

A Project Stage-II Report on

**Advances in Machine Learning Algorithms for  
Hate Speech Detection in Social Media: A Review**

Submitted to

Jawaharlal Nehru Technological University, Hyderabad in partial fulfillment of  
the requirements for the award of the degree

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE & ENGINEERING**

by

**DIWAS KUSHWAHA**

**18831A05J4**

**VARUN JHA**

**18831A05J5**

**SURYA DAMERA**

**18831A05G6**

Under the Esteemed Guidance of

**Dr.B.SANTHOSH KUMAR**

**Professor & Head of the Department**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**GURU NANAK INSTITUTE OF TECHNOLOGY**

(Affiliated to JNTU-Hyderabad)

Ranga Reddy District -501506

2021-2022

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**  
**GURU NANAK INSTITUTE OF TECHNOLOGY**

(Affiliated to JNTU-Hyderabad)

Ranga Reddy District -501506



**CERTIFICATE**

This is to certify that the project entitled “ **ADVANCES IN MACHINE LEARNING ALGORITHMS FOR HATE SPEECH DETECTION IN SOCIAL MEDIA: A REVIEW** ” is being presented with report by **DIWAS KUSHWAHA (18831A05J4), VARUN JHA (18831A05J5), SURYA DAMERA (18831A05G6)** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, to **Jawaharlal Nehru Technological University, Hyderabad**.

**INTERNAL GUIDE**  
**(Dr.B.SANTHOSH KUMAR)**

**PROJECT COORDINATOR**  
**(Mrs.K.SUSHEELA)**

**HEAD OF DEPARTMENT**  
**(Dr.B.SANTHOSH KUMAR)**

**PRINCIPAL**  
**(Dr.S.SREENATHA REDDY)**

**EXTERNAL EXAMINER**



RAM Innovative Infotech  
M : 9195810700  
E : raminnovativeinfotech@gmail.com  
B2, Tili dla  
Flat No. #309, Anurtha Vile  
Opp. YashlOOa Hospital Sonwfigilla

## PROJECT COMPLETION CERTIFICATE

This is to certify that the following students of final year B.Tech, Department of Computer Science and Engineering - Guru Nanak Institute of Technology (GNIT) have completed their training and project at GNIT successfully.

STUDENT NAME:	Roll No:
1. Diwas Kushwaha	18831A05J4
2. Varun Jha	18831A05J5
3. Surya Damera	18831A05G6

The training was conducted on (using Python) Technology for the completion of the project titled Advances in machine learning algorithms For Hate speech detection in social media :A Review in June, 2022. The project has been completed in all aspects.



Min8MajotEEILAYPFile4S For  
M/11, Ied1, IU/B, Ted1, MCA, MS



# **GURU NANAK INSTITUTE OF TECHNOLOGY**

Ibrahimpattanam, R.R. Dist. – 501506.

## **VISION OF GNIT**

To be a world – class educational and research institution in the service of humanity by promoting high quality Engineering and Management Education.

## **MISSION OF GNIT**

- Imbibe soft skills and technical skills.
- Develop the faculty to reach the international standards.
- Maintain high academic standards and teaching quality that promotes the analytical thinking and independent judgment.
- Promote research, innovation and Product development by collaboration with reputed foreign universities.
- Offer collaborative industry programs in emerging areas and spirit of enterprise.



# **GURU NANAK INSTITUTE OF TECHNOLOGY**

Ibrahimpattanam, R.R. Dist. – 501506.

## **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

### **VISION**

To be a premier department of Computer Science and Engineering in the region.

### **MISSION**

- Nurture young individuals into knowledgeable, skill-full and ethical professionals in their pursuit of Computer Science and Engineering.
- Nurture the faculty to expose them to world-class infrastructure.
- Sustain high performance by excellence in teaching, research and innovations.
- Extensive partnerships and collaborations with foreign universities for technology upgradation.
- Develop Industry-Interaction for innovation and product development.



# **GURU NANAK INSTITUTE OF TECHNOLOGY**

Ibrahimpattanam, R.R. Dist. – 501506.

## **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

### **Program Educational Objectives (PEO's)**

**PEO-1:** Graduates shall have the ability to apply knowledge and technical skills in emerging areas of Computer Science and Engineering for higher studies, research, employability, product development and handle realistic problems.

**PEO-2:** Graduates shall maintain ethical conduct, sense of responsibility to serve the society and protect the environment.

**PEO-3:** Graduates shall possess academic excellence with innovative insight, soft skills, managerial skills, leadership qualities, knowledge of contemporary issues for successful professional career.

### **Program Outcomes (PO's)**

**PO-1:** Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

**PO-2:** Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**PO-3:** Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet specified needs with appropriate consideration for public health and safety, and the cultural, societal, and environmental considerations.

**PO-4:** Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO-5:** Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

**PO-6:** The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**PO-7:** Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**PO-8:** Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

**PO-9:** Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**PO-10:** Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**PO-11:** Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**PO-12:** Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.



# **GURU NANAK INSTITUTE OF TECHNOLOGY**

Ibrahimpattanam, R.R. Dist. – 501506.

## **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

### **MAPPING WITH PO'S AND PEO'S**

Sl.No.	Content	PO's	PEO's
1	Introduction	PO1, PO2	PEO1
2	Abstract	PO1, PO2	PEO1
3	System Specifications	PO3, PO5	PEO1
4	Architecture Design	PO3, PO4	PEO1, PEO2
5	HDFS CLI Commands	PO1	PEO3
6	Implementation	P09, PO10, PO11	PEO2, PEO3
7	Results	PO4, PO11, PO12	PEO2, PEO3



## **DECLARATION**

We hereby declare that major project report entitled “**ADVANCES IN MACHINE LEARNING ALGORITHMS FOR HATE SPEECH DETECTION IN SOCIAL MEDIA:A REVIEW**” is the work done by **DIWAS KUSHWAHA, V A R U N J H A , SURYA DAMERA** , bearing the roll no’s **18831A05J4, 18831A05J5, 18831A05G6** towards the fulfillment of the requirement for the award of the Degree of **Bachelor of Technology in Computer Science and Engineering**, to **Jawaharlal Nehru Technological University, Hyderabad**, is the result of the work carried out under the guidance **Dr.B.SANTHOSH KUMAR, Guru Nanak Institute of Technology, Hyderabad**. We further declare that this project report has not been previously submitted before either in part or full for the award of any degree or any diploma by any organization or any universities.

**DIWAS KUSHWAHA (18831A05J4)**

**VARUN JHA (18831A05J5)**

**SURYA DAMERA (18831A05G6)**

## ACKNOWLEDGEMENT

“Task successful” makes everyone happy. But the happiness will be gold without glitter if we didn’t state the people who have supported us to make it a success.

We would like to express our sincere thanks and gratitude to our Principal, **Dr. S. SREENATHA REDDY**, and the Head of the Department- Computer Science and Engineering **Dr. B. SANTHOSH KUMAR**, for having guided us in developing the requisite capabilities for taking up this project.

We thank Project Coordinator, **Mrs.K.SUSHEELA**, Assistant Professor, CSE, GNIT for providing seamless support and the right suggestions that are given in the development of the project.

We especially thank our internal guide, **Dr. B. SANTHOSH KUMAR** for his constant guidance in every stage of the project. We would also like to thank all our lecturers for helping me in every possible way whenever the need arose.

On a more personal note, we thank our beloved parents and friends for their moral support during the course of our project.

## **ABSTRACT**

The aim of this paper is to review machine learning (ML) algorithms and techniques for hate speech detection in social media (SM). Hate speech problem is normally model as a text classification task. In this study, we examined the basic baseline components of hate speech classification using ML algorithms. There are five basic baseline components – data collection and exploration, feature extraction, dimensionality reduction, classifier selection and training, and model evaluation, were reviewed. There have been improvements in ML algorithms that were employed for hate speech detection over time. New datasets and different performance metrics have been proposed in the literature. To keep the researchers informed regarding these trends in the automatic detection of hate speech, it calls for a comprehensive and an updated state-of-the-art. The contributions of this study are three-fold. First to equip the readers with the necessary information on the critical steps involved in hate speech detection using ML algorithms. Secondly, the weaknesses and strengths of each method is critically evaluated to guide researchers in the algorithm choice dilemma. Lastly, some research gaps and open challenges were identified. The different variants of ML techniques were reviewed which include classical ML, ensemble approach and deep learning methods. Researchers and professionals alike will benefit immensely from this study.

## TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	<b>CERTIFICATE</b>	i
	<b>PROJECT COMPLETION CERTIFICATE</b>	ii
	<b>VISION AND MISSION OF GNIT</b>	iii
	<b>VISION AND MISSION OF CSE</b>	iv
	<b>PO'S AND PEO'S</b>	v
	<b>MAPPING WITH PO'S AND PEO'S</b>	vii
	<b>DECLARATION</b>	viii
	<b>ACKNOWLEDGEMENT</b>	ix
	<b>ABSTRACT</b>	x
	<b>LIST OF FIGURES</b>	xv
	<b>LIST OF SYMBOLS</b>	xvi
1.	<b>CHAPTER 1 : INTRODUCTION</b>	1
	➤ GENERAL	1
	➤ OBJECTIVE	2
	➤ EXISTING SYSTEM	3
	1.3.1EXISTINGSYSTEM DISADVANTAGES	3
	1.3.2 LITERATURE SURVEY	4
	1.4 PROPOSED SYSTEM	8
	1.4.1 PROPOSED SYSTEM ADVANTAGES	9
2.	<b>CHAPTER 2 :PROJECT DESCRIPTION</b>	10
	2.1 GENERAL	10
	2.2 METHODOLOGIES	10
	2.2.1 MODULES NAME	10
	2.2.2 MODULES EXPLANATION	11
	2.2.3 MODULE DIAGRAM	12
	2.2.4GIVEN INPUTAND EXPECTED OUTPUT	13
	2.3 TECHNIQUE OR ALGORITHM	13

3.	<b>CHAPTER 3 : REQUIREMENTS</b>	14
	3.1 GENERAL	14
	3.2 HARDWARE REQUIREMENTS	14
	3.3 SOFTWARE REQUIREMENTS	15
	3.4 FUNCTIONAL REQUIREMENTS	15
	3.5 NON-FUNCTIONAL REQUIREMENTS	16

4.	<b>CHAPTER 4 :SYSTEM DESIGN</b>	17
	<b>4.1 GENERAL</b>	17
	4.1.1 ACTIVITY DIAGRAM	18
	4.1.2 USE CASE DIAGRAM	19
	4.1.3 DATA FLOW DIAGRAM	20
	4.1.4SEQUENCE DIAGRAM	21
	4.1.5 COLLABORATION DIAGRAM	22
	4.1.6CLASS DIAGRAM	23
	4.1.7 SYSTEM ARCHITECTURE	24
	4.1.8 OBJECT DIAGRAM	25
	4.1.9 STATE DIAGRAM	26
	4.1.10 COMPONENT DIAGRAM	27
	4.1.11 E-R DIAGRAM	28
	4.2 DATABASE DESIGN (ALL LEVEL	32
5.	<b>CHAPTER 5 :SOFTWARE SPECIFICATION</b>	33
	5.1 GENERAL	33
6.	<b>CHAPTER 6 :IMPLEMENTATION</b>	36
	6.1 GENERAL	36

7.	<b>CHAPTER 7 :SNAPSHOTS</b>	38
	7.1 GENERAL	38
8.	<b>CHAPTER 8 :SOFTWARE TESTING</b>	39
	8.1 GENERAL	39
	8.2 DEVELOPING METHODOLOGIES	40
	8.3 TYPES OF TESTING	41
9.	<b>CHAPTER 9 : APPLICATIONS AND FUTURE ENHANCEMENT</b>	42

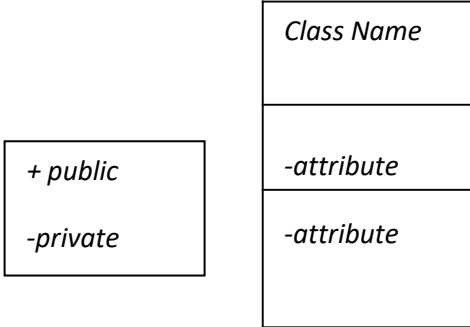
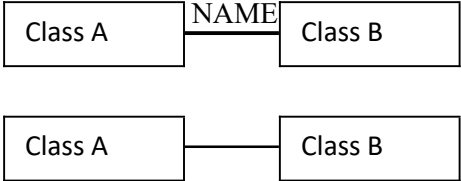
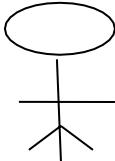
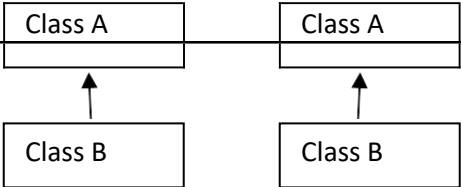
<b>10</b>	<b>CHAPTER 10 :CONCLUSION &amp; REFERENCE</b>	<b>43</b>
	10.1 CONCLUSION	43
	10.2 REFERENCES	44

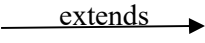
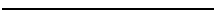
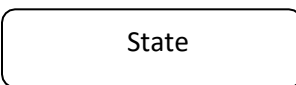
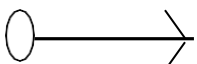
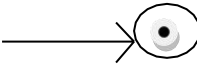
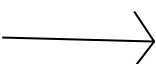
## LIST OF FIGURES

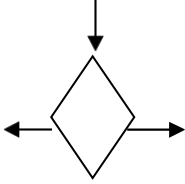
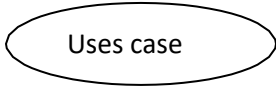
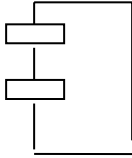
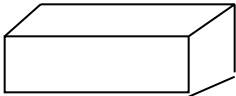
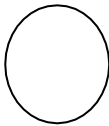

FIGURE NO	NAME OF THE FIGURE	PAGE NO.
2.3.2	Use case Diagram	18
4.2	Class Diagram	20
4.3	Object Diagram	21
4.4	State Diagram	22
4.5	Activity Diagram	24
4.6	Sequence Diagram	26
4.7	Collaboration Diagram	27
4.8	Component Diagram	28
4.9	Data Flow Diagram	29
4.1	Deployment Diagram	31
4.12	System Architecture	32

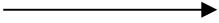
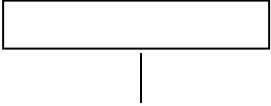
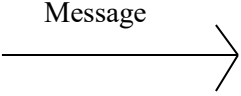


## LIST OF SYMBOLS

S.NO	NOTATION NAME	NOTATION	DESCRIPTION
1.	Class		Represents a collection of similar entities grouped together.
2.	Association		Associations represents static relationships between classes. Roles represents the way the two classes see each other.
3.	Actor		It aggregates several classes into a single classes.
4.	Aggregation		Interaction between the system and external environment

5.	Relation (uses)	uses	Used for additional process communication.
6.	Relation (extends)		Extends relationship is used when one use case is similar to another use case but does a bit more.
7.	Communication		Communication between various use cases.
8.	State		State of the processes.
9.	Initial State		Initial state of the object
10.	Final state		Final state of the object
11.	Control flow		Represents various control flow between the states.

12.	Decision box		Represents decision making process from a constraint
13.	Use case		Interact ion between the system and external environment.
14.	Component		Represents physical modules which are a collection of components.
15.	Node		Represents physical modules which are a collection of components.
16.	Data Process/State		A circle in DFD represents a state or process which has been triggered due to some event or action.
17.	External entity		Represents external entities such as keyboard,sensors,etc.

18.	Transition		Represents communication that occurs between processes.
19.	Object Lifeline		Represents the vertical dimensions that the object communications.
20.	Message		Represents the message exchanged.

# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 GENERAL**

Social media networks (SMNs) are the fastest means of communication as messages are sent and received almost instantaneously [1], [2]. SMNs are the primary media for perpetrating hate speeches nowadays. In line with this, cyber-hate crime has grown significantly in the last few decades [3]. More researches are being conducted to curb with the rising cases of hate speeches in social media (SM). Different calls have been made to SM providers to filter each comment before allowing it into the public domain [4], [5]. The impacts of hate crimes are already overwhelming due to widespread adoption of SM [6] and the anonymity enjoyed by the online users [7]. In this era of big data, it is time-consuming and difficult to manually process and classify massive quantities of text data. Besides, the precision of the categorization of manual text can easily be influenced by human factors, such as exhaustion and competence. To achieve more accurate and less subjective results, it is beneficial to use machine learning (ML) approaches to automate the text classification processes [6]. There have been significant advancements in ML techniques from classical ML, ensemble and deep learning (DL) techniques for hate speech detection. Due to the unprecedented advancement in natural language processing (NLP), several machine learning methods have achieved superior outcomes [8].

## **1.2 OBJECTIVE**

To be able to improve classification of SM texts as hate speech or non-hate speech, researchers and practitioners require an updated understanding of machine learning methodologies, which is fast evolving. Considerable effort has been spent on creating new and effective features that better capture hate speech on SM [9]–[11]. Slangs and new vocabularies are also constantly evolving in the SM space. New and updated datasets are also available across different regions of the world. To bridge the gap, there is a need to review the literature and keep professionals, old and new researchers in the know of the current developments in this research area. On this note, this review becomes necessary to be conducted.

