Social Media Comment Classification and Crime Forecasting

Divya Udayan J[1], Ram Mohan Reddy S[2], Sravan[3], Sai Ganesh[4], Vidyacharan[5]

[12345]Department of Computer Science and Engineering, Amrita School of Computing,

[12345]Amrita Vishwa Vidyapeetham, *Amritapuri, India*

[2] srammohanreddy@am.students.amrita.edu : [3]dvssravan@am.students.amrita.edu,
[4]saiganesh@am.students.amrita.edu : [5]vidyacharanvarma@am.students.amrita.edu

Corresponding Author Mail-Id: divyaudayanj@am.amrita.edu[1]

***Abstract*—The increasing use of social media has led to a vast amount of user-generated content, including comments, which can provide valuable insights for crime forecasting. However, the sheer volume of data makes it difficult to manually analyze and categorize each comment. Once comments are classified, they can be used for crime forecasting. For example, patterns in social media comments related to criminal activity can be analyzed to identify potential hotspots for criminal activity. Additionally, social media comments can provide early warning signals for criminal activity, allowing law enforcement agencies to take preemptive action. Therefore, this paper proposes a social media comment classification model that uses machine learning techniques to identify and classify comments related to criminal activity. The proposed model combines supervised and unsupervised learning algorithms to classify comments into different categories, such as potential criminal activity, crime reporting, and general discussion. Additionally, the paper explores the potential of using the classified comments for crime forecasting by analyzing patterns and trends in the data. The results demonstrate the effectiveness of the proposed model in accurately classifying comments, and the potential for using social media data for crime forecasting. In summary, social media comment classification and crime forecasting is an emerging field of research that has the potential to revolutionize the way law enforcement agencies prevent and respond to criminal activity. By using machine learning algorithms to automatically classify social media comments, law enforcement agencies can gain valuable insights into criminal activity and take proactive measures to prevent it.**

Keywords—intrusion detection, crime, forecasting, machine learning, social media, malignant

I. INTRODUCTION

Machine Learning is a technique that allows machines to learn from previous work without specific instructions [1,2,3]. It is not always possible to identify clear patterns or data after looking at the data [4,5,6]. In these cases, machine learning is used to identify patterns and messages [7, 8]. Machine learning (ML) supports the idea that by giving computers access to accurate data, machines can learn to solve specific problems and complex mathematical problems [9,10, 11,12].

Social media platforms have revolutionized the way people communicate and share information, with millions of users generating vast amounts of content every day. Among this content are comments that can provide valuable insights into criminal activity, which can be used for crime forecasting and prevention. However, the sheer volume of data makes it challenging for law enforcement agencies to manually analyze and categorize each comment. The rise of social media has transformed the way we communicate, share information, and connect with others. Social media platforms such as Twitter, Facebook, and Instagram have become a primary source of information for many people, with millions of users generating vast amounts of content every day. Among this content are comments that can provide valuable insights into criminal activity, which can be used for crime forecasting and prevention. However, the sheer volume of data makes it challenging for law enforcement agencies to manually analyze and categorize each comment. To overcome this challenge, researchers have proposed various machine learning algorithms to automatically analyze and classify social media comments related to criminal activity. The use of natural language processing (NLP) techniques allows for the analysis of the text of comments to identify and categorize comments related to criminal activity. By using machine learning algorithms to classify social media comments, law enforcement agencies can gain valuable insights into criminal activity, enabling them to take proactive measures to prevent it.

The study does not seek to create an entirely reliant on social media data real-time crime prediction system. The proposed framework's effectiveness will be assessed using actual social media data, and its performance will be compared to that of other social media analysis for crime prevention methods already in use. The framework's F1-score, memory, accuracy, and precision in categorizing social media comments and identifying potentially illegal activity will be the main areas of examination. The proposed system would categorize the comments into multiple groups using supervised learning techniques including decision trees, support vector.

