A

MAJOR

PROJECT ON

COMPARISON OF MACHINE LEARNING ALGORITHMS FOR PREDICTING CRIME HOTSPOTS

(Submitted in partial fulfilment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

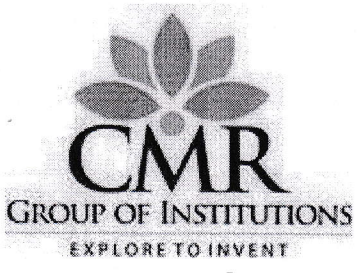
BY

Vodnala Srinidhi	207R1A05C0
Vodnala Saikumar	207R1A05B9
G. Prasanna	207R1A0581

Under the Guidance of

Dr. K. Srujan Raju

(HOD & Professor)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE, New Delhi)
Recognized Under Section 2(f) & 12(B) of the UGC Act.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2020-2024

ABSTRACT:

Crime prediction is of great significance to the formulation of policing strategies and the implementation of crime prevention and control. Machine learning is the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of public property crime from 2015 to 2018 from a section of a large coastal city in the southeast of China as research data to assess the predictive power between several machine learning algorithms.

Results based on the historical crime data alone suggest that the LSTM model outperformed KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks. In addition, the built environment data of points of interests (POIs) and urban road network density are input into LSTM model as covariates. It is found that the model with built environment covariates has better prediction effect compared with the original model that is based on historical crime data alone. Therefore, future crime prediction should take advantage of both historical crime data and covariates associated with criminological theories. Not all machine learning algorithms are equally effective in crime prediction..

EXISTINNG SYSTEM:

It involves collecting historical crime data, preprocessing and engineering features, and splitting the dataset. Various algorithms are chosen and trained on the training set, with hyperparameter tuning to optimize performance.

Evaluation metrics such as precision and recall are employed to assess predictive accuracy on the testing set. Results are compared, and strengths and weaknesses are analyzed, often utilizing visualizations like heat maps. The iterative refinement process may include adjusting features and exploring ensemble methods.

The final findings are documented in a comprehensive report, aiding stakeholders and decision-makers in understanding and utilizing the predictive models effectively.

Drawback of the Existing System:

- Oversimplification of Crime Patterns: The system may oversimplify crime patterns due to the inherent complexity of real-world data.
- Static Environment Assumption: Relying solely on historical data assumes a static environment, neglecting changes in socio-economic factors or law enforcement strategies.
- Limited Evaluation Metrics and Interpretability: The system's reliance on a limited set of evaluation metrics, coupled with challenges in model interpretability, may hinder its adaptability to dynamic crime scenarios.

PROPOSED SYSTEM:

The proposed system for comparing machine learning algorithms in predicting crime hotspots introduces advanced techniques to address existing limitations. It incorporates dynamic data sources, adapting to evolving socio-economic and law enforcement conditions for more accurate predictions. The system employs a broader set of evaluation metrics to comprehensively assess model performance across diverse crime scenarios. Additionally, a focus on enhancing model interpretability ensures transparency for stakeholders, contributing to more informed decision-making in crime prevention and intervention strategies.

Advantages of Proposed System:

- **Adaptability to Dynamic Conditions:** The proposed system integrates dynamic data sources, enabling it to adapt to changing socio-economic and law enforcement conditions, thereby enhancing the accuracy of crime predictions in evolving scenarios.
- **Comprehensive Performance Evaluation:** By employing a broader set of evaluation metrics, the proposed system provides a more comprehensive assessment of model performance across diverse crime scenarios, improving the reliability and applicability of crime prediction models.

REQUIREMENT SPECIFICATION

Hardware Requirements:

For developing the application the following are the Hardware Requirements:

Processor : Pentinum - IV

RAM : 4 GB (min)

Hard Disk : 20 GB

Key Board : Standard Windows Keyboard

Mouse : Two or Three Button Mouse

Monitor : SVGA

Software Requirements:

For developing the application the following are the Software Requirements:

Operating system	:	Windows 7 Ultimate
Coding Language	:	Python
Front-End	:	Python
Back-End	:	Django-ORM
Designing	:	HTML,CSS, Javascript
Data Base	:	MySQL (Wamp Server)

CONCLUSION:

In conclusion, the proposed system for comparing machine learning algorithms in predicting crime hotspots offers a notable leap forward in addressing the limitations of existing methodologies. The incorporation of dynamic data sources allows the system to adapt to changing socio-economic and law enforcement conditions, enhancing the accuracy and relevance of crime predictions in dynamic environments. The comprehensive set of evaluation metrics ensures a thorough assessment of model performance across diverse crime scenarios, providing a more nuanced understanding of predictive capabilities. Moreover, the emphasis on model interpretability promotes transparency, crucial for gaining trust and facilitating informed decision-making among stakeholders. While these advancements hold promise for more effective crime prevention and intervention, ongoing refinement and validation are essential to ensure the system's resilience and reliability in the ever-evolving landscape of crime patterns and societal factors.