## 1.3 Existing System

In Existing System, This paper, therefore, aimed to investigate several neural network models based on convolutional neural network (CNN) and recurrent neural network (RNN) to detect hate speech in Arabic tweets. It also evaluated the recent language representation model bidirectional encoder representations from transformers (BERT) on the task of Arabic hate speech detection. To conduct our experiments, we firstly built a new hate speech dataset that contained 9316 annotated tweets. Then, we conducted a set of experiments on two datasets to evaluate four models: CNN, gated recurrent units (GRU), CNN + GRU, and BERT.

### 13.1 Existing System Disadvantages

- Less accuracy.
- It is significantly slower due to an operation such as maxpool.
- Training of RNN models can be difficult.

### **1.3.2 LITERATURE SURVEY:**

**Title:** The effect of Twitter dissemination on cost of equity: A big data approach,

**Year:**2020

**Authors:** Mohammed S.Albarrak, Marwa Elnahass, Savvas Papagiannidis, Aly Salama.

**Description:** Reducing information asymmetry between investors and a firm can have an impact on the cost of equity, especially in an environment or times of uncertainty. New technologies can potentially help disseminate corporate financial information, reducing such asymmetries. In this paper we analyse firms' dissemination decisions using Twitter, developing a comprehensive measure of the amount of financial information that a company makes available to investors (iDisc) from a big data of firms' tweets (1,197,208 tweets). Using a sample of 4131 firm-year observations for 791 non-financial firms listed on the US NASDAQ stock exchange over the period 2009–2015, we find evidence that iDisc significantly reduces the cost of equity. These results are pronounced for less visible firms which are relatively small in size, have a low analyst following and a small number of investors. Highly visible firms are less likely to benefit from iDisc in influencing their cost of equity as other communication channels may have widely disseminated their financial information. Our investigations encourage managers to consider the benefits of directly spreading a firm's financial information to stakeholders and potential investors using social media in order to reduce firm equity premium (COE).



**Title:** An attention-based friend recommendation model in social network,

**Year:**2020

**Author:** Sundong Kim, C. Cai, H. Xu, J. Wan, B. Zhou, and X. Xie.

**Description:** In social networks, user attention affects the user's decision-making, resulting in a performance alteration of the recommendation systems. Existing systems make recommendations mainly according to users' preferences with a particular focus on items. However, the significance of users' attention and the difference in the influence of different users and items are often ignored. Thus, this paper proposes an attention-based multi-layer friend recommendation model to mitigate information overload in social networks. We first constructed the basic user and item matrix via convolutional neural networks (CNN). Then, we obtained user preferences by using the relationships between users and items, which were later inputted into our model to learn the preferences between friends. The error performance of the proposed method was compared with the traditional solutions based on collaborative filtering. A comprehensive performance evaluation was also conducted using large-scale real-world datasets collected from three popular location-based social networks. The experimental results revealed that our proposal outperforms the traditional methods in terms of recommendation performance.

**Title:** A Survey on Text Classification: From Shallow to Deep Learning

**Year:** 2020

**Author:** Qian Li, Hao Peng, Jianxin Li, Congying Xia.

**Description:** Text classification is the most fundamental and essential task in natural language processing. The last decade has seen a surge of research in this area due to the unprecedented success of deep learning. Numerous methods, datasets, and evaluation metrics have been proposed in the literature, raising the need for a comprehensive and updated survey. This paper fills the gap by reviewing the state of the art approaches from 1961 to 2020, focusing on models from shallow to deep learning. We create a taxonomy for text classification according to the text involved and the models used for feature extraction and classification. We then discuss each of these categories in detail, dealing with both the technical developments and benchmark datasets that support tests of predictions. A comprehensive comparison between different techniques, as well as identifying the pros and cons of various evaluation metrics are also provided in this survey. Finally, we conclude by summarizing key implications, future research directions, and the challenges facing the research area.

**Title:** Task-technology fit and technology acceptance model application to structure and evaluate the adoption of social media in academia

**Author:** Qusay Al-Maatouk; Mohd Shahizan Othman; Ahmed Aldraiwees.

**Year:**2020

**Description:** The purpose of this article was to reduce the dissimilarities in the literature regarding the use of social media for training and its impact on students' academic performance in higher education institutions. The main method of data collection for task-technology fit (TTF) and the technology acceptance model (TAM) was a questionnaire survey. This research hypothesizes that TTF applied to social media for learning will affect technology, task, and social characteristics that in turn improve students' satisfaction and students' academic performance. It also posits that the behavioral intent to use social media for learning will affect comprehension efficiency, ease of use, and enjoyment, all of which also improve students' satisfaction and students' academic performance. The data collection questionnaire was conducted with 162 students familiar with social media. Quantitative structural equation modeling was employed to analyze the results. A significant relationship was found between technology, task, and social features with TTF for utilizing social media for academic purposes, all of which fostered student enjoyment and improved outcomes. Similarly, a clear relationship was found between comprehension efficiency, ease of use, and enjoyment with behavioral intentions to utilize social media for academic purposes that positively affected satisfaction and achievement. Therefore, the study indicates that TTF and behavioral intentions to use social media improve the active learning of students and enable them to efficiently share knowledge, information, and discussions. We recommend that students utilize social media in pursuit of their educational goals. Educators should also be persuaded to incorporate social media into their classes at higher education institutions.

**Title:** A deep learning approach for automatic hate speech detection in the Saudi Twittersphere

**Author:** Raghad Alshalan and Hend Al-Khalifa.

**Year:**2020

**Description:** With the rise of hate speech phenomena in the Twittersphere, significant research efforts have been undertaken in order to provide automatic solutions for detecting hate speech, varying from simple machine learning models to more complex deep neural network models. Despite this, research works investigating hate speech problem in Arabic are still limited. This paper, therefore, aimed to investigate several neural network models based on convolutional neural network (CNN) and recurrent neural network (RNN) to detect hate speech in Arabic tweets. It also evaluated the recent language representation model bidirectional encoder representations from transformers (BERT) on the task of Arabic hate speech detection. To conduct our experiments, we firstly built a new hate speech dataset that contained 9316 annotated tweets. Then, we conducted a set of experiments on two datasets to evaluate four models: CNN, gated recurrent units (GRU), CNN + GRU, and BERT. Our experimental results in our dataset and an out-domain dataset showed that the CNN model gave the best performance, with an F1-score of 0.79 and area under the receiver operating characteristic curve (AUROC) of 0.89.

## 1.4 Proposed System

In proposed system, This article reviewed advances made so far in automatic hate speech detection in social media. Hate speech as a societal problem is an old research area in the arts and humanities, however, it is still a new research area in the computing domain. This study found out that there is more research work in hate speech detection using classical ML than ensemble and deep learning techniques. That means researchers can explore more on hate speech detection using ensemble and deep learning methods.

#### **1.4.1 Proposed System Advantages**

- Easy to predict.
- It is very effective even with high dimensional data.
- It can be used for both regression and classification problem.

## **CHAPTER 2**

### **PROJECT DESCRIPTION**

#### **2.1 GENERAL:**

In this chapter, various supervised machine learning approaches are used. This section provides a general description of these approaches.

#### **1.2 METHODOLOGIES**

##### **MODULE:**

- **Dataset**
- **Pre-Processing**
- **Splitting**
- **Apply Algorithm**
- **Visualization**
- **Accuracy**

## MODULE DESCRIPTION

### ➤ Data set

A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

### ➤ Pre-Processing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way.

### ➤ Splitting

Data splitting is the act of partitioning available data into two portions, usually for cross-validatory purposes. one portion of the data is used to develop a predictive model. and the other to evaluate the model's performance.

- Training Data: Used for train the model or given as input to the to the learning model
- Testing Data: Used for test the model or given as input to the model for prediction.

### ➤ Apply Algorithm

In this we are using support vector machine algorithm to predict accuracy. It is a non-probabilistic supervised machine learning approaches used for classification and regression. It assigns a new data member to one of two possible classes. It defines a hyper plane that separates n-dimensional data into two classes.

## ➤ **Visualization**

Visualization is a technique that uses an array of static and interactive visuals within a specific context to help people understand and make sense of large amounts of data. The data is often displayed in a story format that visualizes patterns, trends and correlations that may otherwise go unnoticed.

## ➤ **Accuracy**

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.



## 2.3 TECHNIQUE USED OR ALGORITHM USED

### ➤ **Algorithm:**

All ML algorithms use mathematical modelling as an integral part of the algorithm. Therefore, the unstructured nature of the texts data must be converted into structured feature space. More researchers are now interested in developing applications that leverage text classification methods, especially with recent advances in NLP and text mining. Generally, hate speech classification leveraging ML can be grouped into five phases: Data collection and exploration, feature extraction, dimensionality reduction, classifiers selection and evaluations.