II. Related Work

Social media is a powerful weapon in today's world for expressing one's emotions. They can instantly make someone popular or even damage someone's reputation [36]. In our digital age, security breaches brought on by malicious software (malware) attacks are on the rise and pose a serious security risk [37]. Large amounts of user-generated content are shared on Twitter every day, and figuring out how people feel about them can be helpful for people, businesses, governments, etc. There is no agreement on which Machine Learning strategy is best for a given application, despite years of research and investigation. Ensemble learners may be able to increase sentiment categorization accuracy, according to recent study [34]. A scalable solution is necessary because harmful attacks are occurring in extremely high rates and changing frequently [38]. Researchers initially concentrated more on subjectivity detection in sentiment analysis. Due to the internet's and the World Wide Web's explosive expansion over the last ten or so years, the web has replaced print media as the primary information source [33]. The increasing malware assaults and security threats are a major concern because numerous gadgets are connected to the Internet without any issues and a lot of data is being gathered. Despite the availability of several virus detection techniques, new techniques are needed to keep up with the size and complexity of such a data-intensive environment [35].

In ref[13], the article discusses the need for effective visual analytics tools and techniques to help researchers and practitioners better analyze and understand complex data sets. It argues that visual analytics, which combines human intuition with computational power, can help bridge this gap by allowing users to interactively explore and analyze data in a more intuitive and efficient manner. It also highlights several challenges and opportunities related to visual analytics, such as data preprocessing, visualization design, and machine learning techniques. Examples of how visual analytics has been applied in various domains are provided. The article proposes several solutions and approaches to address the challenges and opportunities related to visual analytics. These include integrating multiple data sources, preprocessing data, visualization design, and machine learning techniques. Visualization design is essential for creating intuitive and informative visualizations, and using design principles such as simplicity, clarity, and consistency is recommended. Machine learning techniques can be used to automate and improve the visual analytics process. Overall, the article emphasizes the importance of a multidisciplinary approach to visual analytics, involving experts in data science, design, and human-computer interaction.

As per ref[14], Artificial Intelligence/Machine Learning algorithms can be used to predict criminal activities and improve public safety, making crime prediction a challenging task for law enforcement agencies. Artificial Intelligence/Machine Learning approaches can be useful for crime prediction, but more high-quality studies are needed to improve reliability and effectiveness. Future studies should use more diverse datasets, evaluate models using more sophisticated metrics, and consider ethical implications.

The paper in ref [15] explores and reviews machine learning techniques as a potential solution to enhance the accuracy and efficiency of intrusion detection in IoT. The paper reviews machine learning techniques for intrusion detection in the IoT domain, addressing the limitations of traditional methods and exploring the potential of supervised and unsupervised learning approaches.

In ref. [16], the paper highlights the need for good data analytics to improve crime prevention using machine learning. It highlights the need for researchers to obtain crime data and help reduce crime in society by combining existing information and improving existing models.

