Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges

In this paper author is using various machine learning algorithms such as SVM, Random Forest, Naïve Bayes, KNearest Neighbours, and Decision Tree to predict cyberbullying posts from social networks. As additional algorithm we are using ‘Extreme Machine Learning’ algorithm which is an advance algorithm in machine learning area.

Using all algorithms we will build train model with normal and bullying messages and this train model will applied on new posts from users to predict whether new post is normal or contain bullying stuff.

This project consists of following modules

User module: using this module users can create an account. Using account details they can login to application and then send and view posts.

Admin Module: Admin can view all registered user account and then accept or reject new user account. Admin responsible to add new bullying messages to machine learning train dataset. Admin has to run all or at-least one SVM algorithm to perform bullying message detection from user side. Admin can view or monitor all posts send by all users.

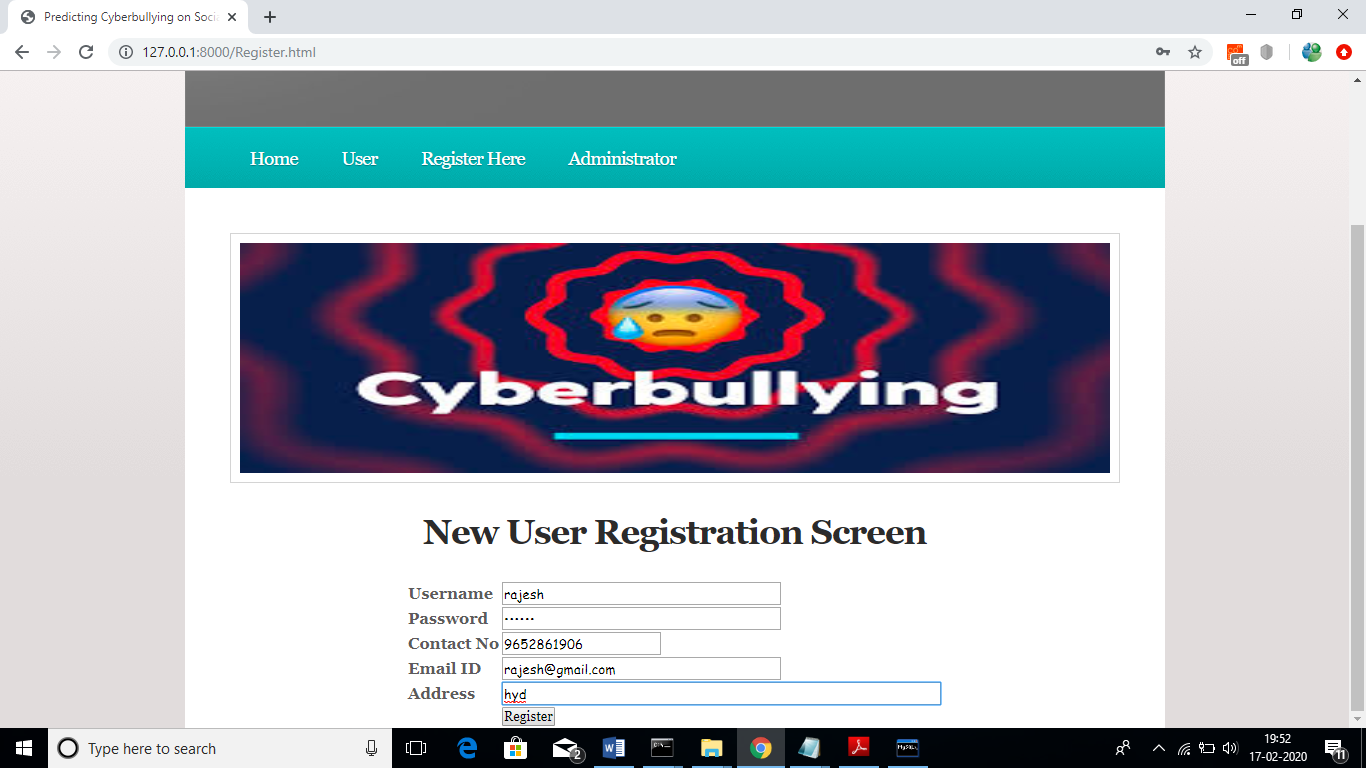
Note: machine learning algorithms will predict bullying or non-bullying message if only some related data available in train model. So you can predict all those messages correctly which are entered by admin. All sample bullying and non-bullying messages are available inside ‘Cyber/dataset.txt’ file.

First create database by copying content from ‘DB.txt’ file and paste in MYSQL console

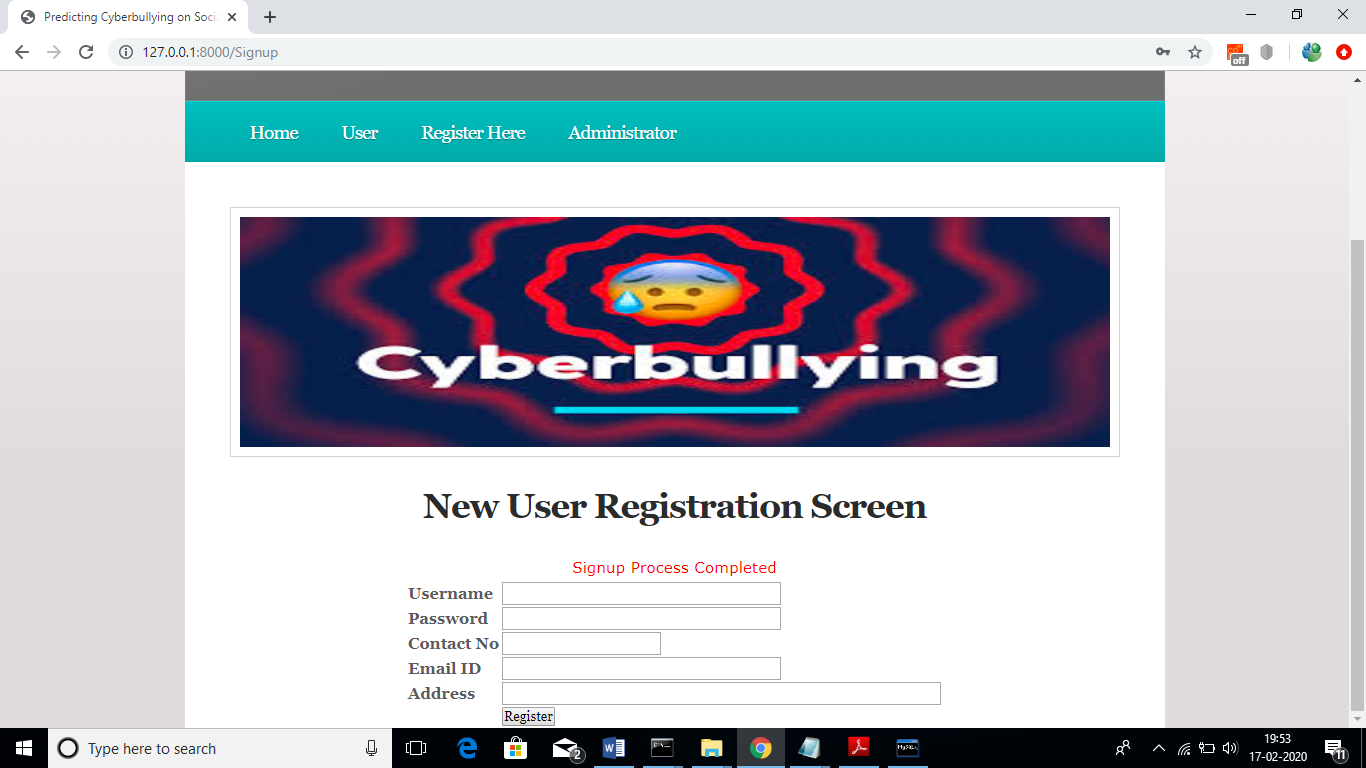
To run this project install DJANGO and then deploy ‘Cyber’ folder in DJANGO and start server and run in browser by entering URL as ‘<http://127.0.0.1:8000/index.html>’. After running above URL will get below screen



