EasyChair Preprint

№ 8312

Hate Text Finder Using Logistic Regression

Ravikanti Vaishnavi, Kumbam Venkatreddy, Vangapati Nagamani and Mettu Sai Madhava Reddy

June 19, 2022

HATE TEXT FINDER USING LOGISTIC REGRESSION

Ravikanti Vaishnavi[1] Kumbam VenkatReddy[2] Vangapati Nagamani[3]

Mettu SaiMadhava Reddy[4]

[1]Department of Information Technology, Vignana Bharathi Institute of Technology, Ghatkesar, Hyderabad, India
vaishnaviravikanti7@gmail.com

[2]Assistant Professor, Vignana Bharathi Institute of Technology,Ghatkesar, Hyderabad, India venkatreddyk231@gmail.com

[3]Department of Information Technology, Vignana Bharathi Institute of Technology, Ghatkesar, Hyderabad, India
nagamanivangapati11@gmail.com

[4]Department of Information Technology, Vignana Bharathi Institute of Technology, Ghatkesar, Hyderabad, India
madhavreddy757@gmail.com

Abstract: Hate Text means a message designed to degrade, intimidate or incite violence or prejudicial action against a person or group of people based on their race, gender, ethnicity, nationality, religion, political affiliation, language, ability or appearance . So, Governments and social media providers put an effort to tackle offensive, abusive, and profanity in social media as an abuse of speech freedom. Considering the number of Internet users in the world and the conflict caused by offensive content involved in posts, there is an urge to develop offensive content detection for posts. This project uses a logistic regression model for classifying the words as (non)offensive words. This model can assist the government in enforcing the information and decreases the number of disputes due to aspiration freedom abuse on social media. In this project, we developed a social blog to demonstrate this entire process and it shows good results. We are testing whether the post contains offensive content or not at the time of posting itself.

Index terms - Logistic Regression , Hostile , Blog Posting and Viewing, User Authentication , Offensive.

1.INTRODUCTION

The hostile substance via the web- predicated networking media destinations might be as stag, defilement, provocation, prejudice, and foul. This hostile substance can make the customer differ from the other individuals allowed through that the misconception between the general populations can do and can prompt mischief different people groups through their ill-bred substance on the web predicated life destinations. According to there's a no confinement for the general population to talk anything they ask and to post whatever they feel, the general population uses these internet grounded life locales in an exceedingly way. It's decreasingly hard to deal with or to characterize that content and to detect the hostile terms regarding their implicit customer who start the application of the hostile terms in the converse.

Lately, Online Social Network has demonstrated to be a feasible vehicle for individuals to uninhibitedly convey what needs to be. The guests can within much of stretch conduct among themselves exercising talk errand people and offer or include posts, film land, textbooks, and soon., on other customer's profile. A portion of these dispatches conceivably allowed to be hostile by certain guests. In the UK, an overview was led. Its perceptivity shows that 28 of the youths progressed nearly in the range of 11 and 16 with a profile on a person-to-person communication web runner have encountered commodity disquieting on that point of which 18 have encountered vicious language and 3 were prompted to hurt themselves. individuals are permitted to banner similar reflections still there's no distinct answer for this issue. As the amounts of guests on informal communication have expanded snappily because of the expanded access to the web, it's demonstrating to be a test for the current fabrics to arrange similar dispatches successfully. In this paper, a frame is proposed, which identifies similar reflections.

Logistic retrogression is used to read a double outgrowth grounded on a set of independent variables. Okay, so what does this mean? A double outgrowth is one where there are only two possible scripts — either the event happens(1) or it doesn't be(0). Independent variables are those variables or factors which may impact the outgrowth(or dependent variable).

2. RELATED WORK

A significant number of scientists have effectively characterized the different methods for distinguishing the hostile language in the online networking organizing destinations. They utilized the current procedure like Natural Language Processing, Blacklist Moderation, and text digging for sifting the substance via web-based networking media. In Data Mining, by utilizing a regulated methodology of order, the hostile terms can be distinguished effectively and intelligently with the constant unique information. Under the administered procedure, different grouping algorithms can be characterized as Naive Bayes, choice tree, K-Nearest neighbor, and bolster vector machine. From this method, the SVM is more supported than different systems. The SVM gives worldwide information arrangement and anticipates a high exactness result than the Naive Bayes and another characterization strategy. The paper by Author [8] portrays the less computational unpredictability of the calculation and expressed that the paper can deal with the enormous dataset than existing scale-up techniques.