## **CHAPTER 3**

### **REQUIREMENTS ENGINEERING**

#### **3.1 GENERAL**

These are the requirements for doing the project. Without using these tools and software's we can't do the project. So we have two requirements to do the project. They are

1. Hardware Requirements.
2. Software Requirements.

#### **3.2 HARDWARE REQUIREMENTS**

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system.

They are used by software engineers as the starting point for the system design. It should state what the system does and not how it should be implemented.

- PROCESSOR : DUAL CORE 2 DUOS.
- RAM : 4GB DD RAM
- HARD DISK : 250 GB

### 3.3 SOFTWARE REQUIREMENTS

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

- OPERATING SYSTEM : WINDOWS 7/8/10
- PLATFORM : SPYDER3
- PROGRAMMING LANGUAGE : PYTHON, HTML
- FRONT END : SPYDER3

### 3.4 FUNCTIONAL REQUIREMENTS

A functional requirement defines a function of a software-system or its component. A function is described as a set of inputs, the behavior, the presented result will help us in identifying the behaviour of employees who can be attired over the next time. Experimental results reveal that the logistic regression approach can reach up to high accuracy over other machine learning approaches.

### 3.5 NON-FUNCTIONAL REQUIREMENTS

The major non-functional Requirements of the system are as follows

➤ **Usability**

The system is designed with completely automated process hence there is no or less user intervention.

➤ **Reliability**

The system is more reliable because of the qualities that are inherited from the chosen platform java. The code built by using java is more reliable.

➤ **Performance**

This system is developing in the high level languages and using the advanced front-end and back-end technologies it will give response to the end user on client system with in very less time.

➤ **Supportability**

The system is designed to be the cross platform supportable. The system is supported on a wide range of hardware and any software platform, which is having JVM, built into the system.

➤ **Implementation**

The system is implemented in web environment using struts framework. The apache tomcat is used as the web server and windows xp professional is used as the platform. Interface the user interface is based on Struts provides HTML Tag

## **CHAPTER 4**

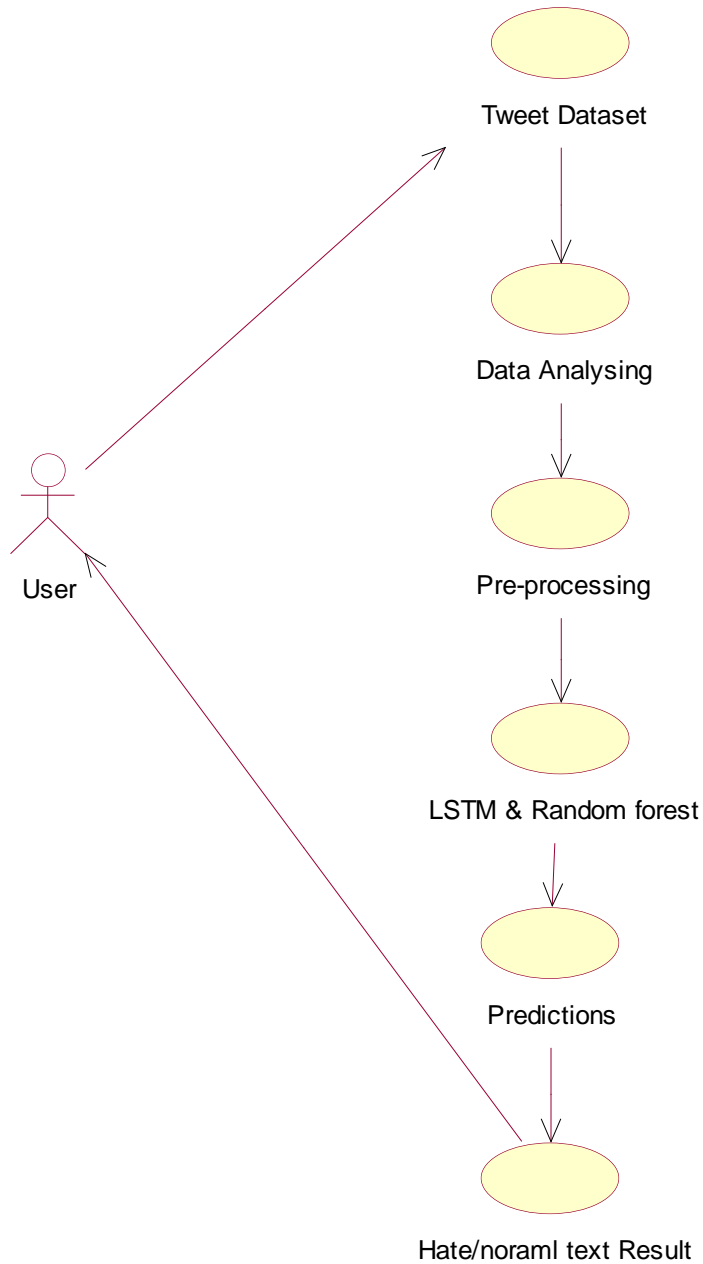
### **DESIGN ENGINEERING**

#### **4.1 GENERAL**

Design Engineering deals with the various UML [Unified Modelling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering. Design is the means to accurately translate customer requirements into finished product.

## UML Diagrams

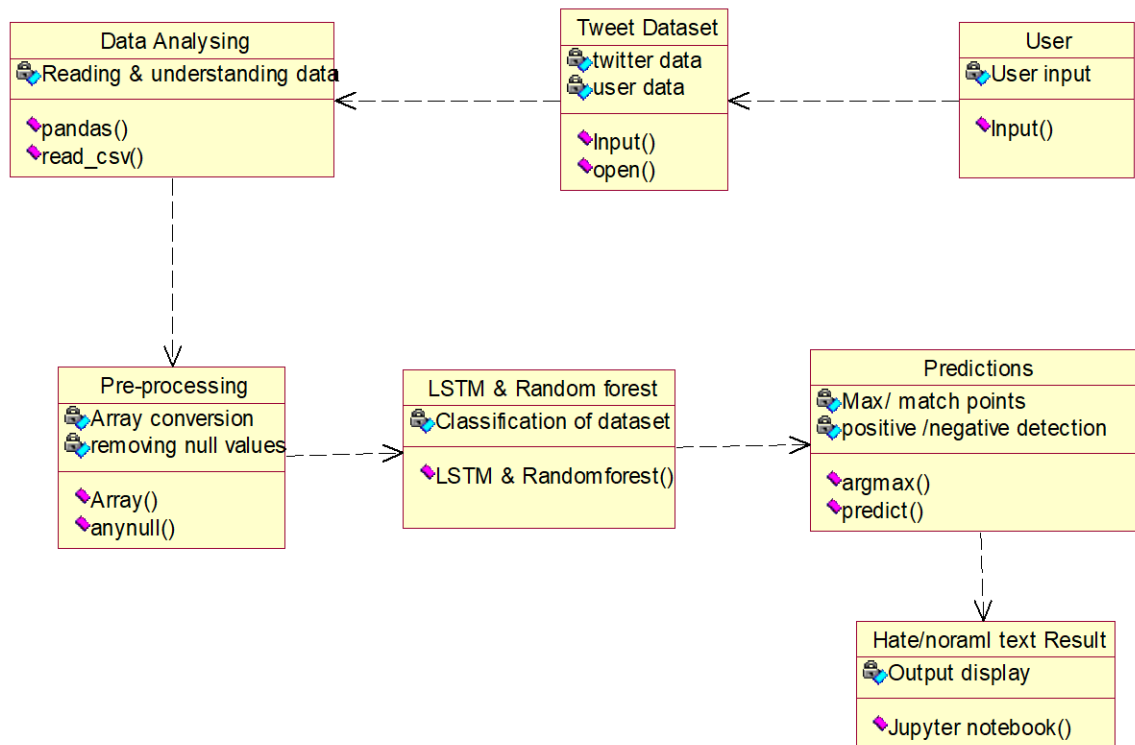
### USE CASE DIAGRAM:



### EXPLANATION:

The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. The above diagram consists of user as actor. Each will play a certain role to achieve the concept.

