Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Harnessing Machine Learning and Deep Learning for Crime Detection in Social Media through Advanced Data Mining Techniques

Suresh V Reddy[1], Dr. Sanjay Bhargava[2]

[1]*Research Scholar, Department of Computer Science & Engineering, Faculty of Engineering and Technology, Mansarovar Global University, Bhopal.*

[2]*Research Guide, Department of Computer Science & Engineering,Faculty of Engineering and Technology, Mansarovar Global University, Bhopal.*

sureshrb27@gmail.com[1], sanjaybhargava78@gmail.com[2]

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

An enormous quantity of user-generated data is created daily as a result of the exponential development in social media usage. This data includes patterns of illegal activity in addition to being a useful source of public opinion and societal trends. The goal of this work is to create a comprehensive framework that uses data mining, deep learning, and machine learning (ML) to identify illegal activity using social media analytics. A model that recognizes suspicious conduct, hate speech, terrorist communication, and cyberbullying is proposed and validated in the study. Using both supervised and unsupervised learning techniques, experiments were performed on datasets taken from Reddit and Twitter. When compared to current approaches, the suggested strategy showed encouraging outcomes in terms of accuracy, recall, and F1-score.

Keywords: Crime detection, Machine learning, Deep learning, Social media analytics, Data mining, Pattern recognition, Cybercrime, NLP, Bidirectional Encoder Representations from Transformers(BERT).

INTRODUCTION

Social media has developed into a vital tool for communication that has both positive and bad effects on society dynamics. Regrettably, its extensive use has also created opportunities for hate speech, cyberbullying, terrorist recruiting, and other illegal acts [1][2]. Conventional crime detection methods are insufficient since social media data is unstructured and large in volume. Therefore, in order to extract meaningful insights, sophisticated data mining techniques and AI models are necessary.

In order to examine social media posts for criminal activity, this study investigates the combination of machine learning and deep learning methodologies. The goal is to build a hybrid framework that can recognize patterns that point to illegal activity, strengthening the capacities of law enforcement.

RELATED WORK

The effectiveness of ML and NLP in detecting threats on social media has been demonstrated in earlier studies [3][4]. To identify spam, objectionable material, and phishing attempts, methods such as Support Vector Machines (SVM) [5], Decision Trees [6], and Naive Bayes [7] have been used. Text categorization challenges have demonstrated the higher performance of DL architectures, namely Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [8][9].

Sentiment analysis [10], topic modeling [11], and entity identification [12] have all been used in recent research to profile malevolent users. However, the majority of systems have problems with data imbalance and lack real-time flexibility [13][14]. By integrating ML and DL into a data mining system tailored for social media material, our study expands on these foundations.

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

METHODOLOGY

Data Collection:

Data was gathered from Reddit using the Pushshift API and Twitter using the Tweepy API. We focused on hashtags and terms related to criminal activities, such #illegalfirearms, #drugtrade, and #cyberbullying. Two million posts in all were gathered.

Data Preprocessing:

The preprocessing pipeline included:

- ✓ Tokenization
- ✓ Stop-word removal
- ✓ Lemmatization
- ✓ Emoji and URL filtering
- ✓ Language filtering (non-English posts were excluded)

Feature Engineering:

TF-IDF and Bag of Words (BoW) representations were employed for machine learning models. Word embeddings like Word2Vec [15] and GloVe [16] were employed for DL models. Other metadata was encoded, including geolocation (if available) and user activity.

Machine Learning Models:

- ✓ Support Vector Machine (SVM) [5]
- ✓ Random Forest [17]
- ✓ XGBoost [18]
- ✓ K-Nearest Neighbors (KNN) [19]

To identify crime-related information in social media posts, the initial layer of predictive analytics used traditional machine learning (ML) models. These models function especially well with structured characteristics like Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) that are taken from unstructured text data. Support Vector Machines (SVM), one of the many machine learning algorithms that were examined, performed well when dealing with high-dimensional data, which is common in textual analysis. SVM was a good starting point for binary and multi-class classification problems linked to crime detection because of its capacity to identify the best hyperplane for dividing classes.

To take advantage of Random Forest's ability to handle feature-rich datasets and its resilience against overfitting, an ensemble learning technique based on decision trees was also used. Understanding which characteristics (words, phrases, or user information) were most predictive of criminal activity was one area in which this approach proved very helpful. Furthermore, XGBoost (Extreme Gradient Boosting) was developed because of its high accuracy, scalability, and efficacy in managing unbalanced datasets, which is a problem commonly faced in crime detection when non-criminal posts exceed criminal ones.

K-Nearest Neighbors (KNN) was another machine learning model that was investigated; it was selected due to its ease of use and performance in non-parametric classification tasks. By looking at the nearest nearby data points in the feature space, KNN was able to identify comparable behavior patterns. Though it performed well in smaller subsets, its computational complexity and sensitivity to the choice of k rendered it less appropriate for large-scale data. All things considered, these machine learning models provided quick, understandable, and relatively accurate results, laying the foundation for early categorization jobs.