Negative results from linear regression, low accuracy for Decision Tree and KNN algorithms, and computational problems with KNN Author suggests Future research will make use of neural networks, Manhattan distance rather than Euclidean distance, a random forest technique with AdaBoost for increased accuracy, and gradient boosting will improve results. Waikato Environment for Know conducted a comparison analysis between violent crime trends from the Communities and Crime Unnormalized Dataset and real crime statistical data. The linear regression method performed best due to its ability to manage some unpredictability in the test data [17]. Knowledge Flow, a new graphical user interface, can replace Internet Explorer and be used to predict crime in an urban environment [18]. Results of twice-weekly forecasts show that it is feasible to make precise predictions. Results can be improved by comparing fortnightly and monthly predictions with a separation between day and night [19]. Crime predictions using ML were examined in Vancouver, Canada, using data gathering, categorization, pattern recognition, forecasting, and visualization [20]. The crime dataset was further examined using boosted decision tree and K-nearest neighbor algorithms. The authors used a crime dataset from Chicago, United States to forecast crime using ML and data science methodologies. The most accurate model was KNN classification, with an accuracy of 0.787 [21]. This paper's goal is to provide how ML can be used to anticipate, identify, and solve crimes more quickly, leading to a decrease in crime. Ref. [22] suggests a feature-level data fusion approach based on a deep neural network (DNN) to accurately forecast the incidence of crime. ML techniques such as regression analysis, kernel density estimation (KDE), and SVM are used to forecast potential crimes based on where they had occurred in the past. A method was proposed to forecast crime using DT and KNN ML algorithms, with the random forest algorithm and adaptive boosting being used to improve accuracy. The accuracy was raised to 99.16% by combining it with under-sampling and oversampling techniques [23]. Ref. [24] provides an analysis of how to classify and forecast crimes using ML and deep learning architectures. Three basic deep learning configurations are presented: spatial and temporal patterns, temporal and spatial patterns, and spatial and temporal patterns simultaneously. Ref. [25] provides a large data and machine learning approach for behavior analysis and criminal prediction. RapidMiner was used to process past crime patterns, while ML-based naive Bayes algorithm can make better predictions using current datasets. Data mining and machine learning (ML) techniques used in criminological investigations are illustrated in ref. [26]. The authors constructed a variety of spatial-temporal variables based on 84 different types of geographic locations for a city in Taiwan and included the idea of a grid-based crime prediction model. The WEKA toolkit was used to create the J48 classifier, which was then trained using a preprocessed crime dataset. The J48 algorithm's testing findings demonstrated 94.25287% accuracy in predicting the unknown category of crime data.

There are various studies conducted on different forecasting methods. The authors of references [27, 28] and Shojaee et al. [28] used the KNNs algorithm to predict crime for the years 2014 and 2013, respectively. A decision tree technique was used in refs. [29, 30] to forecast crime for the years 2015 and 2013, respectively. Naive Bayes, a novel crime detection technique, was used for crime analysis and prediction in references [31, 32]. Although [31] were able to predict crimes with an astounding 87% accuracy, they were unable to use their method on datasets with a lot of features. Wibowo and Oesman [32] neglected to take into account the computing speed, resilience, and scalability and only managed to forecast crimes with an accuracy of 66%.

III. Proposed Methodology

We have implemented the classification and forecasting algorithms in various machine learning algorithms like Logistic Regression, Decision Tree Classifier, Random Forest Classifier, AdaBoost Classifier, K Nearest Neighbors Classifier and XGBoost Classifier. And we also implemented the same using a deep learning algorithm, LSTM, Long Short-Term Memory. From the research of analyzing the toxic comments classification and forecasting, it was said that deep learning algorithms will perform better than the normal machine learning algorithms.
Grid search CV (Cross-Validation), technique used in machine learning to tune the hyperparameters of a model by exhaustively searching over a range of parameter values. For some machine learning algorithms, using Grid search cv makes the algorithm to take less time than the usual algorithm. After the implementation, LSTM performs very well than the normal machine learning algorithms. And in the machine learning algorithms, XGBoost performs well with a comparatively good accuracy score. LSTM performs very well than the normal machine learning algorithms. Recurrent neural network (RNN) architectures such as LSTM are made to address the vanishing gradient issue that plagues conventional RNNs. LSTM incorporates a memory cell that can selectively forget or store information over time, unlike typical RNNs which may struggle to retain long-term dependencies in sequential data, enabling it to learn and recall longer-term patterns. This is accomplished by using gates, which have the ability to regulate the flow of information into and out of the cell. LSTM has grown to be a popular option for a variety of tasks, including time-series forecasting, speech recognition, and natural language processing, among others. It is especially well suited for jobs that include sequences with complex temporal links because of its capacity to manage long-term dependencies.

