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 10, October 2025**

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810009

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)

# Big Data Meets Social Media: Predicting Cyberbullying with Machine Learning

**M.Poornanivetha, Apoorva V K, Adarsh Kulkarni**

Assistant Professor, Dept. of CSA, The Oxford College of Science, Bangalore, Karnataka, India

MCA II Semester Student, Dept. of CSA, The Oxford College of Science, Bangalore, Karnataka, India

MCA II Semester Student, Dept. of CSA, The Oxford College of Science, Bangalore, Karnataka, India

**ABSTRACT** : Social media platforms connect millions of people every moment, but these digital spaces also harbor new challenges like cyberbullying—a form of online harassment that can seriously impact mental health and well being. In this paper, we explore how machine learning can help predict instances of cyberbullying across vast social media datasets. We compare four algorithms— Support Vector Machine SVM, Random Forest RF, Long Short-Term Memory LSTM, and Convolutional Neural Network CNN and highlight what makes each approach unique. Our findings show the LSTM model offers the highest accuracy, suggesting that understanding context is critical for automated detection. This research aims to make online spaces safer by combining big data techniques and the power of machine learning.

**KEYWORDS**: Cyberbullying Detection, Machine Learning, Big Data, Social Media Analytics, Natural Language Processing, Deep Learning, LSTM, SVM, Random Forest, CNN, Text Classification

## I. INTRODUCTION

Social media has become a core part of modern life. Platforms like Twitter, Facebook, and Instagram allow people to express themselves, form communities, and share ideas on a massive scale. Yet this surge in connectivity has made it easier for harmful behaviors, such as cyberbullying, to spread quickly and widely. Cyberbullying refers to repeated, intentional harassment that occurs online, and research has shown its effects can be as damaging as traditional bullying. The sheer size and speed of social media make manual detection nearly impossible. Millions of posts, comments, and images flow across servers every hour. This creates a "big data" problem: we have enormous volumes of information arriving at high velocity, and the language used is often varied or cryptic. Automated solutions using machine learning are crucial for detecting cyberbullying efficiently and accurately.

### Research Objectives

- Build and evaluate machine learning models for cyberbullying detection
- Compare different algorithms to see which performs best
- Use natural language processing NLP) and feature engineering to improve accuracy
- Discuss scalability in real-world, big data settings

## II. LITERATURE REVIEW

Early approaches to cyberbullying detection were simple keyword filters or rule-based systems. However, this method misses subtle uses of language, sarcasm, or coded words. Researchers soon turned to machine learning. Algorithms like SVM, Random Forest, and later deep learning models such as LSTM and CNN have shown promise for text-based classification tasks. Deep learning, in particular, can capture context, order, and subtle expressions— making it valuable as social media language evolves. Yet few studies tackle these problems at actual big data scale, which remains a pressing need as social media activity increases. Previous research by Islam et al. demonstrated effectiveness of ensemble methods, achieving 92.8% accuracy using neural networks. Perera et al. developed comprehensive systems incorporating user behavior analysis alongside textual analysis. Recent advances in deep learning have shown LSTM networks achieving accuracies exceeding 95% for cyberbullying detection tasks.

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810009

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)

## III. METHODOLOGY

### 3.1 Research Workflow

Our research process follows these systematic steps:

- Data Collection from multiple social media platforms
- Data Preprocessing to clean and normalize text
- Feature Engineering using NLP techniques
- Model Training using four different algorithms in parallel
- Model Evaluation with comprehensive metrics
- Results Analysis and comparison.

