

# JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION

TECHNOLOGY

2020-21



## MINOR-II PROJECT REPORT

PANTOMATH

**Name of Students:** Kapil Israni, Ayush Nagar, Akshara Nigam

**Enrollment Nos:** 17104011, 17104012, 17104018

**Name of Supervisor:** Mrs. Sharddha Porwal

Submitted in partial fulfilment of the degree of Bachelor of Technology in Information Technology.

# Table of Contents

	Page No.
<i>Student Declaration</i>	2
<i>Certificate</i>	3
<i>Acknowledgement</i>	4
<i>Summary</i>	5
<i>List of Figures</i>	6
<i>List of Tables</i>	7
<b>Chapter 1: Introduction</b>	
1.1 General Introduction	8
1.2 Problem Statement	8
1.3 Significance/Novelty of the Problem	9
1.4 Brief of the Solution	10
<b>Chapter 2: Literature Survey</b>	12
<b>Chapter 3: Requirement Analysis And Solution Approach</b>	
3.1 Functional and Non-Functional Requirements	20
3.2 High Level Design	20
3.3 Solution approach	21
3.4 Algorithm Details	22
<b>Chapter 4: Implementation</b>	25
<b>Chapter 5: Testing</b>	28
<b>Chapter 6: Findings and Conclusion</b>	31
<i>References</i>	32

## **STUDENT DECLARATION**

We hereby declare that this submission is our own work and that to the best of our knowledge and belief it contains more material previously published or written by another person no material which has been accepted for the reward of any degree or diploma of the university or any other institute of higher learning, except where due acknowledgement has been made in the text.

Date:- 15/05/2020

Kapil Israni (17104011)

Ayush Nagar (17104012)

Akshara Nigam (170104018)

## **CERTIFICATE**

This is to certify that the work entitled “**Pantomath**” submitted by **Kapil Israni, Ayush Nagar** and **Akshara Nigam** of B.Tech (I.T) of Jaypee Institute of Information Technology Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other university or institute for the award of any other degree or diploma.

### **Signature of the Supervisor**

Name of the Supervisor : Mrs. Shraddha Porwal

Date : 15/05/2020

## **ACKNOWLEDGEMENT**

First and foremost we would like to thank our mentor Mrs.Shraddha Porwal of Jaypee Institute of Information Technology, Noida for guiding us thoughtfully and efficiently throughout this project, giving us an opportunity to work at our own pace along our own lines, while providing us with very useful directions whenever necessary.

We would also like to thank our friends and classmates for being great sources of motivation and for providing us encouragement throughout the length of this project. We offer our sincere thanks to other persons who knowingly or unknowingly helped us in this project.

### **Signature of Students**

Kapil Israni (17104011)

Ayush Nagar (17104012)

Akshara Nigam (17104018)

## **SUMMARY**

The main idea of the project is to get to know a person on the basis of his/her twitter posts and retweets. It gives the complete overview of the person which includes the personality type and whether he/she was involved in any hate speech, correspondingly displaying the tweets. Knowing a person that may be a stranger becomes important in situations such as hiring, collaborations and many other unlisted purposes.

There are different methods to predict personality types based on meta programmes. The Myers-Briggs Type Indicator (MBTI) is currently considered as one of the most popular and reliable methods. It describes the preferences of an individual in four dimensions and these basic dimensions combine into one of 16 different personality types. These four dimensions are Extroversion-Introversion (E-I), Sensitive-Intuitive (S-N), Thinking-Feeling (T-F) and Judging-Perceiving (J-P). The above analysis for personality detection is performed using Linear SVC while hate speech detection is done using Logistic Regression as a classifier.

## LIST OF FIGURES

Fig 1	Pg 8	MBTI Type Indicators
Fig 2	Pg 9	16 Personality Types
Fig 3	Pg 20	Project Working
Fig 4	Pg 22	Available Algorithms
Fig 5	Pg 23	Logistic Regression Classification
Fig 6	Pg 23	Linear SVC
Fig 7	Pg 24	Formula of Linear SVC
Fig 8	Pg 25	Plot of Correlation Matrix of the four Personality types
Fig 9	Pg 25	Plot of the count of all the personality types in the dataset
Fig 10	Pg 26	Web Application Product
Fig 11	Pg 26	View of the Application showing the hate speech tweets
Fig 12	Pg 27	Tfidf vs Count Vectorization
Fig 13	Pg 29	Comparing the Accuracies for various Algorithms
Fig 14	Pg 29	Algorithm analysis

## LIST OF TABLES

Table No.	Table Name	Page no.
1	Paper 1	12
2	Paper 2	14
3	Paper 3	16
4	Paper 4	18



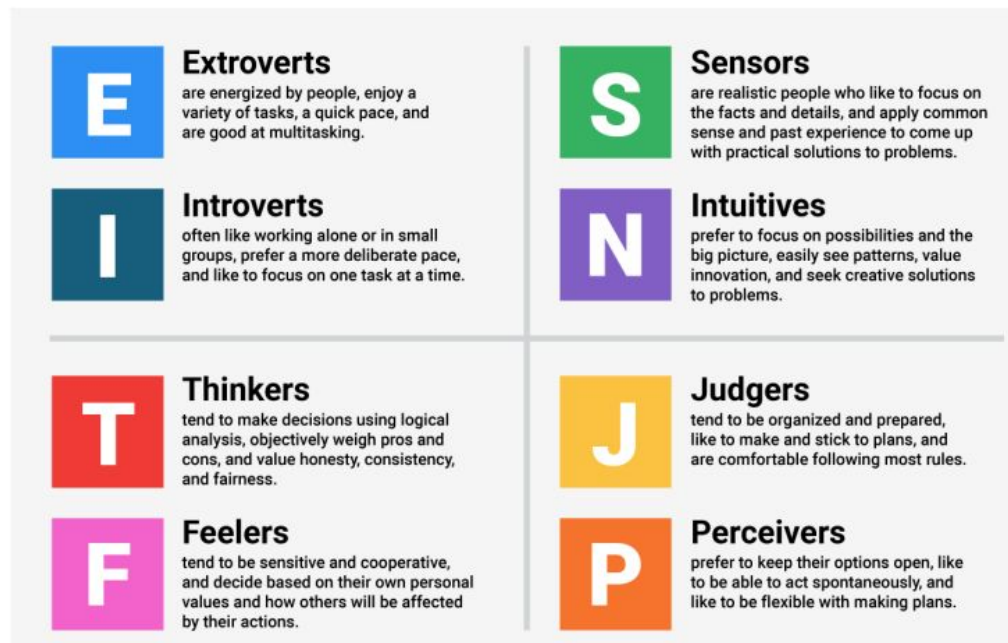
# 1. INTRODUCTION

## 1.1 General Introduction

In the world surrounded by gadgets and applications, social media stands above all, because it is one of the biggest sources of both information and expression. Twitter, being one of the most commonly used social media platforms allows its users to freely communicate, talk about new ideas, start different campaigns, look up to our leaders and idols, follow them and even express angst against the ones we dislike. Thus, it is a platform collecting all kinds of expressions and desires the world has and so we use this information for our purpose of the project to determine the personality type and hate speech for a user to help the ones who wish to know everything (Pantomath) before taking a decision.

