CASCADE is positioned as a trailblazing hybrid model at the top of this developing study field.

Ahuja et al. (2018) gave a summary of earlier work on the subject of sarcasm detection in text data, highlighting the various methodology and approaches employed by various researchers. It covers a wide range of research topics, including the use of various classifiers and feature sets, the value of contextual information, and neural network-based models. Notably, it highlights how crucial it is to continue learning about users' psychological and behavioural makeup in order to improve the automatic detection of sarcasm. This also demonstrated the Gradient Boosting superior performance as a classifier and the persistent effectiveness of ensemble approaches over single classifier algorithms. The paper emphasises the potential significance of extending sarcasm detection to other social media platforms and expanding feature sets.

Agarwal and An (2018) developed a unique model called Affective Word Embeddings for Sarcasm (AWES) to tackle the difficult task of sarcasm detection in text. They emphasise the importance of affective content in sarcasm detection. They investigated how well affective information may be incorporated into word representations. The authors conducted a thorough analysis of several text domains and discovered that emotion-aware representations are better suited for lengthier texts(such product reviews and forum postings), whereas sentiment-aware representations excel at short text sarcasm detection (such as tweets). By presenting a data-driven strategy for enhancing sarcasm identification and illuminating the importance of emotive word embeddings in tackling this difficult problem, this work makes a contribution to the field.

Subramanian et al. (2019) emphasised the widespread use of text-based communication in social media while also recognizing its shortcomings in expressing emotional cues and the frequent occurrence of ambiguity, particularly when sarcasm is involved. It emphasises how important emojis are as a tool for improving communication and communicating emotions. The majority of current research concentrates on text-based sarcasm detection, with little attention paid to the possible insights provided by emojis. In order to improve the precision of sarcasm detection in social media posts,it highlights the need for a complete framework that incorporates both text and emoji signals. It also highlights the need of comprehending the underlying messages and intentions of users.

Liu et al. (2019) emphasised the difficulty of sarcasm recognition without physical signals by introducing the A2Text-Net, a unique deep neural network for text sarcasm detection. The A2Text-Net methodology surpassed conventional machine learning and deep learning techniques by using auxiliary variables like punctuation and part of speech. In addition to highlighting the importance of sarcasm detection in industries like social media, branding, and customer service, the study also provided a flexible framework that might be modified to operate with various deep learning models. This made a substantial contribution to improving our comprehension of natural language and provided insightful information for sentiment analysis studies in the future.

Castro et al. (2019) introduced the Multimodal Sarcasm Detection Dataset (MUStARD) and argues in favour of including multimodal cues to improve the classification of sarcasm. The research stresses the importance of multimodal fusion, underlines the shortcomings of unimodal approaches, and emphasises the demand for sophisticated spatiotemporal fusion techniques. The importance of modelling multi-party discussions and taking into account conversation-specific aspects for context modelling in sarcasm detection is also highlighted. In order to solve issues with dataset size, overfitting, and subtle contextual analysis, future research should consider the prospect of adding speaker localization and cutting-edge neural approaches. This work provides a valuable resource for future investigations into the topic of multimodal sarcasm detection by utilising instances from the MUStARD dataset.

Cai et al. (2019) tackled the increased demand for sarcasm detection in multi-modal social media messaging, particularly on Twitter, where users can mix text with graphics. The authors offer a novel multi-modal hierarchical fusion model that combines text features, image features, and image attributes. Earlier research mainly concentrated on text-based sarcasm detection. The authors show the value of using picture characteristics as a link between text and images by implementing a fusion technique that improves the representation of each modality. A new dataset for multi-modal Twitter sarcasm detection is introduced in the paper, which also quantitatively emphasises the importance of each modality. The findings highlight the need of taking into account multi-modal information for precise sarcasm detection and suggest potential directions for further research, such as including extra common sense to the detection model and including modalities like audio.

Ren et al. (2020) proposed Sentiment Semantics Enhanced Multi-level Memory Network . They focused on Sentiments semantics as many deep learning models have not fully focused on it, but it is very important to know about sentiment semantics for better performance. They developed a multi-level memory network using sentiment semantics to catch the features of sarcasm detection. To capture sentiment semantics the used first level memory network, second level memory network is used to record the contrast between situation in each sentence and sentiment semantics . They used improved CNN to improve the memory network in the absence of local information.