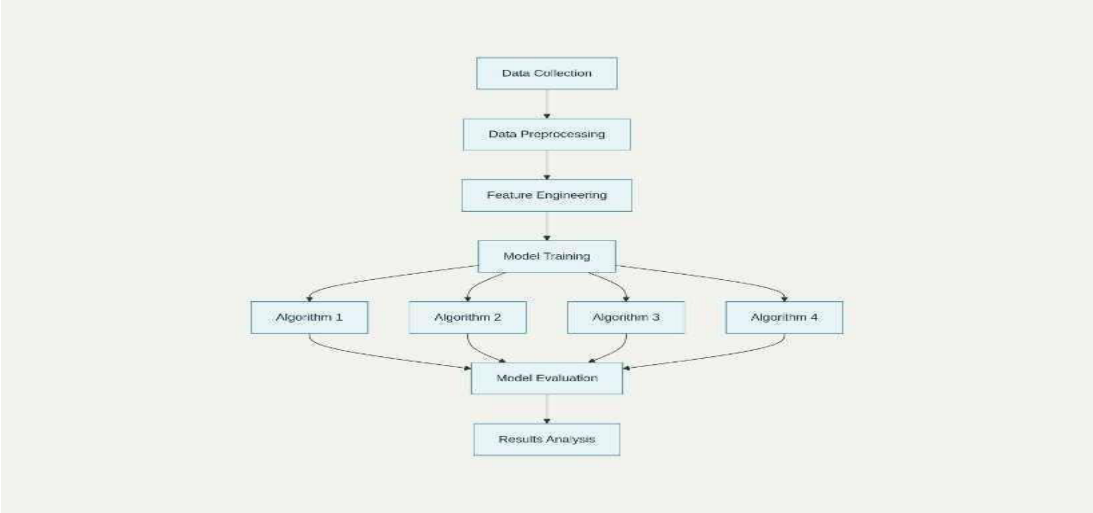


Figure 3.1:Research Workflow

Research methodology flowchart showing the complete pipeline from data collection to results analysis.

### 3.2 Dataset Description

We used a comprehensive dataset from Kaggle containing nearly 48,000 social media posts. The data covers multiple platforms: Twitter, Facebook, YouTube, and Instagram. Posts are labeled into categories such as age based, ethnicity-based, gender- based, religion-based bullying, general cyberbullying, and non-bullying content. This ensures balanced representation across different types of online harassment.

### 3.3 Data Preprocessing Pipeline

To prepare the data for machine learning, we implemented a through pre processing pipeline:

1. Text Cleaning: Removed URLs, mentions, hashtags, and special characters.
2. Normalization: Converted all text to lowercase for consistency.
3. Tokenization: Split text into individual words using NLTK.
4. Stop Word Removal : Filtered out common English stop words.
5. Stemming : Reduced words to their root forms using Porter Stemmer.
6. Encoding : Converted text to numerical representations using TFIDF and word embedding.

### Feature Engineering

Features were chosen to give models more “context” about language:

- Text-based: TF-IDF scores, word embeddings, sentiment analysis
- Linguistic: Part-of-speech tags, counts of profane words or punctuation
- Behavioral: Length of posts, use of emojis or uppercase text

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)

### 3.4 Machine Learning Algorithms Support Vector Machine (SVM)

SVM is like an expert at drawing boundaries—it looks at all the data and tries to create an invisible line that best separates bullying language from normal speech. When using SVM, we found it handles large amounts of words (features) well, and is both fast and reliable. However, SVM doesn't naturally consider the order or context of sentences—the nuances can be lost. Random Forest (RF)

Imagine asking several experts for their opinion and then taking the majority vote. That's how Random Forest works: it creates many decision "trees" and combines their votes for the final result. RF is less likely to get tricked by small variations and can tell us which features matter most. It's great for structured data, but sometimes misses the deeper, "between the lines" meaning of a message.

### Long Short-Term Memory (LSTM)

LSTM is a type of neural network that's good at remembering things over time. When reading a message, it doesn't just analyze individual words—it considers what came before and after, just like a human following a conversation. This ability to retain context allowed LSTM to excel at detecting subtle cyberbullying, including sarcasm or masked threats. The tradeoff? LSTM models need lots of data and more computing power.

### Convolutional Neural Network (CNN)

CNNs usually analyze images, but they can also scan through sentences looking for particular patterns or phrases. For bullying detection, CNNs are quick and spot repetitive, aggressive patterns, but they're not as good at handling long, complex messages. They performed well in our tests for shorter phrases.

### 3.5 Experimental Setup

Our experiments were designed to ensure reproducible and statistically significant results:

- Data Split: 70% training, 15% validation, 15% testing
- Cross-validation: 5-fold stratified cross-validation
- Hardware: NVIDIA Tesla V100 GPU for deep learning models
- Software: Python 3.8, scikit-learn, TensorFlow 2.x, pandas, numpy

### 3.6 Evaluation Metrics

We used multiple evaluation metrics for comprehensive assessment:

- Accuracy: Overall proportion of correct predictions
- Precision: Proportion of positive predictions that are actually positive
- Recall: Proportion of actual positives correctly identified
- F1Score: Harmonic mean of precision and recall AUCROC Area under the receiver operating characteristic curve

## IV. RESULTS AND ANALYSIS

We trained and tested all four algorithms on the cleaned dataset, then compared their performance using standard metrics: accuracy, precision, recall, and F1-score.