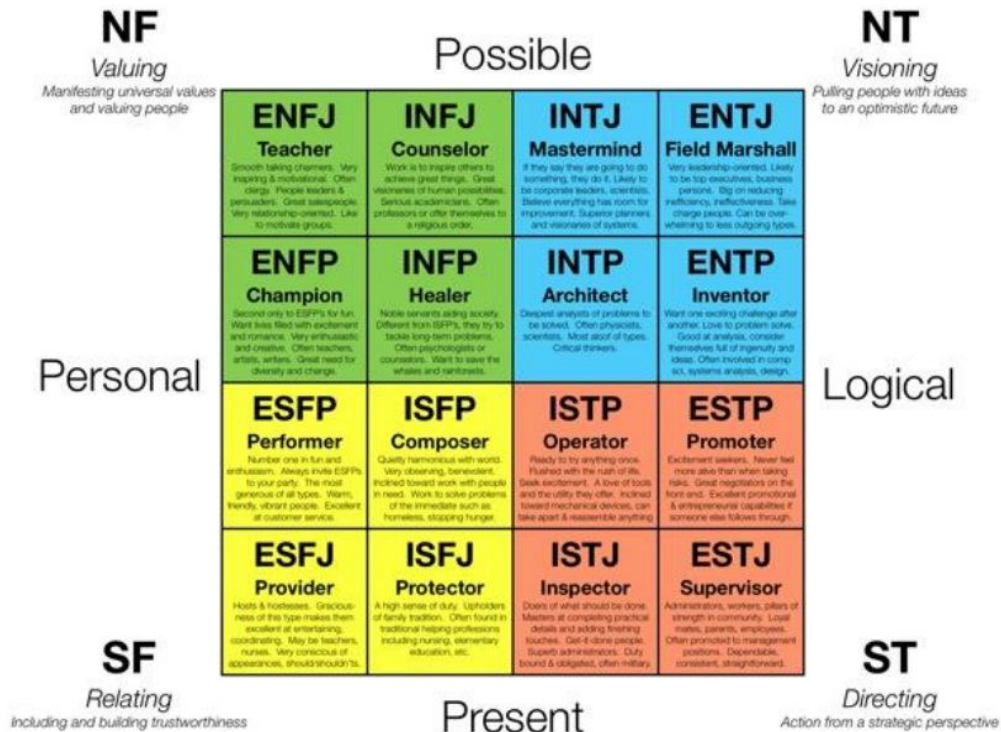
## 1.2 Problem Statement

Firstly, we aim to categorize the twitter users based on their personality types as given by the Myers–Briggs Type Indicator (shown in Fig 1 )



(Fig 1 : MBTI Personality Types)

The sixteen categories as shown in Fig 2 help to determine the kind of person one is. For example, if a recruiter who wishes to hire a person who is logical and able to design new things, he would want to hire a person of the type ENTP.



(Fig 2 : Conclusion of Each Personality)

Secondly, we also analyze whether the person is involved in hate speech or not, because everybody wishes for a good and healthy working environment and doesn't want any ill-behavioural experience around.

### 1.3 Significance/Novelty of the Problem

The project aims to solve the problems of the people who wish to hire the right kind of staff but are often misled by the outer appearance. Through this, they will be able to give the right kind of work to every member and also know their mind while communicating with everybody.

## 1.4 Brief Description of the Solution Approach

The tweets are first cleaned and then fed to the training model. Pre-processing of the text is done by :-

- **Stemming** : Words are reduced to their word stems. A word stem need not be the same root as a dictionary-based morphological root, it just is an equal to or smaller form of the word. There are different algorithms that can be used in the stemming process, but the most common in English is Porter stemmer.
- **Lemmatization** : It involves resolving words to their dictionary form. It requires a lot more knowledge about the structure of a language, it's a much more intensive process than just trying to set up a heuristic stemming algorithm.
- **Count Vectorization** : It involves counting the number of occurrences each word appearing in a document (i.e distinct text such as an article, book, even a paragraph!). The parameter “max\_df”, sets how many features or words you want CountVectorizer to count. Setting this parameter helps in dealing with a large number of documents to avoid speed issues with computation. Also, we remove stop words that could be irrelevant or unnecessary in order to focus on more relevant words in the analysis.
- **Tf-Idf** : The term frequency refers to how much a term (i.e. a word) appears in a document. Inverse document frequency refers to how common or rare a term appears in a document.

The hate speech analysis is done using **Logistic Regression**. The dataset used for analysis is from Kaggle and various other sources. It consists of Tweet and Label

columns. Label 0 means No Hate Speech while Label 1 means Hate speech is detected.

Personality Detection is done using **Linear Support Vector Classifier** and the dataset is taken from Kaggle. It consists of tweet and type, for some of the tweets we had to extract the type from the uneven dataset.

## 2. LITERATURE SURVEY

### 2.1 Paper I

Title of the paper	Linguistic Features Based Personality Recognition Using Social Media Data
Authors	Dilini Sewwandi, Kusal Perera, Sajith Sandaruwan, Anupiya Nugaliyadde, Samantha Thelijagoda
Year of publication	2017
Publishing Details	6th National Conference on Technology and Management (NCTM), Malabe, Sri Lanka

Objective	Results
<p>Social Media has become a prominent platform for opinion and thoughts. This stated that the characteristics of a person can be assessed through social media status updates. The purpose of the research article is to provide a web application in order to detect one's personality using linguistic feature analysis. The personality of the person is classified according to Eysenck's Three-factor personality model. The proposed technique is based on ontology based text classification, linguistic feature-vector matrix using LIWC features including semantic</p>	<p>Fig. 1. High Level Architectural Diagram of the Proposed System</p>

analysis using supervised machine learning algorithms and questionnaire-based personality detection. This is vital for HR management system when recruiting and promoting employees, R&D Psychologists can use dynamic ontology for storage purposes and other API users including universities and sports club.