The following processes are included in the proposed method for categorizing social media comments and forecasting the crime:
Data gathering: Twitter, Facebook, and Instagram are just a few examples of the social networking sites from which information is gathered in the first stage. User comments and postings on criminal behavior or public safety will be included in this material.
Preprocessing data entails removing noise, duplicate information, and superfluous data from the collected data. This stage involves the use of methods like data purification, tokenization, stemming, and lemmatization.
Feature Extraction: Feature extraction is the process of transforming the preprocessed data into a feature space that can be used for analysis. In this stage, methods like bag-of-words, TF-IDF, or word embeddings are employed.
Comment classification: The next step is to classify the comments according to their substance into several groups. In this step, methods including named entity recognition, topic modelling, and sentiment analysis are applied.
Forecasting the likelihood that a crime will occur at a specific location or at a specific time is possible using the categorized remarks. In this step, methods such as time series analysis, geographic analysis, or machine learning techniques are applied.
Reporting and Visualization: Examples of tools that can be used to represent and communicate the results of the study include charts, graphs, and maps. This step allows the stakeholders to gain additional knowledge about the patterns and trends of crime in a certain area.

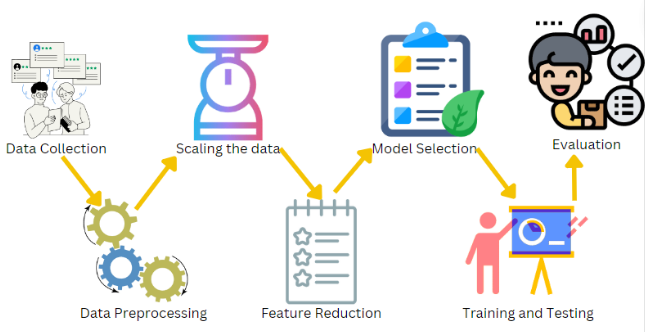

Fig. 1. *Proposed Methodology*

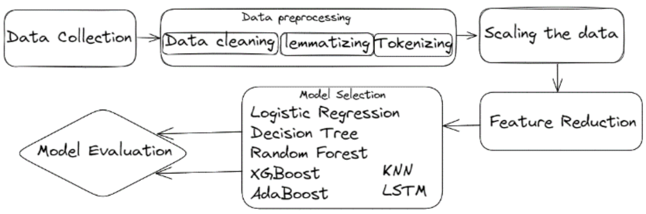

Fig. 2. *Pipeline of the proposed methodology*

Operating LSTM:
New inputs and hidden state — Before sending copies of the input through various gates, the input at the current timestep (x_t) and the hidden state from the previous timestep (h_t-1) are merged. **Forget gate --** Information that should be forgotten is determined by the forget gate. The sigmoid function, which has a range of 0 to 1, determines whether

values in the cell state should be completely forgotten (multiplied by 1), partially remembered (multiplied by any value between 0 and 1), or discarded (multiplied by 0). The **input gate** aids in determining crucial components that should be added to the cell state. The cell state candidate multiplies the input gate results, and only the data that the input gate deems essential is added to the cell state. First, the previous cell state (c_t-1) is multiplied by the forget gate results before **updating the cell state**. The most recent cell state (c_t) is then obtained by adding fresh data from [input gate cell state candidate]. It's time to **update** the concealed state, which is the final step. The output gate values are multiplied by the most recent cell state (c_t), which is then passed via the tanh activation function.

h_{t-1} - hidden state at previous timestep t-1 (short-term memory)
c_{t-1} - cell state at previous timestep t-1 (long-term memory)
x_t - input vector at current timestep t
h_t - hidden state at current timestep t
c_t - cell state at current timestep t

Fig. 3. *Working of LSTM and its abbrevations*

The suggested system aims to provide law enforcement and public safety personnel with a tool for early crime detection and prevention. By looking at social media comments, the technology could collect insightful information on how individuals feel about crime and public safety. This information can be used to improve policing strategies and allocate resources more effectively.