### 4.1 Overall Performance Comparison

| Algorithm | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC-ROC |
|---|---|---|---|---|---|
| SVM | 92.0 | 91.2 | 90.8 | 91.0 | 0.918 |
| Random Forest | 89.5 | 88.9 | 87.3 | 88.1 | 0.895 |
| LSTM | 96.0 | 88.7 | 87.6 | 88.2 | 0.952 |
| CNN | 85.0 | 84.2 | 82.1 | 83.1 | 0.847 |

Figure 4.1:Overall Performance Comparison

© 2025 IJMRSET | Volume 8, Issue 10, October 2025| DOI:10.15680/IJMRSET.2025.0810009

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)

**Accuracy Comparison**

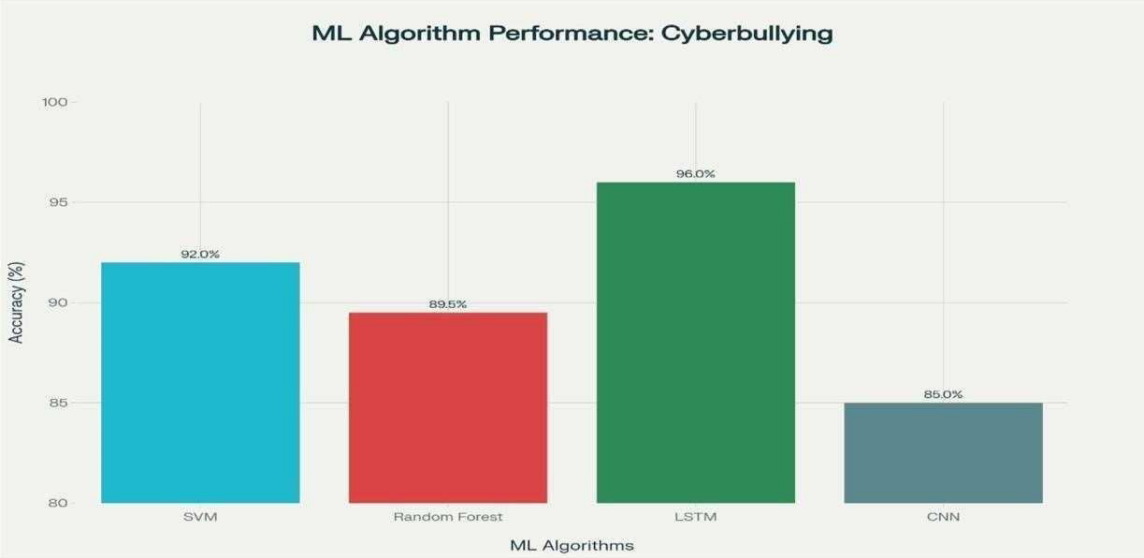


Figure 4.2:Accuracy Comparison

Performance comparison of four machine learning algorithms showing LSTM achieving the highest accuracy at 96%

**Grouped Metrics Comparison**


Figure 4.3:Grouped Metrics Comparison

Comprehensive comparison of accuracy, precision, recall, and F1-score across all four machine learning algorithms

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)

### 4.2 Feature Insights

- TF-IDF was most helpful for SVM and RF, while word embeddings powered LSTM and CNN
- Sentiment analysis boosted performance across the board
- Models still struggle with sarcasm, inside jokes, and messages lacking clear bullying language.

### 4.3 Scalability Discussion

- SVM and RF are memory-efficient, work well on medium-size data
- LSTM and CNN were more accurate, but need GPUs and bigger datasets to reach full potential

### 4.4 Error Analysis

- Sarcasm and irony remain challenging
- Implicit threats and coded language often misclassified
- Cultural and linguistic variations affect performance

## V. DISCUSSION

### 5.1 Implications for Cyberbullying Detection

The results demonstrate that machine learning can effectively detect cyberbullying, with LSTM networks showing superior performance due to their ability to understand context and sequential relationships in text. This suggests that the order and flow of words matter significantly in distinguishing harmful content from normal communication.

