Twitter Spam Detection

Index

Welcome

Team Member :

Ajay Walke
112015006
Group Leader

Rohan Khavale
112016031
Researcher

Manas Agarwal
112015005
Researcher

Sahil Thakare
112015153
Researcher

SUPERVISOR - PROF. RITU TIWARI

1. INTRODUCTION

In recent times, the use of microblogging platforms has seen huge growth, one of them being Twitter. As a result of this growth, businesses and media outlets are increasingly looking for methods to use Twitter to gather information on how people perceive their products and services. Although there has been researched on how sentiments are communicated in genres such as news articles and online reviews, there has been far less research on how sentiments are expressed in microblogging and informal language due to message length limits. In recent years, many businesses have used Twitter data and have obtained upside potential for businesses venturing into various fields. On the other hand, scammers and spambots have been actively spamming Twitter with malicious links and false information, causing real users to be misled. Our goal is to gather an arbitrary amount of data from a prominent social media site, namely, Twitter, and perform spam detection.

2. MOTIVATION

During lockdown we all were continuing our life in online mode weather it be work or talking to our dear ones we all we dependent on Internet but also we all were coming across many spam messages which were misleading and spreading rumors about the ongoing pandemic which made it more difficult to handle the situation. So we as a team decide to target this issue and control the spammers. To do this we did a lot of intensive literature survey and selected twitter as the main social networking platform to test our spam detection model.

3.Problem Statement

Detection of Social Network Spam using Machine Learning

4. Literature Survey

No	Research Paper	Authors	Date Of Publication
1	Twitter Spam detection Based on Machine Learning	Tingmin Wu, Shigang Liu	January 2017
2	Twitter Spam Detection	Ashwini Bhangare, smita Ghodke, Kamini Walunj, Utkarsha Yewale.	March 2018
3	Detection Of Social Network Spam Based on Improved Extreme Learning Machine	ZHIJIE ZHANG ,RUI HOU, JIN YANG	June 2020
4	Real-time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Technique	Anisha P Rodrigues,Roshan Fernandes,Aakash A,Abhishek B,Adarsh Shetty,Atul K,Kuruva Lakshmanna and R. Mahammad Shafi.	January 2022

5. Research Gap

In previous work, more variables needed to be added in the framework to enhance the accuracy of the model and classification rate. Need to improve text similarity for extracted new strange words from tweets. In previous research [3], data mining algorithms were applied on small amounts of collected dataset and limited tweets. So, large amounts of data sets need to be tested for the accuracy of previous algorithms. In Future, we can collect the dataset of tweets in different languages. We can apply data mining algorithms on other social media platforms like Facebook, Instagram, LinkedIn, YouTube, and WhatsApp. More classifiers can be added that can make Twitter spam detection more valuable for users. Research will help to solve model scalability without performing comparative accuracy. Can use the characteristics of spammers at different levels of granularity have been used by some interesting patterns released by spammers.

