|  |  |  |  |
| --- | --- | --- | --- |
| **Synopsis Report**  **on**  **Hate Speech Detection using machine Learning.**  **Submitted as partial fulfillment for the award of**  **BACHELOR OF TECHNOLOGY**  **DEGREE**  **Session 2022-23 in**  **CSE-Data Science**  **By:**  **Ayush Singh (2000321540021)**  **Under the guidance of:**  **Mrs. Neha Yadav Assistant Professor**  **DEPARTMENT OF CSE-DS**  **ABES ENGINEERING COLLEGE, GHAZIABAD** | | | |
|  |  |  | naac A Grade Engineering College |
| **AFFILIATED TO**  **DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW**  **(Formerly UPTU)** | | | |

# Student’s Declaration

I / we hereby declare that the work being presented in this report entitled **“Hate Speech Detection in Machine Learning.”** is an authentic record of my/ our own work carried out under the supervision of Mrs. Neha Yadav **Assistant Professor.** The matter embodied in this report has not been submitted by us for the award of any other degree.

**Date:**

**Signature of student**

**Ayush Singh**

**(2000321540021)**

This is to certify that the above statement made by the candidate(s) is correct to the best of my knowledge.

**Signature of HOD Signature of Supervisor**

**Mr. Prabhat Singh Mrs. Neha Yadav**

**CSE-DS Assistant Professor**

**Date: 02.01.2023 CSE-DS**

**Acknowledgement**

We would like to convey our sincere thanks to **Mrs. Neha Yadav** for giving the motivation, knowledge and support throughout the course of the project. The continuous support helps in a successful completion of project. The knowledge provided is very useful for us.

We also like to give a special thanks to the department of Information and Technology for giving us the continuous support and opportunities for fulfilling our project.

We would also like to extend our sincere obligation to **Mr. Prabhat Singh**, Head of Department, Data Science for providing this opportunity to us.

**Signature of student**

**Ayush Singh**

**(Roll No. 2000321540021)**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Contents** | **Page No.** |
|  | Student’s Declaration | i |
|  | Acknowledgement | ii |
|  | List of Figures | iv |
|  | List of Tables | v |
|  | Abstract | vi |
| Chapter 1: | Introduction | 1 |
| Chapter 2: | Related Work/Methodology | 2 |
| 2.1: | Existing Approaches | 2 |
| 2.2: | Comparative Analysis of Existing Works | 2 |
| Chapter 3: | Project Objective | 3 |
| Chapter 4: | Proposed Methodology | 4 |
| Chapter 5: | Design and Implementation | 5 |
| 5.1: | Work Flow Diagram | 5 |
| Chapter 6: | Results and Discussion | 6 |
| Chapter 7: | Conclusion and Future Scope | 7 |
|  | References | 8 |

**List of Figures**

|  |  |
| --- | --- |
| **Figure No and name** | **Page No.** |
| **Fig.1.** Proposed Approach | 4 |
| **Fig.2.** Work Flow Diagram | 5 |
| **Fig.3.** Declaration of tweet offensive or not. | 6 |

**ABSTRACT**

## In recent years, the increasing propagation of hate speech on social media and the urgent need for effective countermeasures have drawn significant investment from governments, companies, and researchers. A large number of methods have been developed for automated hate speech detection online. This aims to classify textual content into non-hate or hate speech, in which case the method may also identify the targeting characteristics (i.e., types of hate, such as race, and religion) in the hate speech. However, we notice significant difference between the performance of the two (i.e., non-hate v.s. hate). In this work, we argue for a focus on the latter problem for practical reasons. We show that it is a much more challenging task, as our analysis of the language in the typical datasets shows that hate speech lacks unique, discriminative features and therefore is found in the ‘long tail’ in a dataset that is difficult to discover.

## 

## Chapter 1

**Introduction**

 In the current era of the Internet, it is obvious that almost everyone has social media apps to connect and interact with people around the world. At the same time, social media is a place where a lot of personal opinions have been shared about anyone. And most of the time those opinions are offensive and hateful.

It is the model which identifies and detects hateful and offensive speech being poured on the internet. Social media is a place for many people to make hateful and offensive comments about others. So hate speech detection has become an important solution to problems in today’s online world.

Social media platform is a place where everyone can express themselves without any hesitation [6,8]. Twitter is a popular social media platform on which people express themselves in the form of tweets. These tweets are studied to find out the offensive or hatred of a speech present in that tweet.