## CLASS DIAGRAM:

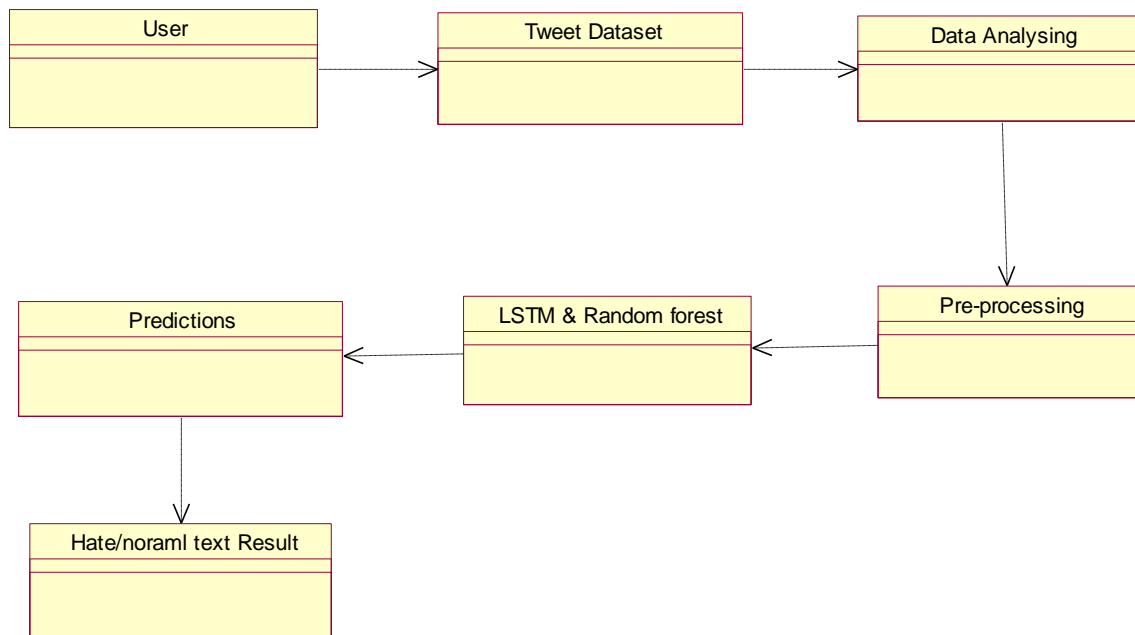


## EXPLANATION:

In this class diagram represents how the classes with attributes and methods are linked together to perform the verification with security. From the above diagram shown the various classes involved in our project.



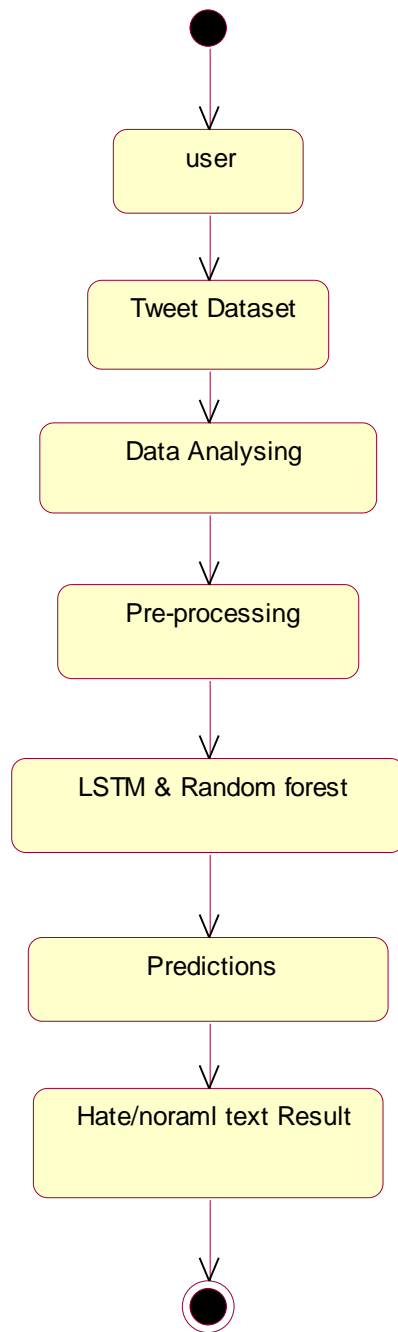
## OBJECT DIAGRAM:



## EXPLANATION:

In the above diagram tells about the flow of objects between the classes. It is a diagram that shows a complete or partial view of the structure of a modeled system. In this object diagram represents how the classes with attributes and methods are linked together to perform the verification with security.

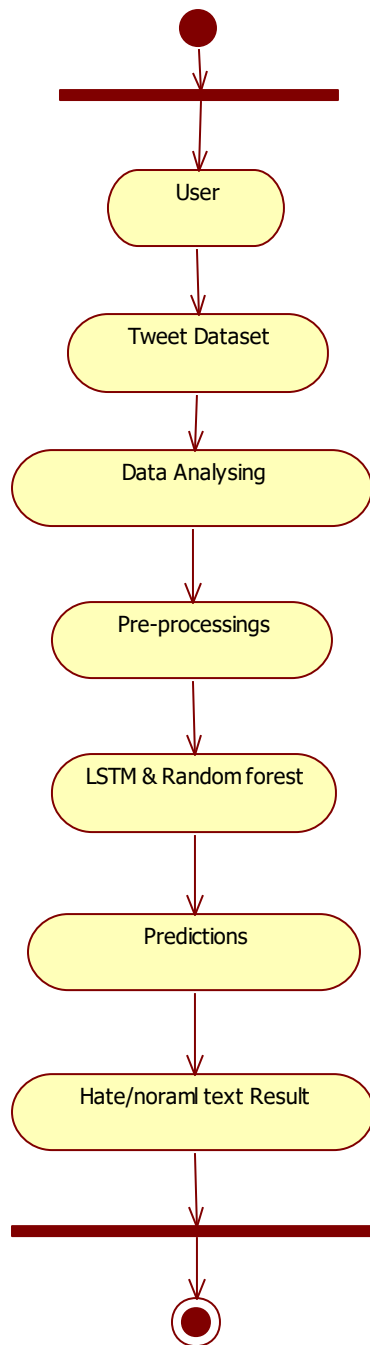
## STATE DIAGRAM:



## EXPLANATION:

State diagram are a loosely defined diagram to show workflows of stepwise activities and actions, with support for choice, iteration and concurrency. UML, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. UML activity diagrams could potentially model the internal logic of a complex operation. In many ways UML activity diagrams are the object-oriented equivalent of flow charts and data flow diagrams (DFDs) from structural development.