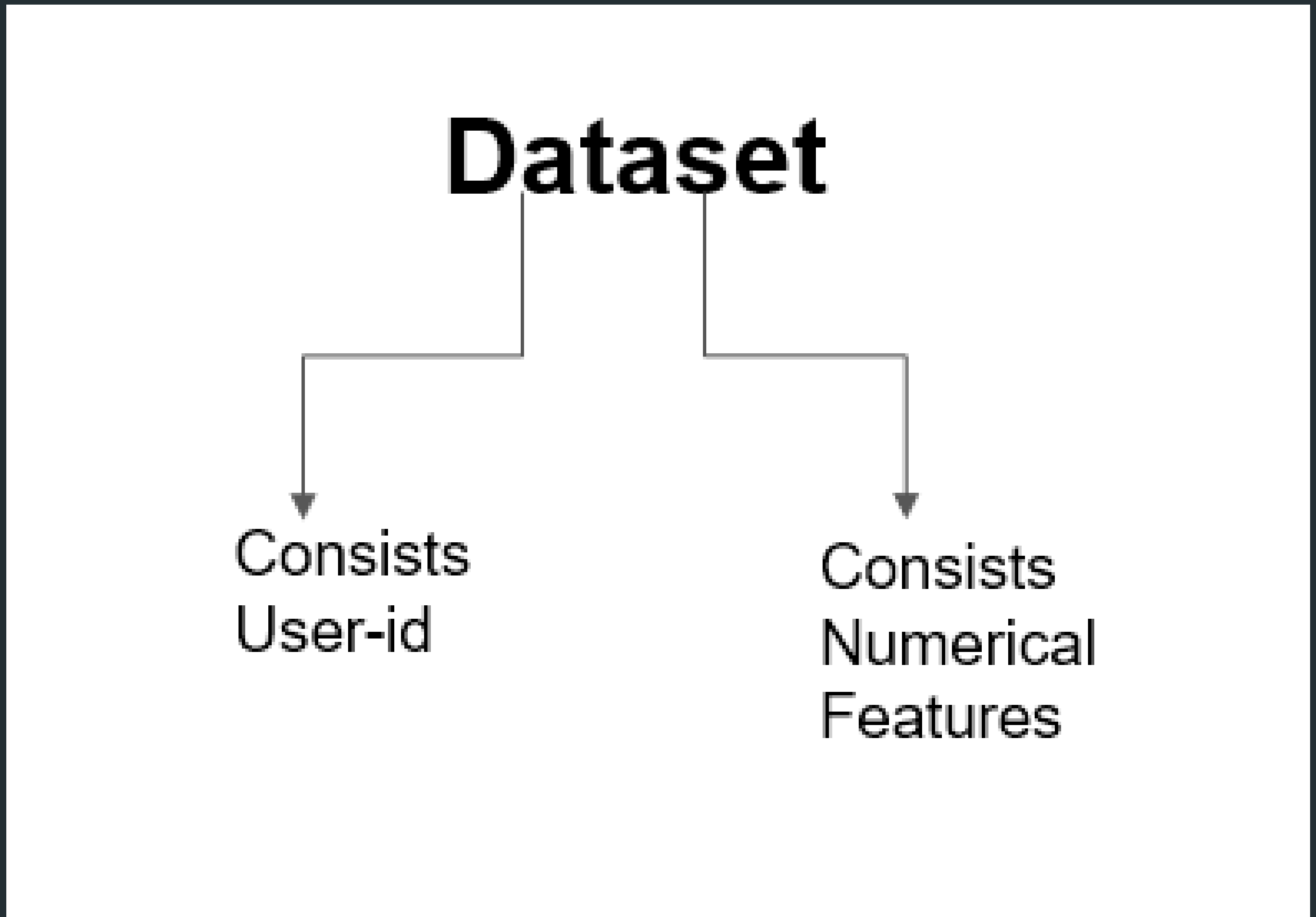
The performances for all four metrics on for datasets are better than other all the time. As shown in Figure 6[4], the F-measure is much higher than others, with averagely 30% higher than Random Forest and almost nine times of Naive Bayes in Dataset 2 and 4. Even the Decision Tree method achieves almost the same as our method at Dataset 1, it only remains half when testing on Dataset 4[4].

6. Objectives

- To judge authenticity of a user.
- To prevent unwanted, malicious, unsolicited content or behaviour, manifested in various ways including microblogs, messages, malicious links, fraudulent reviews, etc.
- To analyse the effectiveness of the research, the work has been computed with existing work and it has been concluded that the value of precision, recall and F-measure of the research work has been increased.

7. Dataset

DataSet



DataSet Preprocessing

1 active_tweeting_frequency_per_day
2 adjusted_nb_of_uses_of_hashtag
3 adjusted_nb_of_uses_of_mention
4 adjusted_nb_of_uses_of_sources
5 adjusted_nb_of_uses_of_url
6 age
7 avg_intertweet_times
8 avg_intertweet_times_seconds
9 content_duration_days
10 date_newest_tweet
11 date_oldest_tweet
12 default_profile
13 default_profile_image
14 diversity_index_of_hashtags
15 diversity_index_of_mentions
16 diversity_index_of_sources
17 diversity_index_of_urls
18 favourites_count
19 followees_per_followers_sq
20 followers_count
21 followers_count_minus_2002
22 followers_per_followees
23 friends_count
24 friends_count_minus_2002
25 hashtags_used_on_average

