IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

SARCASM Detection on Twitter Data

Shaikh Mohammad Sofiyan Vasim, Rohit Vilas Wahatule, Aditya Hanumant Vibhute,

Ameya Arjun Patil, Prof. Suvarna Bahir

Department of Computer Engineering, STES'S Sinhgad Academy of Engineering Kondhwa, Pune, University of Pune, Maharashtra, India

ABSTRACT: Sarcasm is a type of evaluation that consists of words that generally mean the opposite of what you truly want to say and can be used in a variety of situations, such as being upfront, displaying disturbance, or being clever. Sarcasm is occasionally used within communication to convey known material as an individual transmits. Sarcasm, like analysis or a joke, can be used in a variety of situations. Sarcasm, in any case, is extremely harmful for people to perceive. As a result, detecting sarcasm leads to a better understanding of the user's sentiment analysis, which is based on data obtained from websites such as Twitter and Facebook. The advantages of recognising sarcasm for opinion mining have fueled interest in automated sarcasm detection. Identification of sarcasm as an exploratory subject Automated sarcasm recognition software attempts to determine whether a text is sarcastic or neutral. The study article is divided into two parts. In the first phase, it extracts characteristics related to feelings and punctuation, and then the chi-square test is used to identify the most useful features. The top 200 tf-idf characteristics are retrieved and paired with sentiment and punctuation features to detect sarcastic material within the tweet in the second step. The first strategy achieves the highest level of accuracy by utilising .

I. INTRODUCTION

The research article is divided into two sections. It extracts characteristics related to feelings and punctuation in the first step, and then uses the chi-square test to identify the most useful features. In the second step, 200 top tf-idf features are retrieved and combined with sentiment and punctuation features to detect sarcastic material within the tweet. The Support Vector Machine algorithm, with a value of 74.59, yields the best accuracy in the first method.

II. LITERATURE SURVEY

"Machine Learning-Based Sarcasm Detection on Twitter Data," Neha Pawar and Sukhada Bhingarkar. [1] Sarcasm is a subtle kind of irony that is extensively employed in social media. It is typically used to send concealed information in order to criticise and insult a person as well as to recognise them. The sarcastic rearrangement approach is extremely beneficial for improving automated sentiment analysis obtained from various social networks and microblogging sites. Sentiment analysis refers to the expressed attitudes and views of internet users in a certain community, as well as their identification and aggregation. A pattern-based technique employing Twitter data is suggested in this study to identify sarcasm. Four sets of traits are presented that include a lot of particular sarcasm and classify tweets as sarcastic or non-sarcastic. The proposed feature sets are investigated and their extra cost classes are evaluated

Jitendra Kumar, Rahul Gupta "A Statistical Approach for Detecting Sarcasm Using Twitter Data." [2] Sarcasm is a different type of evaluation that consists of words that often mean something opposite to what you genuinely want to say and may be used in a variety of situations, such as to offend, to cause upset, or to appear witty. Sarcasm is occasionally used inside the communication as an individual transmits to convey known material. Sarcasm, like analysis or a joke, may be employed in a variety of contexts. Sarcasm, in any case, is incredibly detrimental for individuals to perceive. As a result, the identification of sarcasm leads to a better comprehension of the user's sentiment analysis, based on data acquired from websites such as Twitter and Facebook. The benefits of recognising sarcasm for opinion mining have fueled interest in automated sarcasm identification as an investigation topic. The software in Automated sarcasm recognition attempts to determine if a text is sarcastic or neutral. There are two parts to the study article. It extracts characteristics linked to feelings and punctuation in the first phase, and then uses the chi-square test to identify the most useful features in the second phase. In the second step, the top 200 tf-idf characteristics are retrieved and paired with sentiment and punctuation features to detect sarcastic material inside the tweet. The first strategy achieves the highest accuracy by employing the Support Vector Machine algorithm, which has a value of 74.59

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

Rafeeque P C, Sreelakshmi K "A Reliable Method for Detecting Sarcasm in Tweets" [3] Sarcasm detection systems are useful for applications such as sentiment analyzers, review processing systems, and natural language processing systems. The suggested system is a sarcasm detection model that uses lexical, pragmatic, context incongruity, subject, and sentiment variables to identify sarcasm in text. Despite the fact that we employ context incongruity as the primary classification feature, our system can recognise sarcastic tweets with and without context incongruity. The suggested system is modelled using support vector machines (SVM) and a decision tree, and both yielded encouraging results.