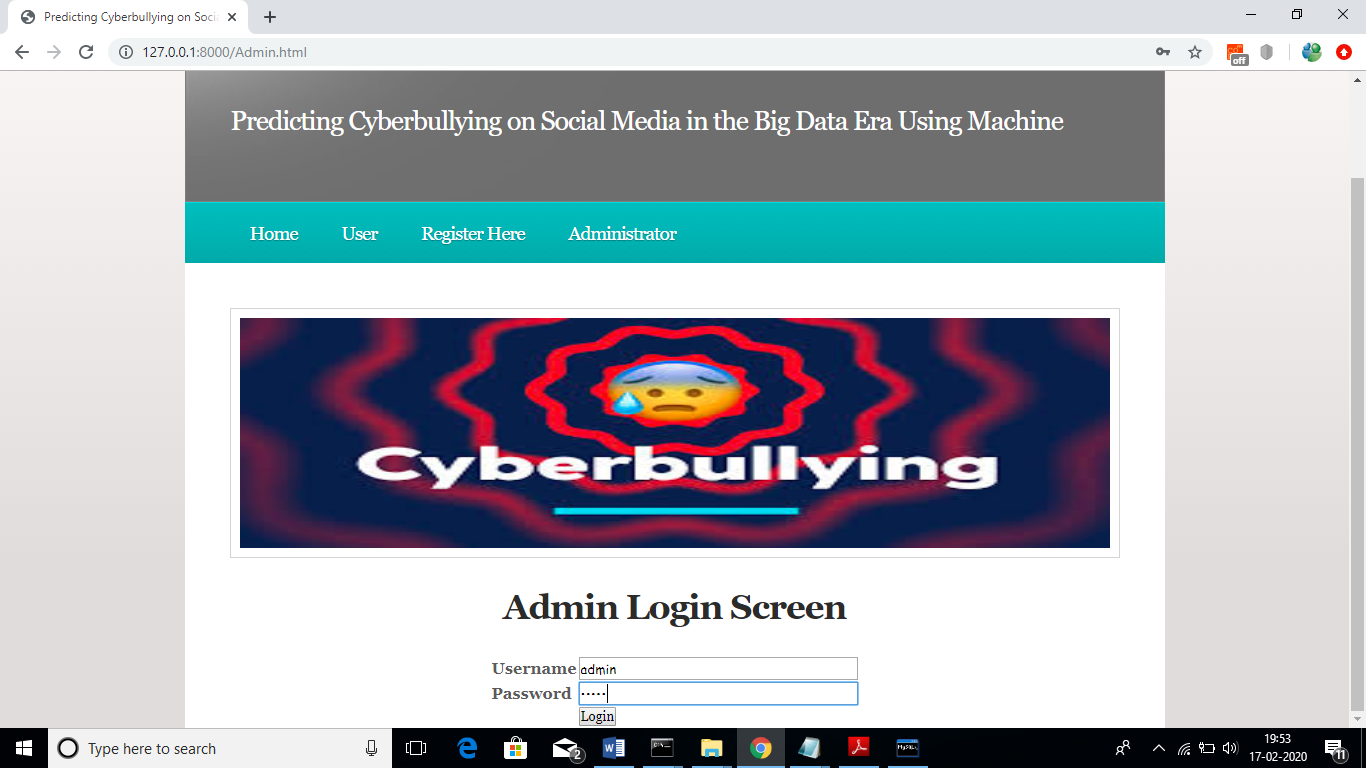
In above screen click on ‘Register Here’ link and add new user to create account



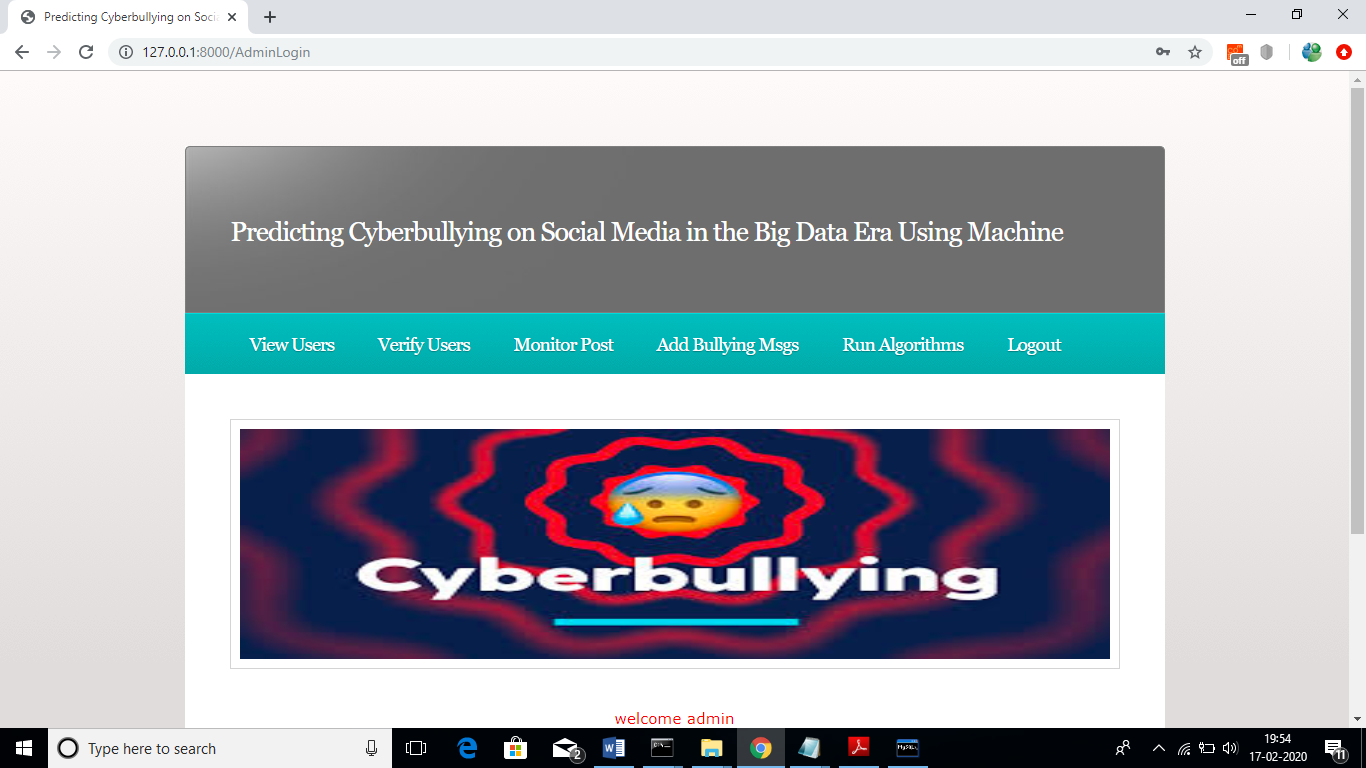
In above screen now click on ‘Register’ button to add details



In above screen sign up process completed. Now click on ‘Administrator’ link to login as admin and give permission to new user



