International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Predictive Analysis of Social Unrest in India: A Machine Learning Perspective on 2016-2023 Data Trends

Supekar Unmesh Mahesh[1]

School of Data Science, Symbiosis Skills and Professional University, Pune, Maharashtra, India
Email: *unmeshsupekar3[at]gmail.com*
https://orcid.org/0000-0001-8134-2755

Abstract: ***Riots and social unrest pose serious threats to public safety and societal stability. For efficient prevention and management, it is essential to comprehend the patterns and dynamics of such incidents. This study focuses on utilizing machine learning techniques to analyze, visualize, forecast, and classifyevents of social unrest in India from 2016 to 2023. The study makes use of a large dataset that includes a variety of information, such as geographic data, event characteristics, connected actors, and 28 different parameters. Two machine learning models are employed: a decision tree classifier and a random forest classifier for the classification and forecasting of social unrest and events. These models are trained and evaluated using the Armed Conflict Location and Event Data (ACLED)dataset, which is split into training and testing sets.***

Keywords: Social unrest; Riot prediction; Machine learning; Armed conflict analysis; Riots analysis India

1. Introduction

Riots and social unrest have become a constant problem in societies around the world, including in India. These events disrupt peace, create enmity, and pose serious risks to public security and social cohesion. Understanding the root causes, patterns and dynamics of social unrest is essential for prevention, mitigation, and policy making. The recent events in India, including the violence in Manipur, have highlighted the continued relevance and importance of studying and responding to social unrest[1]. Violence and communal clashes erupted in Manipur, a state in Northeast India, leading to the death of hundreds and the destruction of property. The clashes erupted over a land dispute, which was further exacerbated by underlying social and political tensions. The Manipur violence is a vivid reminder of the intricate dynamics and multi-faceted causes of social unrest, which necessitates a thorough understanding of the root causes and proactive measures to avoid similar events in the future [2].

Against this backdrop, the use of cutting-edge data analysis tools, such as machine learning, is an excellent way to investigate and forecast riots in India. The availability of vast datasets and computational capabilities enables researchers to delve into historical data, identify influential variables, and develop predictive models to anticipate the likelihood of riots. By harnessing the power of data-driven insights, policymakers, law enforcement agencies, and community leaders can make informed decisions, allocate resources effectively, and implement targeted interventions to mitigate the risk of riots and foster social cohesion [3]. Thus, this research aims to visualize, analyze, classify, and predict events of social unrest in India.

Statistics of the National Crime Records Bureau (NCRB) in its report 'Crime in India 2021, Volume-I' states that, 857 communal or religious riot cases were registered in 2020, 438 in 2019, 512 in 2018, 723 in 2017, and 869 in 2016. In addition to that, 51,606 rioting cases were registered in 2020, 45,985 in 2019, 57,828 in 2018, 58,880 in 2017, and 61,974 in 2016[4].

2. Related Work

Accurate and timely prediction of riots, events of social unrest, and armed conflicts is crucial for societal integrity. Hence it becomes imperative to understand the patterns and dynamics of such events. The authors of[5] attempted to predict disruptive events using Twitter data and proposed a unique temporal Term Frequency–Inverse Document Frequency (TFIDF) and benchmarked the Middle East 2015 and England Riots 2015 datasets[5]. Along similar lines, the authors of [6] have surveyedthe prediction of riots using various machine learning algorithms like K Nearest Neighbours, Support Vector Machine, and Decision Tree [6]. While having various alternatives for algorithms, authors of [7] presented a comparative study for the choice of data sources, methods, and approaches to predict various political instability taking into consideration the Armed Conflict Location and Event Data (ACLED), Uppsala Conflict Data Project Georeferenced Event Data (UCDP-GED) and others [7]. The authors of [8] in their paper discussed techniques to predict crime rates using machine learning and data mining techniques on text data. The data used for their research had 3 types of crimes namely felony, misdemeanor, and infraction and wobblers where the decision tree and random forest along with Gated recurrent unit (GRU) were used widely for the purpose[8]. The authors of [9] in their research have used Indian crime data to build regression models like simple linear regression, multiple linear regression, and support vector regression to predict crime under Indian Penal Code with a vision to help the Police and law enforcement bodies [9]. Authors of [10] have used Real crime data for their analytical study and have used WEKA for building the J48 Decision tree classifier which contributed to 59.15% of accuracy with a very quick training time of 0.76 seconds[10]. Using Deep learning techniques for the prediction of future terrorist attacks was the aim of the researchers [11] for which they utilized the Global Terrorism database containing data from 1970 to 2018 and 34 different data dimensions. In their study, they achieved a maximum accuracy of 94.8% using Deep Neural networks [11]. The research [12] used google trends to detect and forecast protests using data from 2017 to 2019 providing a general approach for the research in the domain of riots and protest prediction [12]. The authors of [13] have used random forest, K nearest neighbors, bagging boosting, support vector classifiers, and others to predict armed conflicts with an average precision of 0.65 for 250 using 250 trees in the algorithm and a training time of 13.24 seconds[13]. Taking into consideration of all the related work this paper proposes the utilization of Decision Trees and Random Forest algorithms for the analysis, classification, and prediction of events in the Armed Conflict Location and Event Data (ACLED).