TABLE II. EXPERIMENTAL RESULT TABLE  
FOR DIFFERENT STUDENT CATEGORIES

Category	Student Name	Extrovert	Introvert	Neuroticism	Emotional Stability	Psychoticism	Tender
SLIIT 1 <sup>st</sup> Year Students	Test01	Agree	Agree	Agree	Agree	Agree	Agree
	Test02	Disagree	Disagree	Agree	Agree	Agree	Agree
	Test03	Agree	Agree	Disagree	Disagree	Agree	Agree
	Test04	Disagree	Disagree	Agree	Agree	Disagree	Disagree
	Test05	Agree	Agree	Agree	Agree	Disagree	Disagree
SLIIT Curtin Students	Test06	Agree	Agree	Agree	Agree	Disagree	Disagree
	Test07	Agree	Agree	Agree	Agree	Agree	Agree
	Test08	Disagree	Disagree	Disagree	Disagree	Agree	Agree
	Test09	Agree	Agree	Agree	Agree	Disagree	Disagree
	Test10	Disagree	Disagree	Agree	Agree	Agree	Agree
A/L Students	Test11	Agree	Agree	Agree	Agree	Disagree	Disagree
	Test12	Disagree	Disagree	Agree	Agree	Disagree	Disagree
	Test13	Disagree	Disagree	Disagree	Disagree	Agree	Agree
	Test14	Agree	Agree	Disagree	Disagree	Agree	Agree
	Test15	Agree	Agree	Agree	Agree	Agree	Agree

## 2.2 Paper II

Title of the paper	Twitter Personality based Influential Communities Extraction System
Authors	Eleanna Kafeza, Andreas Kanavos, Christos Makris, Pantelis Vikatos
Year of publication	2014
Publishing Details	IEEE International Congress on Big Data

Objective	Results																																								
<p>The identification of influential users in social media communities has been recent of major concern since these users can contribute to viral marketing campaigns. In our approach, we extend the notion of influence from users to networks and consider personality as a key characteristic for identifying influential networks. We describe the Twitter Personality based Influential Communities Extraction (T-PICE) system that creates the best influential communities in a Twitter network graph considering users' personality. We then expand existing approaches in users' personality extraction by aggregating data that</p>	<div>Table VI</div> <div>NORMALIZED METRIC FOR RATING INFLUENTIAL COMMUNITIES</div> <table><tr><th>Communities Decomposition</th><th>Tweets / Size</th><th>Followers / Size</th><th>Borda Count / Size</th></tr><tr><td>Simple Blondel</td><td>1,704</td><td>2,201</td><td>1,150</td></tr><tr><td>EL</td><td>2,065</td><td>2,794</td><td>1,467</td></tr><tr><td>DL</td><td>1,651</td><td>2,584</td><td>1,054</td></tr><tr><td>AEO</td><td>1,322</td><td>1,892</td><td>1,111</td></tr></table> <div>Table VII</div> <div>AVERAGE OF DISSIMILARITY RATES OF USERS' PERSONALITY OF THE TOP COMMUNITIES</div> <table><tr><th>Communities Decomposition</th><th>Ranking Tweets</th><th>Ranking Followers</th><th>Ranking Borda Count</th></tr><tr><td>Simple Blondel</td><td>58,7%</td><td>66,3%</td><td>66,3%</td></tr><tr><td>EL</td><td>69,3%</td><td>70,2%</td><td>67,8%</td></tr><tr><td>DL</td><td>54,1%</td><td>61,1%</td><td>54,2%</td></tr><tr><td>AEO</td><td>63,4%</td><td>60,2%</td><td>60,2%</td></tr></table>	Communities Decomposition	Tweets / Size	Followers / Size	Borda Count / Size	Simple Blondel	1,704	2,201	1,150	EL	2,065	2,794	1,467	DL	1,651	2,584	1,054	AEO	1,322	1,892	1,111	Communities Decomposition	Ranking Tweets	Ranking Followers	Ranking Borda Count	Simple Blondel	58,7%	66,3%	66,3%	EL	69,3%	70,2%	67,8%	DL	54,1%	61,1%	54,2%	AEO	63,4%	60,2%	60,2%
Communities Decomposition	Tweets / Size	Followers / Size	Borda Count / Size																																						
Simple Blondel	1,704	2,201	1,150																																						
EL	2,065	2,794	1,467																																						
DL	1,651	2,584	1,054																																						
AEO	1,322	1,892	1,111																																						
Communities Decomposition	Ranking Tweets	Ranking Followers	Ranking Borda Count																																						
Simple Blondel	58,7%	66,3%	66,3%																																						
EL	69,3%	70,2%	67,8%																																						
DL	54,1%	61,1%	54,2%																																						
AEO	63,4%	60,2%	60,2%																																						

represent several aspects of user behaviour using machine learning techniques. We use an existing modularity based community detection algorithm and we extend it by inserting a pre-processing step that eliminates graph edges based on users' personality. The effectiveness of our approach

is demonstrated by sampling the twitter graph and comparing the influence of the created communities with and without considering the personality factor. We define several metrics to count the influence of communities. Our results show that the T-PICE system creates the most influential communities.

Table IV  
10-FOLD CROSS-VALIDATION

Classifiers	A	C	E	N	O
AdaBoost	0.7	<b>0.719</b>	0.581	0.481	0.67
BayesNet	0.726	0.47	<b>0.747</b>	0.517	0.617
IBK	0.476	0.671	0.517	0.469	0.587
J48	0.6	0.7	0.76	0.359	0.52
JRip	<b>0.824</b>	0.525	0.517	0.474	<b>0.695</b>
Multilayer Perceptron	0.473	0.504	0.333	0.408	0.679
Naive Bayes Classifier	0.476	0.678	0.46	0.407	0.605
PART	0.626	0.702	0.669	0.282	0.541
Ridor	0.624	0.52	0.467	<b>0.606</b>	0.585
RotationForest	0.523	0.543	0.594	0.43	0.658
SMO	0.45	0.577	0.367	0.469	0.664

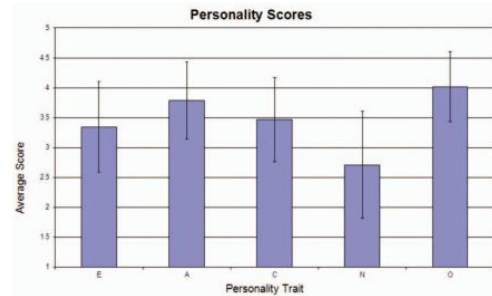
Table V  
LEAVE-ONE-OUT CROSS-VALIDATION

Classifiers	A	C	E	N	O
AdaBoost	0.62	<b>0.726</b>	0.581	0.307	0.605
BayesNet	<b>0.726</b>	0.426	<b>0.803</b>	0.457	0.544
IBK	0.426	0.671	0.452	<b>0.506</b>	0.587
J48	0.65	0.579	0.758	0.469	<b>0.65</b>
JRip	0.724	0.435	0.556	0.428	0.648
Multilayer Perceptron	0.423	0.504	0.273	0.43	0.561
Naive Bayes Classifier	0.476	0.645	0.43	0.344	0.61
PART	0.65	<b>0.726</b>	0.664	0.452	<b>0.65</b>
Ridor	0.65	0.47	0.452	0.343	0.64
RotationForest	0.597	0.629	0.493	0.407	0.601
SMO	0.423	0.55	0.367	0.407	0.561