Various endeavors have been made to recognize and channel content in online informal organizations. Razavi et al utilizes a lexicon of terrible words alongside Bag of Words with the end goal of identification of affront. The nonappearance of specific words in the lexicon would yield mistaken outcomes. Kontosta this et al uses guideline based correspondence to follow and sort online predators. The framework marks and breaks down talk transcripts to recognize ruthless and non savage discourse. Spertus proposed a fire acknowledgment framework called Smokey that includes syntactic builds. It fabricates highlight vectors dependent on the language structure and semantics of each sentence inside each message. McEnergy et al utilized Bag of Words for hostile text location. The disadvantage of this framework is that it has a high false-positive rate. N-gram perhaps utilized as an elective way to deal with Bag of Words and it yields better outcomes. Climbed et al identify hostile tweets utilizing machine learning algorithms which accomplish a positive pace of 75.1% utilizing strategic relapse.

3. PROPOSED SYSTEM

The proposed framework can utilize the Logistic Regression to precisely order and distinguish the offensive and the protective sentence with high exactness. The proposed framework can recognize the potential client by methods for whom the offensive language is utilized. we direct the principal near investigation of different learning models on Hate and Abusive Speech on Twitter and talk

about the likelihood of utilizing extra highlights and context information for upgrades. This task applies machine learning methods to perform computerized hostile language identification. Hostile language can be characterized as communicating outrageous subjectivity and this investigation for the most part centers around two classes 'sensual' and 'bigot'.

ADVANTAGES OF PROPOSED SYSTEM

1. More accurate and high performance
2. It works on huge data sets and online social networks like Twitter, Facebook, etc
3. Its is easy to implement and detect offensive language..
4. Logistic regression is less inclined to over-fitting but it can over fit in high-dimensional data sets.
5. It performs well when the data set is linearly separable.

4.WORKING OF PROPOSED SYSTEM

1. User Authentication.
2. Blog Posting and Viewing.
3. Implement the ML Algorithm i.e Logistic Regression to find the offensive words.
4. Validate The results.

Step 1:

User Authentication or Logging in is usually used to enter a specific page, website, or application, that trespassers cannot see. The user credentials are typically some form of a username and a matching password,[1] and these credentials themselves are sometimes referred to as a login (or logon, sign-in, sign-on).

Step 2:

In Blog Posting and Viewing a user can create a post and user can just view the other users post.But user cannot have access to perform the actions on others posts like update,delete..etc. Only Super User(admin)has right to perform actions on the other users posts.

Step 3:

Logistic regression is used to predict a binary outcome based on a set of independent variables.Logistic regression is used to classify the probability of a binary event occurring, and to dealwith issues of classification.Logistic regression is the correct type of analysis to use when you're

working with binary data.

Step 4:

Validation of result is performed using the output from step 3. Usually , In this step it stops user topost on blog if any offensive or hostile text exists. Otherwise , the user can post it successfully.

User can post any of the content he likes and also he can comment on others content until and unless itcontains some abusive or hate-spreading content. This developed algorithm will help users to stop posting negative or abusive comments if they are willing to do so.

5.RESULTS :

Below are few captured images of the output screen when different inputs are Passed:

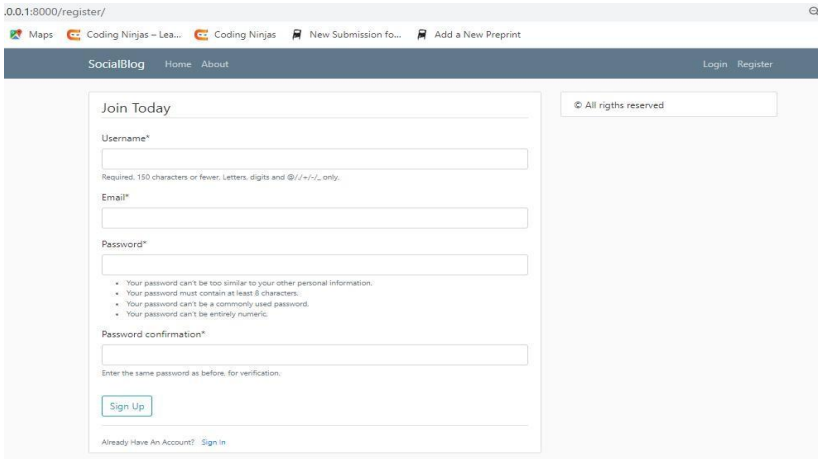

Figure 1: Sign up / Registration Page for User

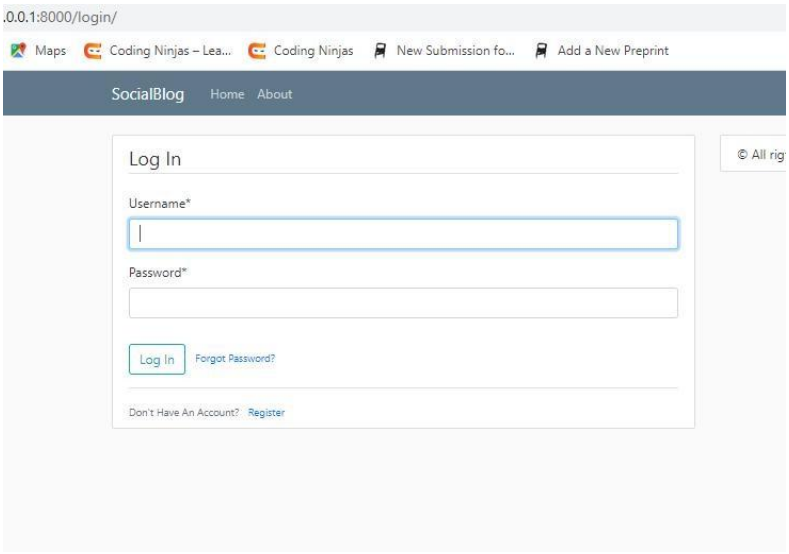

Figure 2: Login Page for User

