Sarcasm Detection

Submitted in the partial fulfillment of the requirements

for the degree of B.Tech in Computer Engineering

by

Atharva Badhe(21CE1306)

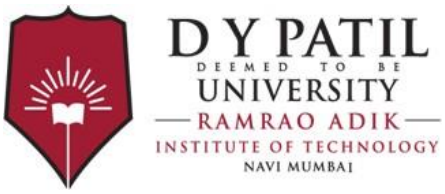
Yashodeep Sonar (21CE1409)

Sairaj Panwalkar (21CE1419)

Varun Rane (21CE1382)

Supervisor

Ms. Sakshi Somani



Department of Computer Engineering

Ramrao Adik Institute of Technology

Sector 7, Nerul, Navi Mumbai

(Under the ambit of D. Y. Patil Deemed to be University)

May 2024

Ramrao Adik Institute of Technology

(Under the ambit of D. Y. Patil Deemed to be University)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706

CERTIFICATE

This is to certify that, the Mini Project-IV report entitled

Sarcasm Detection

is a bonafide work done by

Atharva Badhe(21CE1306)

Yashodeep Sonar (21CE1409)

Sairaj Panwalkar (21CE1419)

Varun Rane (21CE1382)

and is submitted in the partial fulfillment of the requirement for the degree of

B.Tech in Computer Engineering

to the

D. Y. Patil Deemed to be University

Supervisor	Project Co-ordinator
(Ms. Sakshi Somani)	**(Dr. Ekta Sarda)**
Head of Department	Principal
(Dr. Amarsinh V. Vidhate)	**(Dr. Mukesh D. Patil)**

Mini Project Report - IV Approval

This is to certify that the Mini Project - IV entitled ***"Sarcasm Detection using deep learning "*** is a bonafide work done by ***Atharva Badhe (21CE1306)***, ***Yashodeep Sonar (21CE1409)***, ***Sairaj Panwalkar (21CE1419)***, and ***Varun Rane (21CE1382)*** under the supervision of ***Ms. Sakshi Somani***. This Mini Project is approved in the partial fulfillment of the requirement for the degree of ***B.tech in Computer Engineering***

Internal Examiner :

1. ……

2. ……

External Examiners :

1. ……

2. ……

Date : …/…/……

Place : ……

DECLARATION

I certify that the content of this written work is original to me and free of any instances of plagiarism. Wherever I have included someone else's thoughts or words, I have properly cited and referenced the original sources. Additionally, I affirm that we have followed all standards of academic honesty and integrity and that I have not created or faked any idea, data, fact, or source in my work. I am aware that any infraction of the aforementioned will result in the Institute taking disciplinary action against me. It may also result in penalties from the sources who were improperly cited or from whose sufficient permission was not obtained when required.

Date: _____

Atharva Badhe(21CE1306)

Yashodeep Sonar (21CE1409)

Sairaj Panwalkar (21CE1419)

Varun Rane (21CE1382)

Abstract

Sarcasm is the use of words that are intended to convey the exact opposite meaning, usually to offend, irritate, or just for laughs. This project aims to recognize sarcasm in plain text. It is challenging to automatically identify sarcasm in text for a variety of reasons, including the absence of the statement's context, the speaker's tone, their expressions and body language, and their character and personality, among other factors. Any of these factors could have contributed to the perception of a statement as sardonic. But the idea intends to take use of the general sarcastic statement's ability to express opposing feelings by using Natural Language Interpretation. The project's goal is to teach a machine learning model to distinguish between regular and sarcastic sentences.

Contents

List of Figures

Chapter 1

Introduction

Sentiment analysis is the process of examining and identifying the sentiment that underlies a text. To do this, it combines the ideas of natural language processing (NLP) with machine learning. Sentiment analysis allows us to identify whether a piece of text contains positive, negative, or neutral sentiment. It is an effective artificial intelligence (AI) approach with a wide range of commercial applications.