Joseph Herve Balanke ,Haripriya V, "EXTENSION OF THE LEXICON ALGORITHM FOR SARCASM DETECTION. "[4] Lexicon algorithm is used to determine the sentiment expressed by a textual content. This sentiment might be negative, neutral or positive. It is possible to be sarcastic using only positive or neutral sentiment textual contents. Hence, lexicon algorithm can be useful but yet insufficient for sarcasm detection. It is necessary to extend the lexicon algorithm in order to come out with systems that would be proven efficient for sarcasm detection on neutral and positive sentiment textual contents. In this paper, two sarcasm analysis systems both obtained from the extension of the lexicon algorithm have been proposed for that sake. The first system consists of the combination of a lexicon algorithm and a pure sarcasm analysis algorithm. The second system consists of the combination of a lexicon algorithm and a

"EXTENSION OF THE LEXICON ALGORITHM FOR SARCASM DETECTION," Joseph Herve Balanke and Haripriya V. [4] To determine the sentiment indicated by textual material, the Lexicon algorithm is utilised. This feeling might be unpleasant, neutral, or favourable. Only positive or neutral feeling textual elements can be used to be sarcastic. As a result, while the lexical approach might be beneficial, it is insufficient for detecting sarcasm. It is required to enhance the lexicon algorithm in order to develop systems that are effective at detecting sarcasm in neutral and positive sentiment textual materials. For that purpose, two sarcasm analysis systems derived from the lexicon algorithm are proposed in this study. The first system is made up of a lexicon algorithm and a pure sarcasm analysis algorithm. The second system is made up of a lexicon algorithm and a sentiment prediction algorithm..

[5] Jihad aboobaker and E.Ilavarasan, "A Survey on Sarcasm Detection and Challenges." Sarcasm is a form of emotional expression in which people say or write something that is radically different from what they intended. Sarcasm is extremely difficult to identify due to its obscurity. Irony is a sort of sarcasm. One of the primary purposes of sarcasm is to criticise. Sarcasm is commonly used to communicate one's ideas or sentiments, particularly on social networking platforms such as Twitter and Facebook. Perfect analysis and comprehension of sarcastic phrases can increase sentiment analysis accuracy. Understanding the attitude or opinions of individuals or society about a specific event or issue is referred to as sentiment analysis. We attempted to discuss the overall architecture of sarcasm detection, existing approaches, and distinct forms of sarcasm, concerns, challenges, and future scope in this study.

"Multi-modal Sarcasm Detection and Humor Classification in Code Mixed Conversations," Manjot Bedi, Shivani Kumar, Md Shad Akhtar, and Tanmoy Chakraborty. [6] Sarcasm detection and comedy categorization are inherently delicate problems, owing to their reliance on contextual and nonverbal cues. Furthermore, prior research in these two areas are mostly limited to non-English languages such as Hindi due to a lack of qualitative and notated datasets. Given the limits mentioned above, we make two key contributions in this work: (1) We provide the first Hindi-English code-mixed dataset, MaSaC1, for multi-modal sarcasm detection and humour classification in conversational dialogue; (2) We propose MSH COMICS2, an attention-rich neural architecture for utterance classification. We develop efficient utterance representation by employing a hierarchical attention mechanism that attends to only a tiny fraction of the input phrase at a time. In addition, we provide a dialogue level contextual attention technique to use the dialogue history for multi-modal categorization. Extensive trials are carried out for both tasks by altering multi-modal inputs and various MSH-COMICS submodules. We also compare new techniques to old ones. MSH-COMICS outperforms previous models by 1 F1-score point in sarcasm detection and 10 F1-score points in humour categorization. To understand the superiority and dangers of our approach, we diagnose it and conduct a detailed study of the outcomes.