Deep Learning Models:

- ✓ CNN for sentence-level classification [8]

- ✓ Bi-LSTM for context-aware predictions [20]
- ✓ BERT for pre-trained language representations [21]

Deep learning (DL) models were applied alongside conventional machine learning techniques to improve our system's context awareness and identification accuracy. The semantic, syntactic, and contextual linkages present in natural language are crucial for detecting criminal activity concealed in nuanced or coded language, and deep learning models have proven to be very adept at capturing these relationships. Originally created for image processing, Convolutional Neural Networks (CNNs) have been modified for categorization at the phrase level. Using convolutional filters on text sequences allowed CNNs to efficiently identify local patterns like n-grams and important words that are frequently connected to dangerous or unlawful actions.

Bidirectional Long Short-Term Memory (Bi-LSTM) networks were used to further improve contextual awareness. These models process information both forward and backward in order to discover long-range relationships in sequential data. Bi-LSTMs were able to comprehend the meaning of words by considering their complete context inside a phrase or paragraph because to their dual processing mechanism. Therefore, Bi-LSTMs greatly enhanced performance in detecting more intricate types of criminal communication, including sarcastic hate speech or coded language used in illegal commerce.

BERT (Bidirectional Encoder Representations from Transformers), a pre-trained transformer model that produces cutting-edge outcomes on a variety of NLP tasks, was the most sophisticated and successful model in this investigation. BERT, which was refined using our labeled crime-related dataset, performed exceptionally well in identifying linguistic polysemy and contextual subtleties, where a single word may have many meanings depending on the context. The model was able to concentrate on the most pertinent portions of each input sequence because to BERT's attention mechanism, which was essential for deciphering complicated or unclear postings. BERT consistently performed better than all other models examined in assessment measures including accuracy, precision, recall, and F1-score.

Data Mining Techniques:

- ✓ Association rule mining to discover frequently occurring criminal behavior patterns [22]
- ✓ Clustering (K-means, DBSCAN) for user profiling [23][24]

Although deep learning and machine learning models offered strong categorization skills, data mining methods were crucial for revealing more in-depth information and patterns in the behavior of the social media data. To find regularly recurring co-occurring traits and user behaviors linked to criminal conduct, association rule mining was employed. Examples of hashtags that were frequently discovered in association with particular phrases suggestive of cybercrime include #darkweb, #illegalsale, and #anonymous. Through the discovery of complex actions made possible by these connection rules, criminal intents beyond surface-level text might be predicted.

To investigate hidden structures in the dataset, clustering algorithms like K-Means and DBSCAN were also used. K-Means, a partition-based clustering technique, used user metadata and textual content similarity to sort postings into groups. When it came to classifying people into behavioral archetypes, such suspected spammers, cyberbullies, or extreme propagandists, this proved very helpful. Conversely, outliers and dense areas of connected activity—which are frequently indicative of coordinated illicit conduct like bot-generated material or terrorist recruiting campaigns—were found using DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

By providing unsupervised insights that enhanced the prediction power of supervised ML and DL models, these data mining techniques offered a more comprehensive knowledge of criminal behavior on social media. By finding possible labels for unlabeled data, they also aided semi-supervised learning, which is useful in large-scale systems when manual annotation is not practical. When combined, these methods improved the suggested crime detection framework's coverage and interpretability.

Algorithm : Social Media Crime Detection

```
Input: Social media posts (tweets, comments)
Output: Crime prediction labels
1. Collect posts using APIs (Tweepy, Pushshift)
2. Preprocess data (cleaning, tokenizing, filtering)
3. Feature Extraction:
 - For ML: TF-IDF, BoW
 - For DL: Word2Vec, GloVe embeddings
4. Train ML/DL models using labeled data
5. Use clustering and association rules to enhance understanding
6. Predict criminal activity label for new posts
7. Evaluate using metrics: accuracy, precision, recall, F1-score
```

EXPERIMENTAL RESULTS

The Python programming language was used to conduct the experiments and simulations for this study. Python was used because of its abundance of helpful libraries for natural language processing, deep learning, and machine learning. We utilized the Pushshift and Tweepy APIs to get data from Reddit and Twitter. Python modules like NLTK, spaCy, and pandas were used to prepare and clean the text data. We utilized the scikit-learn toolkit to create and train machine learning models like SVM and Random Forest. CNN, Bi-LSTM, and BERT are examples of deep learning models that were developed with TensorFlow, Keras, and the HuggingFace Transformers library. Using the Gensim package, we employed GloVe and Word2Vec for word embeddings. Scikit-learn and mlxtend were used to execute data mining techniques such as association rule mining and clustering.