You can use it, for instance, to examine client comments. Following the acquisition of feedback or reviews via a variety of channels, you may apply sentiment analysis algorithms to those text segments in order to get the sentiment score and, consequently, comprehend your target audience's perspective regarding the product.

When there is a lot of unstructured data, sentiment analysis can be used to automatically tag and classify the data. The method of Net Promoter Score (NPS) surveys is widely employed to acquire insights into consumer perceptions of products and services. Sarcasm plays an important function in human social interaction and can be both adversely insulting and constructively entertaining.

Sarcasm is often used in social media to express unfavorable opinions with positive or exaggeratedly positive wording. As a result, unless sentiment analysis models are specifically designed to account for this possibility, they can easily be fooled by sarcasm. Because of this intentional ambiguity, sarcasm identification is an essential part of sentiment analysis. Sarcasm detection is said to be a binary classification problem.

Sarcasm identification in sentiment analysis is very difficult if the situation's context, the targeted issue, and the surrounding environment are not well understood. It could be difficult for a machine or a human to understand.

1.1 Overview

Deep learning sarcasm detection is a multi-step procedure that begins with the collection of a heterogeneous dataset that comprises utterances that are tagged as sardonic or not. Following the capture of the data, common preprocessing methods like as tokenization, lowercase conversion, and punctuation removal are used to standardize the textual data. After that, the text is converted into numerical representations using techniques such as word embeddings to aid in the understanding of deep learning models.

After it has been trained, the model might go through additional fine-tuning procedures to improve its performance. This refinement could include experimenting with new architectures, modifying hyperparameters, or applying transfer learning strategies to modify previously trained models for sarcasm detection tasks. After the model reaches acceptable performance thresholds, real-world applications can use it to detect sarcasm in a variety of settings, such as sentiment analysis, social media monitoring, and customer feedback analysis. Even with these improvements, sarcasm detection is still difficult since it depends on context, cultural quirks, and subtle language cues—domains where continuous research aims to improve detection systems' precision and resilience.

1.2 Motivation

The impetus behind deep learning sarcasm detection is the requirement to precisely decipher the subtleties of human communication in digital contexts. Sarcasm has become more common as social media platforms and internet forums have grown exponentially. Unfortunately, sarcasm is frequently misinterpreted because of its nuance and reliance on context, which can cause misunderstandings and even confrontations. A viable approach is provided by deep learning models, which use massive datasets to identify the intricate patterns and linguistic clues that indicate sarcasm. These models can improve sentiment analysis, customer feedback analysis, and contextually relevant responses for chatbots and virtual assistants by reliably detecting sarcastic remarks.

Deep learning's use in sarcasm detection is also in line with the larger objectives of improving communication technology and natural language comprehension. Researchers hope to stretch the limits of language comprehension by teaching models to detect sarcasm, which would help machines better understand the subtleties of human emotion.

1.3 Objectives

The main goal of deep learning-based sarcasm detection is to build reliable computer models that can recognize sarcasm in textual input. Researchers are working to create systems that can accurately and automatically distinguish between sardonic and non-sarcastic utterances by utilizing deep learning techniques like neural networks and natural language processing. Improved sentiment analysis in social media monitoring, improved customer feedback analysis in businesses, and more contextually aware replies in conversational AI systems like chatbots and virtual assistants are just a few of the many uses for this goal across multiple domains. The main objective is ultimately to improve understanding and communication in digital contexts by giving machines the ability to detect and react to sarcasm.

Sarcastic language is subtle and context-dependent, therefore achieving this goal requires overcoming a number of sarcasm detection-related obstacles. For deep learning models to effectively identify the complex verbal clues that indicate sarcasm, they need to be trained on huge and diverse datasets. To guarantee generalizability across many domains and applications, these models must also demonstrate robustness to changes in language use, cultural contexts, and individual communication styles. To raise the precision and dependability of sarcasm detection systems, research endeavors concentrate on optimizing algorithms, investigating novel architectures, and capitalizing on advances in natural language processing.[1]

The effective application of deep learning for sarcasm detection has important ramifications for many different domains. Making correct identifications of sarcastic remarks in social media analysis can yield more nuanced insights into public opinion, which in turn can help corporations and policymakers make more educated decisions. Comparably, in the examination of customer feedback, organizations can enhance overall satisfaction by successfully addressing consumer problems by differentiating between real and sarcastic criticism. Furthermore, by enabling more sympathetic and contextually relevant responses, sarcasm detection integrated into conversational AI systems improves user experience and eventually expands the potential of AI-driven communication solutions.