## 2.3 Paper III

Title of the paper	Predicting Personality from Twitter
Authors	Jennifer Golbeck, Cristina Robles, Michon Edmondson, Karen Turner
Year of publication	2011
Publishing Details	IEEE International Conference on Privacy, Security, Risk, and Trust and IEEE International Conference on Social Computing

Objective	Results												
<p>Social media is a place where users present themselves to the world, revealing personal details and insights into their lives. We are beginning to understand how some of this information can be utilized to improve the users' experiences with interfaces and with one another. In this paper, we are interested in the personality of users. Personality has been shown to be relevant to many types of interactions; it has been shown to be useful in predicting job satisfaction, professional and romantic relationship success, and even preference for different interfaces. Until now, to accurately gauge users' personalities, they needed to take a personality test. This made it impractical to use personality</p>	 <p>Fig. 2: Average scores on each personality trait shown with standard deviation bars.</p> <table border="1"> <caption>Personality Scores Data</caption> <thead> <tr> <th>Personality Trait</th> <th>Average Score</th> </tr> </thead> <tbody> <tr> <td>E</td> <td>3.4</td> </tr> <tr> <td>A</td> <td>3.8</td> </tr> <tr> <td>C</td> <td>3.5</td> </tr> <tr> <td>N</td> <td>2.7</td> </tr> <tr> <td>O</td> <td>4.0</td> </tr> </tbody> </table>	Personality Trait	Average Score	E	3.4	A	3.8	C	3.5	N	2.7	O	4.0
Personality Trait	Average Score												
E	3.4												
A	3.8												
C	3.5												
N	2.7												
O	4.0												

analysis in many social media domains. In this paper, we present a method by which a user's personality can be accurately predicted through the publicly available information on their Twitter profile. We will describe the type of data collected, our methods of analysis, and the machine learning techniques that allow us to successfully predict personality. We then discuss the implications this has for social media design, interface design, and broader domains.

TABLE I: Average scores on each personality factor on a normalized 0-1 scale

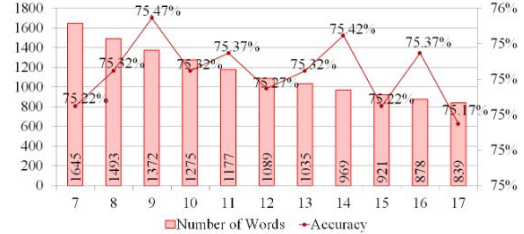
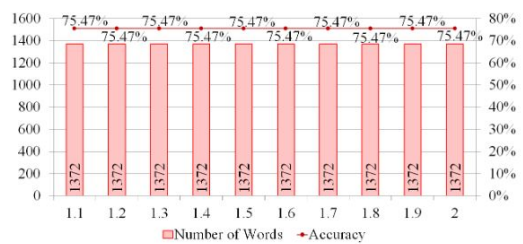
	Agree.	Consc.	Extra.	Neuro.	Open.
Average	0.697	0.617	0.586	0.428	0.755
Stdev	0.162	0.176	0.190	0.224	0.147

TABLE III: Mean Absolute Error on a normalized scale for each algorithm and personality trait.

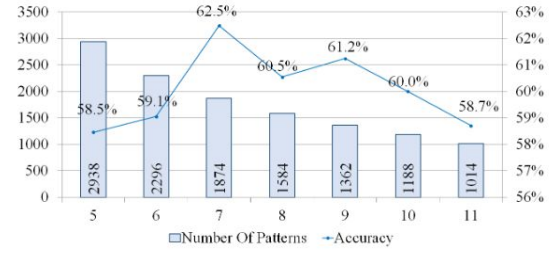
	Agree.	Consc.	Extra.	Neuro.	Open.
ZeroR	0.129980265	0.146204953	0.160241663	0.182122225	0.11923333
GaussianProcess	0.130675423	0.14599073	0.160315335	0.18205923	0.11922558

## 2.4 Paper IV

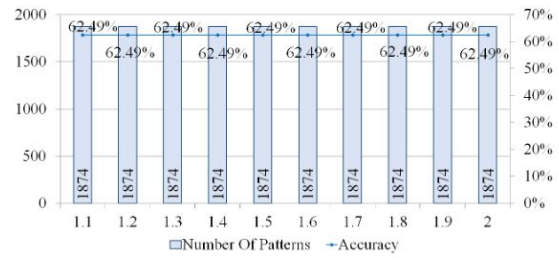
Title of the paper	Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection
Authors	Hajime Watanabe, Mondher Bouazizi , And Tomoaki Ohtsuki
Year of publication	2018
Publishing Details	IEEE

Objective	Results																																																																					
<p>With the rapid growth of social networks and microblogging websites, communication between people from different cultural and psychological backgrounds has become more direct, resulting in more and more “cyber” conflicts between these people. Consequently, hate speech is used more and more, to the point where it has become a serious problem invading these open spaces. Hate speech refers to the use of aggressive, violent or offensive language, targeting a specific group of people sharing a common property, whether this property is their gender (i.e., sexism), their ethnic group or race (i.e., racism) or their believes and religion. While most of the online social networks and microblogging websites forbid the use of hate speech, the size of these networks and websites</p>	<div><table><thead><tr><th>min<sub>occ</sub><sup>u</sup></th><th>Number of Words</th><th>Accuracy</th></tr></thead><tbody><tr><td>7</td><td>1645</td><td>75.22%</td></tr><tr><td>8</td><td>1493</td><td>75.32%</td></tr><tr><td>9</td><td>1372</td><td>75.47%</td></tr><tr><td>10</td><td>1275</td><td>75.32%</td></tr><tr><td>11</td><td>1177</td><td>75.37%</td></tr><tr><td>12</td><td>1089</td><td>75.27%</td></tr><tr><td>13</td><td>1035</td><td>75.32%</td></tr><tr><td>14</td><td>969</td><td>75.42%</td></tr><tr><td>15</td><td>921</td><td>75.22%</td></tr><tr><td>16</td><td>878</td><td>75.37%</td></tr><tr><td>17</td><td>839</td><td>75.17%</td></tr></tbody></table><p><b>FIGURE 4.</b> Classification accuracy (right axis) and number of words collected (left axis) for different values of the parameter <math>min_{occ}^u</math>.</p></div> <div><table><thead><tr><th>Th<sub>u</sub></th><th>Number of Words</th><th>Accuracy</th></tr></thead><tbody><tr><td>1.1</td><td>1372</td><td>75.47%</td></tr><tr><td>1.2</td><td>1372</td><td>75.47%</td></tr><tr><td>1.3</td><td>1372</td><td>75.47%</td></tr><tr><td>1.4</td><td>1372</td><td>75.47%</td></tr><tr><td>1.5</td><td>1372</td><td>75.47%</td></tr><tr><td>1.6</td><td>1372</td><td>75.47%</td></tr><tr><td>1.7</td><td>1372</td><td>75.47%</td></tr><tr><td>1.8</td><td>1372</td><td>75.47%</td></tr><tr><td>1.9</td><td>1372</td><td>75.47%</td></tr><tr><td>2</td><td>1372</td><td>75.47%</td></tr></tbody></table><p><b>FIGURE 5.</b> Classification accuracy (right axis) and number of words collected (left axis) for different values of the parameter <math>Th_u</math>.</p></div>	min <sub>occ</sub> <sup>u</sup>	Number of Words	Accuracy	7	1645	75.22%	8	1493	75.32%	9	1372	75.47%	10	1275	75.32%	11	1177	75.37%	12	1089	75.27%	13	1035	75.32%	14	969	75.42%	15	921	75.22%	16	878	75.37%	17	839	75.17%	Th <sub>u</sub>	Number of Words	Accuracy	1.1	1372	75.47%	1.2	1372	75.47%	1.3	1372	75.47%	1.4	1372	75.47%	1.5	1372	75.47%	1.6	1372	75.47%	1.7	1372	75.47%	1.8	1372	75.47%	1.9	1372	75.47%	2	1372	75.47%
min <sub>occ</sub> <sup>u</sup>	Number of Words	Accuracy																																																																				
7	1645	75.22%																																																																				
8	1493	75.32%																																																																				
9	1372	75.47%																																																																				
10	1275	75.32%																																																																				
11	1177	75.37%																																																																				
12	1089	75.27%																																																																				
13	1035	75.32%																																																																				
14	969	75.42%																																																																				
15	921	75.22%																																																																				
16	878	75.37%																																																																				
17	839	75.17%																																																																				
Th <sub>u</sub>	Number of Words	Accuracy																																																																				
1.1	1372	75.47%																																																																				
1.2	1372	75.47%																																																																				
1.3	1372	75.47%																																																																				
1.4	1372	75.47%																																																																				
1.5	1372	75.47%																																																																				
1.6	1372	75.47%																																																																				
1.7	1372	75.47%																																																																				
1.8	1372	75.47%																																																																				
1.9	1372	75.47%																																																																				
2	1372	75.47%																																																																				