NUMERICAL FEATURES IN DATASET

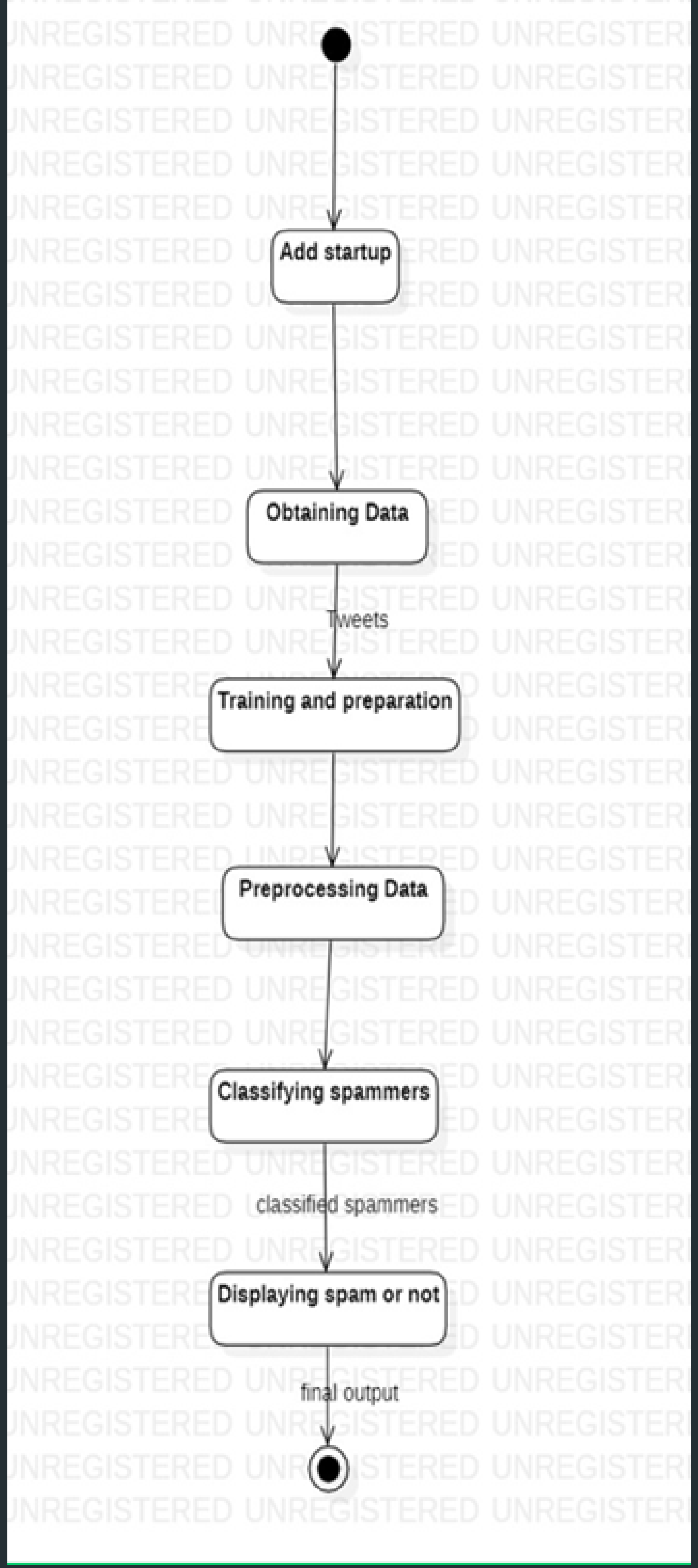
26 lang
27 len_description
28 len_screen_name
29 max_intertweet_times
30 max_intertweet_times_seconds
31 max_nb_characters_per_tweet
32 max_nb_favourites_per_tweet
33 max_nb_hashtags_per_tweet
34 max_nb_hashtags_per_word_in_the_tweet
35 max_nb_mentions_per_tweet
36 max_nb_mentions_per_word_in_the_tweet
37 max_nb_retweets_per_tweet
38 max_nb_symbols_per_tweet
39 max_nb_symbols_per_word_in_the_tweet
40 max_nb_urls_per_tweet
41 max_nb_urls_per_word_in_the_tweet
42 max_nb_words_per_tweet
43 mean_nb_characters_per_tweet
44 mean_nb_favourites_per_tweet
45 mean_nb_hashtags_per_tweet
46 mean_nb_hashtags_per_word_in_the_tweet
47 mean_nb_mentions_per_tweet
48 mean_nb_mentions_per_word_in_the_tweet
49 mean_nb_retweets_per_tweet
50 mean_nb_symbols_per_tweet

8. Methodology

ACTIVITY DIAGRAM :

