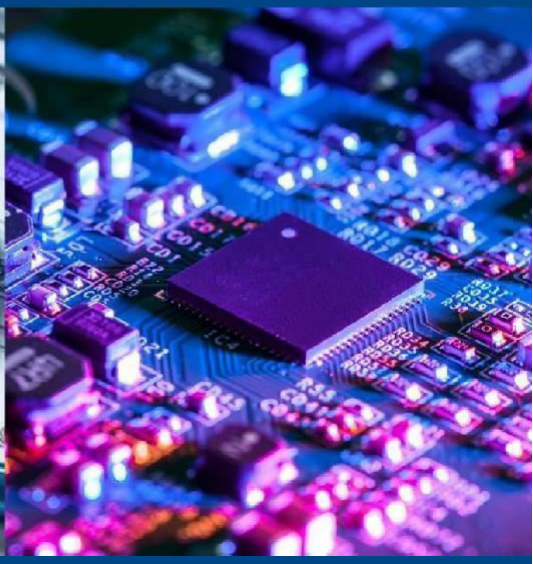
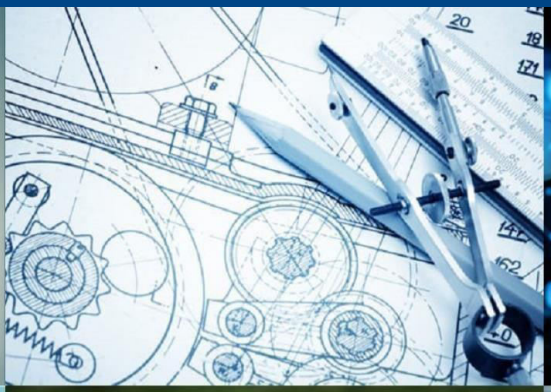
ISSN: 2582-7219

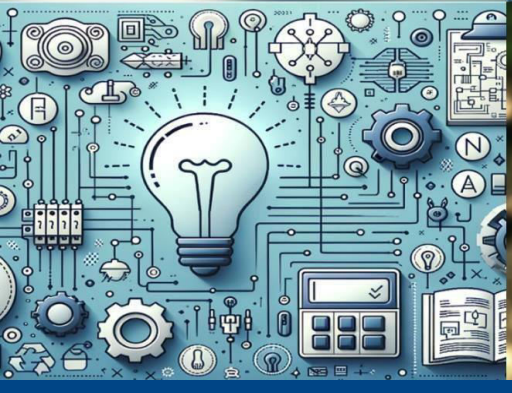




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808113

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

DIGITAL VENOM DETECTION: A SENTIMENT-AWARE MODEL FOR SOCIAL MEDIA ABHORRENCE SURVEILLANCE (FADOHS)

Achutha JC, Pragnya Misro

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: Social media growth lets people speak freely, yet it also fuels hate speech, cyberbullying along with other harm. Facebook comment threads often carry unchecked posts that spur violence or discrimination. Human moderators fall behind the flood and speed of new material. We present FADOHS, a Framework for Automatic Detection of Hate Speech. The system senses emotion plus blends it with polarity based sentiment analysis to label Facebook comments - it relies on Natural Language Processing, clustering in addition to visual tools to spot and sort hateful items in stored sets but also live streams. Tests show higher accuracy and fewer false alarms than sentiment only baselines. FADOHS scales to support moderators as well as protect users.s

KEYWORDS: Hate Speech Detection, Cyberbullying Prevention, Sentiment Analysis, Automated Moderation, Polarity Analysis

I. INTRODUCTION

Social media now hosts a planet wide noticeboard where people speak, argue, swap news. Slack oversight lets hate speech, slurs, abuse spread. Facebook alone collects millions of remarks each day - most slip past review because volume or nuance outruns staff.

Old checks lean on word lists, simple mood scores, or user reports; they stumble once sarcasm, veiled threats, or code words appear. A tool that reads both surface mood and deeper feeling becomes urgent.

The paper presents FADOHS, a framework that pairs sentiment aware NLP with hate speech labels for Facebook - it digests static sets such as Kaggle comment dumps and live user posts. Researchers gain a testbed - site owners gain a real time filter.