makes it almost impossible to control all of their content. Therefore, arises the necessity to detect such speech automatically and filter any content that presents hateful language or language inciting to hatred. In this paper, we propose an approach to detect hate expressions on Twitter. Our approach is based on unigrams and patterns that are automatically collected from the training set. These patterns and unigrams are later used, among others, as features to train a machine learning algorithm.



**FIGURE 7.** Classification accuracy (right axis) and number of patterns collected (left axis) for different values of the parameter  $\min_{occ}^p$ .



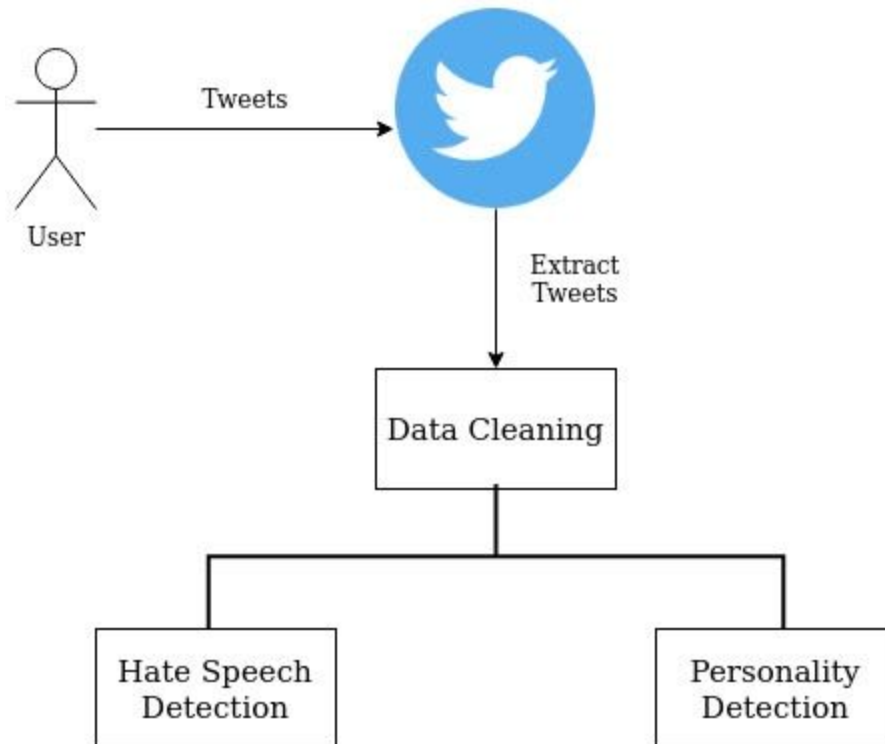
**FIGURE 8.** Classification accuracy (right axis) and number of patterns collected (left axis) for different values of the parameter  $Th_p$ .

### 3. REQUIREMENT ANALYSIS AND SOLUTION APPROACH

#### 3.1 Functional and Non-Functional Requirements

- **ReactJS** : It is a Javascript frontend framework that creates virtual dom which reduces the loading time of a page. The division which has changed only gets reloaded and the rest of the divisions remain unchanged.
- **Flask** : It is a Python framework that helps in creating RestFull API and makes it easy to integrate it with the other functionalities.

#### 3.2 High Level Design :



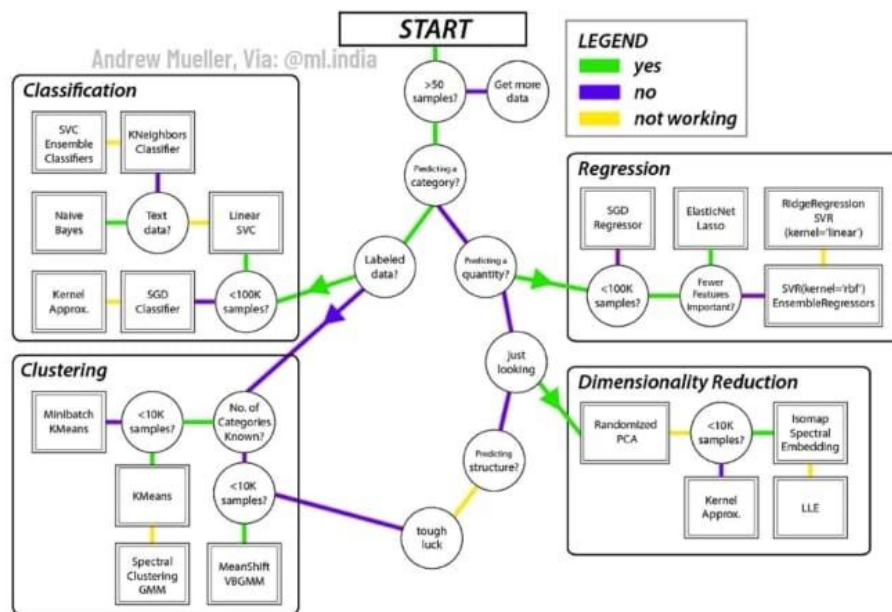
(Fig 3 : Overview)

The above figure gives an overview of the working of the project model and how we use the data for analysis purposes and hence make predictions.

