IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

SARCASM Classifier

Dr. Saroja T. V[1], Prathamesh Wagh[2], Aditi Veer[3], Srushti Thombre[4]
Faculty, Department of Computer Engineering[1]
Student, Department of Computer Engineering[2,3,4]
Shivajirao S Jondhale College of Engineering, Dombivli (E),Thane, Maharashtra, India

Abstract: *Sarcasm is a form of verbal irony where the intended meaning of a statement differs from its literal meaning. Detecting sarcasm is crucial for understanding sentiments and opinions, especially in social media analysis. However, sarcasm detection is challenging due to its reliance on context and tone. In this paper, we propose a sarcasm detection model using a Long Short-Term Memory (LSTM) network, a deep learning-based Recurrent Neural Network (RNN) variant. Our model is trained on social media datasets containing sarcastic and non-sarcastic statements. We compare its performance with the Word2Vec algorithm, analyzing accuracy and effectiveness. The proposed system aims to improve sarcasm classification in text-based data.*

Keywords*:* Sarcasm Detection, Machine Learning, LSTM, Word2Vec, Social Media Analysis

I. INTRODUCTION

1.1 Overview

Sarcasm is often used to express humor, criticism, or irony, making it difficult for traditional text-processing algorithms to detect. Identifying sarcasm in text is crucial for improving sentiment analysis, customer feedback systems, and opinion mining. Existing sarcasm detection models struggle with context-based sarcasm, requiring advanced Natural Language Processing (NLP) techniques. In this study, we employ LSTM-based deep learning to detect sarcasm in textual data, leveraging word embeddings for enhanced performance.

1.2 Problem Statement

Traditional sarcasm detection models struggle with ambiguous expressions, lack contextual awareness, and have high false-negative rates. This research aims to develop a machine learning model that enhances sarcasm classification in textual data using LSTM networks

1.3 Objectives

- Improving Sarcasm Detection Accuracy: The primary objective of the Sarcasm Classifier is to enhance the precision of sarcasm detection by leveraging deep learning models such as Long Short-Term Memory (LSTM) networks.
- Enhancing Natural Language Processing (NLP) Capabilities: The project seeks to advance NLP techniques by integrating Word2Vec embeddings, text preprocessing, and contextual analysis.
- Providing Real-Time Sarcasm Detection and Insights: The Sarcasm Classifier aims to offer real-time analysis of text inputs, enabling users to instantly determine whether a given statement contains sarcasm.
- Optimizing Text Processing and Feature Extraction: The project focuses on efficient data preprocessing techniques, such as removing stopwords, punctuation, and special characters.

II. LITERATURE REVIEW

Author	Date of Publish	Disadvantages	Advantages
Sunusi Kabir, Adamu Habu	March 2023	The used algorithm is not suitable for large data sets. It does not	SVM works well when there is a clear margin-of separation between

,Badamasi Imam Abdullahi Madaki		perform very well when the data has more noise.	classes
Hafsa Ahmad, Wasif Akbar Naeem Aslam, Azka and Mohsin	July 2023	It does not capture position in text, semantics, cooccurrences in different documents	- Effective handling of complex queries. - Reduced response time. - Scalability in system architecture.
Tay Y. , Tuan L.A, Hui S , C	January 2019	Helps better in analyse temporal and sequential data. Helps better in analyse temporal and sequential data.	- Context-aware responses lead to better user satisfaction. - Broader applicability with multi-language - Efficient conversation flow.
Amir S., Wallace B.C	November 2016	Scaling to new languages requires new embedding matrices.	It jointly learns and exploits embeddings for the content and users
Parmar et al	April 2018	Multiple iterations and intermediate results are not supported.	It has characters and special text which includes Interjections and intervening text in the sentence.
Ren Y., Ji D	August 2018	Accuracy is very low	It can capture more sarcastic model