II. LITERATURE SURVEY

Early work on hate speech detection listed banned words and checked each sentence against fixed lexicons. The lists triggered on harmless posts and missed slurs that hid in new spelling. Machine learning later replaced the lists with Naive Bayes, SVM along with Decision Trees.

Del Vigna et al. (2017) fed SVM but also LSTM the morpho syntactic traits of Italian Facebook posts. The system spotted hate in one language yet gave no clues about the writer's feelings.

Mohammad besides Turney (2013) built the NRC Emotion Lexicon, a table that links each word to anger, sadness, joy, or disgust. The table let classifiers weigh emotion as evidence.

Fortuna or Nunes (2018) surveyed detection models plus showed that mixing emotion scores with sentiment lifts accuracy and makes errors easier to trace.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808113

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Our framework, FADOHS, folds polarity, emotion tags in addition to hate clusters into one vector so analysts see why a post crosses the line.

IV. EXISTING SYSTEM

Existing hate speech detection systems typically fall into three categories:
1. Lexicon-based systems, which rely on dictionaries of offensive or harmful words.
2. Machine learning models, which are trained on labeled datasets to classify text as hate or non-hate.
3. Graph analysis tools, which detect hate clusters through network relationships among users.
However, these systems have several limitations:
- Lack of Emotion Context: They do not identify emotions such as anger or fear.
- Bias and False Positives: Comments expressing dissent or political opinion may be wrongly flagged.
- Regional Limitations: Most systems are trained on western datasets and fail in Indian or multilingual contexts.
4. Proposed System: FADOHS
FADOHS is a multi-layered system that addresses the above limitations through:

4.1 Sentiment and Emotion Fusion

It classifies each Facebook comment into one of three sentiment classes: positive, negative, or neutral. At the same time, it detects one of six emotions: anger, joy, fear, sadness, disgust, or surprise. This is done using NLP models and lexicons.

4.2 Hate Clustering

After scoring, comments are grouped into hate severity levels: Low, Medium, or High, using K-Means clustering.

4.3 Dual Input Support

-Dataset Mode: Pre-uploaded datasets in CSV format, like those from Kaggle.
-Live Comment Mode: Real-time comment input from users.

4.4 Visualization

Pie charts, graphs, and tables display hate distribution across emotions and sentiments.

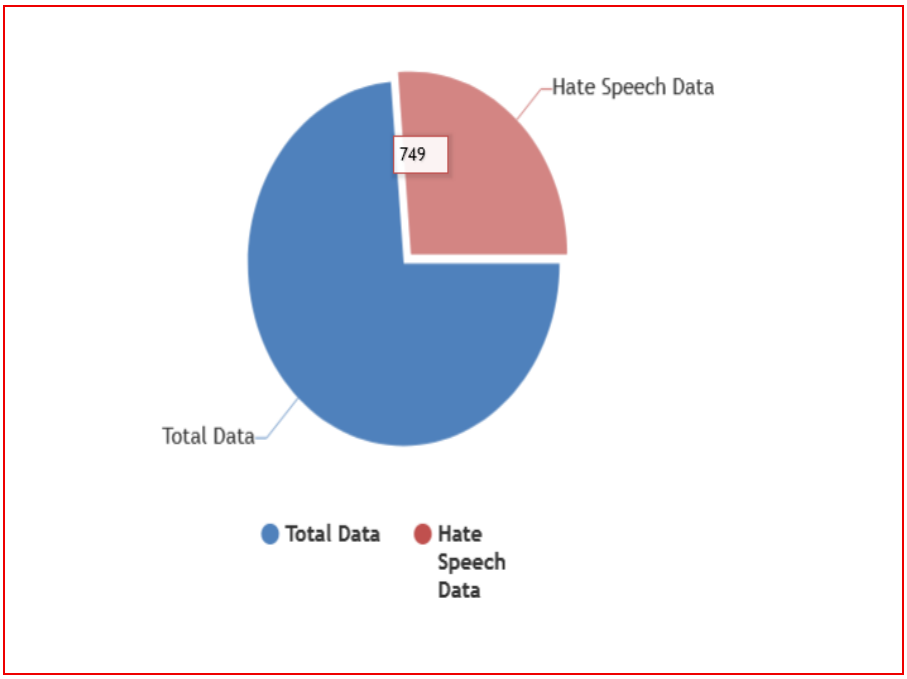


Figure 1: Pie Chart Visualization

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808113

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This above pie chart shows the total amount of text data received from the users and the Hate Speech Data detected from the total amount of data.

