

CRIME ANALYSIS AND PREDICTION USING OPTIMIZED KNN ALGORITHM

Ashmin Angel A, Aishwarya T M, Srinidhi Iswarya B K, Devi R

Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India

Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India

Student, Dept of Computer Science Engineering, PEC College, Tamil Nadu, India

Associate Professor, Dept of Computer Science Engineering, PEC College,

Tamil Nadu, India

----- * * * -----

Abstract-- The objective of this project is to tackle a vital issue in the society - Crimes. Analyzing and examining of crimes happening in the world will give us a Broadview in understanding the crime regions and can be used to take necessary precautions to mitigate the crime rates. Identifying Crime patterns will allow us to tackle problems with unique approaches in specific crime category regions and improve more security measures in society. Current studies show the reason of increase in crime rates is more in areas that are economically backward. In few decades' property crime will be a target. The following approach involves predicting crimes classifying, pattern detection and visualization with effective tools and technologies. Use of past crime data trends helps us to correlate factors which might help understanding the future scope of crimes. In this day and age security is a perspective which is given higher need by all political and government worldwide and intending to decrease wrongdoing frequency. As information mining is the proper field to apply on high volume wrongdoing dataset and information

picked up from information mining approaches will be helpful and bolster police power. So In this paper wrongdoing investigation is finished by performing k-implies bunching on wrongdoing dataset utilizing quick digger apparatus.

1.INTRODUCTION

In present situation hoodlums are getting mechanically complex in carrying out wrongdoing and one test looked by knowledge and law requirement offices is trouble in breaking down enormous volume of information associated with wrongdoing and psychological militant exercises along these lines offices need to realize procedure to get criminal and stay ahead in the interminable race between the crooks and the law implementation. So suitable field need to picked to perform wrongdoing

examination and as information mining alludes to extricating or mining information from a lot of information, information mining is utilized here on high volume wrongdoing dataset and information picked up from information mining approaches is helpful and bolster police powers.

2.LITERATURE SURVEY

1. Machine Learning Based Anomaly Detection for Load Forecasting Under Cyberattacks

They used techniques like Anomaly detection, cyberattack, dynamic programming, load forecasting, machine learning. Accurate load forecasting can make both economic and reliability benefits for power system operators. However, the cyberattack on load forecasting may mislead operators to make unsuitable operational decisions for the electricity delivery. The predicted load data is first used to reconstruct the benchmark and scaling data by using the k-means clustering. The naive Bayes classification is then used to determine the specific attack template.

2. Process-induced deformation of L-shaped variable-stiffness composite structures during cure

They used techniques like Composite materials; L-shape; Variable

stiffness; Process-induced deformation; Spring-in. It is the parameterized L-shaped parts were modeled to investigate the effects of the radius of curvature, the angle, the laminate thickness and the length of flange on spring-in angle of the parts. In this paper the parameterized investigation on the process-induced spring-in of L-shaped variable-stiffness composites was presented.

3. Framework For Image Forgery Detection And Classification Using Machine Learning

They used algorithms like Artificial Neural Networks; GLCM features; Graphical User Interface; Machine Learning ; Support Vector Machine. Future work will continue to refine the methodology so that there are lesser loopholes in the analysis and will hopefully come up with a better method in future. The advancement in the field of science and technology, the introduction of various advance images editing tools are also surging up.

4. Cluster Analysis for Anomaly Detection in Accounting Data

An Audit Approach¹, In the future, this research can also be further improved by analyzing the sensitivities of different load forecasting methods with the load data tampered with cyber attacks.

5. Analysis of Counterfeit Currency Detection Techniques for Classification Model

Counterfeit currency is one of the threats which creates vice to nation's economy and hence impacts the growth worldwide. Producing forge currency or fabricating fake features in the currency considered to be a crime. Over the past few years many researchers have proposed various techniques to identify and detect forged currency. They used techniques like fake currency, classification model, statistical techniques, machine learning, Logistic regression etc. The effort is also made to analyze and compare the prediction and classification statistical technique i.e. logistic regression and LDA.

3.PROPOSED SYSTEM

In this we accept that wrongdoing information mining has a promising future for expanding the adequacy and proficiency of criminal and knowledge examination. Visual and instinctive lawbreaker and knowledge examination methods can be created for wrongdoing design. As we have applied bunching procedure of information digging for wrongdoing investigation, we can likewise perform different strategies of information mining, for example, arrangement. Additionally, we can perform investigation

on different dataset, for example, undertaking review dataset, neediness dataset, help viability dataset, and so forth.

SYSTEM ARCHITECTURE



4.MODULES,

DESIGN AND SPECIFICATION:

4.1.1.DATA PRE-PROCESSING:

DATA COLLECTION

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data we collect to develop practical artificial intelligence (AI) and machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand. Data for which you already know the target answer is called *labelled data*.

DATA PRE-PROCESSING

Organize your selected data by formatting, cleaning and sampling from it.