1.4 Organization of the report

The format of the report is as follows: The Chapter 2 reviews the literature. In Chapter 3, the problem with the system is defined. This covers the classification of problems, suggested technologies, device architecture, and necessary hardware and software. The project is covered in Chapter 4, along with an analysis of the results and screenshots of the operational model. However, Chapter 5 outlines the inference and upcoming research on the method to be applied as a better model. The report concludes with a plagiarism report and a weekly progress report.

Chapter 2

Literature Survey

An increasing amount of research and development is being put into the difficult task of sarcasm detection by deep learning, as seen by the rising corpus of literature in this field. Prominent research works cover many aspects of sarcasm detection, from model architectures and assessment techniques to dataset generation. While Barbieri et al. (2014) suggested a deep learning technique using convolutional neural networks (CNNs) specifically designed for tweet sarcasm detection, Bamman et al. (2015) investigated behavioral modeling approaches on Twitter. In order to capture the complex structure of text data, Ghosh et al. (2017) created hierarchical attention networks. Meanwhile, Su et al. (2020) built on the success of BERT by developing BERTweet, a customized pre-trained language model for Twitter data. Furthermore, Joshi et al. (2016) investigated frameworks for multi-task learning that included auxiliary activities like as Sentiment analysis to improve the performance of sarcasm detection. [2]

Literature Survey of Existing Systems

T1	i) Contrast between positive sentiment and negative situation ii) Contrast between negative sentiment and positive situation iii) Contrasting connotations iv) Verity negation v) Temporal facts extraction
T2	i) Wit ii) Snivel iii) Prevarication iv) Rampant
T3	i) Prosodic variations ii) Structural variations iii) Lexical analysis
T4	i) Language expertise ii) Environment expertise
T5	i) Likes and dislikes prediction

2.1 Survey of Existing System

An analysis of current deep learning systems for sarcasm detection demonstrates the variety of strategies and techniques used to address this challenging issue. To create efficient sarcasm detection systems, researchers have used a variety of deep learning architectures, such as transformer-based models like BERT and GPT, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Using large-scale annotated datasets with sardonic and non-sarcastic text examples for model training and evaluation is a common feature shared by these approaches.

Evaluation criteria including accuracy, precision, recall, F1-score, and area under the ROC curve are frequently used to gauge how well these systems operate. Despite the notable advancements in deep learning-based sarcasm detection, issues including cultural disparities, domain adaptation, and the ambiguity of sarcasm still exist. In order to increase sarcasm detection accuracy, future research efforts might concentrate on resolving these issues, creating more interpretable models, and investigating multimodal techniques that combine text with other modalities including audio and visual inputs.

2.2 Limitations of Existing System

Limitations of the existing sysytem are :

1. **Contextual Understanding** : Deep learning models could have trouble picking up on the subtleties of sarcasm in context, which could result in misunderstandings in some circumstances.
2. **Dataset Bias** : The capacity of the models to generalize may be impacted by biases present in the sarcasm detection datasets, such as a bias towards particular domains or cultural situations.
3. **Ambiguity** : Deep learning algorithms find it difficult to detect sarcasm effectively, particularly when the intended meaning is not evident, because sarcasm frequently contains nuanced language clues and ambiguity.
4. **Class Imbalance** : When sarcastic and non-sarcastic samples in a dataset are not balanced, it can be difficult to train a model, which can result in predictions that are biased and poor performance on minority classes.

5. **Domain Adaptation** : To guarantee reliable performance in a variety of scenarios, models trained in one domain might not generalize well to other domains. This calls for the use of domain adaptation approaches.