Potamias et al. (2020) developed a transformer-based approach to irony and sarcasm detection. Figurative language (FL) is present all over the internet on all the social media platforms. Short texts come with greater challenges as it is difficult to understand their literal meaning due to its contrasting and metaphorical content; this remains the unsolved issue in the field of Natural Language Processing. They proposed neural network technology that builds on recently proposed pre-trained transformer-based network architecture which is further enhanced with the employment and devise of a recurrent convolutional neural network. They kept data preprocessing as minimum. They presented the first transformer-based methodology, supporting the pre-trained RoBERTa model combined with a recurrent convolutional neural network, to handle fanciful language in social media. They also wanted to cut down on preprocessing and engineered feature extraction steps which are, according to them, unnecessary when overly trained deep learning methods such as transformers are used.

Yaghoobian et al. (2021) focused on automatically finding sarcasm in written text. Sarcasm detection means figuring out when someone is saying the opposite of what they mean in sentences that express feelings or opinions. In the history of sarcasm research, three major methods stand out: Hashtag-Driven Learning, Semi-Supervised Pattern Finding and Using Extra Information. The paper also talks about a rule-based approach, where computers look for specific signs of sarcasm using preset rules. The rules rely on finding words or ideas that have hidden emotions and label them as "positive" or “negative”. For studying sarcasm, three types of text are used: short text with just one sentence, long text with sarcastic sentences mixed in with regular ones, and transcripts of conversations(TV Shows). Deep learning techniques were also used, where computers learn more about the people writing the text and the topics they discuss to get better at spotting sarcasm.

Razali et al. (2021) focused on spotting sarcasm in tweets using a combination of deep learning features and carefully crafted features related to the tweet's context. They extracted a feature set from a Convolutional Neural Network (CNN) and blended it with handcrafted features to find the best combination. They tested various machine learning techniques to classify these features and found that Logistic Regression worked best for identifying sarcasm. They used five different feature sets: deep features, incongruity features, hyperbolic features, temporality features, and dislike features. However, the temporality and dislike features didn't perform well because they were not very common in the dataset. The informal language used in tweets, including words related to time, events, nouns, verbs, and pronouns, made them difficult to detect. Despite this, these features still contributed significantly to sarcasm detection, and when combined with other features, they could be even more valuable.

Lou et al. (2021) explored a unique way to understand sarcasm. They create two special types of graphs for each sentence: an "affective graph" that captured emotional information and a "dependency graph" based on sentence structure. These graphs helped in identifying how words in a sentence affected each other emotionally. They introduce a framework called Affective Dependency Graph Convolutional Network (ADGCN) to leverage these graphs for sarcasm detection. The ADGCN has three main parts: Constructing Affective and Dependency Graphs: This step creates the emotional graph and a syntax-aware dependency graph for each sentence using both emotional knowledge and sentence structure. Learning Context Representation: They used bidirectional LSTMs (Bi-LSTM) to understand the context of the sentence and create vector representations for it. Learning Graph Representation: they used multi-layer Graph Convolutional Networks (GCNs) to make sense of the emotional connections between words in the sentence for sarcasm detection.

Handoyo et al. (2021) proposed a smart way to spot sarcasm on Twitter. They used a powerful model called RoBERTa and boosted their dataset with Global Vector representations (GloVe) to improve word understanding and context. This helped them create more data and balance their dataset. They checked how well their model did on four different datasets. Two of them had tweets with specific hashtags (Ptacek and Ghosh), collected automatically. The other two datasets (iSarcasm and SemEval-18) were manually reviewed and included tweets from surveys and tweets marked as sarcastic by third-party reviewers. The interesting part was that when they added more data to the sarcastic tweets, it didn't make the model better at recognizing sarcasm, but it did help it get better at finding non-sarcastic text. They found that RoBERTa worked well for spotting sarcasm in various tasks.

Eke et al. (2021) suggested a way to spot sarcasm in text using a mix of deep learning and traditional machine learning. Sarcasm can make the tone of text seem opposite to what it really means, making it difficult for sentiment analysis. Understanding the context is a big challenge in handling such content. Three benchmark datasets were used which were provided by different researchers. They tried three different models for classifying the data:Bi-LSTM with GloVe Embedding: They used a deep learning approach with a Bidirectional LSTM (Bi-LSTM) and GloVe word embeddings to create context vectors for words. BERT Model: The second model relied on a powerful model called BERT, which is great at understanding context. Feature Fusion Model: This one combined BERT features, sentiment-related data, syntactic (language structure) information, and GloVe word embeddings with traditional machine learning. They evaluated their models using four important measures: precision, recall, accuracy, and F-measure. This research is notable for being the first to blend BERT features, word embeddings, and linguistic and sentiment-related information for sarcasm detection.