3. Methodology

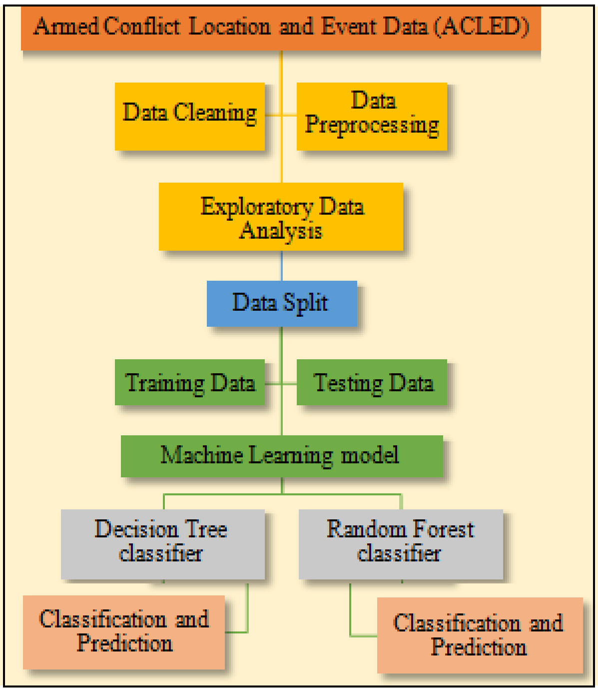

Figure 1: Diagram of the proposed methodology

The paper uses the Armed Conflict Location & Event Data (ACLED) for analyzing and predicting the events. Data used in the analysis range from the year 2016 to 2023. The analysis of armed conflict data using the ACLED dataset follows a sequential process. Firstly, the data is carefully cleaned to eliminate inconsistencies and errors. Secondly, the preprocessed data undergoes necessary transformations to ensure compatibility with machine learning algorithms. Further, the Exploratory data analysis techniques are then applied to unveil underlying patterns and relationships within the dataset. The data is subsequently divided into separate training and testing subsets to facilitate model development and evaluation.

Finally, the Machine learning models, specifically the Decision Tree classifier and Random Forest classifier, are chosen to train on the labeled training data. These models learn from the patterns and relationships present in the training data, enabling them to make accurate predictions. To assess the models' performance and generalizability, the testing data is employed. By comparing the models' predictions with the actual labels in the testing data, their accuracy and effectiveness are measured. This comprehensive process flow ensures that armed conflict data is meticulously cleaned, transformed, and analyzed to extract valuable insights. By leveraging machine learning models, accurate predictions can be made regarding event types based on the ACLED dataset. The paper hence utilizes the Decision tree and Random forest classifiers to classify, analyze, and predict types of armed conflict and events into 6 categories namely 'Battles,' 'Strategic developments,' 'Riots,' 'Protests,' 'Violence against civilians,' and 'Explosions/Remote violence' to aide the policymakers, law enforcement agencies, and community leaders to make informed decisions, allocate resources effectively, and implement targeted interventions to mitigate the risk of riots and foster social cohesion[14].