IV. Results and Analysis

Beginning with univariate analysis, count plot was used to discover that the dataset was unbalanced, with more normal comments than negative ones (including malignant, highly malignant, rude, threat, abuse and loathe). Additionally, it was discovered using a distribution plot for comments' length that, after cleaning, the majority of comments' lengths reduce from a range of 0 to 1100 to 0 to 900. Further word cloud analysis revealed that malignant comments frequently contain terms like fuck, nigger, moron, hate, suck, etc. Words like ass, fuck, bitch, shit, die, suck, and faggot are used often in very highly-malignant comments. Words like nigger, ass, fuck, suck, bull shit and bitch are examples of rude remarks. Threatening comments often include phrases like die, must, kill, murder, etc. Comments that are abusive often utilize terms like idiot, nigger, obese, jew, and bitch. Loath comments that use terms like nigga, stupid, nigger, die, gay, cunt, etc.

Among all the algorithms, LSTM performed well on the dataset. Among the machine learning algorithms used, Random Forest Classifier performs good followed by K Nearest Neighbors classifier and AdaBoost Classifier. It has been proved that the deep learning algorithms perform better than the usual Machine Learning algorithms but training the data on other deep learning algorithms will strengthen the conclusion.

The models were trained using a dataset of 159572 rows of different comments. Following training, these models are compared for performance using a variety of criteria. Any model that is trained using data that is divided into testing and training typically produces four different outputs that may be used to assess performance.

• True Positive (TP): Indicates how many positive cases the model properly classified as positive.
• False Positive (FP): This statistic shows how many negative occurrences the model mistook for positive ones.
• True Negative (TN): This statistic shows how many negative occurrences the model accurately classified as negative.
• False Negative (FN): This indicator shows how many positive cases the model mistook for negative ones.

The metrics used to assess a machine learning model's performance are described below.
Accuracy: A prominent assessment metric in machine learning is accuracy, which quantifies how well a model performs in terms of accurately predicting the result of a classification problem. It is determined as the total number of predictions produced by the model divided by the number of right forecasts.

$$Accuracy = \frac{(TP + FN)}{TP + FN + TN + FP}$$

S. No	Algorithm	Accuracy
1	Logistic Regression	89.71
2	Decision Tree Classifier	87.32
3	Random Forest Classifier	91.78
4	AdaBoost Classifier	90.92
5	KNN Classifier	91.21
6	XGBoost Classifier	91.61
7	LSTM	99.45

Table. 1. *Accuracy Scores of Various Machine Learning Algorithms*

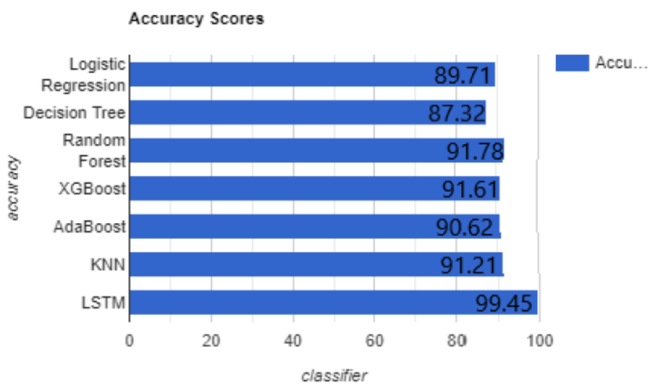

Fig. 4. *Accuracy Scores of Various Machine Learning Algorithms.*

Precision measures the percentage of accurate positive predictions among all the positive predictions generated by a model in machine learning. When the cost of a false positive is large, precision is helpful.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ \ Positive}$$

Recall is a machine learning metric that assesses the percentage of accurate positive predictions among all the real positive examples in the dataset. When the cost of a false negative is substantial, the recall is helpful.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

F1 Score: For binary classification tasks, machine learning uses the F1 score as a performance statistic. A higher number indicates greater performance. It is a method of integrating accuracy and recall into a single score that goes from 0 to 1.