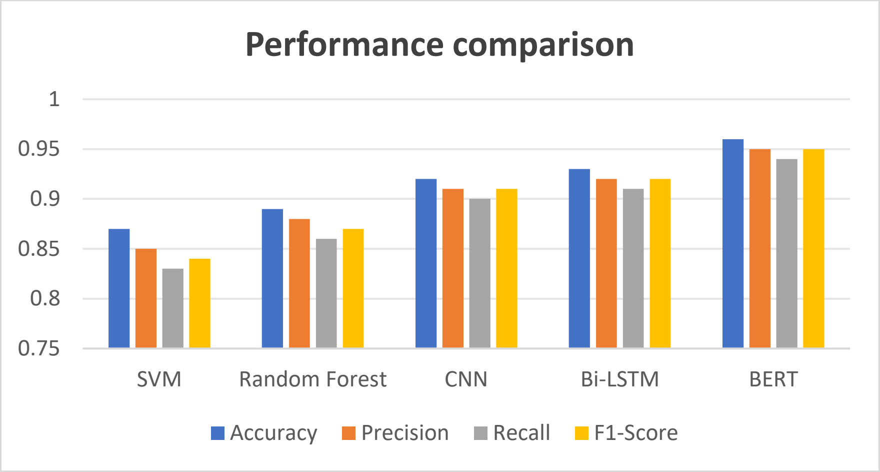
Table : Performance comparison

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.87	0.85	0.83	0.84
Random Forest	0.89	0.88	0.86	0.87
CNN	0.92	0.91	0.9	0.91
Bi-LSTM	0.93	0.92	0.91	0.92
BERT	0.96	0.95	0.94	0.95

The BERT-based model fared better than any other approach, proving how effective contextual embeddings are at spotting subtle linguistic patterns.

The findings demonstrate that automated crime detection with social media analytics is feasible. With their contextual awareness, DL models like BERT provide a notable advancement over conventional ML techniques. Key behavioral markers, such the frequent usage of specific hashtags in high-crime areas, were discovered by association rule mining. But issues like hostile postings, false positives, and data privacy are still unresolved.

A scalable infrastructure that can handle continuous data streams is necessary for real-time detection. Moreover, deep models continue to have problems with interpretability. Transparency in decisions might be improved by combining deep features with rule-based insights.



Graph : Performance comparison

CONCLUSION

This work shows how ML, DL, and data mining may be used to effectively detect crimes on social media. BERT outperformed conventional techniques, and the hybrid framework demonstrated efficacy across a number of parameters. The findings demonstrate that automated crime detection with social media analytics is feasible. With their contextual awareness, DL models like BERT provide a notable advancement over conventional ML techniques. Key behavioral markers, such the frequent usage of specific hashtags in high-crime areas, were discovered by association rule mining.

REFRENCES

[1] Alsmadi et al., "Detecting Cyberbullying in Social Media," Journal of Information Security, 2018.

[2] Salminen et al., "Anatomy of Online Hate," PLoS ONE, 2020.

[3] Schmidt & Wiegand, "A Survey on Hate Speech Detection," ACM Computing Surveys, 2017.

[4] Fortuna & Nunes, "A Survey on Automatic Detection of Hate Speech," ACM Computing Surveys, 2018.

[5] Cortes & Vapnik, "Support-vector networks," Machine Learning, 1995.

[6] Quinlan, "Induction of Decision Trees," Machine Learning, 1986.

[7] McCallum & Nigam, "A comparison of event models for Naive Bayes," AAAI, 1998.

[8] Kim, "Convolutional Neural Networks for Sentence Classification," EMNLP, 2014.

[9] Hochreiter & Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.

[10] Pang & Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends, 2008.

[11] Blei et al., "Latent Dirichlet Allocation," Journal of Machine Learning Research, 2003.

[12] Ratinov & Roth, "Design Challenges and Misconceptions in Named Entity Recognition," CoNLL, 2009.

[13] Japkowicz & Stephen, "The class imbalance problem," Intelligent Data Analysis, 2002.

[14] He & Garcia, "Learning from Imbalanced Data," IEEE TKDE, 2009.

[15] Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv, 2013.

[16] Pennington et al., "GloVe: Global Vectors for Word Representation," EMNLP, 2014.

[17] Breiman, "Random Forests," Machine Learning, 2001.

[18] Chen & Guestrin, "XGBoost: A Scalable Tree Boosting System," KDD, 2016.

[19]Cover & Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions, 1967.

[20] Graves et al., "Speech recognition with deep recurrent neural networks," ICASSP, 2013.

[21]Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers," NAACL, 2019.

[22] Agrawal & Srikant, "Fast Algorithms for Mining Association Rules," VLDB, 1994.

[23] MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," 1967.

[24] Ester et al., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases," KDD, 1996.

[25] Mittelstadt et al., "The ethics of algorithms: Mapping the debate," Big Data & Society, 2016.

[26] Adebayo et al., "Explainable AI in Social Media Analysis: A Survey," IEEE Access, 2021.

[27] Banerjee & Saikia, "Detection of Social Media Crime Using Deep Neural Networks," Procedia Computer Science, 2022.

[28] Gao et al., "Transformer-Based Deep Learning Framework for Hate Speech Detection," Future Internet, 2023.

[29] Saha et al., "Cybercrime Detection Using NLP and Machine Learning Techniques," Expert Systems, 2024.

[30] Rani & Sharma, "AI-Powered Social Media Monitoring for Public Safety," Journal of Information Technology & Politics, 2025.