3.1. Decision Tree classifier

Decision trees are powerful and intuitive classification models used extensively in machine learning for various applications, including the classification of complex datasets such as the Armed Conflict Location & Event Data (ACLED). Decision trees make decisions based on a hierarchical structure resembling a tree, where each internal node represents a feature or attribute, and each leaf node represents a class or outcome. In the context of ACLED classification, decision trees can be trained to analyze different features like location, type of conflict, actors involved, and other relevant variables. By recursively splitting the dataset based on these features, decision trees create a set of rules or conditions that guide the classification process. The final result is a tree-like structure that allows for straightforward interpretation and understanding of how the classification decisions are made. Decision trees are particularly useful in handling mixed and heterogeneous datasets, making them well-suited for the classification of various types of armed conflicts, including battles, strategic developments, riots, protests, violence against civilians, and explosions/remote violence[15].

3.2. Random Forest classifier

Random forest classifiers are a highly effective ensemble learning method used for the classification of Armed Conflict Location & Event Data (ACLED) into various categories, including battles, strategic developments, riots, protests, violence against civilians, and explosions/remote violence. Random forests harness the collective power of multiple decision trees to achieve accurate and reliable classification results. Each decision tree within the random forest is trained on a randomly sampled subset of the original dataset, ensuring diversity in the training process. At each split, a random subset of features is considered, further enhancing the robustness and generalization capability of the model. By aggregating the predictions of individual decision trees, such as through majority voting, the random forest produces a final classification output[16]. This ensemble approach mitigates the risk of overfitting and improves the

Paper ID: SR24102233348 DOI: https://dx.doi.org/10.21275/SR24102233348

model's ability to handle complex and high-dimensional datasets like ACLED. Additionally, random forests offer valuable insights into feature importance, enabling analysts to understand the most influential factors driving the classification outcomes. With their versatility, resilience to noise and outliers, and ability to handle missing values, random forest classifiers are a well-suited and effective solution for accurately classifying armed conflicts and identifying distinct conflict types within ACLED[16].

3.3. The Data

The data used for analysis is Armed Conflict Location & Event Data (ACLED) for the region of South Asia specific to India. The data is merged and combined to form a single data source with a time range for combined final data used is 01 January 2016 to 07 July 2023. The data has 31 different columns and 132966 distinct data points [14].

General	Event Type	Sub-Event Type
Violent events	Battles	*Armed clash*
		Government regains territory
		Non-state actor overtakes territory
	Explosions/Remote violence	*Chemical weapon*
		Air/drone strike
		Suicide bomb
		Shelling/artillery/missile attack
		Remote explosive/landmine/IED
		Grenade
	Violence against civilians	*Sexual violence*
		Attack
		Abduction/forced disappearance
Demonstrations	Protests	*Peaceful protest*
		Protest with intervention
		Excessive force against protesters
	Riots	*Violent demonstration*
		Mob violence
Non-violent actions	Strategic developments	*Agreement*
		Arrests
		Change to group/activity
		Disrupted weapons use
		Headquarters or base established
		Looting/property destruction
		Non-violent transfer of territory
		Other

Figure 2: Types of events in Armed conflict location event data

Fig. shows various classes of event types addressed in the ACLED and their subevent type. The paper proposes a solution for the analysis of ACLED and predicts the corresponding Event type for the inputs provided.

4. Results and Analysis

The results obtained after the analysis are stated below in two distinct sections for Decision Tree and Random Forest classifier respectively.

4.1. The Decision Tree classifier

The Decision Tree classifier was evaluated using different criteria (Entropy and Gini) and maximum depths (2, 4, and 8) to assess its classification performance on ACLED. The criterion represents the measure used to assess the quality of the splits in the decision tree, while the maximum depth restricts the depth of the tree.

Table 1: Performance of the Decision Tree Classifier

Decision Tree Classifier				
Criterion	Max depth	CPU time	Wall time	Accuracy
Entropy	2	1 m 15 s	1 m 23 s	0.9038
Entropy	4	1 m 20 s	1 m 26 s	0.9640
Entropy	8	1 m 14 s	1 m 28 s	0.9966
Gini	2	1 m 04 s	1 m 20 s	0.9038
Gini	4	0 m 59 s	1 m 07 s	0.9640
Gini	8	0 m 50 s	0 m 53 s	0.9921