## ACTIVITY DIAGRAM:

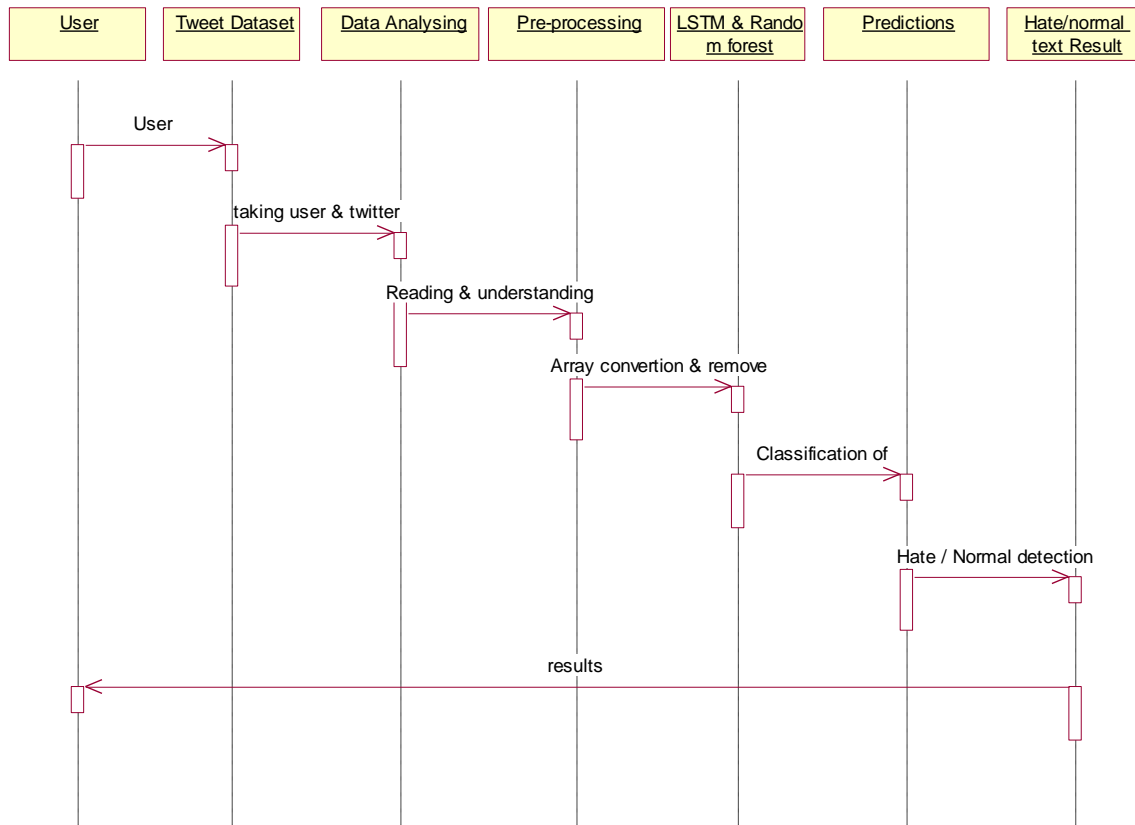


## EXPLANATION:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language,

activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

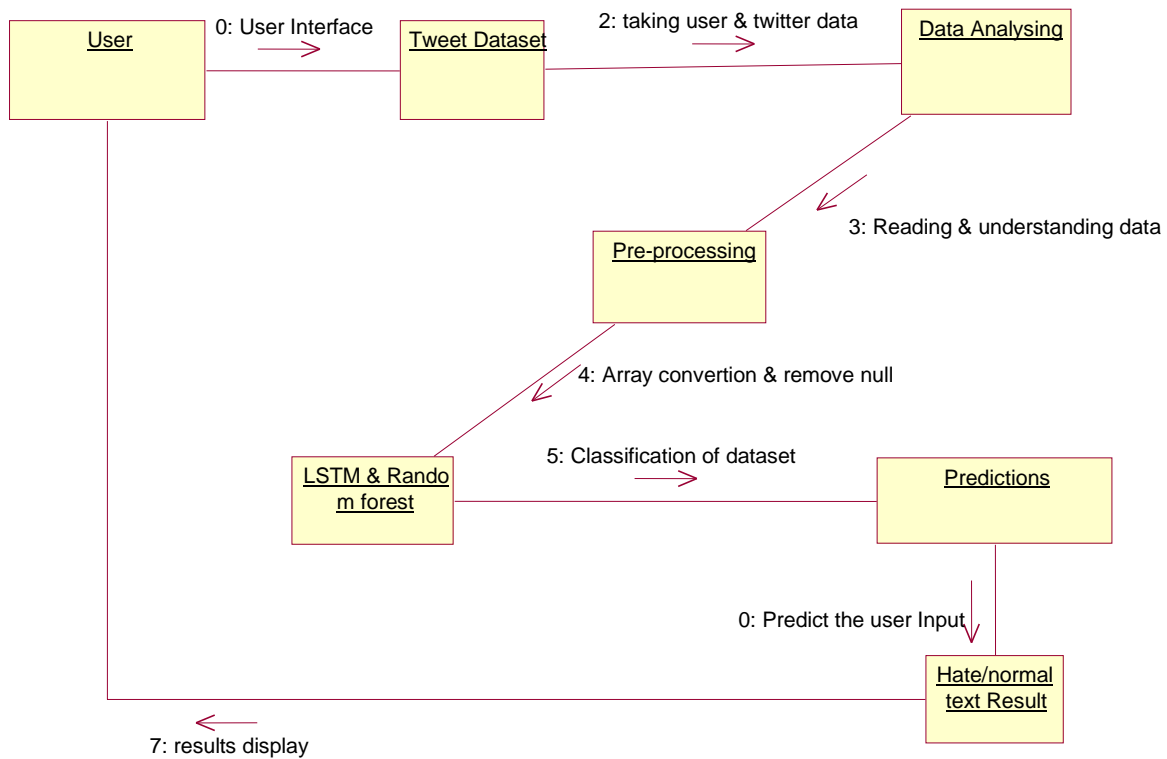
## SEQUENCE DIAGRAM:



## EXPLANATION:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.

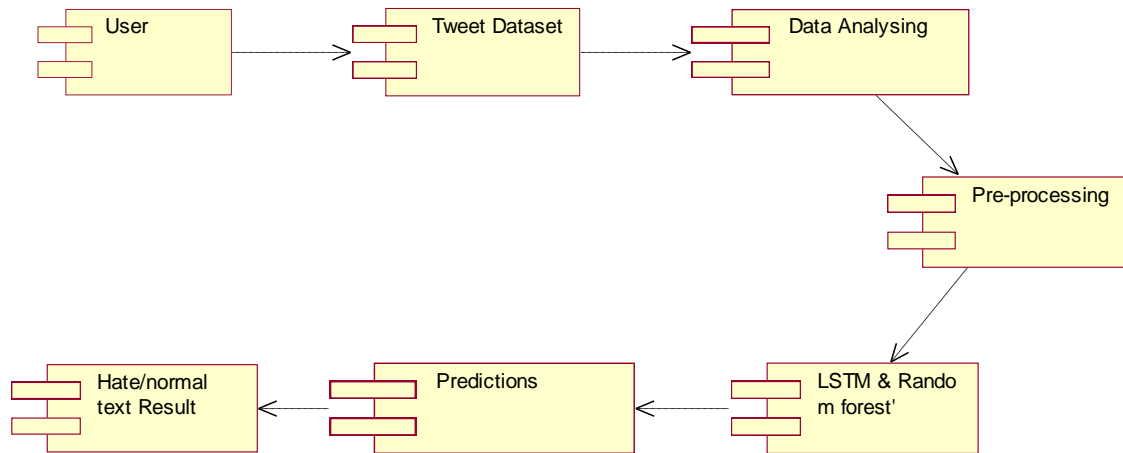
## COLLABORATION DIAGRAM:



## EXPLANATION:

A collaboration diagram, also called a communication diagram or interaction diagram, is an illustration of the relationships and interactions among software objects in the Unified Modeling Language (UML). The concept is more than a decade old although it has been refined as modeling paradigms have evolved.

## COMPONENT DIAGRAM:



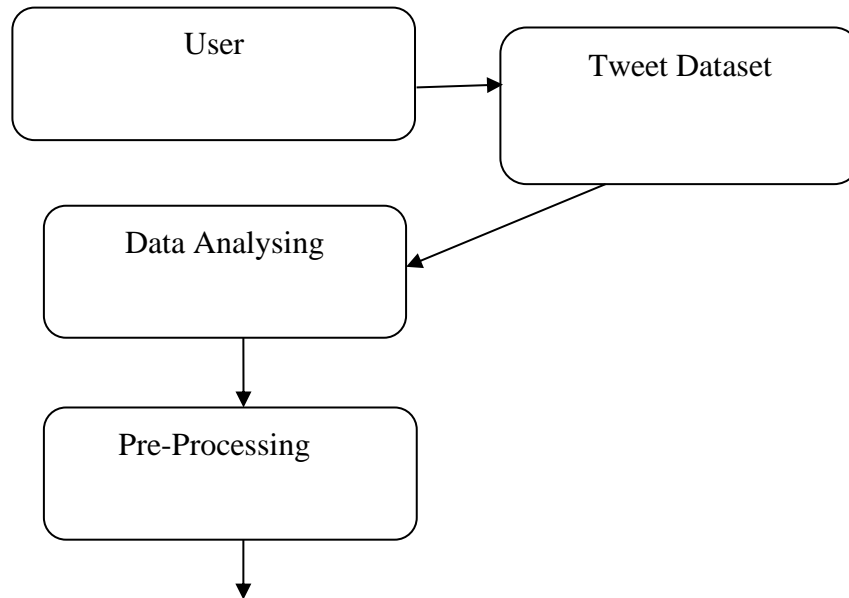
## EXPLANATION:

In the Unified Modeling Language, a component diagram depicts how components are wired together to form larger components and or software systems. They are used to illustrate the structure of arbitrarily complex systems. User gives main query and it converted into sub queries and sends through data dissemination to data aggregators. Results are to be showed to user by data aggregators. All boxes are components and arrow indicates dependencies.

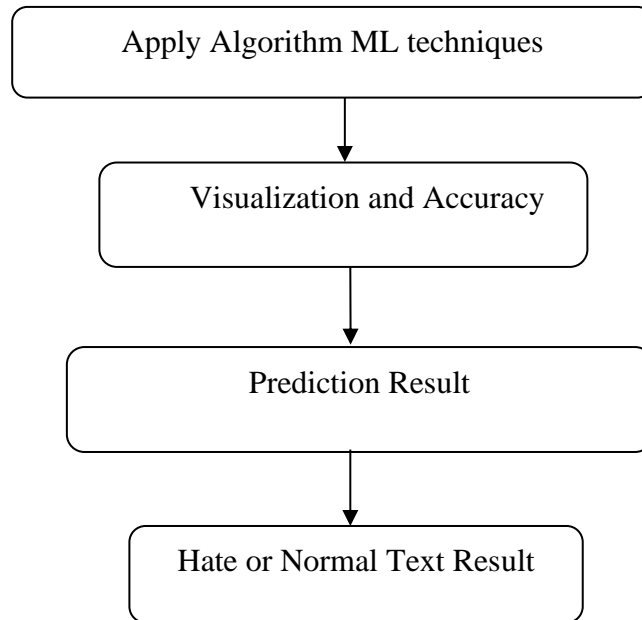


## DATA FLOW DIAGRAM:

Level-0:



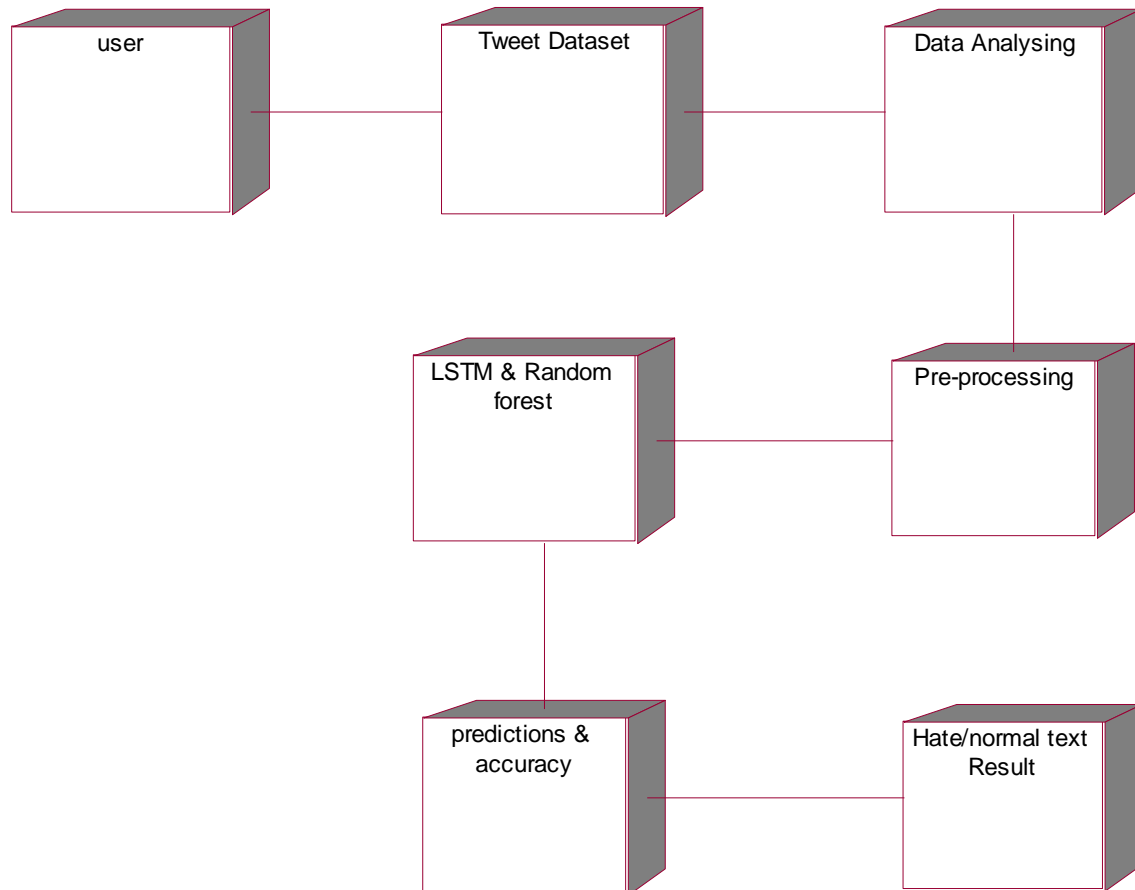
## Level-1



## EXPLANATION:

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. Often, they are a preliminary step used to create an overview of the system which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design). A DFD shows what kinds of data will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of processes, or information about whether processes will operate in sequence or in parallel.

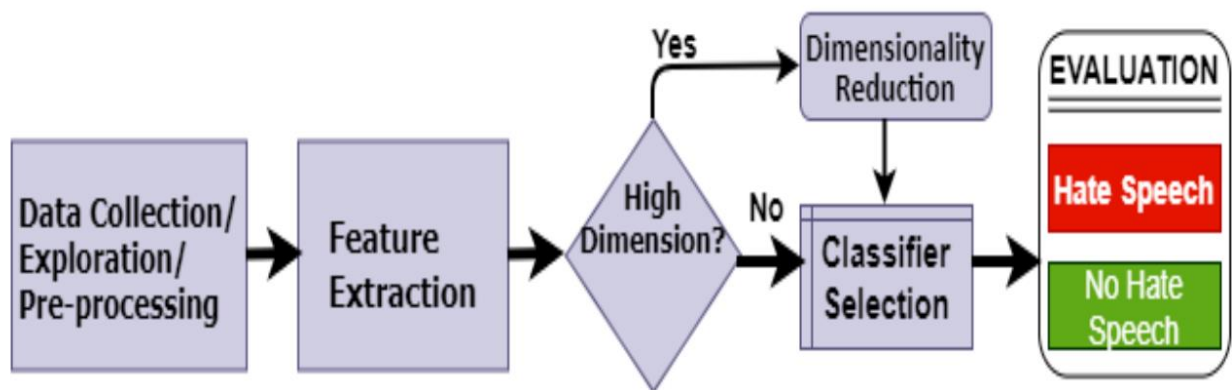
## Deployment Diagram



### EXPLANATION:

In the Unified Modeling Language, a component diagram depicts how components are wired together to form larger components and or software systems. They are used to illustrate the structure of arbitrarily complex systems. User gives main query and it converted into sub queries and sends through data dissemination to data aggregators. Results are to be showed to user by data aggregators. All boxes are components and arrow indicates dependencies.

## SYSTEM ARCHITECTURE



**Fig. Proposed Methodology**

# CHAPTER 5

## DEVELOPMENT TOOLS

### Python

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

### History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

### Importance of Python

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

## Features of Python

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to-maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** – Python provides interfaces to all major commercial databases.
- **GUI Programming** – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** – Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below –

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- IT supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

### **Libraries used in python:**

- numpy - mainly useful for its N-dimensional array objects.
- pandas - Python data analysis library, including structures such as dataframes.
- matplotlib - 2D plotting library producing publication quality figures.
- scikit-learn - the machine learning algorithms used for data analysis and data mining tasks.



Figure : NumPy, Pandas, Matplotlib, Scikit-learn

# **CHAPTER 6**

## **IMPLEMENTATION**

### **6.1 GENERAL**

Hate speech refers to any kind of communication in speech, writing or behaviour, which attacks or uses pejorative or discriminatory language regarding a person or a group based on some sensitive information or protected characteristics [5], [30]. These protected characteristics include religion, ethnicity, nationality, marital status, health status, race, colour, disability, sexual orientation, descent, gender or other identity factors [31]. Hate speech is a widespread phenomenon and has become an accepted reality as a common enemy of all law-abiding citizens across the world. This is a dangerous and illegal act that needs to be discouraged! Most of the hate speech messages on SM are constructed through texts [32]. However, images and sounds are also used in the dissemination of hate speeches [32]. Therefore, any attempt to address this problem through Computer perspective, text classification is the best bet. There is no universally accepted definition of hate speech, no consensus agreement on an individual definition [33]. It has been observed that a clearer and precise definition of hate speech can simplify the annotators work and consequently increase the annotators' agreement rate [34]. Although, it can be difficult in some countries to differentiate between appropriate speech and hate speech. Hence, giving a precise and universal definition of hate speech become more difficult and complicated. For example, there is a thin line between hate speech and normal speech under the First Amendment in the US. However, any speech that contributes to a criminal act is punishable as part of a hate crime. The debate on what can be classified as hate speech is not new, but there are conscious and renewed efforts as the world experience the Black Lives Matter (BLM) movement across the world. The BLM movement came up after the death of George Floyd. Beside hate speech, there are other abusive online behaviours which are worthy of clarification, such as cyberbullying. Cyberbullying as a kind of cyber harassment [35] means repetitive hostile behaviour through SM in an attempt to deliberately and consistently threaten or hurt individuals who cannot defend themselves easily [36], [37] and is common among youth [38], [39]. Cyber-hate or Hate speech and Cyberbullying are all different forms of abusive online behaviour [17], [36]. Cyberbullying can be considered as



Hate speech when sensitive or protected feature of a victim is the target of the attack. Hate speech is distinguished from cyber-bullying such that hate speech will affect not just a person but does have consequences for the entire group or society [18]. Hate speech is a complicated and multi-faceted concept that has been difficult to understand, by both human beings and computer systems [40].

## **CHAPTER 7**

### **SNAPSHOTS**

#### **GENERAL**

This project is implements like application using python and the Server process is maintained using the SOCKET & SERVERSOCKET and the Design part is played by Cascading Style Sheet.

#### **SNAPSHOTS**

# **CHAPTER 8**

## **SOFTWARE TESTING**

### **8.1 GENERAL**

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

### **8.2 DEVELOPING METHODOLOGIES**

The test process is initiated by developing a comprehensive plan to test the general functionality and special features on a variety of platform combinations. Strict quality control procedures are used. The process verifies that the application meets the requirements specified in the system requirements document and is bug free. The following are the considerations used to develop the framework from developing the testing methodologies.

### **8.3 Types of Tests**

#### **8.3.1 Unit testing**

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program input produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

### **8.3.2 Functional test**

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems/Procedures: interfacing systems or procedures must be invoked.