III. METHODOLOGY

1. Dataset: - First get the dataset from the Kaggle dataset and upload it in the system. Total two dataset in the form of .json file is uploaded and both the datasets are concatenated.
2. Cleaning the Data : Second step is the processing and cleaning of the data. Cleaning and data processing is implemented because it will be difficult for the machine learning algorithm to do further classifications. Cleaning the data includes removing all the links, flags and emojis from the following tweets, comments and headlines. Converting all the upper cases into lower cases because machine learning algorithms works well if it is provided smaller case letters. Remove abbreviations and punctuations and symbols too. Also remove all the stop-words from the English language. Examples of stop-words are “to”, “ the ”, “ a” etc. After all the data cleaning the system will be left with a proper cleaned list of words..
3. Train-Test split :In train-test split, the validation split will be equal to 0.2 which means 20% of the data will be reserved for testing purpose and the remaining 80% of the data will be used for training purpose.
4. Loading the LSTM model and building the Embedding layer : Now using LSTM model the cleaned data in the form of words will be converted into vectors containing fixed dimensions. So it will result into sequence of vectors rather than sentences. This is achieved by mapping words into meaningful space where the distance between words is related to semantic similarity. This is nothing but the LSTM model.
5. Building the Recurrent Neural Network (RNN) :Now once the words are converted into sequence of vectors will pass these sequences to the Recurrent Neural Network. The category of recurrent neural network used in the proposed system is LSTM(long short term memory). The main reason to select LSTM model is because the LSTM itself a type of RNN and recurrent neural networks are very good at learning sequences. This model will be helpful in producing the predictions.
6. Training and visualizing :The training doesn’t take long time. It will carry out total of 25 epochs throughout the training process. An epoch is the number of passes a training dataset takes around an algorithm . Here epoch will be useful to get the validation loss and accuracy . Once the validation loss and accuracy are obtained we visualize the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

trainings which will compare both training loss and validation loss and will also compare training accuracy and validation accuracy.

7. Sarcasm Detection :A function will be made and used to check whether the model that is created is able to detect sarcasm or not. In this function we just need the pass the sentence as a parameter and check whether the following sentence is sarcastic or not

IV. SYSTEM DESIGN

Processor (CPU):
Minimum: Intel i5 (10th Gen) / AMD Ryzen 5.
Recommended: Intel i7/i9 (12th Gen) / AMD Ryzen 7/9.
RAM:
Minimum: 16GB
Recommended: 32GB
Storage:
Minimum: 512GB SSD
Recommended: 1TB SSD
Graphic card:
Minimum: NVIDIA GTX 1650
Recommended: NVIDIA RTX 3060/3070 (CUDA-enabled)
Operating System (Windows):
Minimum: Windows 10 (64-bit)
Recommended: Windows 11 (latest version)
Additional Requirements:
WSL 2 (Windows Subsystem for Linux) for compatibility
Python, TensorFlow/PyTorch, Gensim, NumPy, Pandas
Jupyter Notebook or Google Collab for development

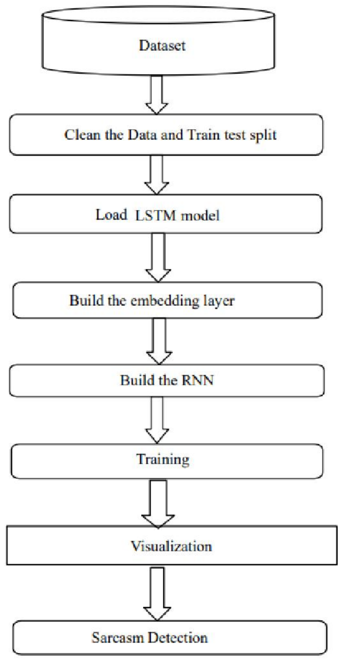
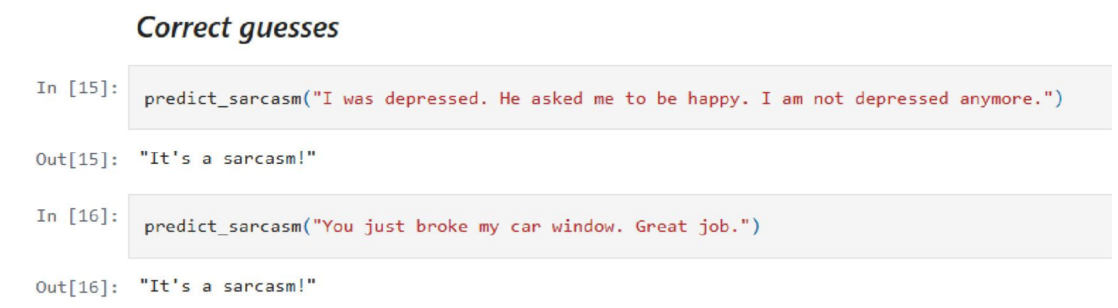


Figure 4.2 Block Diagram

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

V. RESULTS & ANALYSIS

Experiments were conducted over multiple epochs to evaluate model performance. LSTM models achieved higher accuracy compared to traditional machine learning classifiers.