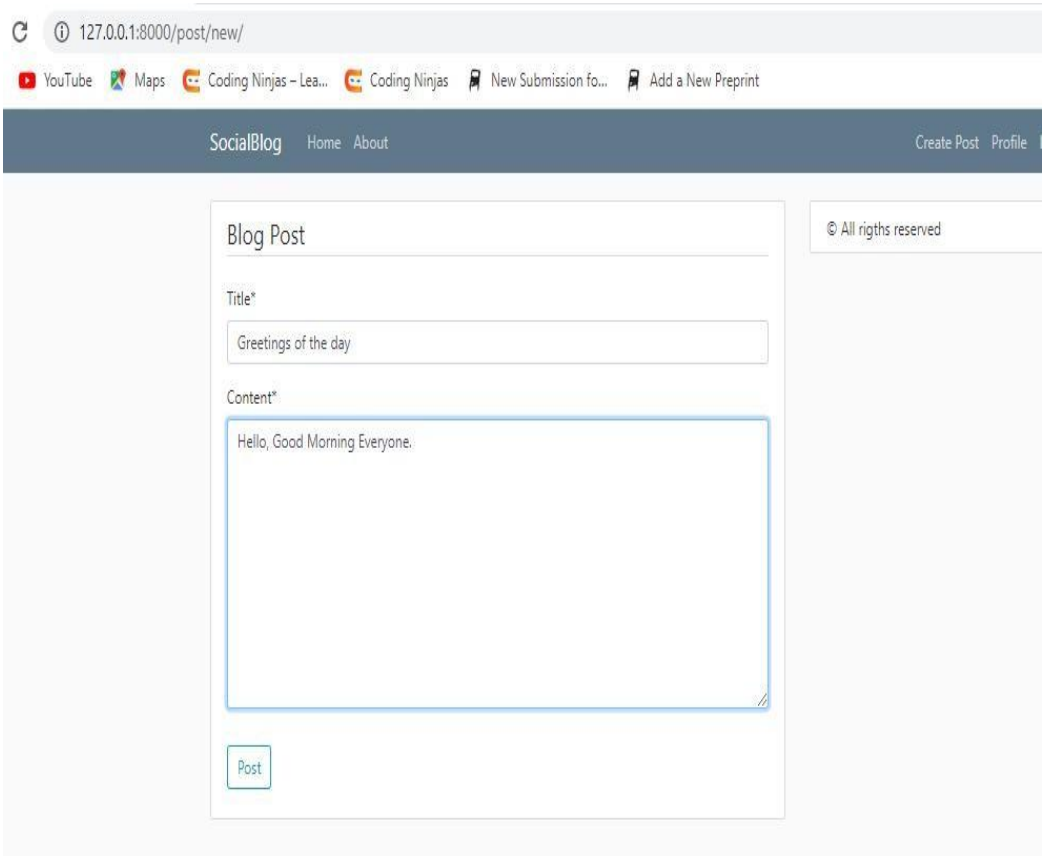

Figure 3: User Created a post

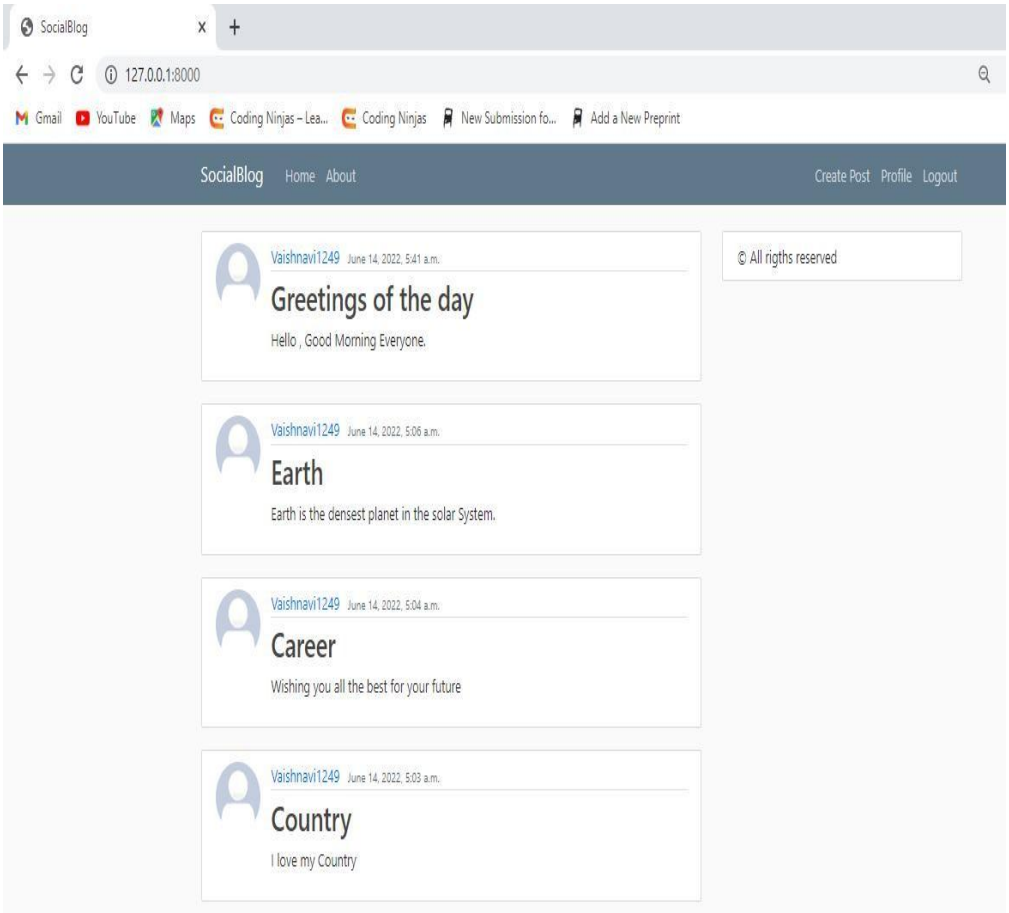

Figure 4: Posted Successfully (non - Offensive)