**3.3 Solution Approach :** Firstly, we analyze the dataset to establish what kind of machine learning technique would be used. Since the dataset consists of both input and expected output, we use **Supervised Learning (Classification)**.

- For personality detection, the classification task was divided into 16 sub-classes and further into four binary classification tasks, since each MBTI type is made of four binary classes. Each one of these binary classes represents an aspect of personality according to the MBTI personality model. As a result, four different binary classifiers were trained, whereby each one specializes in one of the aspects of personality. Thus, in this step, a model for each type indicator was built individually. Term Frequency-Inverse Document Frequency (TF-IDF) was performed and MBTI type indicators were binarised. Variable X was used for posts in TF-IDF representation and variable Y was used for the binarised MBTI type indicator. After extracting the features, we then train and test for different types of classification techniques and then compare the accuracy and select the most suitable model. The trained model is pickled and stored to avoid training over and over again and decrease the computation time.
- For hate speech detection, we have used Logistic Regression to classify the tweets. Count Vectorisation is used to form the bag of words. We merged three different datasets and performed the cleaning of data. We had 2 classes 0 for not a hate speech and 1 for hate speech. We compare the accuracy score and F1 score with other algorithms and select the best model. The trained model is pickled and saved to increase the responsiveness.
- For extracting tweets, the user handle is provided and correspondingly, through the twitter API we extract all the posts/tweets and retweets of the user and pass it to the personality detection model and hate speech detection model and the analysis of the result is shown.

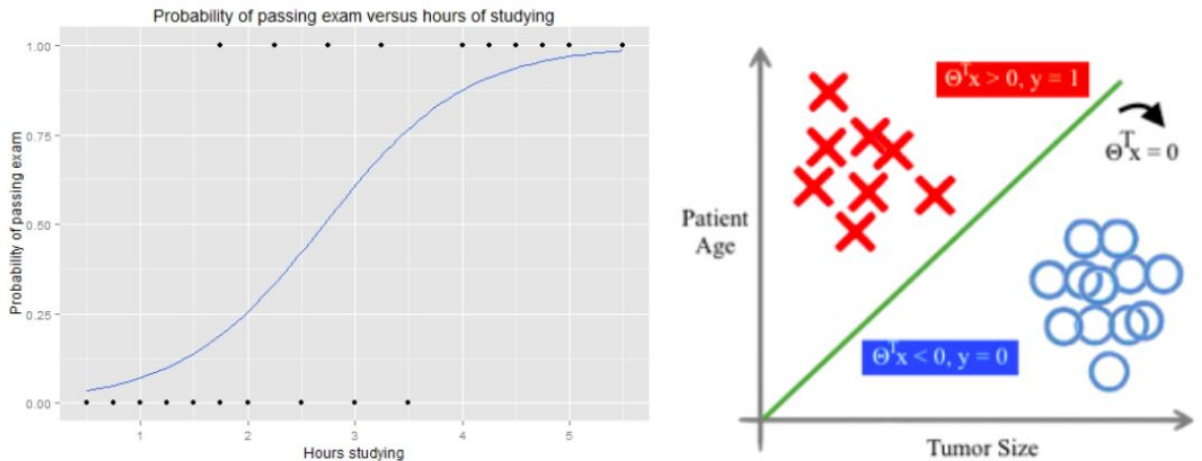
### 3.4 Algorithm Details



(Fig 4 : Available Algorithms)

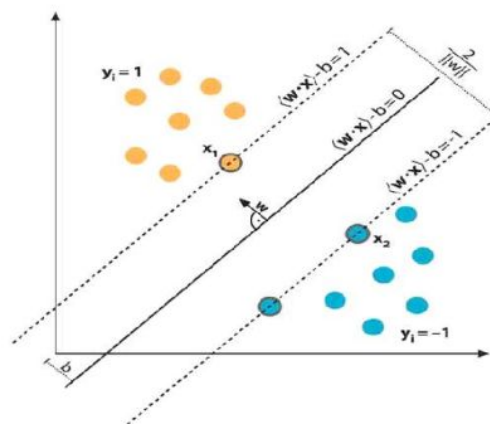
The above diagram explains in brief the kind of algorithms we should use for our purpose.

**Logistic Regression Classifier :-** Logistic Regression is a 'Statistical Learning' technique dedicated to 'Classification' tasks. As a 'Supervised-Classification' method, Logistic Regression helps us to converge those 'uncertain' posteriors with a differentiable 'decision function'. Here  $C=100$  where  $C$  is a control variable that retains strength modification of Regularization by being inversely positioned to the Lambda regulator.



(Fig 5 : Logistic Regression Classifier)

**Linear Support Vector Classifier :-** It's a Support Vector Machine, which is used for classification purposes. The goal of the SVM algorithm is to create the best boundary that can segregate the points into respective classes so that we can easily put the new data point in the correct category. This decision boundary is called a hyperplane. So, we look to maximize the margin between the data points and the hyperplane (in this case a Kernel='linear').



(Fig 6 : Linear SVC )

The 'hinge loss' is used to maximize the loss function value. When our model correctly predicts the class of the data point, we only update the gradient from the regularization parameter.



When there is a misclassification, we include the loss along with the regularization parameter to perform gradient update.

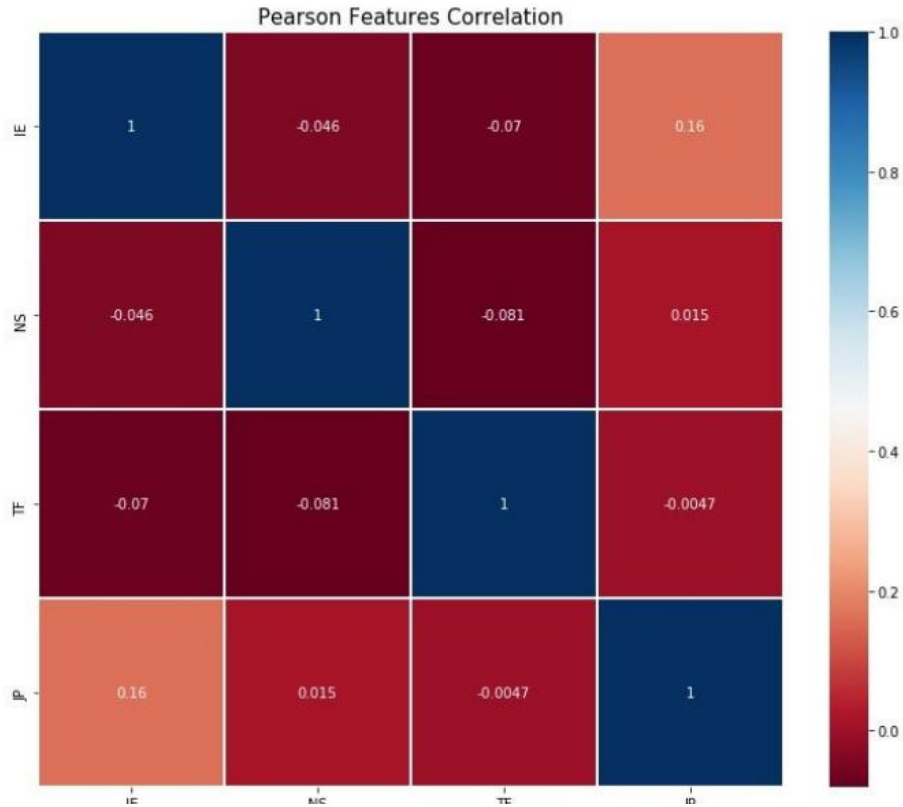
$$\begin{array}{ll}
 w \cdot x_i + b \geq 1 & \text{for } y_i = +1 \\
 w \cdot x_i + b \leq -1 & \text{for } y_i = -1
 \end{array}$$