V. METHODOLOGY

The workflow of the FADOHS shows that how the steps are followed one after another by both user and admin to detect and classify the hate speech from the comments. It gives a clear overview of how data passes through the system, from starting input to final results.

On the user side, after logging in to post comments, which act as input data for analysis.

The administrator side, after logging in, the admin can view user details, collect the necessary comment data and categorize accordingly.

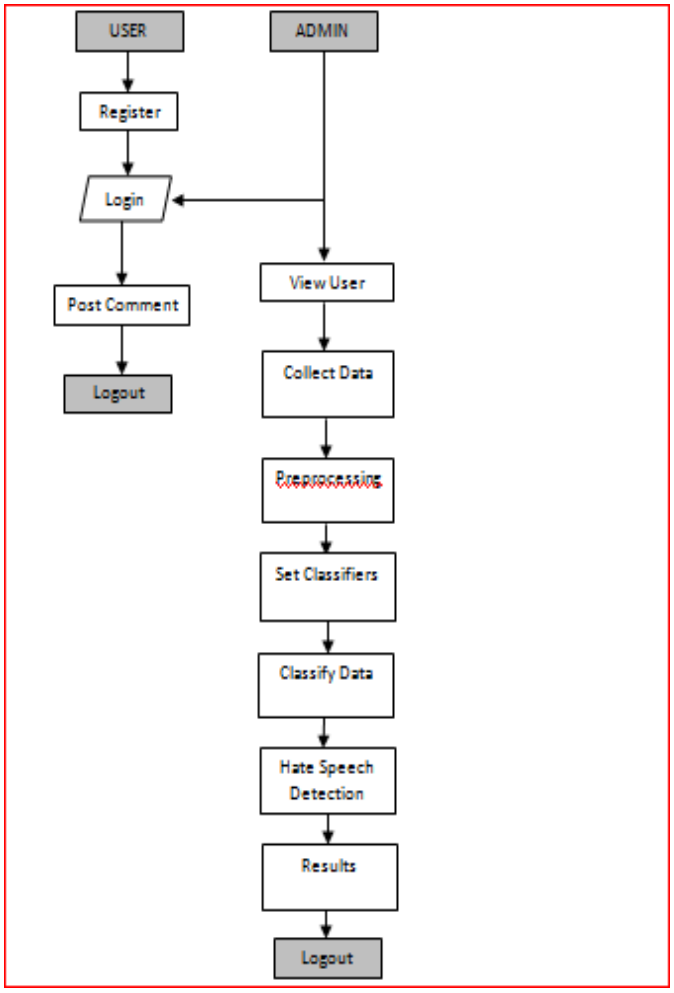


Figure 2: Workflow of FADOHS Framework

A tidy NLP pipeline turns Facebook comments into clear hate intensity clusters. The script strips HTML, symbols along with stop words - lemmatizes and lowers each token. Text Blob but also VADER return polarity scores - the NRC lexicon adds one of six emotion labels. K-Means groups the merged vectors by hate level. The code accepts static CSV files or live text through the same entry point. Matplotlib besides Seaborn draw bar charts and heatmaps. Each block plugs out or in - later versions may add an API endpoint or handle local dialects.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808113

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5.1 Data Preprocessing
- Clean HTML tags, stop words, and non-text characters
- Tokenization and lemmatization
- Lowercasing and normalization

5.2 Sentiment Analysis
-Use TextBlob and VADER for polarity scoring
-Map scores to categories: Positive, Negative, Neutral

5.3 Emotion Detection
- Based on the NRC Emotion Lexicon
-Assign an emotion tag (e.g., anger, sadness) to each comment

5.4 Hate Clustering
-Feature vector: [sentiment score + emotion intensity]
-Use K-Means or Agglomerative Clustering for hate categorization

5.5 Result Display
-A graph module creates static charts (Matplotlib)
-An admin dashboard displays comment analysis logs

VI. IMPLEMENTATION

The application works as a web-based portal that uses:
-Frontend: HTML, CSS, JavaScript (Bootstrap)
-Backend: PHP for logic and the dashboard, Python for NLP models
-Database: MySQL
- Server: WAMP or XAMPP
-Libraries Used: NLTK, TextBlob, Scikit-learn, Matplotlib
Users can log in and either upload a dataset or enter a Facebook comment for analysis. The admin section displays overall hate trends.