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Monitoring the model's accuracy throughout several epochs is frequently done alongside the training and validation stages of machine learning. A whole pass of the entire training dataset is referred to as an epoch during the training stage. Insights into the model's learning process and possible problems like overfitting may be obtained by charting the accuracy attained on both the training and validation datasets over epochs.

In machine learning, measures like training and validation loss are frequently employed to assess model performance during training in addition to accuracy monitoring. The difference between the model's anticipated output and the actual labels in the training and validation sets is measured as loss. The model's learning progress, convergence, and possible problems like overfitting may be understood by charting the training and validation loss over epochs.

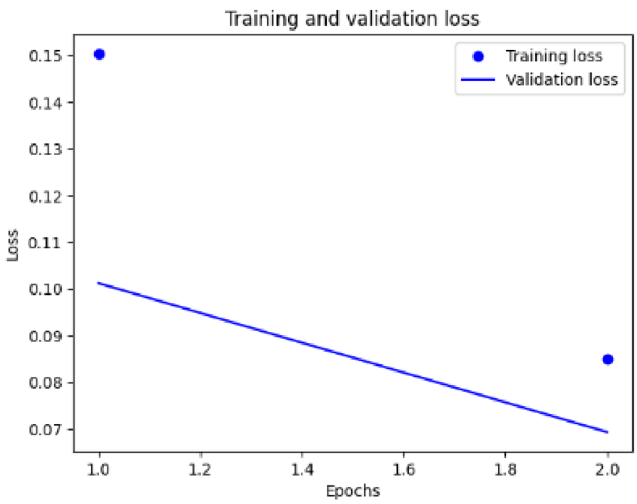

Fig. 5. *Training and Validation loss*

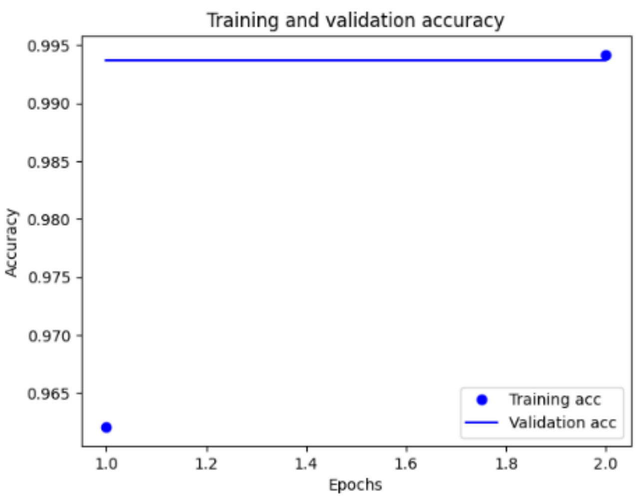

Fig. 6. *Training and Validation accuracy*

V. Conclusion and Future Work

A. *Conclusion*

Although the idea of predicting crimes before they occur is straightforward, putting it into practice requires much more. This piece of writing was meant to support academics working to use such cutting-edge technologies in the real world and make crime prediction a reality. Every few years, police deploy new technology like biometrics and face recognition, but using such software has the potential to profoundly alter how police function for the better. This research provided a framework for imagining how machine and deep learning components may contribute to the development of a system that is significantly more beneficial to the police. The artificial intelligence (AI) in our suggested system can do everything from track down criminal hotspots to identify persons based on voice notes. Making the system itself will be the initial challenge, followed by issues with its usage and implementation, among other things. To put it another way, imagine a scenario where we implement such a technology into a police force, achieving tips or leads that are much more dependable and potentially eradicating crime at a much faster rate.

Among all the algorithms, LSTM performed well on the dataset. Among the machine learning algorithms used, Random Forest Classifier performs good followed by K Nearest Neighbors classifier and AdaBoost Classifier.