The results demonstrated that, with the Entropy criterion, the Decision Tree classifier achieved an accuracy of 90.38% at a maximum depth of 2, 96.40% at a maximum depth of 4, and an impressive accuracy of 99.66% at a maximum depth of 8. Training and evaluation times varied, with the longest time recorded as 1 minute and 28 seconds for a maximum depth of 8. Similarly, when using the Gini criterion, the Decision Tree classifier yielded comparable accuracy results. It attained an accuracy of 90.38% at a maximum depth of 2, 96.40% at a maximum depth of 4, and a notable accuracy of 99.21% at a maximum depth of 8. The training and evaluation times were generally shorter for the Gini criterion, with the fastest being 50 seconds at a maximum depth of 8.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	0.00	0.00	0.00	0
2	1.00	1.00	1.00	10667
3	1.00	0.52	0.68	2661
4	0.00	0.00	0.00	0
5	0.00	0.00	0.00	0
micro avg	0.90	0.90	0.90	13328
macro avg	0.33	0.25	0.28	13328
weighted avg	1.00	0.90	0.94	13328
samples avg	0.90	0.90	0.90	13328

Figure 3: Decision tree classification report for depth 2

Overall, the Decision Tree classifier exhibited strong performance across different criteria and maximum depths, with higher depths generally leading to improved accuracy. The choice between the Entropy and Gini criteria had minimal impact on accuracy, although the Gini criterion offered faster training and evaluation times. These findings underscore the effectiveness of the Decision Tree classifier in accurately classifying the data and its sensitivity to the maximum depth parameter for the classification of ACLED.

```
---Decision Tree Results---
Prediction number: 10736
Actual target classname: Riots
Actual class: [0 0 1 0 0 0]
Predicted target classname: Riots
Predicted class: [0 0 1 0 0 0]
```

Figure 4: Decision Tree Prediction results

4.2. The Random Forest Classifier

The Random Forest classifier was assessed using different numbers of estimators (10, 20, 40, and 50) to evaluate its classification performance on ACLED. The number of estimators refers to the number of decision trees in the random forest ensemble.

Table 2: Performance of the Random Forest Classifier

Random Forest Classifier			
N_estimators	CPU time	Wall time	Accuracy
10	1 m 45 s	2 m 00 s	0.9915
20	1 m 24 s	1 m 27 s	0.9945
40	2 m 32 s	2 m 51 s	0.9949
50	2 m 48 s	2 m 52 s	0.9948

The results demonstrated consistently high accuracy for the Random Forest classifier across the different numbers of estimators. With 10 estimators, it achieved an accuracy of 99.15%. The training and evaluation times for this configuration were 1 minute and 45 seconds of CPU time and 2 minutes of wall time. Increasing the number of estimators to 20 improved accuracy, yielding a classification accuracy of 99.45%. The training and evaluation times decreased to 1 minute and 24 seconds of CPU time and 1 minute and 27 seconds of wall time. With 40 and 50 estimators, the accuracy remained consistently high, with both configurations achieving an accuracy of approximately 99.49% and 99.48% respectively. However, the training and evaluation times increased to 2 minutes and 32 seconds (CPU time) and 2 minutes and 51 seconds (wall time) for 40 estimators, and 2 minutes and 48 seconds (CPU time) and 2 minutes and 52 seconds (wall time) for 50 estimators.

	precision	recall	f1-score	support
0	0.98	1.00	0.99	433
1	0.60	1.00	0.75	75
2	1.00	1.00	1.00	10660
3	1.00	1.00	1.00	1381
4	0.93	1.00	0.97	296
5	0.95	0.99	0.97	379
micro avg	0.99	1.00	1.00	13224
macro avg	0.91	1.00	0.95	13224
weighted avg	0.99	1.00	1.00	13224
samples avg	0.99	0.99	0.99	13224

Figure 5: Random Forest classification report

Overall, the Random Forest classifier demonstrated strong classification performance on ACLED across different numbers of estimators. Increasing the number of estimators generally improved accuracy, although the gains were marginal. It is important to consider the computational resources and trade-offs, as higher numbers of estimators resulted in longer training and evaluation times on ACLED.

```
---Random Forest Results---
Prediction number: 8526
Actual target classname: Riots
Actual class: [0 0 1 0 0 0]
Predicted target classname: Riots
Predicted class: [0 0 1 0 0 0]
```