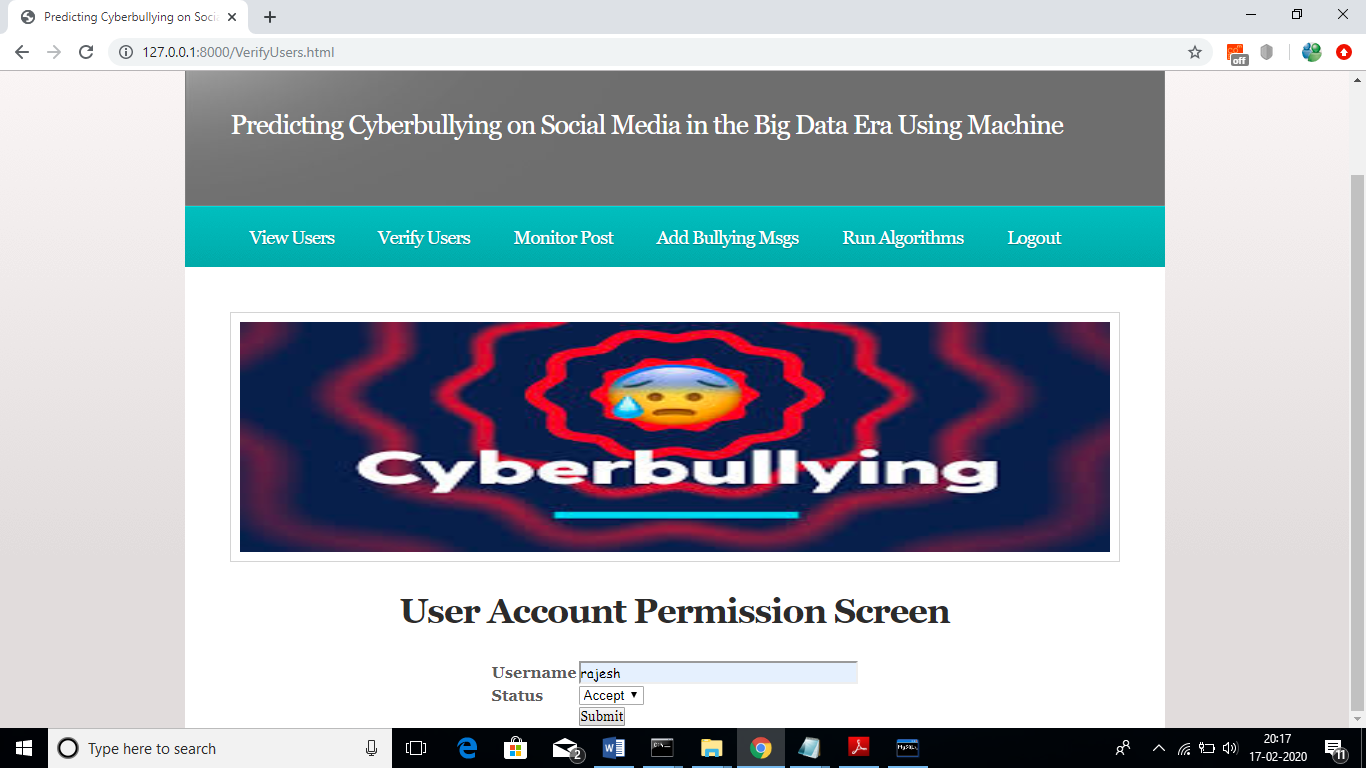
In above screen login as ‘admin’ by giving username as ‘admin’ and password as ‘admin’. After login will get below screen



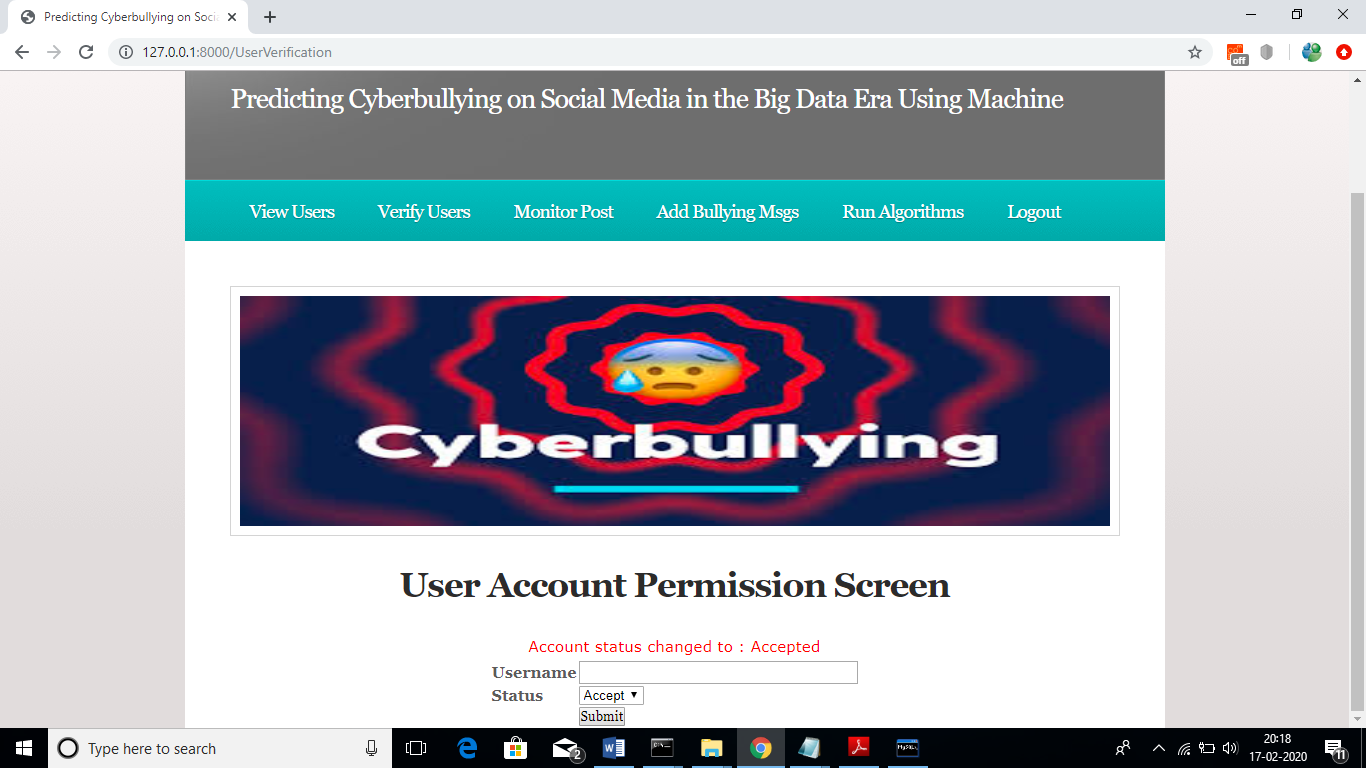
Now admin can click on ‘View Users’ link to view all users list