Bedi et al. (2021) researched on the Hindi-English code-mixed conversational dialog. Two major contributions were made by them: First, a dataset called MaSaC, which has content in both Hindi and English, for detecting sarcasm in a multi-modal way. Second, a new neural network model called MSH-COMICS that was good at classifying sentences. The model used a special attention mechanism that helped it focus on small parts of a sentence at a time, making it efficient. A context layer was also added that looked at the conversation history to make better sense of the text. Sometimes, understanding sarcasm or humour depends on the context. For code-mixed content (mixing languages like Hindi and English), first it is figured out which language each word is in, then only a text can be processed more effectively for different tasks. MaSaC is a fresh dataset that combines both languages in a multi-modal setting, and then MSH-COMICS is used to classify the sentences. The tests showed that this approach with enhanced textual representation worked well and improved the performance in identifying sarcasm and humour.

AL-An et al. (2021) aimed to improve the detection of sarcasm using two techniques: LSTM and Auto-Encoder. They used a standard dataset and applied some preprocessing steps to the text. The key to their success was the Auto-Encoder. The tool took the document's information and tried to create the same information as output. This process helped the model learn the essential details that strongly influenced sarcasm. They also used Long Short-Term Memory (LSTM) to feed the embedding of each word into the model, which helped in creating an overall understanding of the document. The Auto-Encoder can dig deep into the data, uncovering hidden patterns and relationships through a process of encoding and decoding. This helped the model understand how different features are connected, improving sarcasm detection.

Baroiu and Matu (2022) conducted a systematic literature review on the evolution of automatic sarcasm detection from its inception in 2010 to the present day. They explored the interdisciplinary nature of this work, drawing from Artificial Intelligence, specifically Natural Language Processing (NLP). NLP has made significant progress over the past decade, with sentiment analysis being a key area of focus in both academia and industry. The challenge they addressed was how to handle figurative language, like sarcasm and irony, within sentiment analysis. Sarcasm, defined as a form of irony expressing a negative or critical attitude, relies heavily on contextual understanding. Sarcasm can introduce noise into data analysis, leading researchers to develop models for sarcasm detection, which has become a subfield within sentiment analysis and NLP. The paper's core aim was to provide a systematic literature review of this subfield, and offer valuable insights to NLP researchers. This review assisted researchers in staying current with the latest advancements in sarcasm detection, aiding them in selecting appropriate approaches for their specific tasks.

Ren et al. (2023) proposed a knowledge-augmented neural network model for sarcasm detection.They investigated a knowledge-augmented [neural network](https://www.sciencedirect.com/topics/social-sciences/neural-network) model that leverages the contextual information of the original text from external knowledge source, for sarcasm detection. Firstly, from the external knowledge source context is extracted for the original text. After that original text and contest fetched at the initial step are processed sequentially through the embedding and encoding layers ,to automatically learn high-level [semantic representation](https://www.sciencedirect.com/topics/computer-science/semantic-representation). After these steps the softmax layer is used to output the classification probability. Evaluating this model on Semeval-2018 Task 3 dataset,82.79% F1 score was achieved. Experimental analysis also indicated the importance of contextual information in sarcasm detection.

# **RESEARCH GAPS**

* Enhancing the performances of sentiment analysis and opinion mining is still one of the goals.
* Research on automatic annotation of models to detect more sarcastic sentences in different kinds of text is needed.
* Integration of linguistic and cognitive features is needed to improve the performance of models.
* Deep learning approaches were less explored.
* Explore new ways to identify stereotypical negative activity because this type of world knowledge is essential to analyse sarcasm.
* Recognizing the intensity of the negativity may also be useful to distinguish strong contrast from weak contrast.

# **OBJECTIVES**

We have found the following objectives from the above research gaps:

1. Corpus creation.
2. To extract features based on linguistic and cognitive factors.
3. To manually train the deep learning model.
4. To evaluate the model on the basis of precision, recall, f-measures and accuracy.