Three common data pre-processing steps are:

- **Formatting:** The data you have selected may not be in a format that is suitable for you to work with. The data may be in a relational database and you would like it in a flat file, or the data may be in a proprietary file format and you would like it in a relational database or a text file.
- **Cleaning:** Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data you believe you need to address the problem. These instances may need to be removed. Additionally, there may be sensitive information in some of the attributes and these attributes may need to be anonymized or removed from the data entirely.
- **Sampling:** There may be far more selected data available than you need to work with. More data can result in much longer running times for algorithms and larger computational and memory requirements. You can take a smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

4.1.2. FEATURE EXTRACTION:

Features selection is done which can be used to build the model. The attributes used for feature selection are Block, Location, District, Community area, like dates, crime description, day of week, X, Y, Location.

4.1.3. EVALUATION MODEL:

In this module we implement the algorithm which one give best accuracy value. Based on that applying the algorithm it should calculate and give the accuracy. Here we are applying the algorithm and get the result on the form of graphical format.

5.RESULT AND DISCUSSION:

First, we take wrongdoing dataset, Channel dataset as indicated by prerequisite and make new dataset which has ascribe as per investigation to be done. Open quick digger device and read exceed expectations record of wrongdoing dataset and apply "Supplant Missing worth administrator" on it and execute activity. Perform "Standardize administrator" on resultant dataset and execute activity. Perform k implies grouping on resultant dataset framed after standardization and execute activity. From plot perspective on result plot information among wrongdoings and get required

group. Examination should be possible on bunch framed.

6.CONCLUSION AND FUTURE ENHANCEMENT:

In the present study, ML models with machine learning algorithms (ensemble and simile), SVM, SVM-Random forest, SVM-stacking and Naive Bias, were designed and were implemented. Each predetermined factor was feed into a violent crime training dataset (murder, rape, robbery, etc.). We discovered a major conclusion after successfully training and validating simodels.

In the future, the whole model will be converted into an opensource library and connected to the crime site, allowing it to function at the highest level of expertise Function on a framework where the threshold for class crime rates can be set. Instead of a limited crime, the highest crime rate can be measured. A small grouping is used to assess the success of all of the criminal figures examined in the proposal. A Calculation can be measured on a sufficiently broad scale of local or cloud-based crime with heavy datasets will pay forecast of multi-label charging, expand more possibilities, and realistically increase our research

REFERENCES:

- [1] [1] M. Cahill and G. Mulligan, “Using geographically weighted regression to explore local crime patterns,” *Social Sci. Comput. Rev.*, vol. 25, no. 2, pp. 174–193, May 2007, doi: 10.1177/0894439307298925.
- [2] [2] J. M. Caplan, L. W. Kennedy, and J. Miller, “Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting,” *Justice Quart.*, vol. 28, no. 2, pp. 360–381, Apr. 2011, doi: 10.1080/07418825.2010.486037.
- [3] [3] A. Almeahmadi, Z. Joudaki, and R. Jalali, “Language usage on Twitter predicts crime rates,” in *Proc. 10th Int. Conf. Secur. Inf. Netw.*, Oct. 2017, pp. 307–310, doi: 10.1145/3136825.3136854.
- [4] [4] V. K. Borooah and N. Ireland, “Deprivation, violence, and conflict: An analysis of Naxalite activity in the districts of India,” *Int. J. Conf. Violence*, vol. 2, no. 2, pp. 317–333, 2008, doi: 10.4119/UNIBI/ijcv.42.
- [5] [5] A. Babakura, M. N. Sulaiman, and M. A. Yusuf, “Improved method of classification algorithms for crime prediction,” in *Proc. Int. Symp. Biometrics Secur. Technol. (ISBAST)*, Aug. 2014, pp. 250–255, doi: 10.1109/ISBAST.2014.7013130.
- [6] [6] P. Pławiak, M. Abdar, and U. R. Acharya, “Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring,” *Appl. Soft Comput.*,

- vol. 84, Nov. 2019, Art. no. 105740, doi: 10.1016/j.asoc.2019.105740.
- [7] [7] Z. Li, T. Zhang, Z. Yuan, Z. Wu, and Z. Du, “Spatio-temporal pattern analysis and prediction for urban crime,” in Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD), Aug. 2018, pp. 177–182, doi: 10.1109/CBD.2018.00040.
- [8] [8] A. Almaw and K. Kadam, “Survey paper on crime prediction using ensemble approach,” Int. J. Pure Appl. Math., vol. 118, no. 8, pp. 133–139, 2018. [Online]. Available: <https://internal-pdf://107.93.182.66/18.pdf%0Ahttp://www.ijpam.eu>
- [9] [9] T. B. Hyde, H. Dentz, S. A. Wang, H. E. Burchett, S. Mounier-Jack, and C. F. Mantel, “The impact of new vaccine introduction on immunization and health systems: A review of the published literature,” Vaccine, vol. 30, no. 45, pp. 6347–6358, 2015, doi: 10.1016/j.vaccine.2012.08.029
- [10] [10] S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma, and N. Yadav, “Crime pattern detection, analysis & prediction,” in Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Apr. 2017, pp. 225–230, doi: 10.1109/ICECA.2017.8203676.