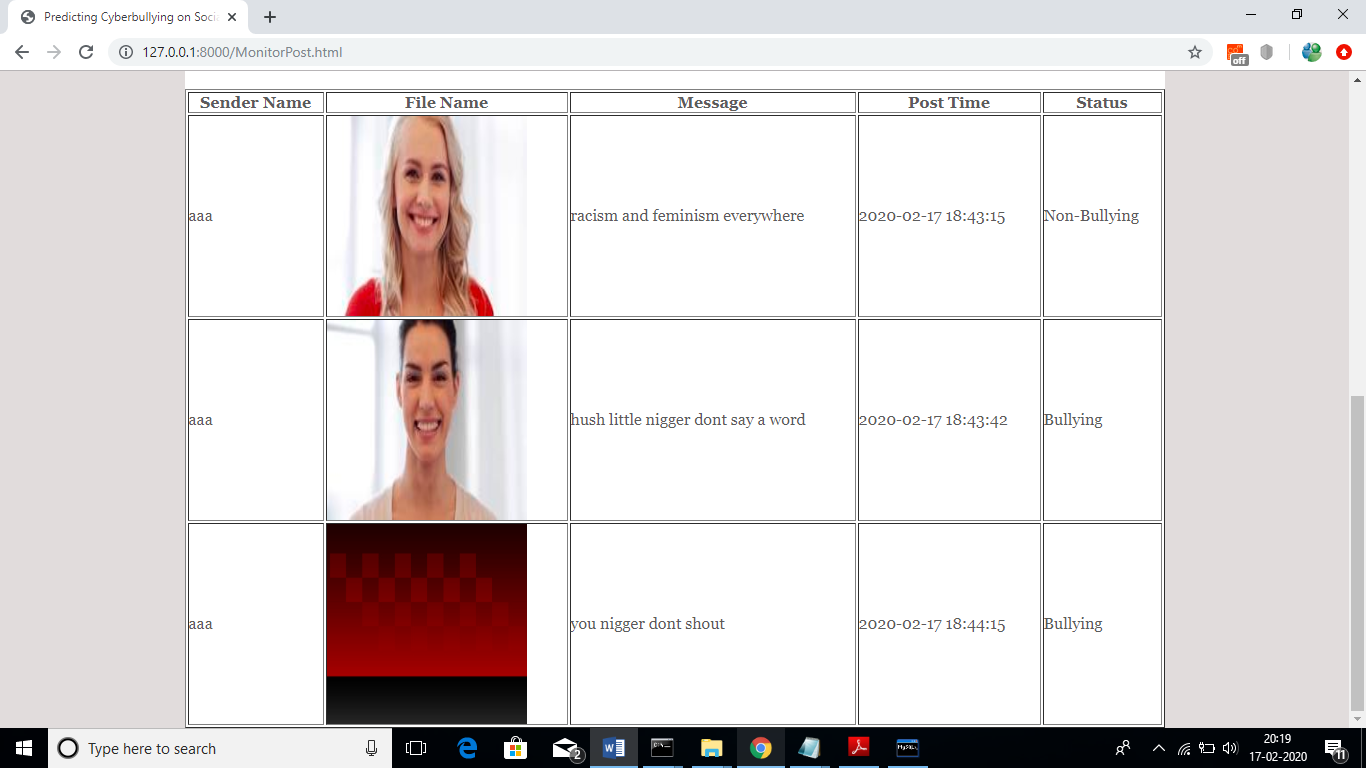
In above screen we can see ‘rajesh’ account is in pending state and to give permission to rajesh. Now admin will click on ‘Verify Users’ link to get below screen and to give permission



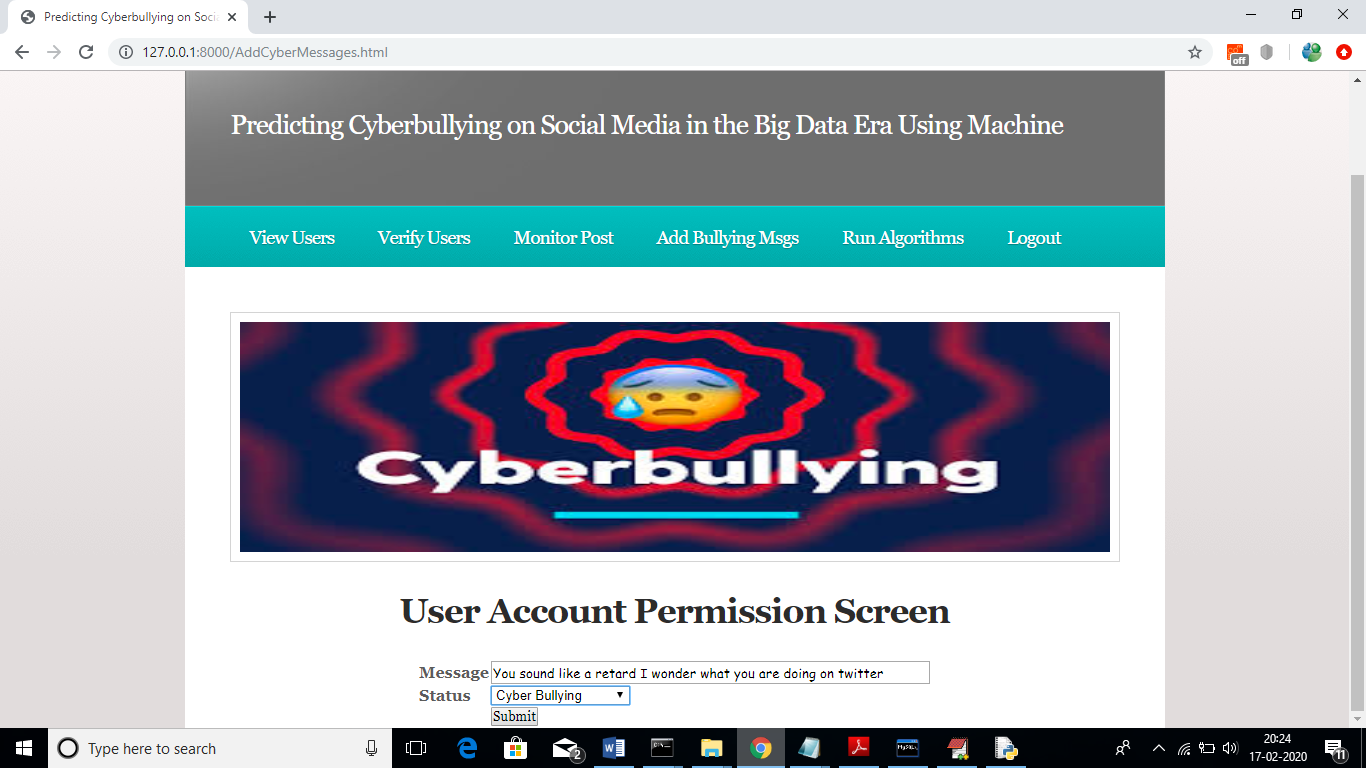
In above screen admin will enter username and then select ‘Accept’ or ‘Reject’ option to give permission.