6. **Interpretability** : The absence of interpretability in deep learning models makes it challenging to comprehend how they make their predictions, which can undermine their usefulness and credibility in practical applications.

7. **Computational Complexity** : Deep learning models for sarcasm detection might be difficult to scale and make available due to their high computational demands during training.

8. **Cultural Variations** : Models trained on datasets from particular cultural backgrounds may find it difficult to generalize their findings to other cultures because sarcastic expressions can differ between cultures and languages.

Chapter 3

Proposed System

3.1 Problem Statement

The challenge lies in developing precise deep learning models capable of identifying sarcasm within text. The most appropriate deep learning architecture must be chosen, a variety of labeled datasets must be obtained, and complex language cues and cultural variances must be captured. The creation of efficient feature extraction techniques, attention mechanisms, and transfer learning tactics are important goals since they will boost model performance in a variety of applications and domains, eventually improving digital environment communication.

3.2 Proposed Methodology

Proposed Methodology are :

1. **Data Collection** : The initial stage entails compiling an extensive dataset of text examples that are both satirical and non-satirical. To guarantee diversity and representativeness, this dataset should come from a variety of sources and disciplines.

2. **Preprocessing** : To standardize the format and eliminate noise, the data is first collected and then goes through preprocessing. This entails deleting stopwords (often used words like ”the,” ”is,” etc.) and extraneous features like punctuation. It also entails tokenization, which divides text into individual terms or tokens.[3]

3. **Feature extraction** : Is the process of turning the preprocessed text data into numerical vectors that the deep learning model can comprehend. Word embeddings, which depict words as dense vectors in a continuous space and capture the semantic links between words, are

commonly used to do this.

4. **Model Selection** : The best model for sarcasm detection is found by examining various deep learning architectures. Recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based models such as BERT or GPT are popular options, and each has advantages and disadvantages of its own.

5. **Model Training** : Using the preprocessed and feature-extracted dataset, the chosen model is trained. Using gradient descent and backpropagation to modify its internal parameters, the model learns to map the input text to the appropriate sarcasm label (sarcastic or non-sarcastic) during training.

6. **Evaluation** : Following training, a different test dataset is used to assess the trained model's performance using a variety of metrics, including accuracy, precision, recall, and F1-score. This stage aids in evaluating the model's ability to generalize to fresh, untested data.

7. **Fine-tuning** : To increase the model's accuracy even more, fine-tuning approaches might be used if its performance isn't up to par. This could entail utilizing transfer learning from previously trained models, adjusting hyperparameters, or testing out other architectures.

8. **Deployment** : The model can be used for sarcasm detection in practical applications if it reaches an acceptable level of performance. This can entail keeping an eye on social media discussions, reviewing client testimonials, or improving chatbot and virtual assistant functionality to better comprehend and react to ironic comments.

3.3 System Design

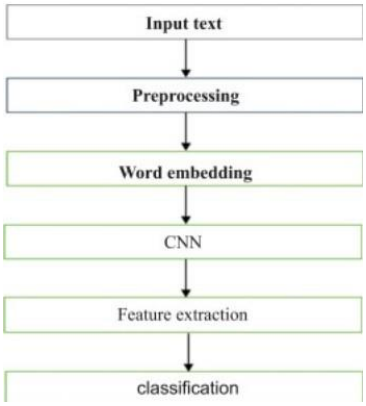


Figure 3.1: System Design

3.4 Details of Hardware/Software Requirement

Software Requirement :

1. OS : Windows OS
2. Programming Language : Python
3. Software : VS Code
4. Libraries : Libraries for Sarcasm Detection are as follows :

i. Streamlit : Is a Python toolkit that makes it simple to create and share data-driven web apps, dashboards, and machine learning applications using straightforward Python scripts.

ii. Pandas : An effective Python data manipulation and analysis toolkit that offers data structures like DataFrame and Series for working with structured data in an efficient manner. It can be used to load, clean, preprocess, and analyze datasets.

iii. re (Regular Expressions) : Facilitates the manipulation of strings and matches patterns, enabling operations like text replacement, pattern extraction, and input data validation using predefined patterns.

iv. string : A Python module that offers standard string formatting, case conversion, and character set access functions, among other string operations and constants.

v. nltk (Natural Language Toolkit): The Natural Language Toolkit, or nltk, is a framework that allows Python applications to be built to work with data from human languages. It provides text processing modules for tasks like tokenization, stemming, lemmatization, and part-of-speech tagging, along with interfaces to corpora and lexical resources.