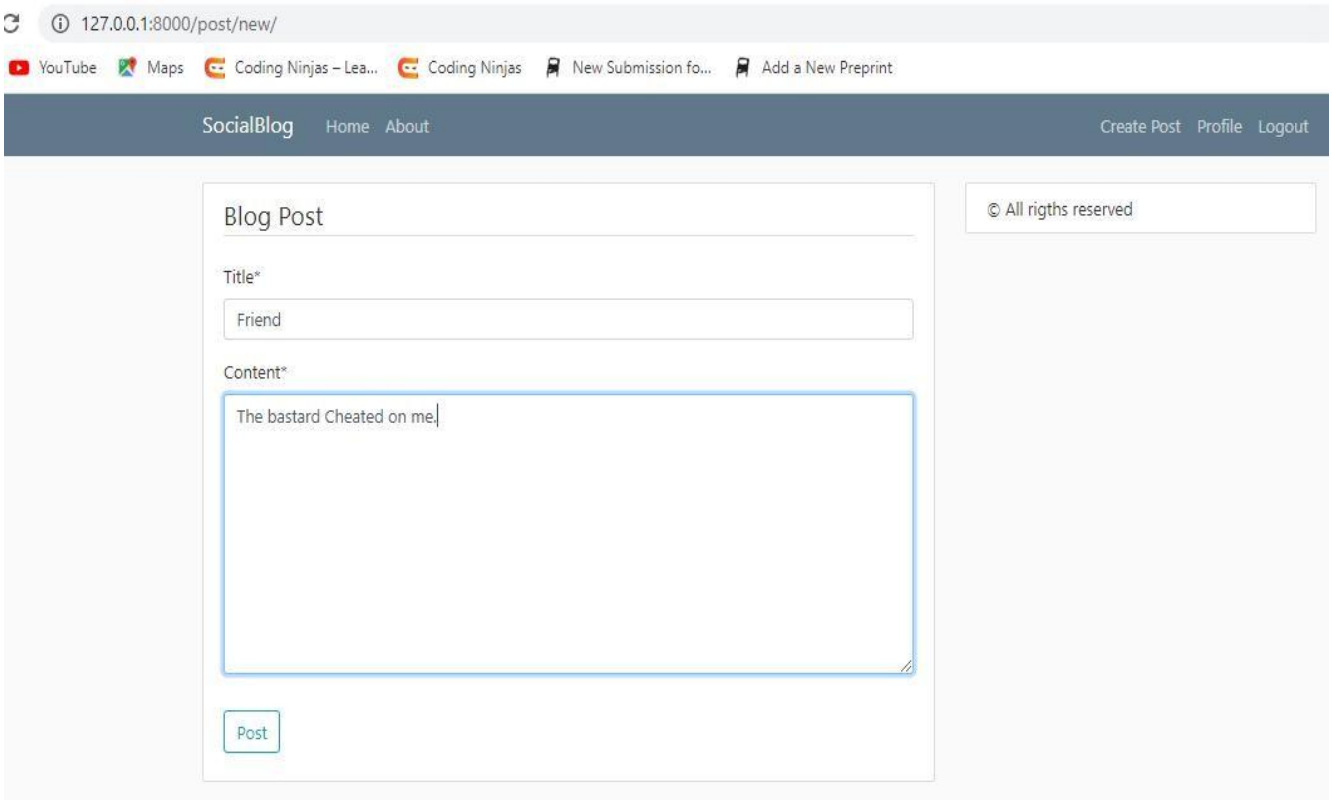

Figure 5: User Trying to post Offensive words

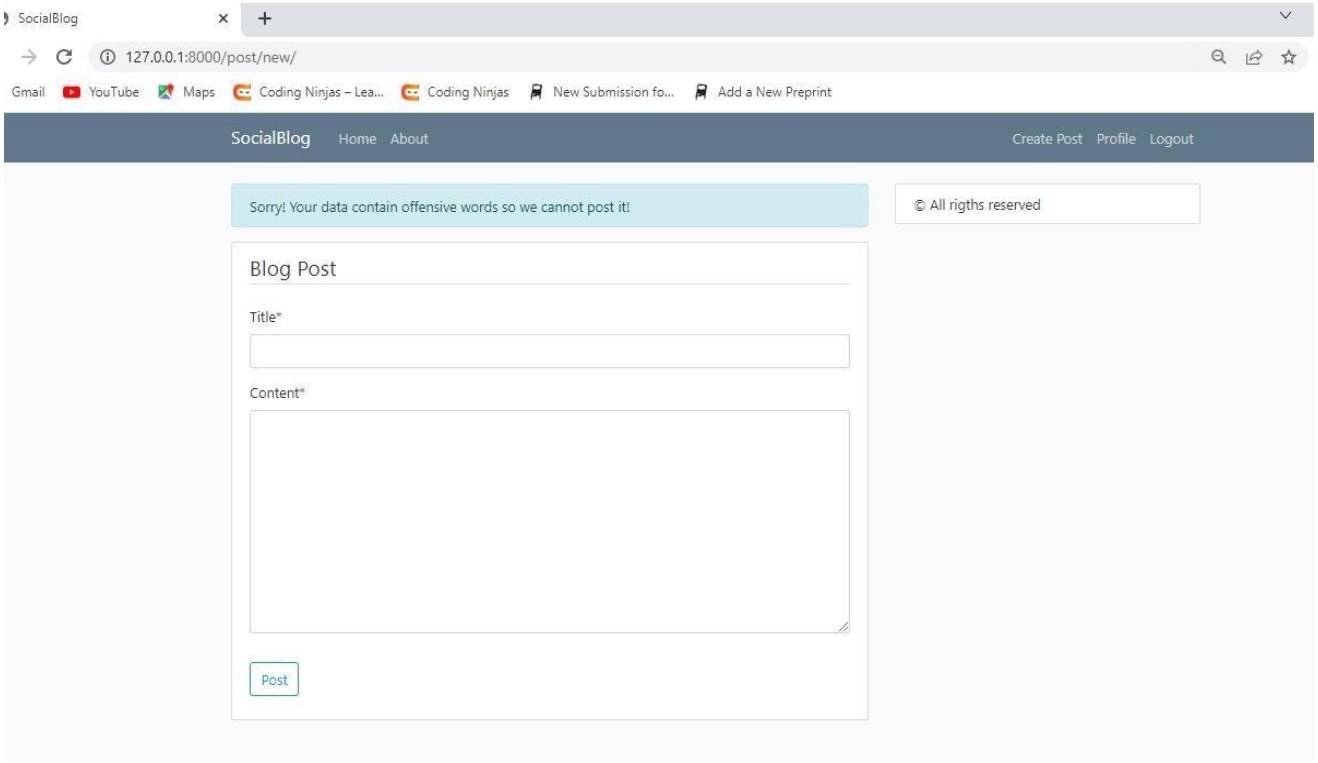

Figure 6: Cannot be Posted due to offensive

6.CONCLUSION

In this examination, we research existing text-mining strategies in recognizing hostile substances for securing juvenile online well-being. Explicitly to distinguish hostile substance in online networking, and further, foresee a client's probability to convey hostile substance. "Our examination has a few commitments. In the first place, we for all intents and purposes concept the thought of online hostile substance, and further recognize the commitment of pejoratives/obscenities and obscenities in deciding hostile substance, and present hand creating syntactic principles in distinguishing verbally abusing badgering". Second, we improved the customary ML techniques by not just utilizing lexical highlights to recognize hostile dialects, yet additionally fusing style highlights, structure highlights, and context-explicit highlights to all the more likely to foresee a client's probability to convey hostile substance in internet-based life

7.FUTURE SCOPE

In future , we can implement more classifications of hate speech. And also we can implement it in other languages like Hindi, Tamil, etc. We can still improve the performance of the algorithm and can implement artificial intelligence for the automatic detection of hate speech.

8.REFERENCES

