IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Cloud Crime Insights Via Machine Learning: A Cloud-Native Predictive Analytics Framework For Crime Forecasting And Safety Zone Classification

Eswaraiah Billu and C. Jyothsna
PG Scholar, Dept; of CSE, CREC (Autonomous), Tirupati
Associate Professor, Chadalawada Ramanamma Engineering College (Autonomous), Tirupati

Abstract: *Crime prediction and spatial safety analysis have become critical areas in modern smart policing and public safety management. Traditional crime analysis systems rely heavily on manual interpretation, fragmented databases, and reactive approaches, which often result in delayed decision-making and inefficient resource allocation. This paper presents a cloud-native predictive analytics framework titled "Cloud Crime Insights via Machine Learning," designed to forecast crime trends and classify geographical safety zones using machine learning techniques and cloud infrastructure. The proposed system integrates K-Means Clustering, Linear Regression, and Random Forest Classifier algorithms to analyze historical crime datasets and generate predictive insights. The system is deployed on Amazon Web Services (AWS) infrastructure using EC2 and S3 services, while Flask serves as the backend framework for handling prediction requests and visualization generation. Interactive dashboards and Plotly visualizations provide intuitive analysis for authorized users. Experimental analysis demonstrates improved scalability, proactive forecasting capability, and efficient crime hotspot classification.*

Keywords: Crime Prediction, Machine Learning, K-Means Clustering, Random Forest, AWS Cloud

I. INTRODUCTION

The rapid increase in crime incidents and the growing complexity of criminal activities have created significant challenges for law enforcement agencies worldwide. Traditional crime analysis systems largely depend on manual investigation, localized databases, and historical reviews, making it difficult to identify emerging crime patterns efficiently. These systems are often reactive rather than proactive, causing delays in resource deployment and strategic planning. Machine Learning (ML) and cloud computing technologies offer promising solutions to these limitations by enabling large-scale data processing, predictive forecasting, and automated classification of crime-proneregions. Predictive policing systems can analyze historical datasets, detect spatial and temporal patterns, and forecast future crime risks with higher accuracy. This paper proposes a cloud-native crime analytics platform called "Cloud Crime Insights via Machine Learning," which integrates multiple machine learning models with scalable AWS cloud infrastructure. The framework utilizes K-Means Clustering for spatial zone classification, Linear Regression for crime forecasting, and Random Forest Classifier for multidimensional risk assessment. Several researchers have explored the use of machine learning and data mining techniques in crime prediction systems. Traditional statistical approaches mainly focused on retrospective crime analysis rather than predictive intelligence. Recent research introduced machine learning algorithms such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Neural Networks for crime forecasting. These techniques improved predictive performance but often lacked cloud integration and real-time accessibility. Cloud-based predictive systems have recently gained popularity because of their scalability, distributed processing capabilities, and reduced infrastructure cost. AWS services such as EC2, S3, and Cognition enable secure and elastic deployment of analytical applications.

DOI: 10.48175/IJARSCT-35516

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

II. EXISTING SYSTEM

The existing crime analysis systems primarily rely on localized databases and manual processing methods. Crime records are typically stored independently within different police stations or regional departments, making unified analysis difficult.

Disadvantages of Existing System

- Manual and time-consuming workflows
- Fragmented geographical data storage
- Reactive rather than predictive analysis
- Poor scalability for large datasets
- Human bias in hotspot identification
- Limited real-time accessibility

III. PROPOSED SYSTEM

The proposed system, "Cloud Crime Insights via Machine Learning," is a comprehensive predictive analytics platform hosted entirely on AWS cloud infrastructure.

The framework integrates:

- Amazon EC2 for computation and hosting
- Amazon S3 for centralized data storage
- Flask backend for prediction APIs
- Machine learning models for forecasting and classification
- Plotly dashboards for interactive visualization

Advantages of Proposed System

- Proactive predictive intelligence
- Cloud scalability and elasticity
- Automated bias-free clustering
- Centralized secure data lake
- Interactive visual analytics
- Faster decision-making support

IV. SYSTEM ARCHITECTURE

The architecture of the proposed system consists of multiple interconnected modules that work together to perform data processing, prediction, visualization, and cloud deployment.

Workflow

1. Data collection from NCRB and Kaggle datasets
2. Data cleaning and preprocessing
3. Feature engineering and encoding
4. Model training and validation
5. Cloud deployment using AWS services
6. User interaction through Flask web interface
7. Real-time prediction and visualization

Data Preprocessing

Raw datasets often contain inconsistent values, missing records, and formatting issues. The preprocessing phase includes:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

- Null value handling
- Data cleaning
- Duplicate removal
- Label encoding
- Feature scaling
- Standardization

V. MACHINE LEARNING ALGORITHMS

The proposed framework uses multiple machine learning algorithms to perform prediction and classification tasks.

A. K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm used for spatial crime zone classification. Objective Function: J = sum of squared distances between data points and cluster centroids.The algorithm classifies regions into Red Zone, Orange Zone, and Green Zone categories.

B. Linear Regression

Linear Regression is used for temporal crime forecasting. Linear Equation: $y = mx + b$
The model predicts future crime rates based on historical population and crime statistics.

C. Random Forest Classifier

Random Forest is an ensemble learning algorithm used for multidimensional risk assessment. The classifier combines multiple decision trees to improve prediction accuracy and reduce over fitting.
The system is implemented using Python-based technologies and AWS cloud services.

Technologies Used

Frontend
- HTML
- CSS
- JavaScript
- Plotly.js

Backend

- Flask Framework
- Python
- Pandas
- Scikit-learn

Cloud Services

- Amazon EC2
- Amazon S3
- AWS Cognition

VI. CONCLUSION

This paper presented a cloud-native predictive analytics framework for crime forecasting and safety-zone classification using machine learning algorithms and AWS cloud infrastructure. The proposed system integrates K-Means Clustering, Linear Regression, and Random Forest Classifier to analyze historical crime datasets and generate proactive

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

intelligence. The framework successfully addresses the limitations of traditional crime analysis systems by providing scalable cloud deployment, automated classification, predictive forecasting, and interactive visualization.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science and Engineering, SV College of Engineering, Tirupati, for providing guidance and support throughout the development of this research work.

REFERENCES

[1]. Breiman, L., "Random Forests," Machine Learning Journal, vol. 45, no. 1, pp. 5–32, 2001.
[2]. MacQueen, J., "Some Methods for Classification and Analysis of Multivariate Observations," Proceedings of the Fifth Berkeley Symposium, 1967.
[3]. Han, J., Kamber, M., and Pei, J., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers. [4]. Géron, A., Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media.
[5]. AWS Documentation, Amazon Web Services Cloud Architecture Guidelines.
[6]. Pedregosa, F. et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research. [7]. National Crime Records Bureau (NCRB), Government of India Crime Statistics Reports.
[8]. Plotly Technologies Inc., "Interactive Data Visualization for Python."
[9]. Flask Documentation, Python Web Framework.
[10]. Kaggle Public Crime Datasets Repository