The main objective of the project is : -

1. To analyze the tweets from the twitter and to decide whether a speech of that tweet is offensive , hate or neither.

## Chapter 2 Related Work

The related work associated with our project is given below:

### Existing Approaches

* Twitter Sentiment Analysis using Python:
  + To do the analysis of twitter data using python and find the positive and negative hatred percentage [5].
* Word frequency and sentiment analysis of twitter messages during Elections or some worldwide events.
  + To find the frequency of each word and do the sentiment analysis of the dataset [2].
* Elections /National Events: a sentiment analysis
  + To perform the analysis of sentiments of Elections dataset [4].

### Comparative Analysis of Existing Works

* + - In the existing projects, the words with positive or negative polarity are obtained but in our project we are considering only negative polarity of dataset.

## Chapter 3

**Project Objective**

* This project will analyze whether a tweet is offensive or not.
* To implement an algorithm for declaration of tweet offensive , hate or neither.
* This project will Machine Learning Algorithm and gives the output along with the accuracy.

## Chapter 4 Proposed Methodology

The proposed methodology related to our project is given below:

Step 1: Identify the abusive words in the tweet. Tweets are extracted from the dataset present on the Kaggle.

Step 2: The preprocessing of the dataset is done. It involves the following steps:

* Removal of capitalization
* Remove whitespace with a single space
* Removal of punctuations and numbers
* Removal of extra spaces
* Removal of links[https://abc.com]

Step 3: Analyzing the polarity of the dataset.

Step 4: Giving the step 3 output in machine learning algorithm and analyze it to find the algorithm with best accuracy.

Step 5: The result is obtained.

* Extraction of Dataset from Twitter API
* Pre-processing of Data to remove special characters, punctuations, Stop Words and Images
* Processing of Data to analyze the polarity of the Dataset
* To use Machine Learning Algorithm
* Result is displayed.

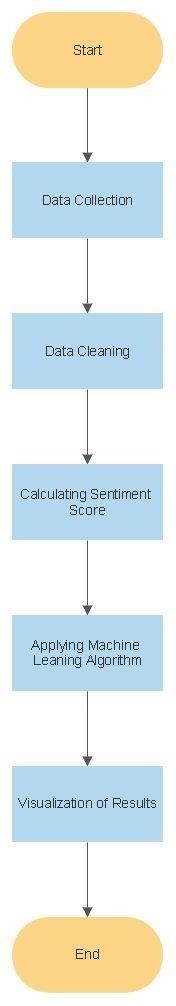
**Fig.1.** Proposed Approach

## Chapter 5

**Design and Implementation**

The design and implementation of our project is as follows:

### 5.1. Work Flow Diagram



**Fig.2.** Work Flow Diagram

The dataset for building our hate speech detection model is available on www.kaggle.com . The dataset consist of Twitter hate speech detection data , used to research hate-speech detection . The text in the data is classified as hate speech , offensive language and neither. Due to nature of study ,it’s important that abstract data contains text that can be considered racist ,sexist homophobic or generally offensive.

## Chapter 6 Results and Discussion

The result we got from analyzing the tweet is given below in Fig.3.