# **PROPOSED METHODOLOGY**

## **Corpus Creation:**

Corpus creation is a major step in sarcasm detection because it is necessary to select the proper data for training. There are three types of datasets available on the internet viz short text, long text and other datasets according to Joshi et al. (2017). Short texts are like tweets and threads, which have length limitations set by the platforms. Long texts are like discussion forums and posts and other datasets are like text from novels or tv shows.

We will be using a dataset from the series ‘The Big Bang Theory’ and ‘F.R.I.E.N.D.S.’ which are American sit-com dramas, having abundance of sarcastic dialogues which will be helpful in training and testing the model. We took reference for these datasets from MUStARD[[2]](#footnote-2). It is a dataset of 11424 lines containing sarcastic and non-sarcastic dialogues. We will be training data manually by marking them sarcastic or non-sarcastic.

For Example:

**Sarcastic sentence:** *I love working on Sundays.*

**Non-sarcastic sentence:** *I love Fridays.*

Example of Dataset:

1. SHELDON: “If you're compiling a mix CD for a double suicide. Oh, I hope that scratching post is for you.”,

CONTEXT: “Oh, good Lord.”,

“God, that's a good song.”

SHOW: “BBT”,

SARCASM: True

1. SHELDON: “Not to mention you'd have to power down on Saturdays.”,

CONTEXT: “I'm with you. I just have to make sure if I'm a synthetic human I'd still be Jewish.”,

“I promised my mother.”,

“I suppose you could have your android penis circumcised.”,

“But that's something your rabbi would have to discuss with the manufacturer.”

SHOW: “BBT”,

SARCASM: False

## **Extraction of Linguistic and Cognitive factors:**

Linguistic features refer to the characteristic or element of the language that can be analysed and studied in communication. These features are essential for understanding how language works and how it is used for various purposes. Linguistic factors play a crucial role in sarcasm detection. Sarcasm is a form of figurative language where the intended meaning is opposite to the literal meaning of the words used. They are used to identify the cues and markers that may indicate the presence of sarcasm in a statement. Some key linguistic factors on we will be working on, are word choice, tone of voice, contextual incongruity, hyperbole, negation, situation awareness, emoji, punctuation, repetition and echoing. Cognitive function is a broad term that refers to mental processes involved in the acquisition of knowledge, manipulation of information, and reasoning. These features are central to human cognition and have significant implications for psychology, education, and other fields. Cognitive features are relevant in sarcasm detection, as understanding sarcasm often requires more than just linguistic analysis; it involves grasping the speaker's or writer's intended meaning and the underlying cognitive processes. Some cognitive features that play an important role in sarcasm detection are contextual awareness, emotion recognition, emotional intelligence, common sense reasoning, incongruity Detection and theory of mind. Sarcasm detection often combines linguistic and cognitive features, as well as machine learning techniques, to make accurate assessments. For instance, machine learning models can be trained on large datasets to recognize patterns in language and context that indicate sarcasm, and cognitive elements such as theory of mind and context comprehension can be integrated into these models to improve their accuracy.

## **Deep Learning Model:**

Deep learning, a subset of machine learning, demonstrates superior performance when dealing with data that lacks a predefined structure. Deep learning enables computational models to perform categorization tasks using the supplied text, voice, or images while continuously learning from the given data. Deep learning models can achieve cutting-edge precision that occasionally even surpasses that of a human being. Neural network topologies with numerous layers and enormous amounts of labelled data are used to train models. However, recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer architecture-based deep learning models, in particular, have shown promise in collecting the contextual information and patterns required for comprehending sarcasm.

Despite the promise of these deep learning techniques, sarcasm recognition is still an unsolved research issue. Even with well-developed deep learning models, it can be difficult to create models that can comprehend and interpret humour, context, and linguistic nuance. Additionally, as adequately labelled ironic data is frequently scarce and arbitrary, gathering and labelling datasets for training such algorithms can be a challenge.

The BERT (Bidirectional Encoder Representations from Transformers) model due to its capacity to recognize complex contextual links inside language, is frequently used for sarcasm detection. In particular, BERT's bidirectional nature enables it to comprehend linguistic nuances, such as the subtleties and complexities that are frequently distinctive to sarcastic phrases. BERT may successfully spot patterns and contextual indicators that suggest the existence of sarcasm in text by thoroughly evaluating the surrounding words and their relationships. BERT is also well suited for tasks that need a thorough understanding of language semantics, such as recognizing sarcasm, because it underwent pre-training on a variety of large-scale datasets. It has the ability to parse text using "common sense" that is largely human-like.