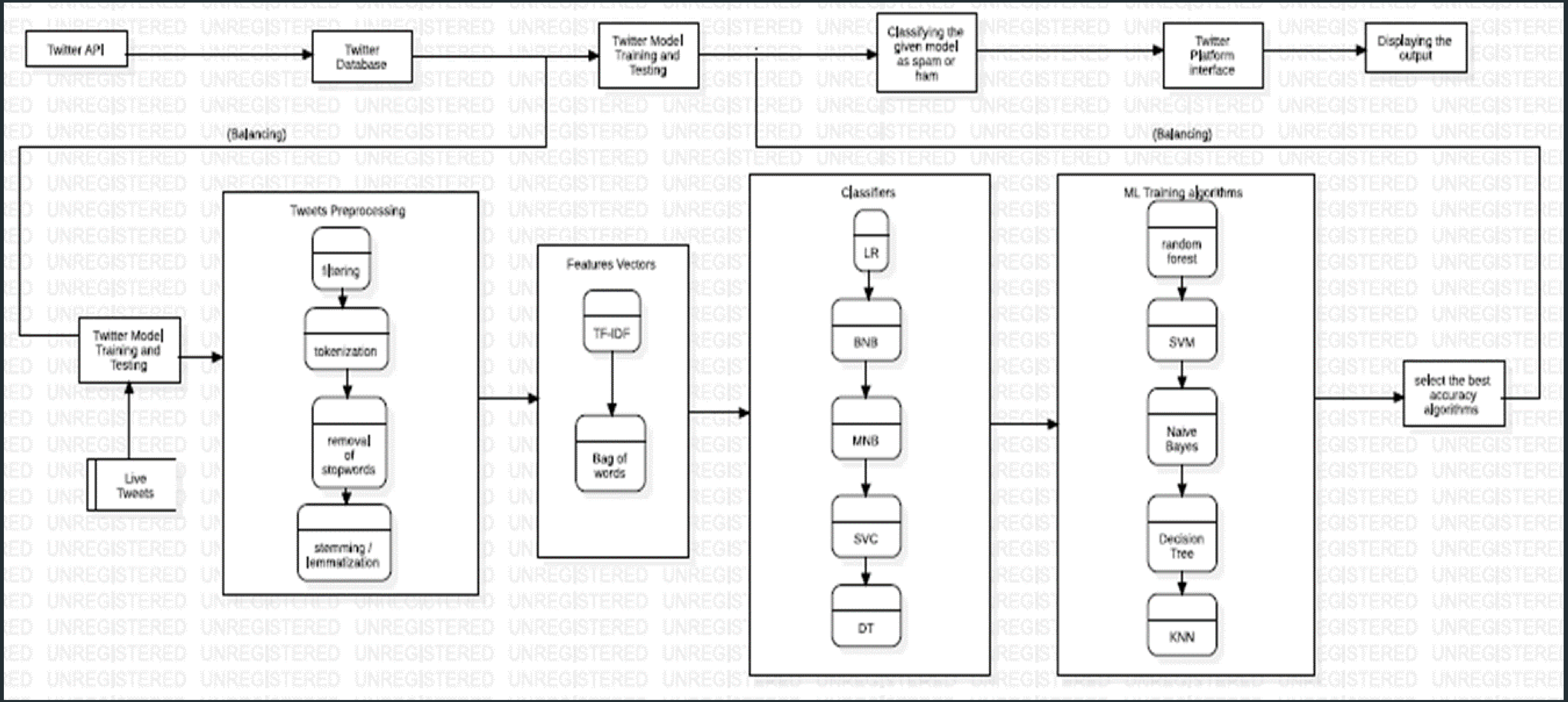
THE FLOW OF THE MODEL IS TOTALLY BASED ON THE DATASET.

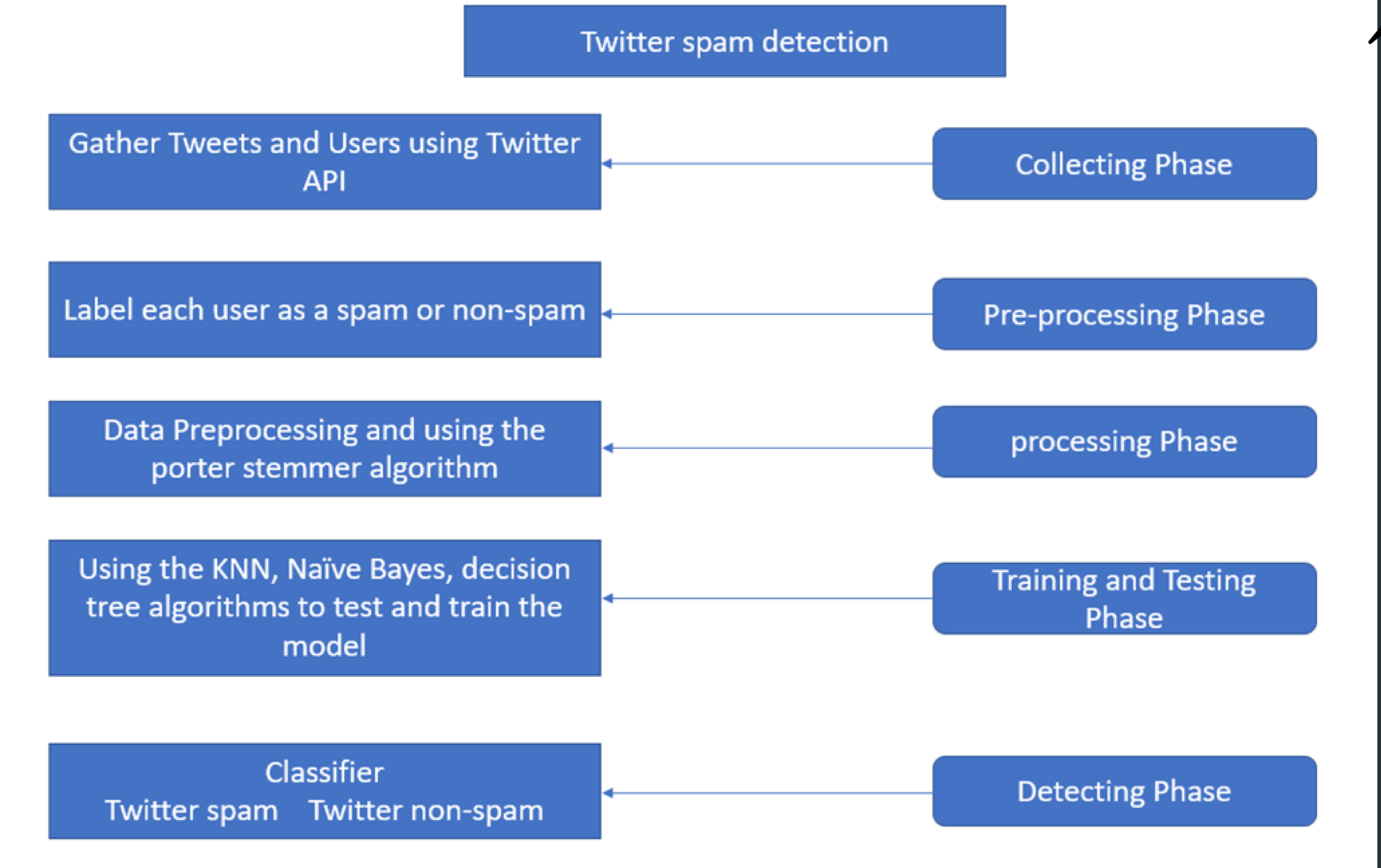
THE SUBSEQUENT PHASES ARE USED FOR MANIPULATION OF THE MODEL.