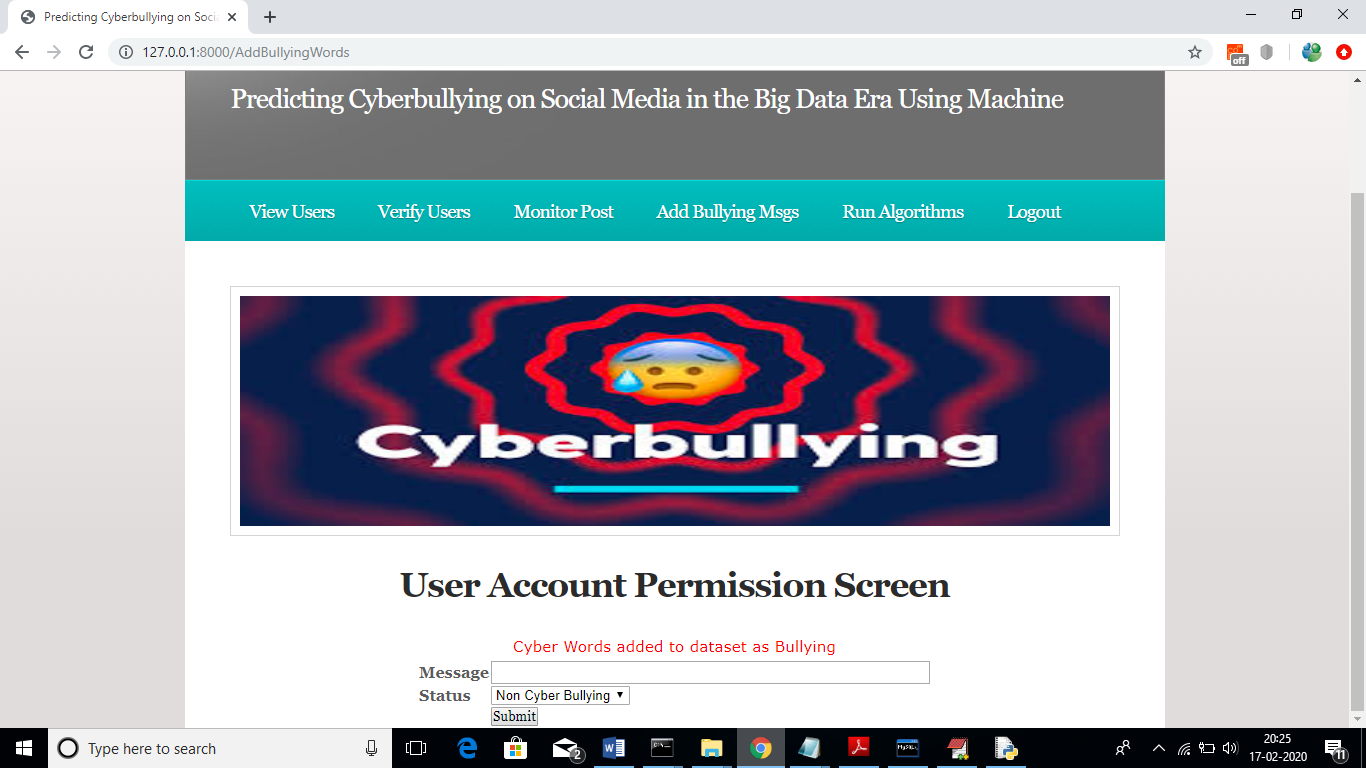
In above screen account state changed to ‘Accepted’. Now admin can click on ‘Monitor Post’ to view all post from past users



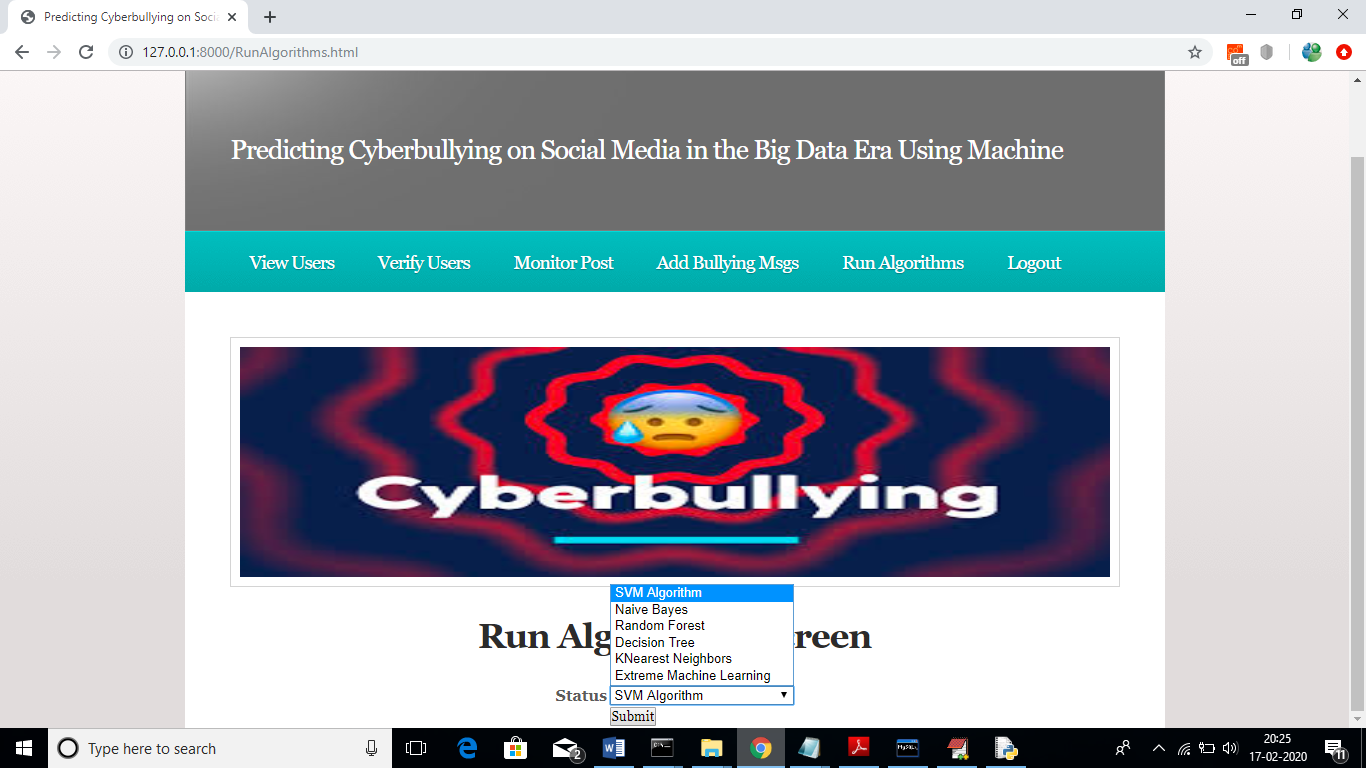
In above screen application will automatically detect whether message is non-bullying or bullying from machine learning algorithms. Now admin can click on ‘Add Bullying Msgs’ link to add words



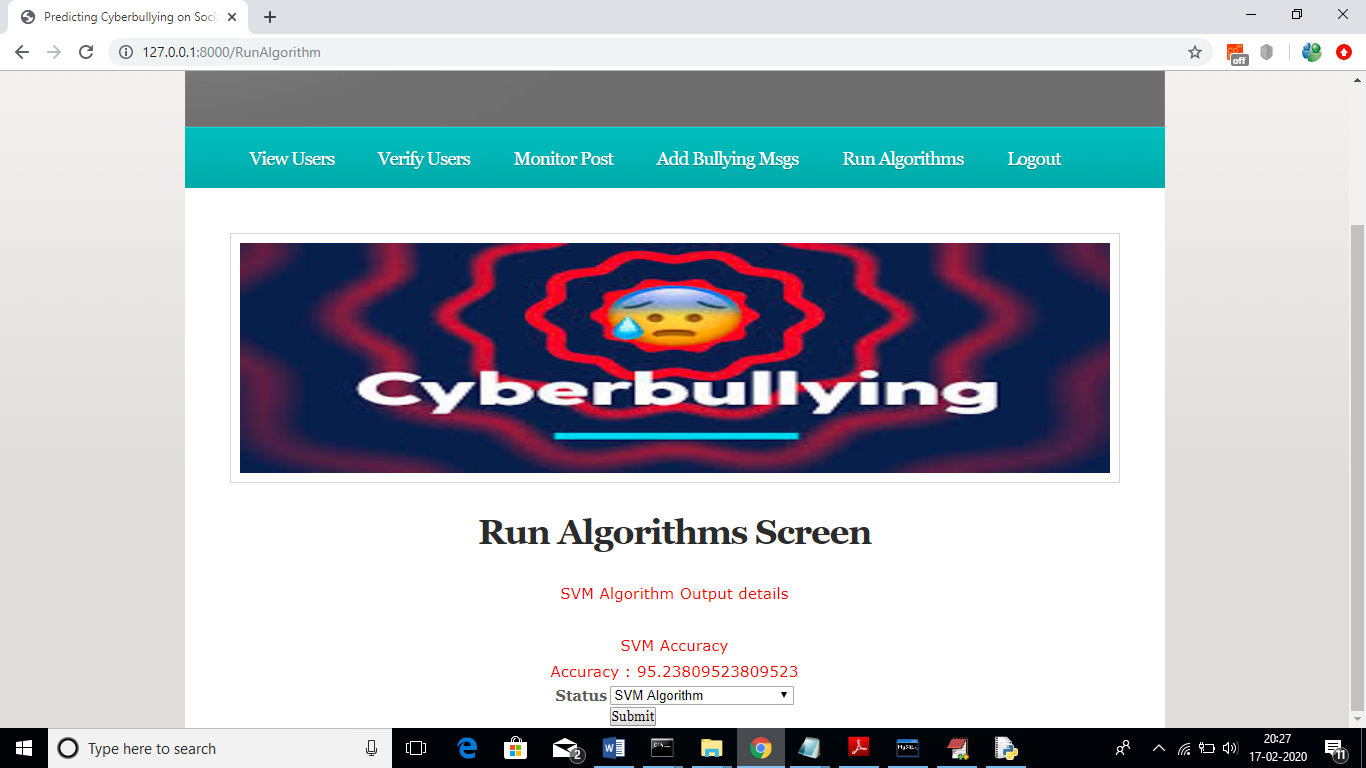
In above screen admin adding one sentence as ‘Cyber Bullying’ and similarly he can add all possible bullying and non-bullying messages. After adding messages will get below screen