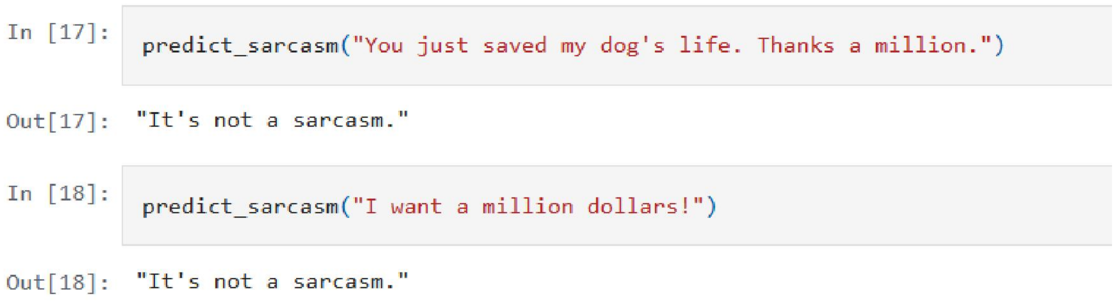
Correct guesses

```
In [15]:  predict_sarcasm("I was depressed. He asked me to be happy. I am not depressed anymore.")

Out[15]:  "It's a sarcasm!"

In [16]:  predict_sarcasm("You just broke my car window. Great job.")

Out[16]:  "It's a sarcasm!"
```



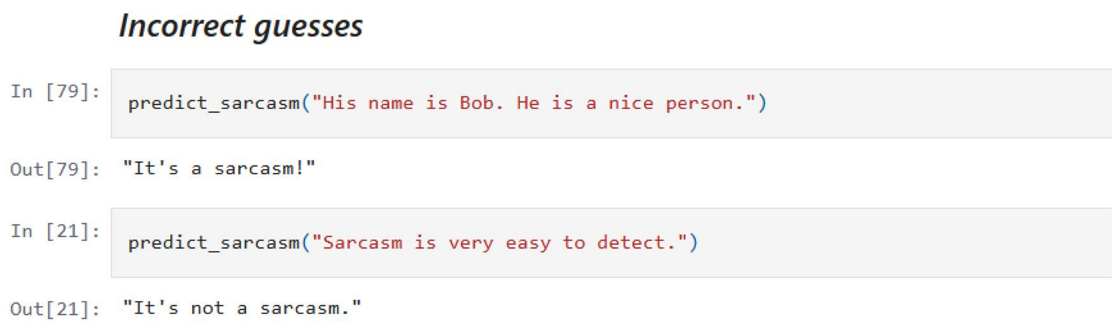
```
In [17]:  predict_sarcasm("You just saved my dog's life. Thanks a million.")

Out[17]:  "It's not a sarcasm."

In [18]:  predict_sarcasm("I want a million dollars!")

Out[18]:  "It's not a sarcasm."
```

Fig 5.1 Correct guesses



Incorrect guesses

```
In [79]:  predict_sarcasm("His name is Bob. He is a nice person.")

Out[79]:  "It's a sarcasm!"

In [21]:  predict_sarcasm("Sarcasm is very easy to detect.")

Out[21]:  "It's not a sarcasm."
```

Fig 5.2 Incorrect guesses

VI. CONCLUSION

Conclusion

In conclusion, sarcasm detection using algorithms like LSTM in NLP and machine learning projects holds significant potential across various applications. Whether it's improving customer service interactions, analyzing social media sentiment, moderating online forums, or enhancing virtual assistants, the ability to accurately detect sarcasm in text can lead to more effective communication and decision-making. However, developing robust sarcasm detection models requires careful consideration of dataset quality, feature engineering, algorithm selection, and model evaluation. Overall, sarcasm detection represents an exciting and challenging area within the field of NLP and machine learning, with numerous opportunities for innovation and practical application in various domains.

Future Work

Future research should focus on refining algorithms, incorporating contextual information, and exploring multi-modal approaches to further improve the accuracy and reliability of sarcasm detection systems. Despite these challenges, the utilization of LSTM offers valuable insights and opportunities for advancing the field of sarcasm detection, ultimately contributing to more sophisticated and nuanced natural language processing capabilities..

REFERENCES

[1] Sunusi Kabir, Adamu Habu, Badamasi Imam, Abdullahi Madaki (2023). Sentiment analysis of sarcasm detection in social media.
[2] Hafsa Ahmad, Wasif Akbar, Naeem Aslam, Azka Mohsin (2023). TF-IDF Feature Extraction based Sarcasm Detection on Social Media.
[3] Parmar K., Limbasiya N., Dhamecha M. (2018). Feature-based Composite Approach for Sarcasm Detection.
[4] Ren Y., Ji D. (2018). Context-augmented Convolutional Neural Networks for Twitter Sarcasm Detection.
[5] Tay Y., Tuan L.A., Hui S. (2018). Reasoning with Sarcasm by Reading In-Between the Lines.
[6] Poria S., Cambria E., Hazarika D., Vij P. (2016). A deeper look into sarcastic tweets using deep convolutional neural networks.