III. PROPOSED SYSTEM

How to leverage the current one's output to improve sentiment analysis and opinion mining performance. The proposal aims to develop an efficient method for detecting sarcastic tweets and to investigate how to leverage this information (whether the tweet is sarcastic or not) to improve the accuracy of sentiment analysis.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

SYSTEM ARCHITECTURE

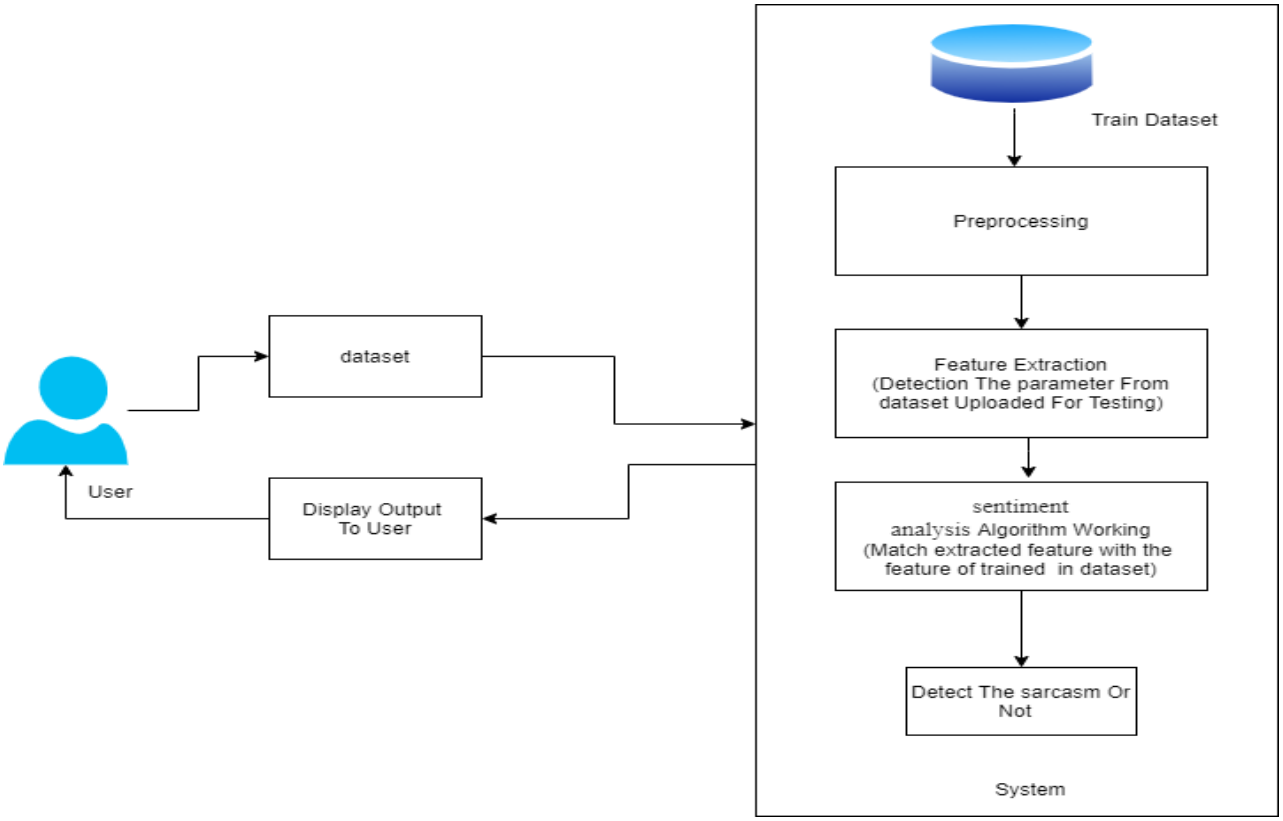


Fig. System Architecture

IV. ALGORITHM

sentiment analysis:-

- • DeepForest appears to be the best option for models that are not based on neural networks. As more research is conducted on both neural network and non-neural network-based models, the accuracy of sentiment analysis and classification tasks is bound to improve. Python sentiment analysis is a way for evaluating text to determine the sentiment buried within it. This is accomplished through the use of a combination of machine learning and natural language processing (NLP).
- Sentiment analysis examines the emotions represented in a piece of writing. Sentiment Analysis is the most often used text categorization technique, analysing an incoming message and determining if the underlying sentiment is favourable, negative, or neutral. By experimenting with the demo, you may enter a statement of your choosing and measure the underlying mood

V. SDLC MODEL

The software development cycle consists of several phases, including design, implementation, and deployment. This section describes the various phases of the software development model. The SDLC model for project development is depicted in the figure below. The waterfall SDLC model was chosen because it is simple to implement and works well for this project.

Requirements Analysis: At this stage, the business requirements, definitions of use cases, and associated documentation are studied.