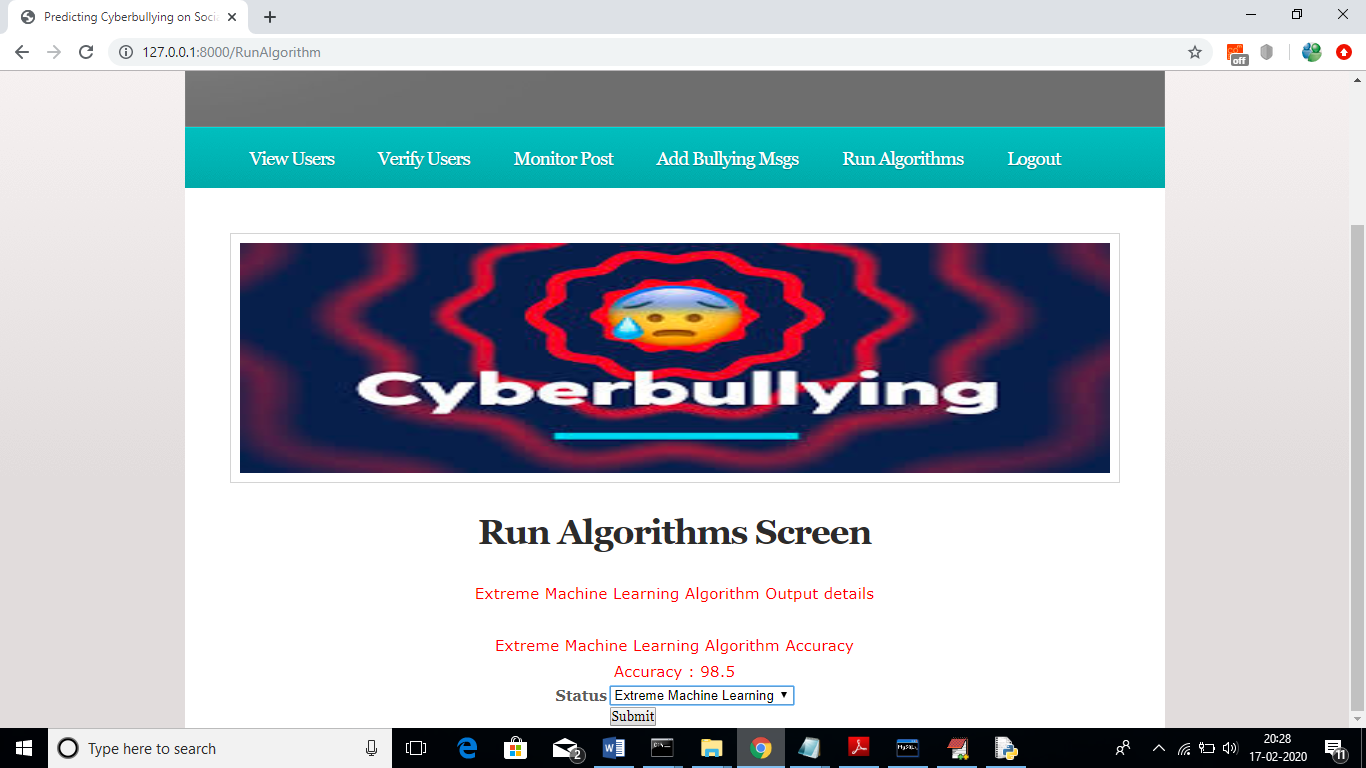
Now admin can click on ‘Run Algorithms’ link to generate train model using entire dataset to predict user posts as normal or bullying



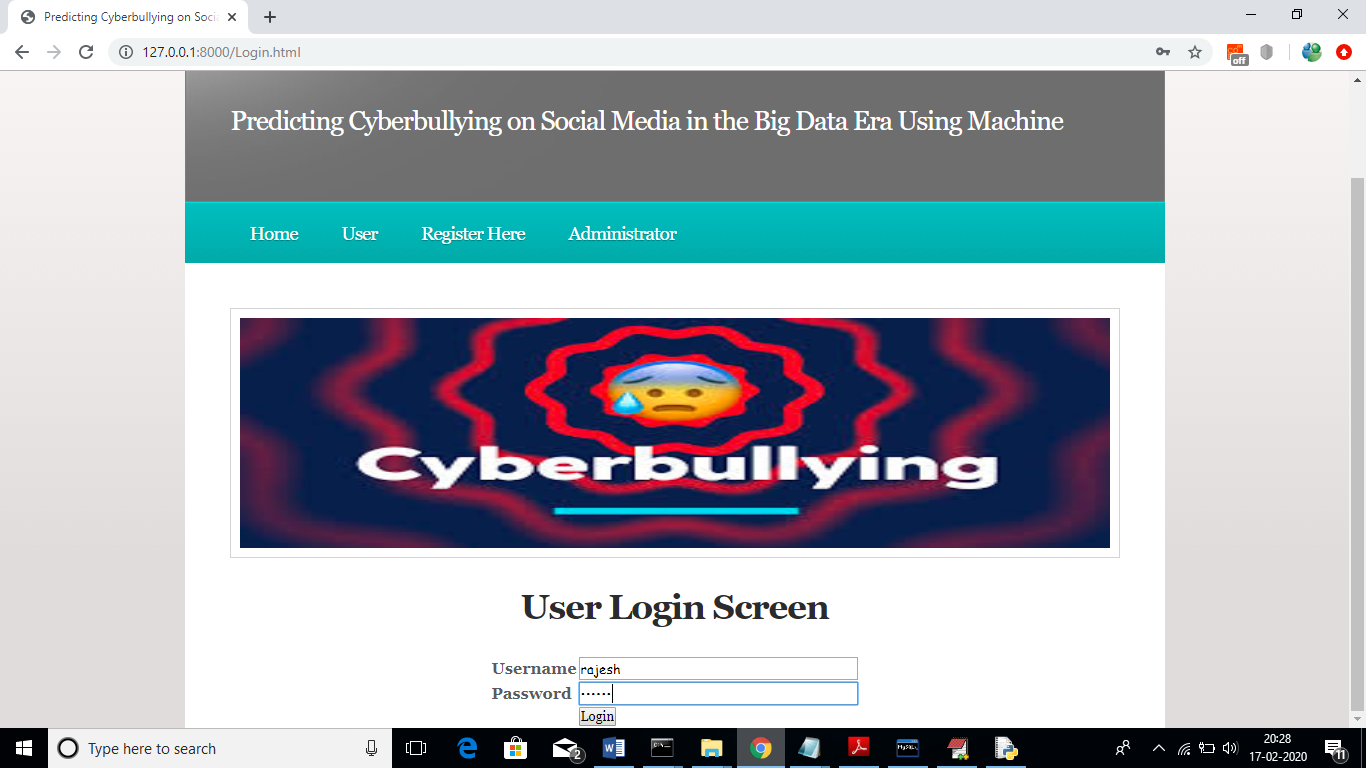
In above screen admin has to select each algorithm and click on ‘Submit’ button to train model and we will get accuracy also for each algorithm. Admin has to repeat this step whenever first time he starts the server or upon adding new bulling messages.