*combining above two equation, it can be written as*

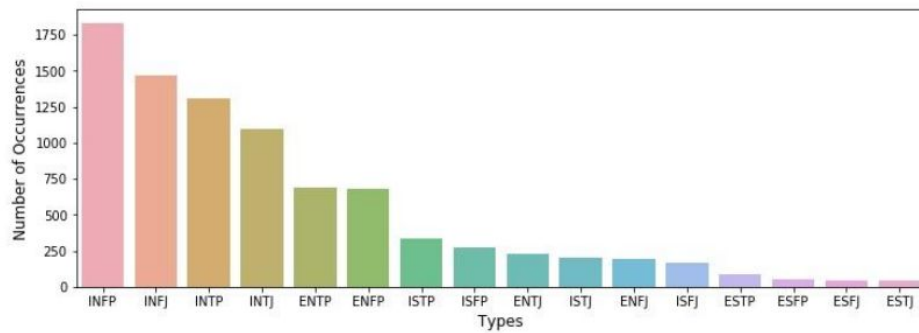
$$y_i(w \cdot x_i + b) - 1 \geq 0 \quad \text{for } y_i = +1, -1$$

(Fig 7 : Formula used by Linear SVC )

## 4. IMPLEMENTATION

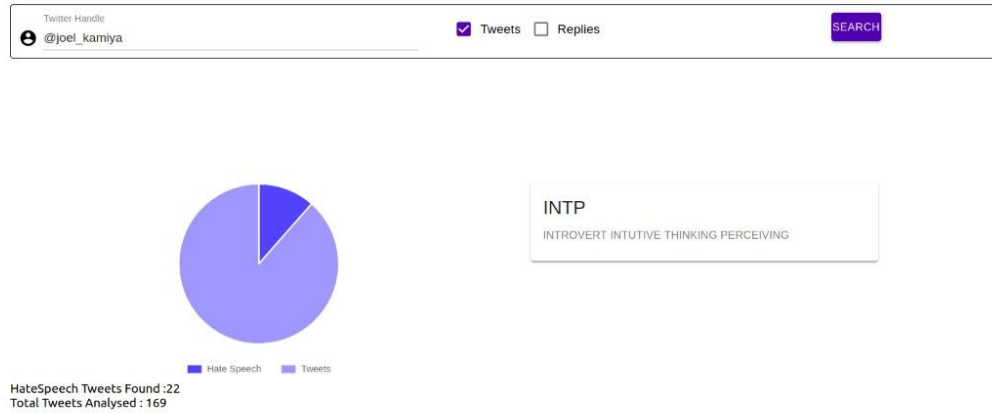


(Fig 8 : Correlation Matrix of the four Personality types IE , NS , TF , JP )




(Fig 9 : Count of all the personality types in the given dataset)

# Pantomath



(Fig 10 : Web application Product )

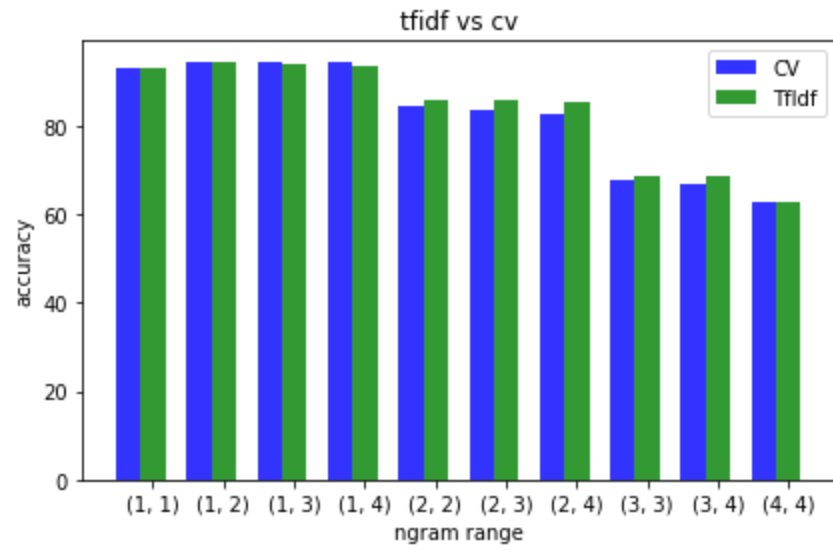


HateSpeech Tweets Found :22  
Total Tweets Analysed : 169

INTP  
INTROVERT INTUTIVE THINKING PERCEIVING

Sno	Potential Hate Speech Tweet
0	RT @michael_lee111: I thought this was a wedding. This nigga seen this bitch all school year & fake crying. I hate prom
1	I was added to this random gc last night and these ill racist kids started talking crazy black twitter do your ting... <a href="https://t.co/57T3TAYrKR">https://t.co/57T3TAYrKR</a>
2	RT @Pacino13_: I keep forgetting this nigga wasn't from the 1700s
3	RT @onlyfanobtainer: I'm gonna make a thread of white people songs that are bangers

(Fig 11 : Table of potential hate speech tweets shown on the web application )



(Fig 12 : TfIdf vs CountVectorization for hate speech analysis using Logistic Regression)

## 5. TESTING

**5.1 Unit Testing :** The dataset was made fit to be trained by the model by cleaning the dataset and dropping the unnecessary columns. Further, it was tested on some students and some political leaders to check whether the personality and hate speech detection is accurate or not.