## **Evaluating The Model:**

The models are evaluated using four important measures: precision, recall, F-measure, and accuracy. There are three important variables that should also be declared: True Positive (TP) is the number of words that have been classified correctly, False Positive (FP) is the number of words that have been classified incorrectly and True Negative (TN) is the number of words that have not been classified

Precision (PRE) provides model accuracy and focuses on how accurate the model's positive predictions are.

Recall (REC) measures the model's ability to identify all the actual positive instances.

F-measure (F-M) is a cumulative factor to test the overall effect of the recall and precision in order to find the overall impact of false negative instances and false positive instances over the whole accuracy.

Accuracy (ACC) provides the percentage ratio of the predicted instances. It measures the overall correctly classified instances.

# **Tools and Techniques**

* Python,
* FastText,
* GloVe,
* Scikit-Learn,
* Tensorflow,
* Keras,
* Streamlit (for UI)

# **Work Plan**

The Gantt Chart below represents our work plan for the project, and the time is in fortnight starting from 15th Sept. 23 to 15th April 23

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **S. No.** | **Objective** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| 1. | Literature Review |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2. | Corpus Development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3. | Extracting Linguistic and Cognitive Features |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4. | Development of Deep Learning Model |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5. | Evaluating the Model |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6. | Report Writing |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Fortnight 1-3:** Literature Review

**Fortnight 2-11:** Corpus development as a continuous process of project.

**Fortnight 7-11:** Extracting Linguistic and Cognitive Factors alongside corpus development.

**Fortnight 10-13:** Developing the deep learning model using BERT while extracting the factors.

**Fortnight 13-14:** Evaluation of the linguistic and cognitive factors, and deep learning model.

**Fortnight 14-15:** Report Writing.

# **Expected Outcomes**

We wish to develop a web-application powered by contextual embeddings that will be capable of distinguishing between non-sarcastic and sarcastic sentences. The application will take an input sentence from the user and will predict the most appropriate class (non-sarcastic or sarcastic) of the sentence.

# **REFERENCES**

Agrawal, A., & An, A. (2018, June). Affective representations for sarcasm detection. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 1029-1032).

Ahuja, R., Bansal, S., Prakash, S., Venkataraman, K., & Banga, A. (2018). Comparative study of different sarcasm detection algorithms based on behavioural approach. Procedia computer science, 143, 411-418.

AL-Ani, M. M., Omar, N., & Nafea, A. A. (2021). A Hybrid Method of Long Short-Term Memory and Auto-Encoder Architectures for Sarcasm Detection. J. Comput. Sci, 17(11), 1093-1098.

Amir, S., Wallace, B. C., Lyu, H., & Silva, P. C. M. J. (2016). Modelling context with user embeddings for sarcasm detection in social media. *arXiv preprint arXiv:1607.00976*.

Bamman, D., & Smith, N. (2015). Contextualized sarcasm detection on twitter. In *proceedings of the international AAAI conference on web and social media* (Vol. 9, No. 1, pp. 574-577).

Băroiu, A. C., & Trăușan-Matu, Ș. (2022). Automatic sarcasm detection: Systematic literature review. *Information*, *13*(8), 399.

Bedi, M., Kumar, S., Akhtar, M. S., & Chakraborty, T. (2021). Multi-modal sarcasm detection and humor classification in code-mixed conversations. IEEE Transactions on Affective Computing.

Bharti, S. K., Pradhan, R., Babu, K. S., & Jena, S. K. (2017). Sarcasm analysis on twitter data using machine learning approaches. Trends in Social Network Analysis: Information Propagation, User Behavior Modeling, Forecasting, and Vulnerability Assessment, 51-76.

Bouazizi, M., & Ohtsuki, T. O. (2016). A pattern-based approach for sarcasm detection on twitter. *IEEE Access*, *4*, 5477-5488.

Cai, Y., Cai, H., & Wan, X. (2019, July). Multi-modal sarcasm detection in twitter with hierarchical fusion model. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 2506-2515).

Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., & Poria, S. (2019). Towards multimodal sarcasm detection (an \_obviously\_ perfect paper). arXiv preprint arXiv:1906.01815.

Chaudhari, P., & Chandankhede, C. (2017, March). Literature survey of sarcasm detection. In 2017 International conference on wireless communications, signal processing and networking (WiSPNET) (pp. 2041-2046). IEEE.