Hardware Requirement :

1. Processor : Intel® i3 2.7GHz and above.
2. Ram : 4 GigaByte and above.
3. Hard disk : 256 Gigabyte and above.

Chapter 4

Results and Discussion

The findings and debate on deep learning approaches for sarcasm detection point to encouraging developments in the precise recognition of sarcastic comments in textual data. Recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer-based architectures like BERT and GPT are examples of deep learning models that work well at identifying subtle verbal signals and contextual data that are essential for sarcasm detection. Even with great accuracy in controlled settings, problems including class imbalance, cultural differences, confusing utterances, and biased datasets remain, necessitating further research to address them. In order to overcome these issues and eventually progress the creation of more precise, dependable, and interpretable sarcasm detection models for better communication in digital contexts, efforts are being made to increase interpretability and investigate multimodal techniques.[4]

4.1 Implementation Details

Data collection and preprocessing are usually the first steps in the deep learning implementation of sarcasm detection process. To do this, compile a dataset of text samples—both sardonic and non-sarcastic—from a variety of sources. Then, clean the dataset by tokenizing, lowercasing, and eliminating extraneous features like stopwords and punctuation. Subsequently, the preprocessed text data is transformed into numerical representations by feature extraction techniques, such as word embeddings, which extract the semantic content of words. The best method for sarcasm detection is then determined by examining a variety of deep learning architectures, such as transformer-based models like BERT and GPT, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). [5]

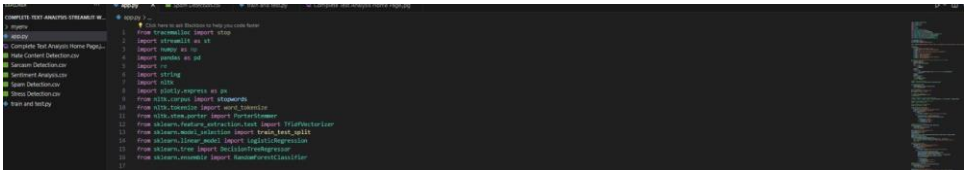


Figure 4.1: Used libraries for Sarcasm Detection

Above are the libraries used in programmer shown in figure 4.1.

Figure 4.2: Web page

This image illustrates how libraries are used in our ”Sarcasm Detection” project and implies that there is another image that offers documentation or visual proof of how libraries have been included in the ”Sarcasm Detection” project. ”Libraries” in this context probably refers to code modules or software libraries that improve the project’s functioning in figure 4.2.