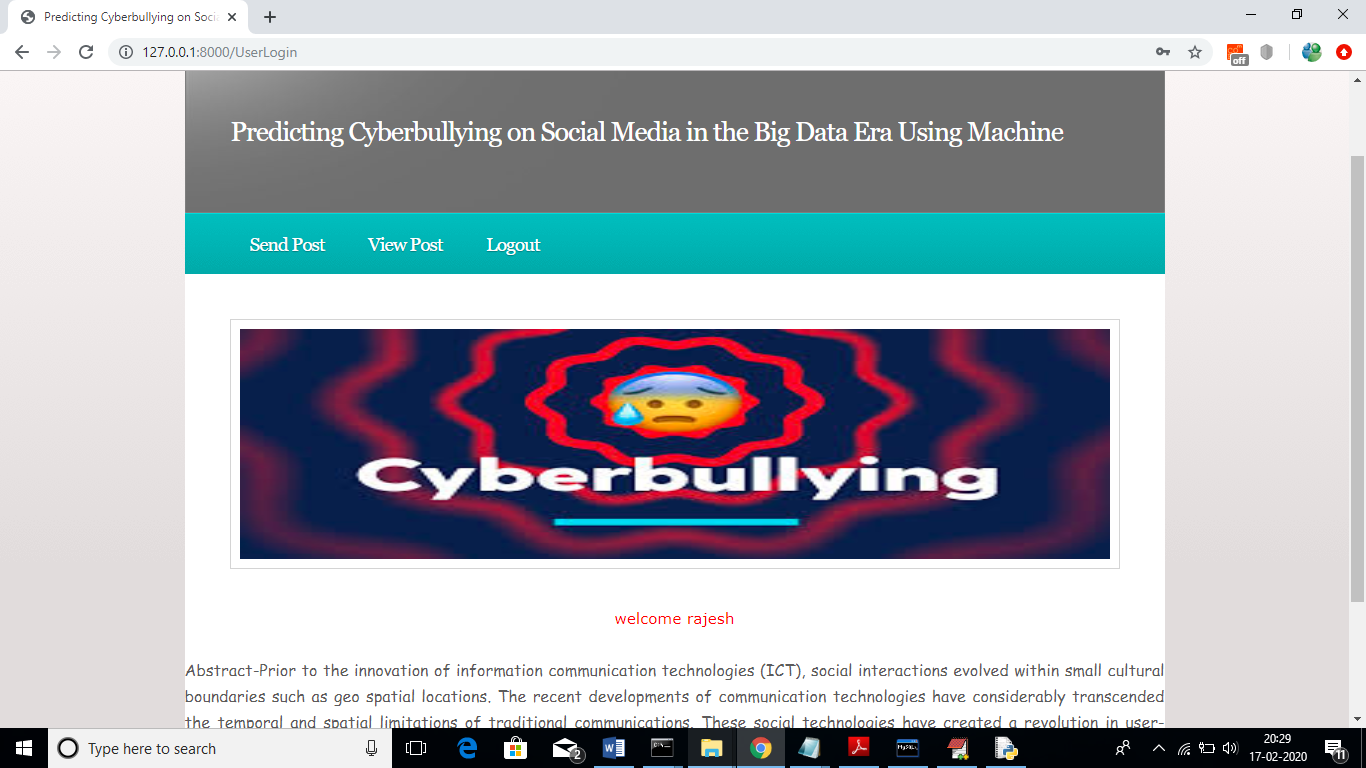
In above screen I ran SVM and got accuracy as 95. Similarly u need to select all algorithms one by one and run it.



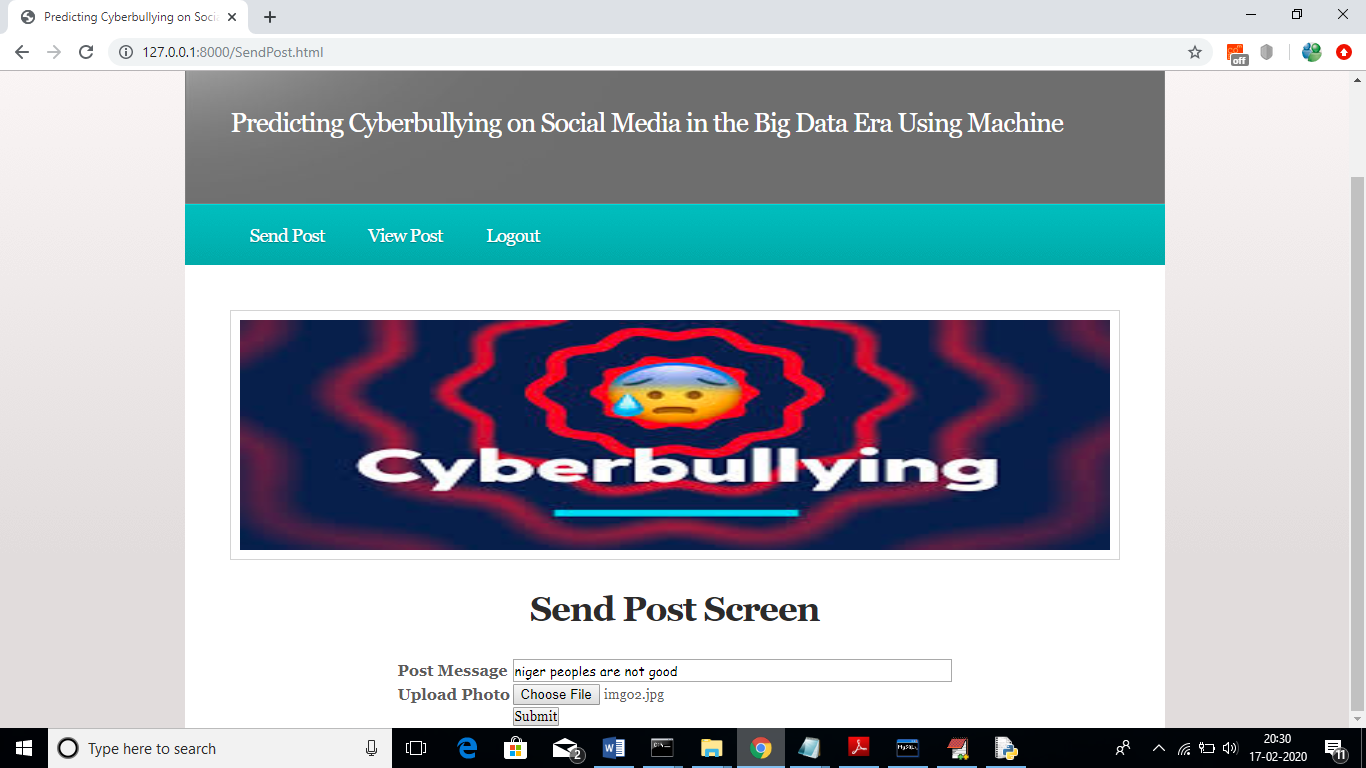
In above screen advance ‘Extreme Machine Learning’ algorithm gave 98% accuracy. Now admin logout and login as user to send posts.