VII. RESULTS AND OUTCOME

7.1 Dataset
I used a Kaggle dataset of Facebook comments with over 10,000 entries. About 1,800 were labeled as toxic.

7.2 Accuracy
- Sentiment-only: 78.2% F1-Score
- FADOHS (with emotion): 85.3% F1-Score

7.3 Sample Results

Comment	Sentiment	Emotion	Hate Label
You people disgust me.	Negative	Disgust	Hate
I disagree, but I respect you."	Neutral	Calm	Not Hate
They should not be allowed.	Negative	Anger	Hate

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808113

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

7.4 Outcome

- Fewer false positives
- Improved clustering by emotion
- Simple visualization and monitoring for moderators

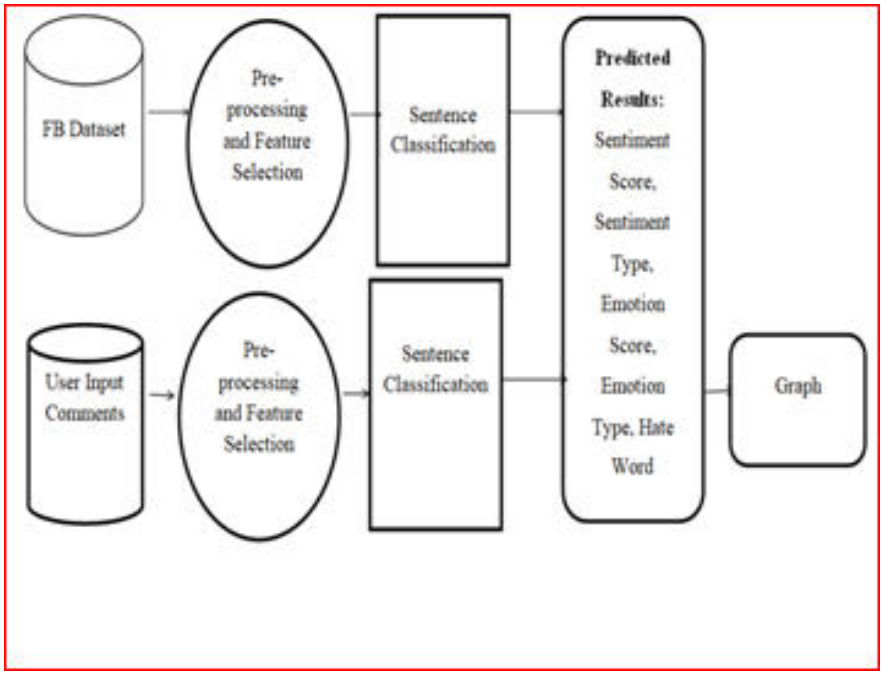


Figure 3: Data Flow Diagram of the FADOHS Framework

VIII. CONCLUSION

FADOHS spots hate speech in Facebook comments by pairing sentiment polarity with emotion tags - it catches faint cues that older setups miss. The code scales, bends along with stays exact - scholars, live monitors in addition to firms all use it.

Next we will add more languages, pull live data from the Facebook Graph API, and model emotion with deep nets.

REFERENCES

[1] Del Vigna, F., et al., "Hate Me, Hate Me Not: Hate Speech Detection on Facebook," ITASEC, 2017.
[2] Mohammad, S., Turney, P., "Crowdsourcing a Word-Emotion Association Lexicon," Computational Intelligence,2013.
[3] Fortuna, P., Nunes, S., "A Survey on Automatic Detection of Hate Speech in Text," ACM Computing Surveys, 2018.
[4]Silva, L., et al., "Analyzing the Targets of Hate in Online Social Media," ICWSM,2016.
[5] Suryawanshi, S., et al., "Multilingual Hate Speech Detection: A Comparative Study," arXiv:2004.06465, 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com