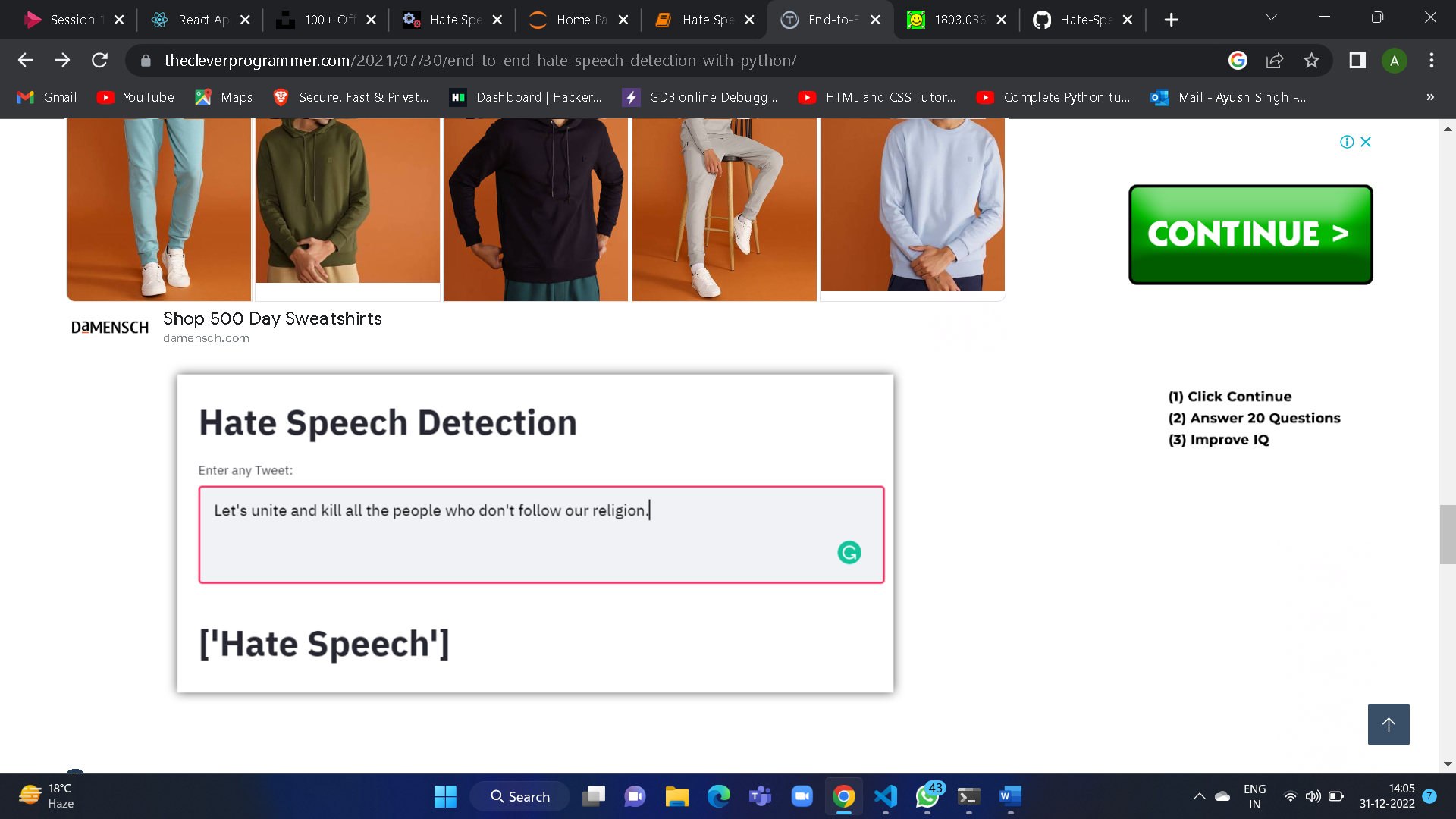


Fig.3. shows that tweet is an hate speech as it contain harsh words.

## Chapter 7

**Conclusion and Future Scope**

* The project will give the overall polarity score of Tweets and will find which is the best Algorithm for hate speech detection.
* From the analyses of the tweets, we observe that mostly from people with political views have offensive and hateful opinions.
* In future we will be planning to perform the analysis on various other social platforms Instagram, Facebook, etc. and also try to identifies hateful and offensive speech.

## References

1. Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. (2020). An" Infodemic": Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak. *medRxiv*.
2. Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word frequency and sentiment analysis of twitter messages during Coronavirus pandemic. *arXiv preprint arXiv:2004.03925*.
3. Samuel, J., Ali, G. G., Rahman, M., Esawi, E., & Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, *11*(6), 314.
4. Kumar, A., Khan, S. U., & Kalra, A. (2020). COVID-19 pandemic: a sentiment analysis. *European Heart Journal*.
5. Ahuja, S., & Dubey, G. (2017, August). Clustering and sentiment analysis on Twitter data. In *2017 2nd International Conference on Telecommunication and Networks (TEL- NET)* (pp. 1-5). IEEE.
6. Suman, C., Saha, S., Bhattacharyya, P., & Chaudhari, R. S. (2020). Emoji Helps! A Multi-modal Siamese Architecture for Tweet User Verification. *Cognitive Computation*, 1- 16
7. Neethu, M. S., & Rajasree, R. (2013, July). Sentiment analysis in twitter using machine learning techniques. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
8. Gupta, S., Singh, A., & Ranjan, J. (2020). Sentiment Analysis: Usage of Text and Emoji for Expressing Sentiments. In *Advances in Data and Information Sciences* (pp. 477-486). Springer, Singapore
9. Rajput, N. K., Grover, B. A., & Rathi, V. K. (2020). Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. *arXiv preprint arXiv:2004.03925*.
10. Medford, R. J., Saleh, S. N., Sumarsono, A., Perl, T. M., & Lehmann, C. U. An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Early Public Sentiment for the COVID-19 Outbreak. In *Open Forum Infectious Diseases*.