Design: At this stage, the designs for the data models will be defined, and various data preparation and analysis will take place. Implementation: At this stage, the model's actual development will take place. Appropriate algorithms, mathematical models, and design patterns will be used to develop the agent's back-end and front-end components based on the data model designs and requirements from previous stages.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

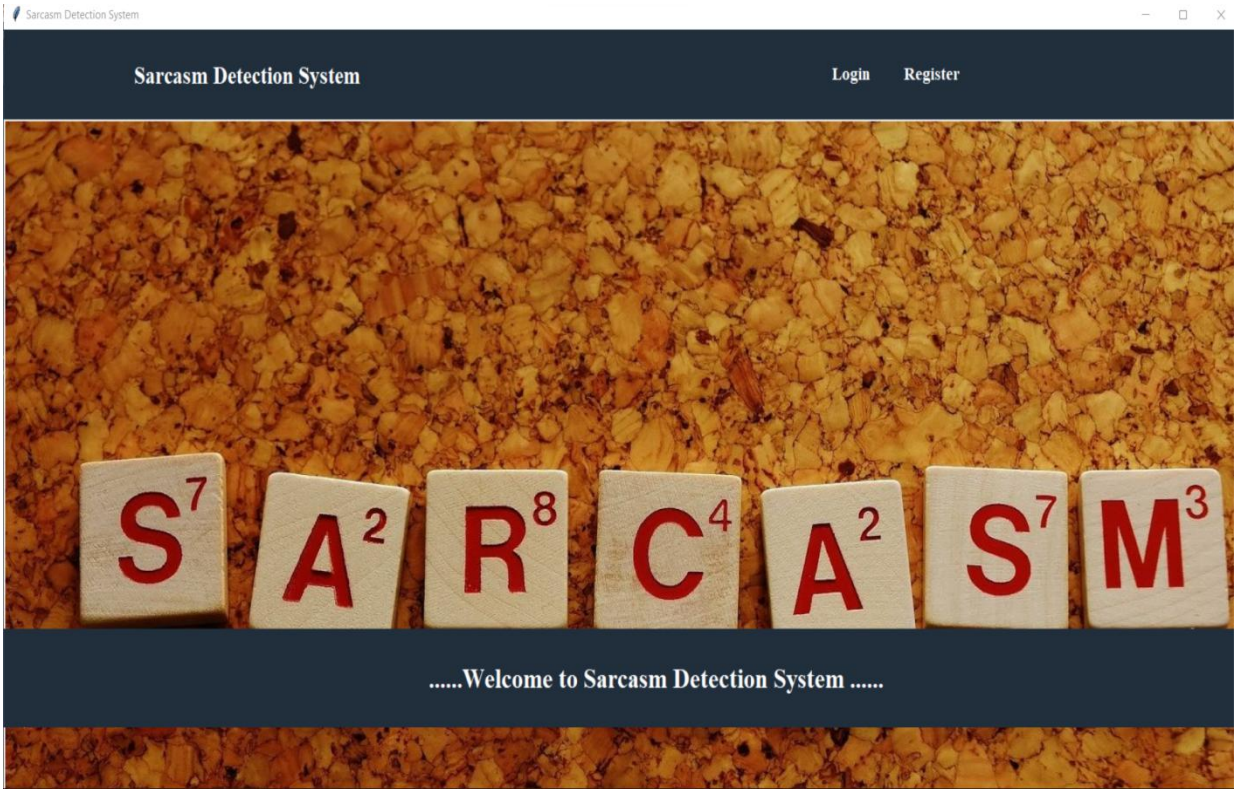
| DOI: 10.15680/IJIRCCE.2022.1004141|

Testing :In this stage, the developed model based on the previous stages will be tested. Several validation t ests will be run on the trained model. Deployment: Once the model has been validated for accuracy scores, it is ready for deployment or use in simulated scenarios.

Maintenance: Various inputs/scenarios will be countered by the model during the use of the developed solution, which may affect the model's overall accuracy. Or, as time passes, the model may no longer meet the needs of the business. As a result, the model must be maintained on a regular basis in order to maintain its desired state of operation.