Data Flow Diagram :

DATA FLOW DIAGRAM REPRESENT THE ALL INCOMING AND OUTGOING FIELDS IN MODEL.




Twitter spam detection
Gather Tweets and Users using Twitter API
Collecting Phase
Label each user as a spam or non-spam
Pre-processing Phase
Data Preprocessing and using the porter stemmer algorithm
processing Phase
Using the KNN, Naïve Bayes, decision tree algorithms to test and train the model
Training and Testing Phase
Classifier
Twitter spam Twitter non-spam
Detecting Phase

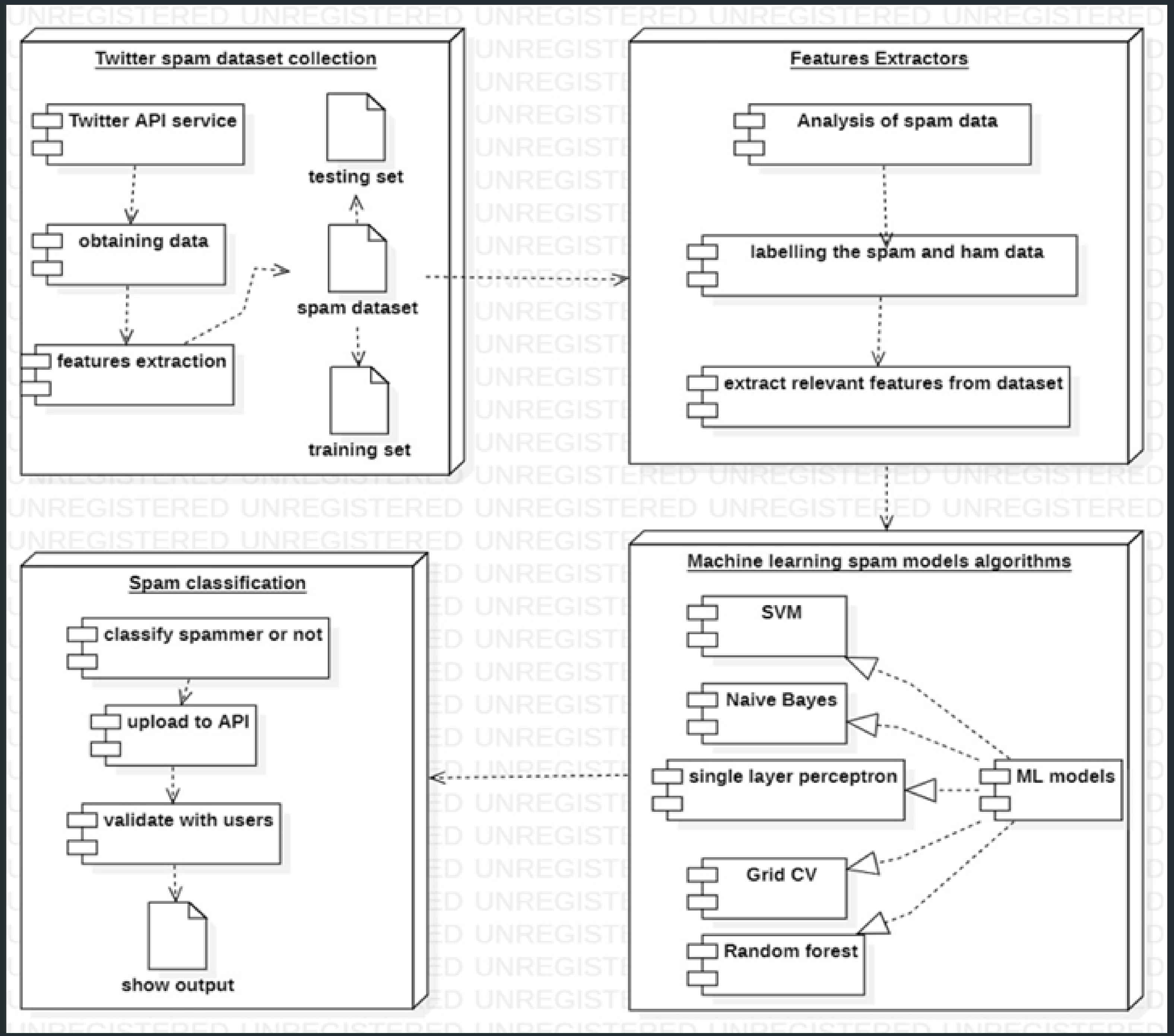
Machine Learning Algorithms

- NAÏVE BAYES
- KNN
- SVM
- RANDOM FOREST
- DECISION TREE
- MULTILAYER PERCEPTRON
- GRID SEARCH CV

9. Analysis And Design

THE MODEL IS CLASSIFIED IN THE FOLLOWING STAGE:

- DATASET COLLECTION
- DATA PRE-PROCESSING
- STANDARDIZATION
- MACHINE LEARNING ALGORITHMS
- REPRESENTING THE OUTPUT



10. Results

NAÏVE BAYES CLASSIFICATION REPORT

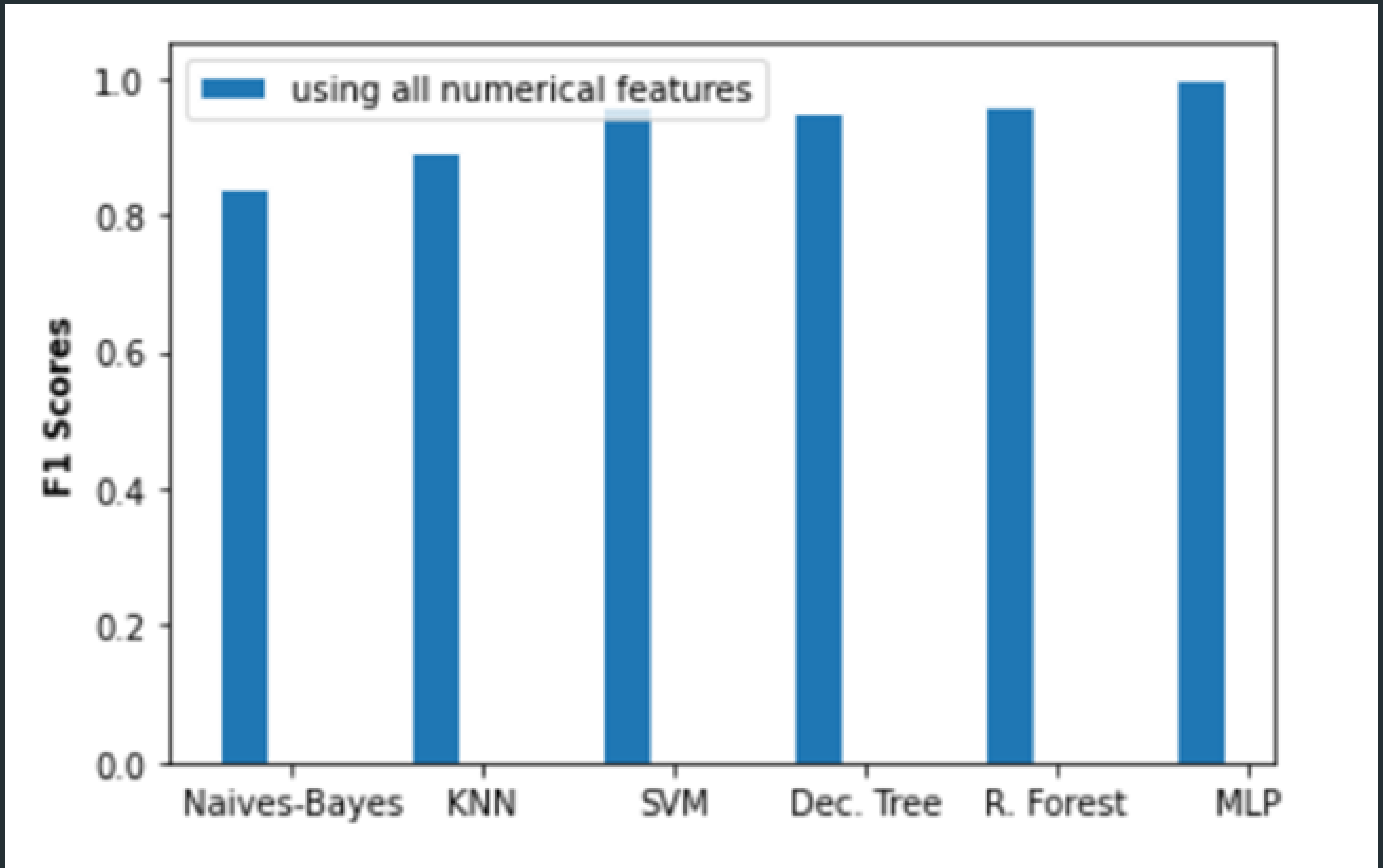
Model train accuracy score: 0.9228

	precision	recall	f1-score	support
0	0.99	0.88	0.93	85
1	0.63	0.94	0.76	18
accuracy			0.89	103
macro avg	0.81	0.91	0.84	103
weighted avg	0.92	0.89	0.90	103

