

**Project Batch ID**

B435

<b>Degree/ program</b>	B.Tech	<b>Specialisation</b>	Computer Science & Engineering
----------------------------	--------	-----------------------	--------------------------------

<b>Name of student</b>	<b>Register Number</b>	<b>Department</b>	<b>Mobile Number</b>	<b>Email ID</b>
Shivam Pandey	RA1911003010383	CSE	9026528477	<a href="mailto:Sp7287@srmist.edu.in">Sp7287@srmist.edu.in</a>
Mayank Sinha	RA1911003010386	CSE	9934996311	<a href="mailto:ms7651@srmist.edu.in">ms7651@srmist.edu.in</a>
<b>Academic Year</b>	2022-2023 (Odd)	<b>Semester</b>	7	
<b>Course Code</b>	18CSP107L	<b>Course Title</b>	Minor Project/Internship	

## Mission Statement

<b>Working Title of the Project:</b>	<b>A comparative study of toxic language detection using NLP algorithm.</b>		
<b>Project Site / Location</b>	Inhouse project		
<b>Name and address of the company / organisation (Applicable for projects with industry or industry support)</b>	SRM University, Kattankulathur, Chengalpattu District-603203		
<b>Supervision Team</b>			
	<b>Supervisor</b>	<b>Co-Supervisor</b>	<b>External Supervisor (If applicable)</b>
<b>Name</b>	Dr. G. Ramya		
<b>Designation</b>	Assistant Professor		
<b>Department</b>	Department of Computing Technologies		
<b>Campus</b>	Kattankulathur		
<b>Telephone</b>	9884655522		
<b>E-mail</b>	ramyag3@srmist.edu.in		

## **PRODUCT OR PROBLEM DESCRIPTION:**

As there are numerous algorithms utilised for toxic language detection in the field of natural language processing now, but they in this algorithm not developed considering production environment. The speed of categorization has a significant impact on how quickly an application is used in a production setting where the algorithm is running in real-time. The technique is only useful for theoretical applications if it performs classification tasks slowly. In the actual world, in this algorithm need algorithms that operates quickly and with excellent accuracy. Through this research, in this algorithm hope to identify an algorithm that works best in the actual world, meaning that there is no compromise in this algorithm speed and accuracy. For the classification job, in this algorithm will make use of the toxic language dataset. After classifiers have been trained, they will be put to the test using a variety of criteria, including accuracy, recall, precision, F1 score, and prediction time. In this algorithm shall categorise the flaws in different NLP algorithms based on the parameters.

## **ASSUMPTIONS AND CONSTRAINTS:**

In this project it is assumed that each and every supervised algorithm tested is capable in detecting the toxic comment. In this project the specific use case of an algorithm is ignored.

The constraints is the specificity of an algorithm in Natural languages processing field.

### Division of work and contributors

Time period		Activities or components of the project	Name/Register Number of the Individual Contributor	Names/Register Number of the Joint Contributors
From Date	To Date			
10/08/22	12/08/22	Finalising the project outline.	Mayank Sinha	
15/08/22	18/08/22	Requirement gathering.	Mayank Sinha	Mayank Sinha/ Shivam Pandey
25/08/22	30/08/22	Data source identification and UML diagram designing.		Mayank Sinha/ Shivam Pandey
10/09/22	17/09/22	Data extraction and cleaning.	Shivam Pandey	
25/09/22	28/09/22	Studying about NLP algorithms		Mayank Sinha/ Shivam Pandey
03/10/22	10/10/22	NLP model selection	Mayank Sinha	
11/10/22	16/10/22	NLP model training and testing	Shivam Pandey	
20/10/22	27/10/22	NLP app development	Shivam Pandey	
28/10/22	01/11/22	NLP app deployment	Shivam Pandey	

### Summary record of major progress meetings with supervisors

Summary record of major progress meetings with supervisors			Working title of dissertation/research project:	
Meeting date & supervisors present	Progress since last meeting	Agreed programme of work and target dates	Other issues, e.g. facilities, supervision, training needs, etc.	Date of next meeting
18.08.2022	Created the PPT for the zeroth review and the finalization of the project title	Creating the ppt for the first review, literature survey, challenges to address, UML BLOCK diagram, Algorithm Used	Needed to work more on the presentation part	21.09.2022
22.09.2022	Created the first review PPT containing all the required needs.  UML diagrams.	To study around 15 to 20 research papers and to conclude the results of different nlp algorithms.	Needed to read and study more research papers.  Architecture diagram  Detailed design diagrams  Research gap	17.10.2022
17.10.2022	Created the ppt for the third review.  Read more than 15 research papers.	To complete the full project paper and to publish the paper on an online forum.	To complete the full final report and to publish it on an online forum.	05.10.2022
05.10.2022	Completed the final report and submitted the paper on an online forum.	This is the final review.		