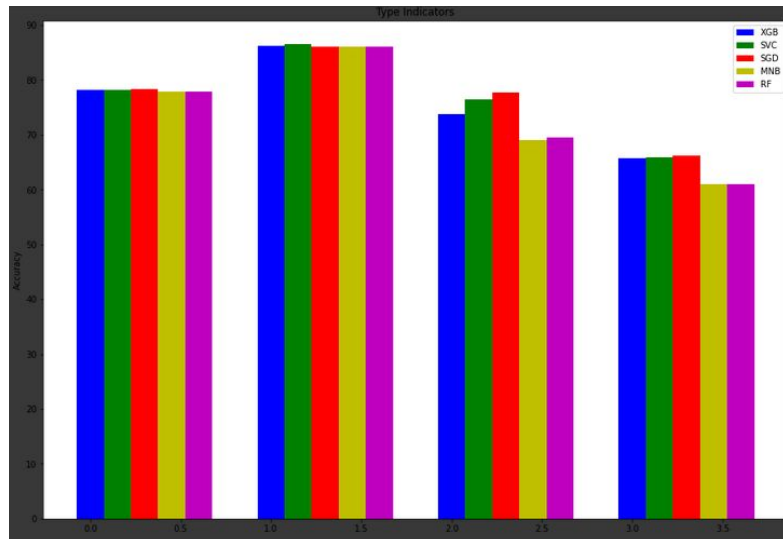
### 5.2 Testing and Results

The project stood upon our expectations while successfully completing the following tasks :-

- a) Using the Twitter API to fetch tweets and retweets of the given user in time.
- b) Accurate predictions with good accuracy and F1 scores.
- c) Making a web application to carry forward the tasks.

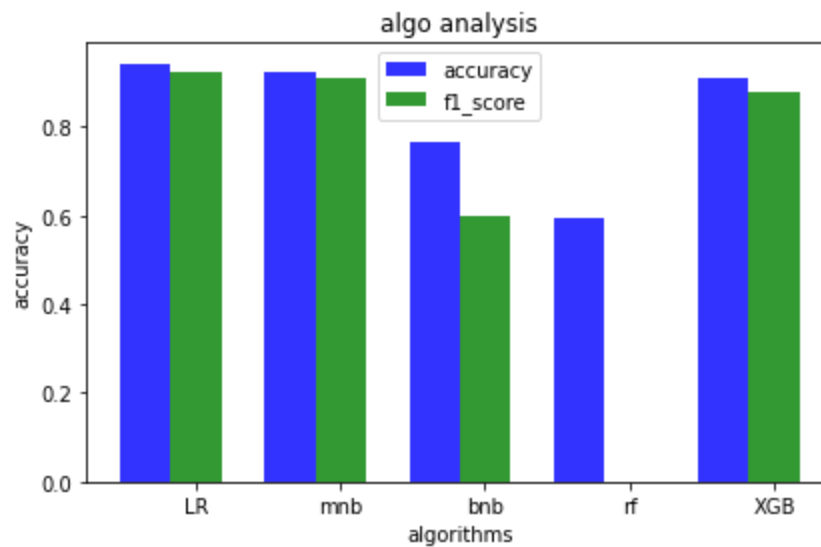
Accuracy for personality detection is compared with five algorithms as shown in Fig 13. On the X-axis is the classification for each of the 4 types of MBTI personality while on the Y-axis is the accuracy score. The coloured bars represent :

Blue	-	XGBoost Classifier (XGB)
Green	-	Linear Support Vector Classifier (SVC)
Red	-	Stochastic Gradient Descent Classifier(SGD)
Yellow	-	Multinomial Naive Bayes Classifier (MNB)
Magenta	-	Random Forest Classifier (RF)



(Fig 13 : Comparing the Accuracies for various Algorithms)

On comparing the results, we find that XGBoost, Linear SVC and SGD classifiers have the best accuracy but Linear SVC has the best F1 score. So the Linear SVC model was chosen for the analysis.



(Fig 14: algorithm analysis for hate speech detection)

- LR - Logistic Regression
- Mnb - Multinomial Naive Bayes
- Bnb - Bernoulli Naive Bayes
- Rf - Random forest
- XGB - XGBoost classifier

## 6. FINDINGS AND CONCLUSION

**6.1 Findings :** The project helped us to delve deep into text classification, using and learning about techniques like word embedding (word2vec) and know details about classification algorithms used and hyper-parameter tuning.

**6.2 Conclusion :** This project has helped us develop a new machine learning method for automating the process of hate speech detection and personality type prediction based on the MBTI personality type indicator. The natural language processing toolkit (NLTK) and Linear SVC and Logistic Regression, has been helpful in carrying out the results. Moreover, Pandas, Numpy, re, Seaborn, Matplotlib and Sklearn were other Python libraries. The accuracy of each of the models was evaluated and the performance was compared to the latest and most successful existing methods which used the same dataset. The results show that the methodology presented in this project has better accuracy and reliability in comparison to other methods. This can effectively assist NLP practitioners and psychologists in regards to the identification of personality types and associated cognitive processes and psychology.

**6.3 Future Scope :** Though the project has stood up to our expectations yet there is always room for improvement. We hope to add another feature of classifying the tweets, whether they belong to Sports, Politics, Technology, Entertainment or Business category.

**6.4 Limitations :** Since we have created a web application, therefore it is necessary for it to be responsive and fast, but the Twitter API we are using is slow to fetch the tweets and retweets of a user. Hence we only fetch 200 tweets for a user and do the analysis on them.

## REFERENCES

- [1] Anzhela Zhusupova, “Characterizing the Personality of Twitter Users based on their Timeline Information” Thesis, ISCTE-University Institute of Lisbon, October 2016.
- [2] Charvet, S.R. Words that Change Minds: Mastering the Language of Influence, 2nd Revised ed.; Kendall/Hunt Publishing Co.: Dubuque, IA, USA, 1997.
- [3] Tieger, P.D.; Barron-Tieger, B. Do What You Are: Discover the Perfect Career for You through the Secrets of Personality Type, 4th ed.; Sphere: London, UK, 2007.
- [4] Myers, I.B.; McCaulley, M. Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator, 15th ed.; Consulting Psychologists Press: Santa Clara, CA, USA, 1989.
- [5] Eleanna Kafeza , Andreas Kanavos , Pantelis Vikatos, Christos Makris, ”T-PICE: Twitter Personality based Influential Communities Extraction System”, IEEE International Congress on Big Data, 2014.
- [6] Dilini Sewwandi, Kusal Perera, Sajith Sandaruwan, Anupiya Nugaliyadde, Samantha Thelijagoda, “Linguistic Features Based Personality Recognition Using Social Media Data”, 6th National Conference on Technology and Management (NCTM), Malabe, Sri Lanka, 2017.
- [7] Jennifer Golbeck, Cristina Robles, Michon Edmondson, Karen Turner, “Predicting Personality from Twitter”, IEEE International Conference on Privacy, Security, Risk, and Trust and IEEE International Conference on Social Computing, 2011.
- [8] Hajime Watanabe, Mondher Bouazizi , And Tomoaki Ohtsuki, “Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection”, IEEE, 2018.