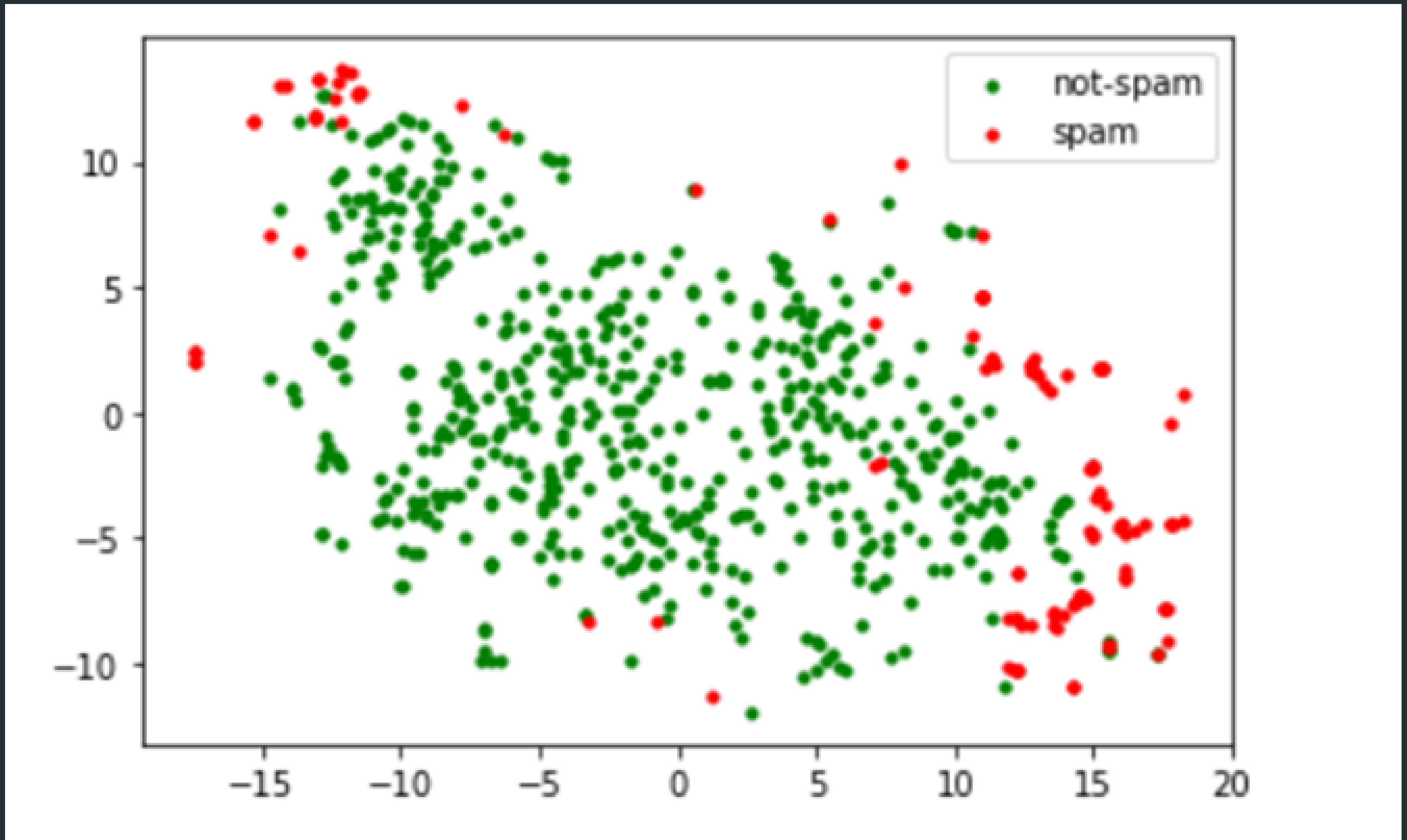
KNN CLASSIFICATION REPORT

	precision	recall	f1-score	support
0	0.94	0.99	0.97	85
1	0.93	0.72	0.81	18
accuracy			0.94	103
macro avg	0.94	0.86	0.89	103
weighted avg	0.94	0.94	0.94	103

Comparative Analysis of Various applied algorithms



Final output of the model :



10. Conclusion And Future Work

THIS STUDY PRESENTS A NOVEL TWITTER SPAM DETECTION METHOD AND NEW INSIGHTS INTO THE SOPHISTICATEDLY EVOLVING TECHNIQUES FOR SPAMMING ON TWITTER. IN WHICH THE FEATURES SET CONSISTS OF USER ATTRIBUTES, CONTENT, ACTIVITY, AND RELATIONSHIPS IN ONLINE SOCIAL NETWORKS FOR IDENTIFYING THE REAL SPAM. IT WAS ALSO SHOWN THAT AUTOMATED SPAM ACCOUNTS FOLLOW A WELL-DEFINED PATTERN WITH SURGES OF INTERMITTENT ACTIVITIES. SPAMMER DETECTION HAS A STRONG COMMERCIAL INTEREST BECAUSE COMPANIES OR INDIVIDUALS WANTS TO IMPROVE THE SECURITY ON SOCIAL MEDIA. DURING THE ANALYSIS OF THE DATA, WE OBSERVED THAT SPAM USERS TEND TO BE SELECTIVE IN FOLLOWING OTHER USERS THEREBY FORMING ENCLAVES OF SPAMMERS. THIS IS A HIGH-LEVEL OBSERVATION THAT WE AIM TO EXPLORE FURTHER IN THE FUTURE. ADDITIONALLY, BOTH THE TWO BROAD USER GROUPS, I.E. HUMAN USERS AND SOCIAL BOT (AUTONOMOUS ENTITY) USERS CONTAIN SPAMMERS, WHOSE SPAMMING BEHAVIOUR