B. *Future Work*

Several tactics may be used to improve the identification of poisonous remarks and crime predicting. Including a more inclusive and varied training sample is a key first step. To guarantee the efficacy of the categorization system, a wide variety of data must be incorporated because toxicity can manifest itself in different ways across cultures and points of view. Addressing concerns with prejudice and fairness is also crucial. Like any AI system, algorithms for classifying toxic comments are subject to biases that can perpetuate unfavorable stereotypes or unjustly target particular groups. To ensure justice and prevent escalating socioeconomic inequities, it is essential to address these issues. Another important improvement is the development of more complex toxicity categories. The existing binary distinction between statements that are harmful and those that are not ignores the many levels and expressions of toxicity. The accuracy of the classification system may be improved by adding more complex categories, allowing for more focused responses to combat harmful behavior. The identification of poisonous remarks and crime predicting can be considerably improved by putting these changes into practice.

VI. References

[1] Ahir K, Govani K, Gajera R, Shah M (2020) Application on virtual reality for enhanced education learning, military training and sports. Augment Hum Res 5(1):7.

[2] Talaviya T, Shah D, Patel N, Yagnik H, Shah M (2020) Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artif Intell Agric 4:58–73.

[3] Jha K, Doshi A, Patel P, Shah M (2019) A comprehensive review on automation in agriculture using artificial intelligence. Artif Intell Agric 2:1–12.

[4] Kakkad V, Patel M, Shah M (2019) Biometric authentication and image encryption for image security in cloud framework. Multiscale Multidiscip Model Exp Des 2(4):233–248.

[5] Pathan M, Patel N, Yagnik H, Shah M (2020) Artificial cognition for applications in smart agriculture: a comprehensive review. Artif Intell Agric 4:81–95.

[6] Pandya R, Nadiadwala S, Shah R, Shah M (2020) Buildout of methodology for meticulous diagnosis of K-complex in EEG for aiding the detection of Alzheimer's by artificial intelligence. Augment Hum Res 5(1):3.

[7] Dey A (2016) Machine learning algorithms: a review. Int J Comput Sci Inf Technol 7(3):1174–1179

[8] Sukhadia A, Upadhyay K, Gundeti M, Shah S, Shah M (2020) Optimization of smart traffic governance system using artificial intelligence. Augment Hum Res 5(1):13.

[9] Musumeci F, Rottondi C, Nag A, Macaluso I, Zibar D, Ruffini M et al (2019) An overview on application of machine learning techniques in optical networks. IEEE Commun Surv Tutorials 21(2):1381–1408.

[10] Patel D, Shah Y, Thakkar N, Shah K, Shah M (2020) Implementation of artificial intelligence techniques for cancer detection. Augment Hum Res 5(1):6.

[11] Kundalia K, Patel Y, Shah M (2020) Multi-label movie genre detection from a movie poster using knowledge transfer learning. Augment Hum Res 5(1):11.

[12] Marsland S (2015) Machine learning: an algorithmic perspective. CRC Press, Boca Raton, pp 1–452.

[13] Shah, N., Bhagat, N. & Shah, M. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Vis. Comput. Ind. Biomed. Art 4, 9 (2021).

[14] Fatima Dakalbab, Manar Abu Talib, Omnia Abu Waraga, Ali Bou Nassif, Sohail Abbas, Qassim Nasir, Artificial intelligence & crime prediction: A systematic literature review, Social Sciences & Humanities Open, Volume 6, Issue 1, 2022, 100342, ISSN 2590-2911

[15] W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[16] Jenga, K., Catal, C. & Kar, G. Machine learning in crime prediction. J Ambient Intell Human Comput 14, 2887–2913 (2023).

[17] McClendon L, Meghanathan N (2015) Using machine learning algorithms to analyze crime data. Mach Lear Appl Int J 2(1):1–12.