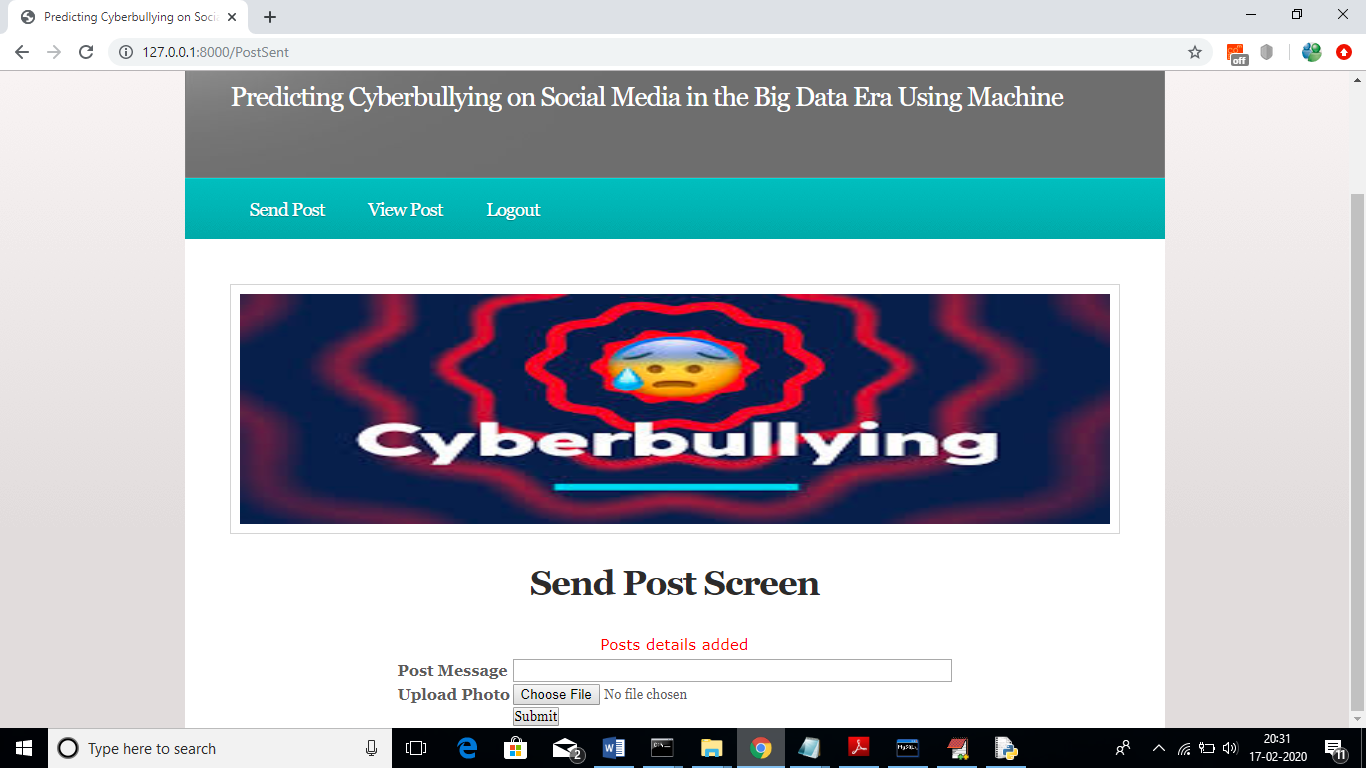
In above screen rajesh user is login and after login will get below screen

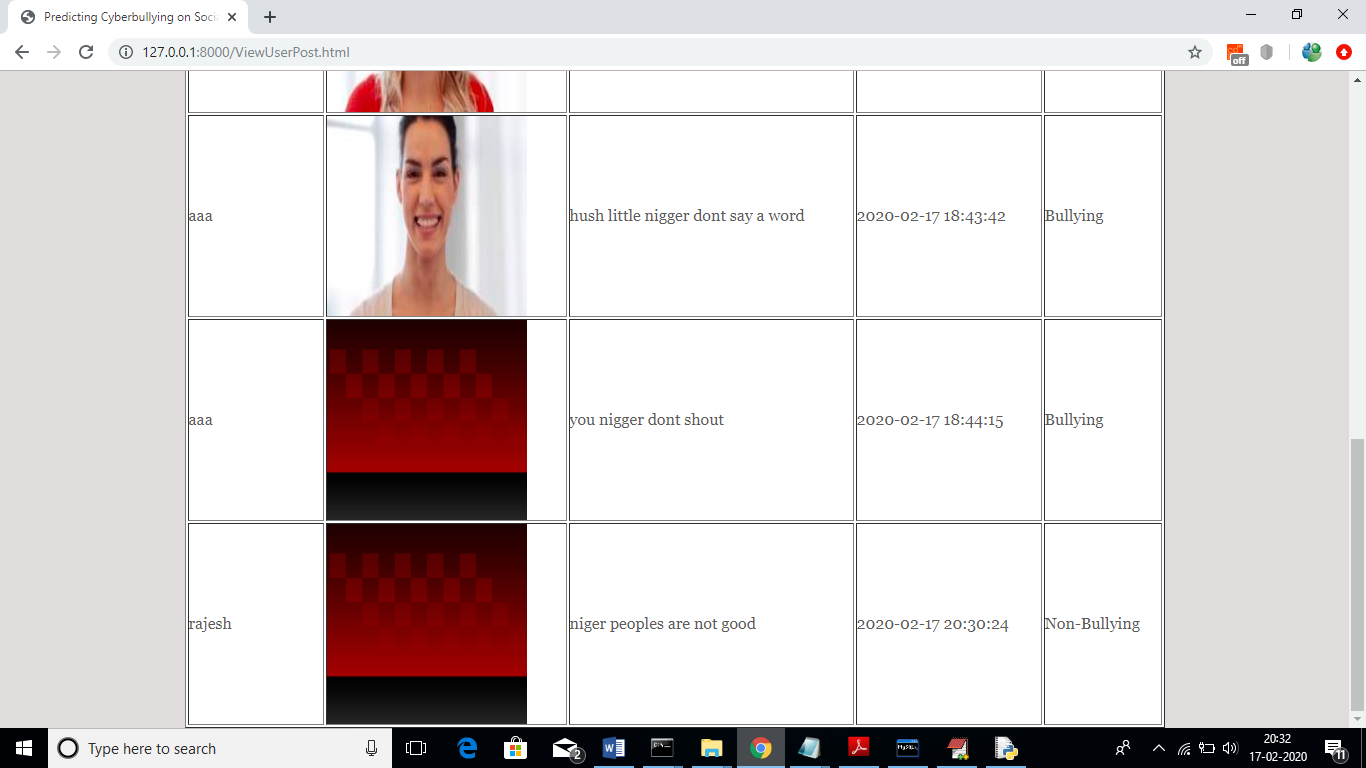


In above screen click on ‘Send Post’ link to get below screen



In above screen as post I added some messages and uploaded a photo also. After posting message will get below screen





In above screen we are seeing posts from all users and rajesh post predicted as ‘Non-Bullying’. Here based on words given in dataset will get prediction as bullying on non-bullying.