VI. RESULT

1. GUI MAIN:

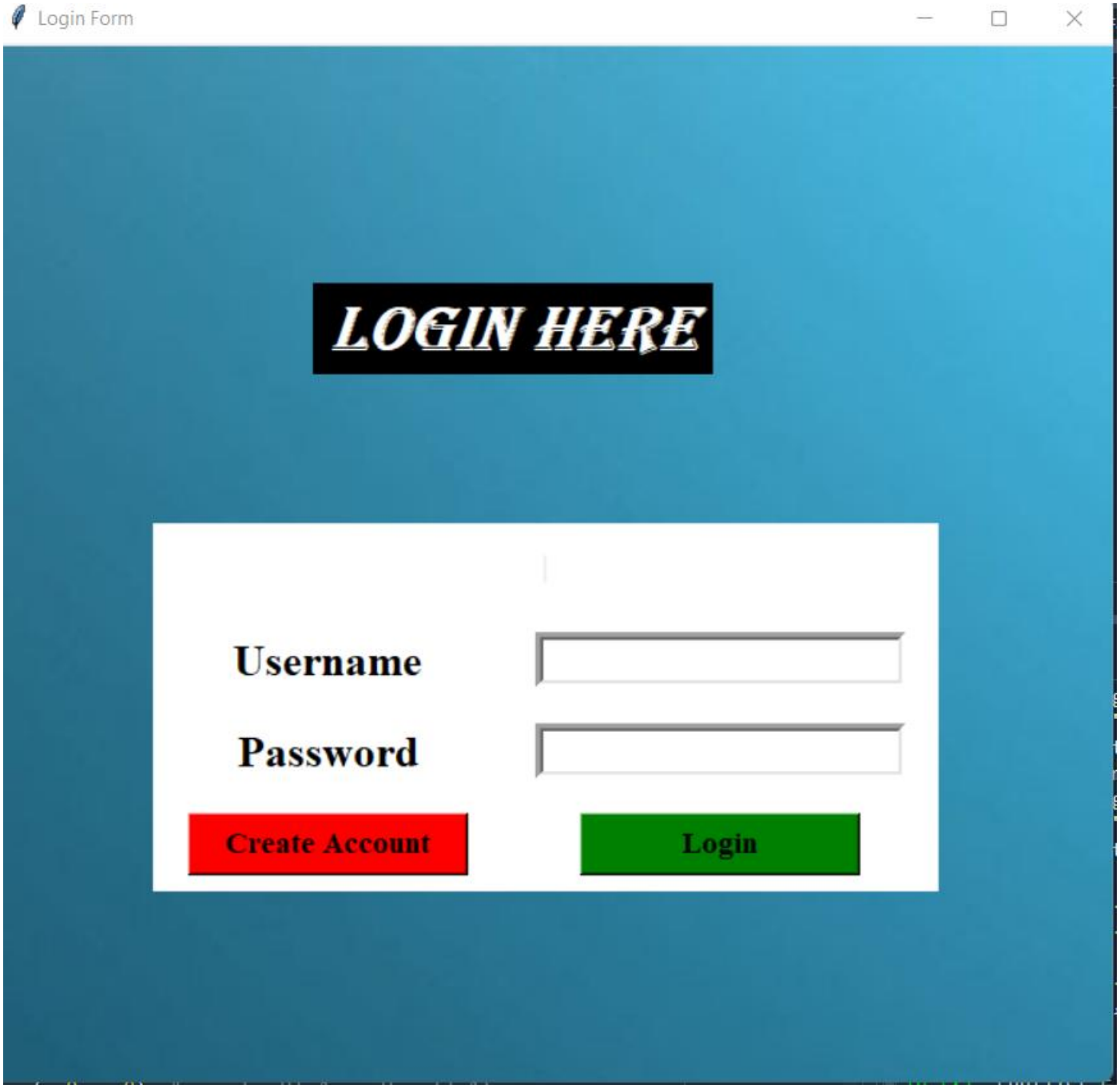


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

2.LOGIN:



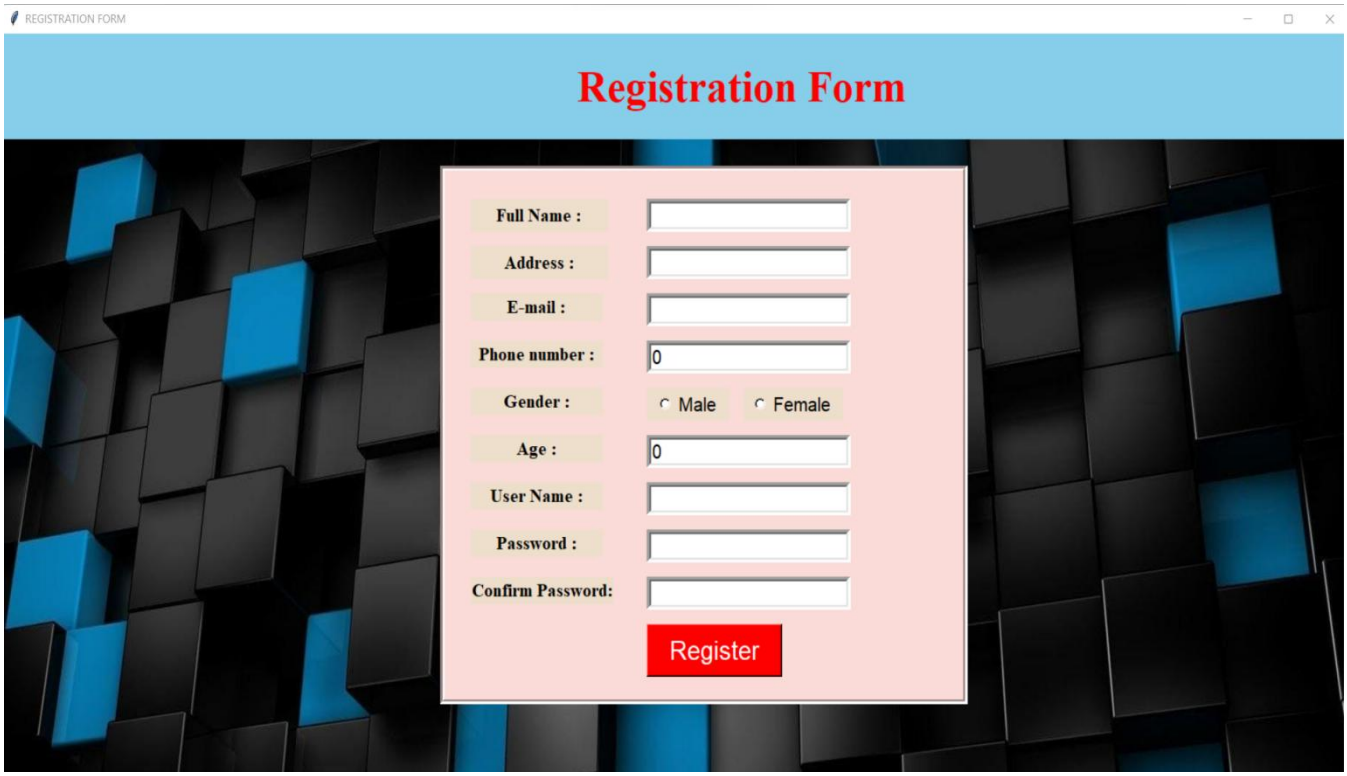
International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