Eke, C. I., Norman, A. A., & Shuib, L. (2021). Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model. IEEE Access, 9, 48501-48518.

Handoyo, A. T., & Suhartono, D. (2021). Sarcasm detection in Twitter--performance impact while using data augmentation: Word embeddings. arXiv preprint arXiv:2108.09924.

Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., & Mihalcea, R. (2018). Cascade: Contextual sarcasm detection in online discussion forums. arXiv preprint arXiv:1805.06413.

Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic sarcasm detection: A survey. ACM Computing Surveys (CSUR), 50(5), 1-22.

Khodak, M., Saunshi, N., & Vodrahalli, K. (2017). A large self-annotated corpus for sarcasm. arXiv preprint arXiv:1704.05579.

Kolchinski, Y. A., & Potts, C. (2018). Representing social media users for sarcasm detection. arXiv preprint arXiv:1808.08470.

Kunneman, F., Liebrecht, C., Van Mulken, M., & Van den Bosch, A. (2015). Signalling sarcasm: From hyperbole to hashtag. *Information Processing & Management*, *51*(4), 500-509.

Liu, L., Priestley, J. L., Zhou, Y., Ray, H. E., & Han, M. (2019, December). A2text-net: A novel deep neural network for sarcasm detection. In 2019 IEEE first international conference on cognitive machine intelligence (CogMI) (pp. 118-126). IEEE.

Liu, P., Chen, W., Ou, G., Wang, T., Yang, D., & Lei, K. (2014). Sarcasm detection in social media based on imbalanced classification. In *Web-Age Information Management: 15th International Conference, WAIM 2014, Macau, China, June 16-18, 2014. Proceedings 15* (pp. 459-471). Springer International Publishing.

Lou, C., Liang, B., Gui, L., He, Y., Dang, Y., & Xu, R. (2021, July). Affective dependency graph for sarcasm detection. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1844-1849).

Maynard, D., Dupplaw, D., & Hare, J. (2013). Multimodal sentiment analysis of social media.

Misra, R., & Arora, P. (2019). Sarcasm detection using hybrid neural networks. arXiv preprint arXiv:1908.07414.

Nagwanshi, P., & Madhavan, C. V. (2014, October). Sarcasm detection using sentiment and semantic features. In *International Conference on Knowledge Discovery and Information Retrieval* (Vol. 2, pp. 418-424). SCITEPRESS.

Potamias, R. A., Siolas, G., & Stafylopatis, A. G. (2020). A transformer-based approach to irony and sarcasm detection. Neural Computing and Applications, 32, 17309-17320.

Ptáček, T., Habernal, I., & Hong, J. (2014, August). Sarcasm detection on Czech and English twitter. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 213-223).

Rajadesingan, A., Zafarani, R., & Liu, H. (2015, February). Sarcasm detection on twitter: A behavioural modelling approach. In *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 97-106).

Razali, M. S., Halin, A. A., Ye, L., Doraisamy, S., & Norowi, N. M. (2021). Sarcasm detection using deep learning with contextual features. IEEE Access, 9, 68609-68618.

Ren, L., Xu, B., Lin, H., Liu, X., & Yang, L. (2020). Sarcasm detection with sentiment semantics enhanced a multi-level memory network. Neurocomputing, 401, 320-326.

Ren, Y., Wang, Z., Peng, Q., & Ji, D. (2023). A knowledge-augmented neural network model for sarcasm detection. Information Processing & Management, 60(6), 103521.

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013, October). Sarcasm as contrast between a positive sentiment and negative situation. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 704-714).

Subramanian, J., Sridharan, V., Shu, K., & Liu, H. (2019). Exploiting emojis for sarcasm detection. In Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12 (pp. 70-80). Springer International Publishing.

Yaghoobian, H., Arabnia, H. R., & Rasheed, K. (2021). Sarcasm detection: A comparative study. arXiv preprint arXiv:2107.02276.

Zhang, M., Zhang, Y., & Fu, G. (2016, December). Tweet sarcasm detection using deep neural networks. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: technical papers* (pp. 2449-2460).

1. <https://www.merriam-webster.com/dictionary/sarcasm> [↑](#footnote-ref-1)
2. https://github.com/soujanyaporia/MUStARD/blob/master/data/sarcasm\_data.json [↑](#footnote-ref-2)