Figure 6: Random Forest Prediction results

5. Conclusion

Decision trees and random forest classifiers are powerful tools for the classification of Armed Conflict Location & Event Data (ACLED) into distinct categories such as battles, strategic developments, riots, protests, violence against civilians, and explosions/remote violence. Decision trees provide an intuitive and interpretable approach by creating a hierarchical structure of conditions that guide the classification process. On the other hand, random forests leverage the collective intelligence of multiple decision trees, improving accuracy, generalization, and robustness. With their ability to handle complex datasets, handle missing values, and identify important features, decision trees, and random forest classifiers are well-suited for analyzing, forecasting, and understanding armed conflicts in ACLED. These techniques contribute to a better understanding of the dynamics and patterns of conflicts, aiding in conflict prevention, management, and decision-making processes. By utilizing these classification methods, researchers, policymakers, and analysts can gain valuable insights into armed conflicts and contribute to fostering peace and stability in conflict-affected regions.

6. Future Scope

The real-time, accurate prediction and forecasting of Armed conflicts, Riots, Battles, and other events of social unrest is a wide domain to explore. Future research may include dimensions like but not limited to areas like usage of Time series, temporal analysis, and use of Natural language processing techniqueson the data to predict the events of social unrest. The research also possible to integrate multiple datasets to create a huge data source for the algorithms to train for multi-faceted, multi-dimensional dynamic data to address events and crises of unique dynamic characteristics.

References

[1] --Pprc Delhi, “A FACT SHEET ON COMMUNAL RIOTS IN INDIA.”

[2] “Unfolding the Manipur Riots | Economic and Political Weekly.” https://www.epw.in/journal/2023/19/comment/unfolding-manipur-riots.html (accessed Jul. 15, 2023).

[3] D. Braha, “Global Civil Unrest: Contagion, Self-Organization, and Prediction,” *PLoS One*, vol. 7, no. 10, Oct. 2012, doi: 10.1371/JOURNAL.PONE.0048596.

[4] “Crime in India 2021 Statistics Volume I National Crime Records Bureau.”

[5] N. Alsaedi, P. Burnap, and O. Rana, “Can we predict a riot? Disruptive event detection using Twitter,” *ACM Trans Internet Technol*, vol. 17, no. 2, Mar. 2017, doi: 10.1145/2996183.

[6] K. Deb, A. Konar, K. Das, and M. Sen, “A Short Survey on Riot Prediction.” [Online]. Available: https://hal.science/hal-03245310

[7] C. Raleigh, R. Kishi, and A. Linke, “Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices,” *Humanit Soc Sci Commun*, vol. 10, no. 1, Dec. 2023, doi: 10.1057/s41599-023-01559-4.

[8] R. M. Saeed and H. A. Abdulmohsin, “A study on predicting crime rates through machine learning and data mining using text,” *Journal of Intelligent Systems*, vol. 32, no. 1. De Gruyter Open Ltd, Jan. 01, 2023. doi: 10.1515/jisys-2022-0223.

[9] R. M. Aziz, P. Sharma, and A. Hussain, “Machine Learning Algorithms for Crime Prediction under Indian Penal Code,” *Annals of Data Science*, Jul. 2022, doi: 10.1007/s40745-022-00424-6.

[10] G. N. Obuandike, A. Isah, and J. Alhasan, “Analytical Study of Some Selected Classification Algorithms in WEKA Using Real Crime Data,” 2015. [Online]. Available: www.ijarai.thesai.org

[11] M. I. Uddin *et al.*, “Prediction of Future Terrorist Activities Using Deep Neural Networks,” *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/1373087.

[12] J. C. Timoneda and E. Wibbels, “Spikes and Variance: Using Google Trends to Detect and Forecast Protests,” *Political Analysis*, vol. 30, no. 1, pp. 1–18, Jan. 2022, doi: 10.1017/PAN.2021.7.

[13] V. Helle Andra-Stefania Negus Jakob Nyberg, “Improving armed conflict prediction using machine learning CivilingenjörsprogrammetCivilingenj¨Civilingenjörsprogrammetiinformationsteknologi Master Programme in Computer and Information Engineering.” [Online]. Available: www.it.uu.se

[14] “Armed Conflict Location & Event Data Project (ACLED) Codebook.”

[15] Y. Y. Song and Y. Lu, “Decision tree methods: applications for classification and prediction,” *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[16] F. Guzzardo, “Spatial Conflict Prediction with Machine Learning Conflict Vulnerability in the Sahel Region.”