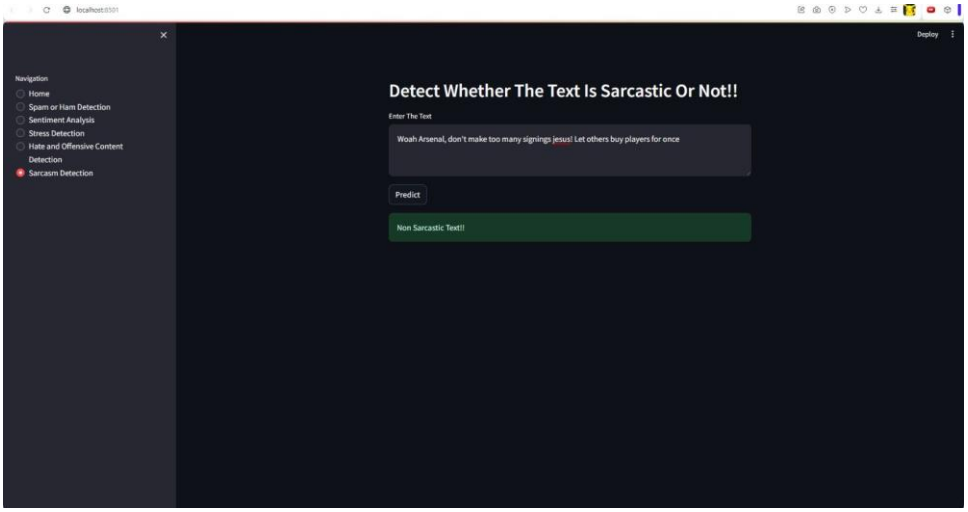


Figure 4.3: Opened web application

The suggested approach makes it easier for users by enabling them to launch several navigation with only in one click shown in figure 4.3.

Chapter 5

Conclusion and Future Work

Identifying sarcasm is one of the main challenges in sentiment analysis. In this paper, we made an effort to give a summary of the several sarcasm detection attempts conducted in the in the past, utilizing different dataset formats, sarcastic identification techniques, and some challenges with the process. Although the results from the multimodal dataset are superior to those from the textual dataset, they are still not optimal. The ability to recognize sarcasm has grown in importance in recent years. Memes have grown in popularity as a means of disseminating sarcastic messages, yet it may still be difficult to determine if a statement is sarcastic or not. [6]

It is challenging to determine whether or not someone is being sarcastic if you don't have any prior information or understanding of their body language or facial expressions. Since most studies on sarcasm detection are done in English, we have also evaluated some studies that were conducted in other languages. Memes are now widely shared on many social media platforms, therefore we also need to take into account typographic and graphical pictures that may include sarcasm. It raises expectations for possible future work to detect sarcasm in typographic and infographic pictures with and without other feature sets.

References

[1] M. S. Razali, A. A. Halin, L. Ye, S. Doraisamy, and N. M. Norowi, "Sarcasm detection using deep learning with contextual features," *IEEE Access*, vol. 9, pp. 68609–68618, 2021.

[2] C. I. Eke, A. A. Norman, and L. Shuib, "Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model," *IEEE Access*, vol. 9, pp. 48501–48518, 2021.

[3] A. Kumar and G. Garg, "Empirical study of shallow and deep learning models for sarcasm detection using context in benchmark datasets," *Journal of ambient intelligence and humanized computing*, vol. 14, no. 5, pp. 5327–5342, 2023.

[4] Y. Y. Tan, C.-O. Chow, J. Kanesan, J. H. Chuah, and Y. Lim, "Sentiment analysis and sarcasm detection using deep multi-task learning," *Wireless personal communications*, vol. 129, no. 3, pp. 2213–2237, 2023.

[5] S. Kedar *et al.*, "Sarcasm detection using deep learning," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 1S, pp. 57–64, 2021.

[6] S. K. Bharti, R. K. Gupta, P. K. Shukla, W. A. Hatamleh, H. Tarazi, and S. J. Nuagah, "Multimodal sarcasm detection: a deep learning approach," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–10, 2022.

Appendix A

Weekly Progress Report

Appendix B

Plagiarism Report

Acknowledgments

We would like to sincerely thank Dr. Mukesh D. Patil, Principal of Ramrao Adik Institute of Technology, and Dr. Amarsinh V. Vidhate, Head of the Computer Engineering Department, for your invaluable assistance in finishing my research, Sarcasm Detection. We would especially want to thank Ms. Sakshi Somani, our project supervisor, and Dr. Ekta Sarda, our mini project coordinator, for their time and efforts during the year. We were able to complete the assignment with the aid of your insightful counsel and recommendations. We will always be appreciative of your help in this regard. We would like to thank that the group members,i.e., Atharva Badhe (21CE1306), Yashodeep Sonar (21CE1409), Sairaj Panwalkar (21CE1419), and Varun Rane (21CE1382), and not by someone else.

Date: _____