### **8.3.3 System Test**

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

### **8.3.4 Performance Test**

The Performance test ensures that the output be produced within the time limits, and the time taken by the system for compiling, giving response to the users and request being send to the system for to retrieve the results.

### **8.3.5 Integration Testing**

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g., components in a software system or – one step up – software applications at the company level – interact without error.

### **8.3.6 Acceptance Testing**

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

#### **Acceptance testing for Data Synchronization:**

- The Acknowledgements will be received by the Sender Node after the Packets are received by the Destination Node
- The Route add operation is done only when there is a Route request in need
- The Status of Nodes information is done automatically in the Cache Updation process

### **8.2.7 Build the test plan**

Any project can be divided into units that can be further performed for detailed processing. Then a testing strategy for each of this unit is carried out. Unit testing helps to identify the possible bugs in the individual component, so the component that has bugs can be identified and can be rectified from errors.

## **CHAPTER 9**

### **FUTURE ENHANCEMENT**

#### **9.1 FUTURE ENHANCEMENT**

In our future work, The application of ML for automatic HS detection on SM needs to be encouraged and supported. The needs to consider the HS variables based on each country is an issue that needs more researchers' attention. Each country or region may have different variables for HS. For example, marital status and health status are commonly used as HS variable in Nigeria but it has not been addressed by any work in the past. This research has found out that special characters and numeral symbols mostly used in Nigeria for constructing HS comments have not been addressed by current state-of-the-art. For example, the use of “419” to mean an unwholesome behaviour is commonplace in Nigeria. No research has covered this. The targeted audience for this research review is mostly newcomers in the domain of hate speech (text) classification in the SM. This review provides all the required steps needed to follow in conducting text classification tasks using ML and some open challenges in the domain.

## **CHAPTER 10**

### **CONCLUSION & REFERENCE**

#### **10.1 CONCLUSION**

This article reviewed advances made so far in automatic hate speech detection in social media. Hate speech as a societal problem is an old research area in the arts and humanities, however, it is still a new research area in the computing domain. Therefore, there is a need to constantly update researchers with the advances or progresses made to keep researchers informed. We analysed the approaches from classical ML, Ensemble and deep learning approaches in detecting hate speech in social media. This study found out that there is more research work in hate speech detection using classical ML than ensemble and deep learning techniques. That means researchers can explore more on hate speech detection using ensemble and deep learning methods. This research also discussed the weaknesses and strengths which can be of help in guiding the researchers' choice of one technique over the other. This article also identified some open challenges in hate speech detection which include: Cultural variations, pandemic or natural disaster, data sparsity, imbalance dataset challenge and dataset availability concern. The application of ML for automatic HS detection on SM needs to be encouraged and supported. The needs to consider the HS variables based on each country is an issue that needs more researchers' attention. Each country or region may have different variables for HS. For example, marital status and health status are commonly used as HS variable in Nigeria but it has not been addressed by any work in the past. This research has found out that special characters and numeral symbols mostly used in Nigeria for constructing HS comments have not been addressed by current state-of-the-art. For example, the use of "419" to mean an unwholesome behaviour is commonplace in Nigeria. No research has covered this. The targeted audience for this research review is mostly newcomers in the domain of hate speech (text) classification in the SM. This review provides all the required steps needed to follow in conducting text classification tasks using ML and some open challenges in the domain.

## REFERENCES

1. [1] M. S. Albarrak, M. Elnahass, S. Papagiannidis, and A. Salama, “The effect of Twitter dissemination on cost of equity: A big data approach,” *Int. J. Inf. Manage.*, vol. 50, pp. 1–16, Feb. 2020.
2. [2] C. Cai, H. Xu, J. Wan, B. Zhou, and X. Xie, “An attention-based friend recommendation model in social network,” *Comput., Mater. Continua*, vol. 65, no. 3, pp. 2475–2488, 2020.
3. [3] H. Watanabe, M. Bouazizi, and T. Ohtsuki, “Hate speech on Twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection,” *IEEE Access*, vol. 6, pp. 13825–13835, 2018.
4. [4] P. Fortuna and S. Nunes, “A survey on automatic detection of hate speech in text,” *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–30, Sep. 2018.
5. [5] A. Guterres, “United nations strategy and plan of action on hate speech,” United Nations, New York, NY, USA, Tech. Rep., 2019.
6. [6] Q. Li et al., *A Survey on Text Classification: From Shallow to Deep Learning*, vol. 37, no. 4. New York, NY, USA: Cornell Univ. Library, 2020.
7. [7] Q. Al-Maatouk, M. S. Othman, A. Aldraiweesh, U. Alturki, W. M. Al-Rahmi, and A. A. Aljeraiwi, “Task-technology fit and technology acceptance model application to structure and evaluate the adoption of social media in academia,” *IEEE Access*, vol. 8, pp. 78427–78440, 2020.
8. [8] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” *Information*, vol. 10, no. 4, pp. 1–68, 2019.
9. [9] T. Davidson, D. Warmesley, M. Macy, and I. Weber, “Automated hate speech detection and the problem of offensive language,” in *Proc. 11<sup>th</sup> Int. Conf. Web Soc. Media (ICWSM)*, 2017, pp. 512–515.
10. [10] Z. Waseem and D. Hovy, “Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter,” in *Proc. NAACL Student Res. Workshop*, 2016, pp. 88–93.



11. [11] P. Burnap and M. L. Williams, “Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making,” *Policy Internet*, vol. 7, no. 2, pp. 223–242, Jun. 2015.
12. [12] S. S. Bodrunova, A. Litvinenko, I. Blekanov, and D. Nepiyushchikh, “Constructive aggression? Multiple roles of aggressive content in political discourse on Russian YouTube,” *Media Commun.*, vol. 9, no. 1, pp. 181–194, Feb. 2021.
13. [13] F. Tulkens, “The hate factor in political speech. Where do responsibilities lie?” *Polish Ministry Admin. Digitization Council Eur.*, Warsaw, Poland, Tech. Rep., 2013.
14. [14] R. Slonje, P. K. Smith, and A. Frisé, “The nature of cyberbullying, and strategies for prevention,” *Comput. Hum. Behav.*, vol. 29, no. 1, pp. 26–32, Jan. 2013.
15. [15] M. A. Al-Garadi, M. R. Hussain, N. Khan, G. Murtaza, H. F. Nweke, I. Ali, G. Mujtaba, H. Chiroma, H. A. Khattak, and A. Gani, “Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges,” *IEEE Access*, vol. 7, pp. 70701–70718, 2019.
16. [16] M. Stegman and M. Loftin, “An essential role for down payment assistance in closing America’s racial homeownership and wealth gaps the price of the homeownership gap,” *Urban Inst.*, Washington, DC, USA, Tech. Rep., 2021.
17. [17] R. Alshalan and H. Al-Khalifa, “A deep learning approach for automatic hate speech detection in the Saudi Twittersphere,” *Appl. Sci.*, vol. 10, no. 23, pp. 1–16, 2020.
18. [18] A. Al-Hassan and H. Al-Dossari, “Detection of hate speech in social networks: A survey on multilingual corpus,” in *Proc. Comput. Sci. Inf. Technol. (CS IT)*, Feb. 2019, pp. 83–100.
19. [19] A. Schmidt and M. Wiegand, “A survey on hate speech detection using natural language processing,” in *Proc. 5th Int. Workshop Natural Lang. Process. Social Media*, 2017, pp. 1–10.
20. [20] A. Alrehili, “Automatic hate speech detection on social media: A brief survey,” in *Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2019, pp. 1–6.
21. [21] A. Rodriguez, C. Argueta, and Y.-L. Chen, “Automatic detection of hate speech on Facebook using sentiment and emotion analysis,” in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIIC)*, Feb. 2019, pp. 169–174.

22. [22] G. Weir, K. Owoeye, A. Oberacker, and H. Alshahrani, “Cloud-based textual analysis as a basis for document classification,” in Proc. Int. Conf. High Perform. Comput. Simul. (HPCS), Jul. 2018, pp. 629–633.
23. [23] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, “Antisocial behavior in online discussion communities,” in Proc. 9th Int. Conf. Web Soc. Media (ICWSM), 2015, pp. 61–70, 2015.
24. [24] T. Granskogen and J. A. Gulla, “Fake news detection: Network data from social media used to predict fakes,” in Proc. CEUR Workshop, vol. 2041, no. 1, 2017, pp. 59–66.
25. [25] L. Tamburino, G. Bravo, Y. Clough, and K. A. Nicholas, “From population to production: 50 years of scientific literature on how to feed the world,” *Global Food Secur.*, vol. 24, Mar. 2020, Art. no. 100346.
26. [26] V. S. Raleigh, “Trends in world population: How will the millennium compare with the past,” *Hum. Reprod. Update*, vol. 5, no. 5, pp. 500–505, 1999.