[1] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, and Vasudeva Varma. 2017. Deep learning for hate speech detection in tweets. In Proceedings of the 26th International Conference on WorldWide Web Companion, pages 759–760.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022.

[3] Peter Burnap and Matthew Leighton Williams. 2014. Hate speech, machine classification and statistical modelling of information flows on twitter: Interpretation and communication for policydecision making.

[4] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Mean birds: Detecting aggression and bullying on twitter. In Proceedings of the 2017 ACM, pages 13–22. ACM.

[5] Kyunghyun Cho, Bart Van Merri¨enboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder forstatistical machine translation.

[6] Nemanja Djuric, Jing Zhou, Robin Morris, Mihajlo Grbovic, Vladan Radosavljevic, and Narayan Bhamidipati. 2015. Hate speech detection with comment embeddings, pages 29–30. ACM.

[7] Maeve Duggan. 2017. Online harassment 2017. Pew Research Center; 2018.

[8] Antigoni Founta, Constantinos Djouvas, Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Gianluca Stringhini, Athena Vakali, Michael Sirivianos, and Nicolas Kourtelli Large scale crowdsourcing and characterization of twitter abusive behavior.

[9] Jennifer Golbeck, Zahra Ashktorab, Rashad O Banjo, Alexandra Berlinger, Siddharth Bhagwan, Cody Buntain, Paul Cheakalos, Alicia A Geller, 2017.

[10]https://www.javatpoint.com/logistic-regression-in-machine-learning