[18] Frank E, Hall M, Trigg L, Holmes G, Witten IH (2004) Data mining in bioinformatics using Weka. Bioinformatics 20(15):2479–2481.

[19] Rummens A, Hardyns W, Pauwels L (2017) The use of predictive analysis in spatiotemporal crime forecasting: building and testing a model in an urban context. Appl Geogr 86:255–261.

[20] Kim S, Joshi P, Kalsi PS, Taheri P (2018) Crime analysis through machine learning. In: Paper presented at the IEEE 9th annual information technology, electronics and mobile communication conference. IEEE, Vancouver 1-3 November 2018.

[21] Bharati A, Sarvanaguru RAK (2018) Crime prediction and analysis using machine learning. Int Res J Eng Technol 5(9):1037–1042

[22] Kang HW, Kang HB (2017) Prediction of crime occurrence from multi-modal data using deep learning. PLoS One 12(4):e0176244.

[23] Hossain S, Abtahee A, Kashem I, Hoque M, Sarker IH (2020) Crime prediction using spatio-temporal data. arXiv preprint arXiv:2003.09322.

[24] Stalidis P, Semertzidis T, Daras P (2018) Examining deep learning architectures for crime classification and prediction. arXiv preprint arXiv: 1812.00602. p. 1–13

[25] Jha P, Jha R, Sharma A (2019) Behavior analysis and crime prediction using big data and machine learning. Int J Recent Technol Eng 8(1):461–468

[26] Tyagi D, Sharma S (2018) An approach to crime data analysis: a systematic review. Int J Eng Technol Manag Res 5(2):67–74.

[27] Sun CC, Yao CL, Li X, Lee K (2014) Detecting crime types using classification algorithms. J Digit Inf Manag 12(8):321–327.

[28] Shojaee S, Mustapha A, Sidi F, Jabar MA (2013) A study on classification learning algorithms to predict crime status. Int J Digital Content Technol Appl 7(9):361–369

[29] Obuandike GN, Isah A, Alhasan J (2015) Analytical study of some selected classification algorithms in WEKA using real crime data. Int J Adv Res Artif Intell 4(12):44–48.

[30] Iqbal R, Murad MAA, Mustapha A, Panahy PHS, Khanahmadliravi N (2013) An experimental study of classification algorithms for crime prediction. Indian J Sci Technol 6(3):4219–4225.

[31] Jangra M, Kalsi S (2019) Crime analysis for multistate network using naive Bayes classifier. Int J Comput Sci Mob Comput 8(6):134–143

[32] Wibowo AH, Oesman TI (2020) The comparative analysis on the accuracy of k-NN, naive Bayes, and decision tree algorithms in predicting crimes and criminal actions in Sleman regency. J Phys Conf Ser 1450:012076.

[33] Jasmine Bhaskar, Sruthi K, Prema Nedungadi(2014) Enhanced Sentiment Analysis of Informal Textual Communication in Social Media By Considering Objective Words And Intensifiers, IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014).

[34] L. K. Devi, Subathra P., Kumar P. N., V., D. S. Ravi, and B.K., P.(2015) Tweet Sentiment Classification Using an Ensemble of Machine Learning Supervised Classifiers Employing Statistical Feature Selection Methods, Advances in Intelligent Systems and Computing, vol. 415. Springer Verlag, pp. 1-13.

[35] Venkatraman, S., Alazab, M., & Vinayakumar, R. (2019). A hybrid deep learning image-based analysis for effective malware detection. Journal of Information Security and Applications, 47, 377-389.

[36] Naveenkumar, K. S., Vinayakumar, R., & Soman, K. P. (2019). Amrita-CEN-SentiDB 1: Improved Twitter Dataset for Sentimental Analysis and Application of Deep learning. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT),pp. 1-5.

[37] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust intelligent malware detection using deep learning. IEEE Access, 7, 46717-46738.

[38] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., AlNemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. IEEE Access, 7, 41525-41550.