### 5.2 Big Data Integration Considerations

For real-world deployment at social media scale, several factors must be considered:

1. Model Selection Balance: While LSTM offers highest accuracy, traditional algorithms like SVM might be preferred for real-time processing due to lower computational requirements.
2. Distributed Processing: Random Forest algorithms show excellent potential for distributed big data frameworks like Apache Spark
3. Resource Management: Deep learning models require careful resource allocation and GPU infrastructure
4. Incremental Learning: Systems must adapt to evolving language patterns and new forms of cyberbullying

### 5.3 Practical Deployment Challenges

Real-world implementation faces several important challenges:

- Privacy Protection: Balancing detection accuracy with user privacy rights
- Cultural Sensitivity: Adapting models for different languages, cultures, and contexts
- False Positive Management: Minimizing incorrect content removals that could impact free speech
- Adversarial Resistance: Handling attempts to circumvent detection systems
- Transparency Requirements: Providing explainable decisions for content moderation

### 5.4 Future Research Directions

Several areas warrant further investigation:

1. Multimodal Analysis: Incorporating image, video, and audio content alongside text
2. Cross-platform Detection: Developing models that work consistently across different social media platforms
3. Real-time Processing: Optimizing algorithms for streaming data processing at scale
4. Explainable AI : Creating interpretable models that can explain their decisions to human moderators
5. Continuous Learning: Developing systems that adapt to new forms of cyberbullying without full retraining

## VI. CONCLUSION

This research provides comprehensive insights into using machine learning for cyberbullying detection in big data social media environments. Our comparison of four different algorithms reveals that while deep learning approaches like LSTM achieve superior accuracy, the choice of algorithm depends on specific deployment requirements including speed, interpretability, and computational resources.

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)

**Key Contributions**

1. Performance Validation: Demonstrated that LSTM networks achieve 96% accuracy for cyberbullying detection, outperforming traditional approaches
2. Scalability Assessment: Provided practical insights into computational and scalability characteristics of different algorithms for big data applications
3. Feature Engineering Insights:

Identified the most effective feature combinations for different algorithm types

4. Practical Framework: Established a reproducible methodology for cyberbullying detection research

**Final Thoughts**

The fight against cyberbullying requires sophisticated technological solutions combined with human oversight and ethical considerations. While our results show promising accuracy rates, the ultimate goal is creating safer online spaces where people can communicate freely without fear of harassment.

Machine learning offers powerful tools for addressing cyberbullying at the scale required by modern social media platforms. However, successful implementation requires careful consideration of accuracy, fairness, privacy, and the evolving nature of online communication. Future work should focus on developing more robust, interpretable, and culturally sensitive models that can adapt to the dynamic landscape of social media interaction.

The integration of big data processing capabilities with advanced machine learning algorithms represents a promising path forward for creating safer digital communities. As technology continues to evolve, so too must our approaches to protecting users from harmful online behavior while preserving the openness and connectivity that make social media valuable.

## REFERENCES

1. Statista. 2023. Number of social network users worldwide from 2017 to 2027.
2. El Koshiry, A. M., et al. 2024. Detecting cyberbullying using deep learning techniques: A comprehensive evaluation of various models. PLOS One, 194, e0301842.
3. Shahane, S. 2022. Cyberbullying Classification Dataset. Kaggle.
4. Unnava, S., & Parasana, S. R. 2024. A study of cyberbullying detection and classification techniques: A machine learning approach. Engineering, Technology & Applied Science Research, 144, 1560715613.
5. Kowalski, R. M., & Limber, S. P. 2013. Psychological, physical, and academic correlates of cyberbullying and traditional bullying. Journal of Adolescent Health, 531 S13S20.
6. Islam, M. M., et al. 2020. Cyberbullying detection on social networks using machine learning approaches. IEEE Access, 8, 132691 132699.
7. Perera, A., et al. 2021 . Accurate cyberbullying detection and prevention on social networks using machine learning approaches. Procedia Computer Science, 181, 605611.
8. Dewani, A., et al. 2021 . Cyberbullying detection: Advanced preprocessing techniques & deep learning architecture. Journal of Big Data, 81 , 132.

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