TENDS TO BE SIMILAR. THE DISTINCTION BETWEEN LEGITIMATE HUMAN USERS VS. LEGITIMATE SOCIAL BOTS AS WELL AS HUMAN SPAMMERS VS. SOCIAL BOT SPAMMERS NEEDS TO BE INVESTIGATED FURTHER. ANOTHER INTERESTING DIMENSION FOR FUTURE WORK IS TO STUDY THE EFFECT OF THE RECENT INCREASE IN THE MAXIMUM LENGTH OF TWEETS ON SPAMMING ACTIVITY. INTUITIVELY, AUTOMATED SPAM ACCOUNTS WILL FACE DIFFICULTIES IN GENERATING LENGTHIER TWEETS INTELLIGENTLY, THEREBY MAKING THESE TWEETS EASIER TO IDENTIFY.

11. References

1. TINGMIN WU, SHIGANG LIU, “TWITTER SPAM DETECTION BASED ON DEEP LEARNING,” IN IEEE ACCESS, 2017
2. ANISHA P RODRIGUES, ROSHAN FERNANDES, AAKASH A, ABHISHEK B, ADARSH SHETTY, ATUL K, KURUVA LAKSHMANNA, “REAL-TIME TWITTER SPAM DETECTION AND SENTIMENT ANALYSIS USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES,” IN IRJET ACCESS, 2017
3. ASHWINI BHANGARE, SMITA GHODKE, KAMINI WALUNJ, UTKARSHA YEWALE, “TWITTER SPAMMER DETECTION,” IRJET ACCESS, 2018
4. Z. ZHANG, R. HOU AND J. YANG, "DETECTION OF SOCIAL NETWORK SPAM BASED ON IMPROVED EXTREME LEARNING MACHINE," IN IEEE ACCESS, VOL. 8, PP. 112003-112014, 2020, DOI: 10.1109/ACCESS.2020.3002940.

CODE

GITHUB LINK OF SOURCE-CODE :

HTTPS://GITHUB.COM/AJAYWALKE/TWITTER-SPAM-DETECTION

Thank You