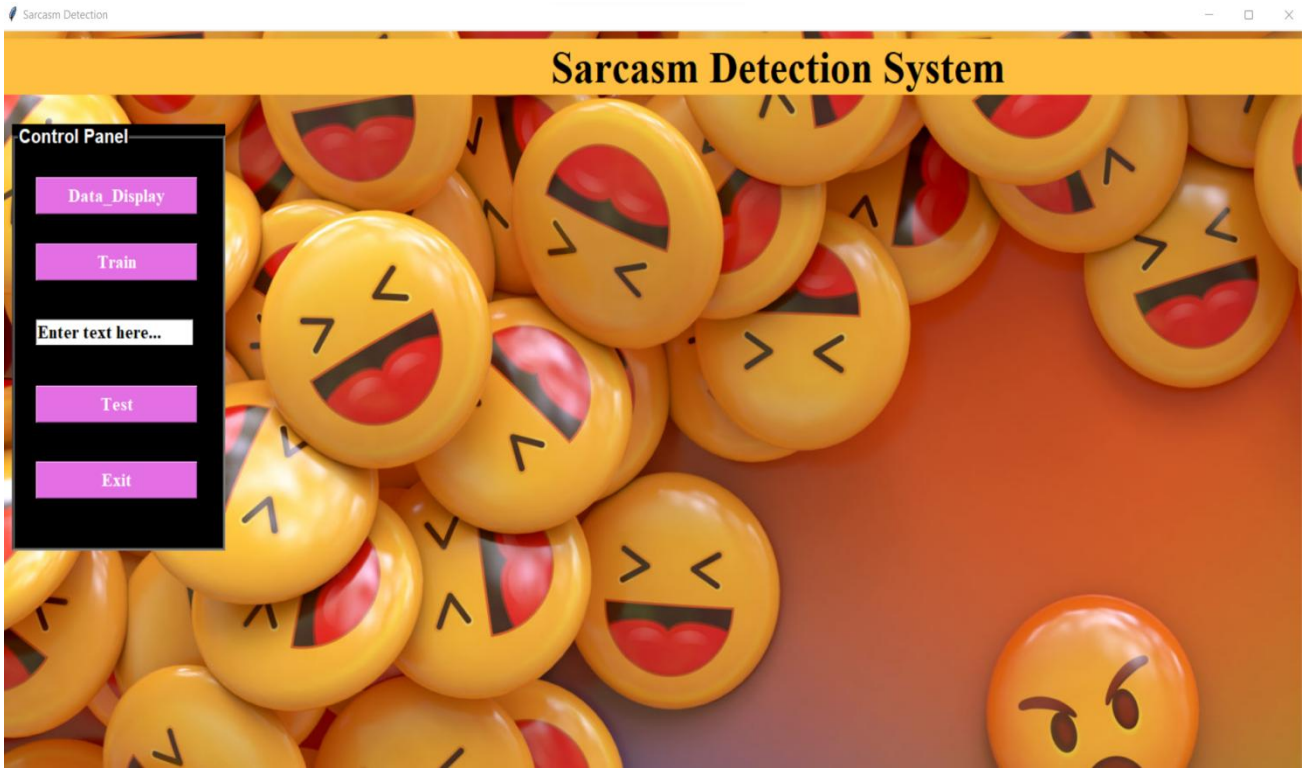
|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

3.REGISTRATION:



4.GUI MASTER:



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004141|

VII. CONCLUSION

This paper proposes a method for detecting sarcasm in Twitter posts in English and Hindi. Because sarcasm is highly dependent and contextual, mood and other contextual signals can help in detecting sarcasm text. Rather than a dataset, the method uses sarcastic tweets, with a total of 9,104 tweets containing sarcasm. The SVM, KNN, and Random forest classifiers are used in the system. The Random forest classifier outperformed the other classifiers in terms of accuracy, demonstrating the effectiveness of the technique. The retrieved patterns do not cover all of the sarcastic detection patterns. As a result, in the future, Neural Network, Genetic Algorithm, and Pattern-based Approach can be combined to improve accuracy.

REFFERENCES

[1] 1 Yi Tay, Mondher Bouazizi And Tomoaki Otsuki (Ohtsuki), "A Pattern-Based Approach For Sarcasm Detection On Twitter" in August 24, 2016

[2] Yi Tay†, Luu Anh Tuan, Siu Cheung Hui, JianSu, "Reasoning with Sarcasm by Reading In-between"arXiv:1805.02856v1 [cs.CL] 8 May 2018

[3] Aditya Joshi, Pushpak Bhattacharyya, Mark J. Carman.2018, "Sarcasm Detection Using Incongruity within Target Text", In: Investigations in Computational Sarcasm",Cognitive Systems Monographs, vol 37. Springer

[4] Aditya Joshi, Pushpak Bhatta charyya,Mark J. Carman.2018, "Sarcasm Detection Using Contextual Incongruity. . . In: Investigations in Computational Sarcasm", Cognitive Systems Monographs, vol 37. Springer

[5] Shubhadeep Mukherjee, Dr. Pradip Kumar Bala, "Sarcasm Detection in MicroblogsUsing Na¨ıve Bayes and Fuzzy Clustering" In Proceedings of Technology in Society,2017 pages 19-27

[6] Z. Nishant Nikhil, Muktabh Mayank Srivastava, "Binarizer at SemEval-2018 Task 3:Parsing dependency and deep learning for irony detection"arXiv preprintarXiv:1805.01112 .

[7] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2010.

[8] C. G. Akcora, M. A. Bayir, M. Demirbas, and H. Ferhatosmanoglu, "Identifying breakpoints in public opinion," in Proc. 1st Workshop Soc. Media Anal.,

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details