International Journal on Applications in Information and Communication Engineering
Volume 9 : Issue 2 : April 2023, pp 53 – 56 www.aetsjournal.com ISSN (Online) : 2394 - 6237

CRIME RATE PREDICTION AND ANALYSIS USING K-MEANS CLUSTERING ALGORITHM

M. HEMALATHA , V.SELVI

***Abstract*—** The project **"CRIME RATE PREDICTION AND ANALYSIS USING K-MEANS CLUSTERING ALGORITHM "is** developed to Currently, Prevention is better that Cure. Preventing a crime from occurring is better than investigating what or how the crime had occurred. Just like vaccination is given to a child to prevent disease, in today's world with such higher crime rate and brutal crime happenings, it has become necessary to have a vaccination systems that prevents from crimes happening. Byvaccinating society against crime it refers to various methods such as educating peoples, creating awareness, increasing efficiency and proactive policing methods and other deterrent techniques. Crime incident prediction has depends mainly on the historical crime record and various geospatial and demographic information.

Data mining is a process of extracting information and discovering patterns from large amount of data. Crime Analysis can be a lucrative application of data mining. Data mining can play an important role in analyzing and predicting crimes using the data stored in repositories. Crime rate in India is increasing day by day, becoming a topic of major concern, hindering good governance in India. Due to awful growth in crime rate, it has become impossible to analyze those crime related data and detect crime patterns or predict future crimes by intelligence agencies or local law enforcement agencies. This project presents a detailed analysis of the various relevant crime patterns and statistical analysis of crime data. This study will be helpful to law enforcing agencies in making strategies and tactics to address crime and disorder

I. INTRODUCTION

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.

The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

II. OBJECT DETECTION- AN OVERVIEW

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

The discipline of machine learning employs various approaches to teach computers to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers

M. Hemalatha, Assistant Professor , Department of Computer Applications , Erode Sengunthar Engineering College (Autonomous), Perundurai , Erode.
(Email : hemasengunthar@gmail.com)

V.Selvi , PG Scholar , Department of Computer Applications, Erode Sengunthar Engineering College (Autonomous), Perundurai , Erode.
(Email : selvigowri1991@gmail.com)

of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST dataset of handwritten digits has often been used.

Machine learning Algorithms Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms

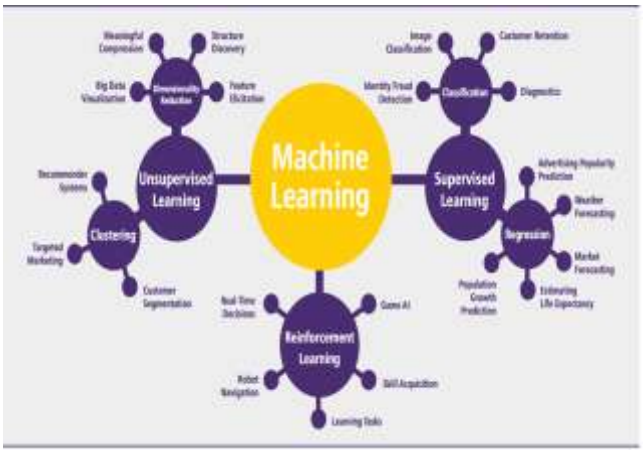

FIGURE 1:MACHINE LEARNING

III. LITERATURE SURVEY

A. ABSTRACT

Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customer, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected). When the model learned how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is a male, and 30% it is a female.

The label can be of two or more classes. The above Machine learning example has only two classes, but if a classifier needs to predict object, it has dozens of classes (e.g., glass, table, shoes, etc. each object represents a class)

IV. SYSTEM IMPLEMENTATION

A. MODULES:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

MODULES DESCSRIPTION:

1)DATA COLLECTION:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.

Comparison of Machine Learning Algorithms for Predicting Crime Hotspots taken from kaggle and some other source

2)DATASET:

The dataset consists of 821 individual data. There are 27 columns in the dataset, which are described below.

STATE:State in India
DISTRICT: District in the state of India.
Year: 2001-2018
MURDER**:** Total number of murder rate
RAPE**:** Total number of rape rate
THEFT: Total number of theft rate
Total crime: Total number of total crime rate

3)DATA PREPARATION:

we will transform the data. By getting rid of missing data and removing some columns. First we will create a list of column names that we want to keep or retain.

Next we drop or remove all columns except for the columns that we want to retain.

Finally we drop or remove the rows that have missing values from the data set.

4)MODEL SELECTION:

While creating a machine learning model, we need two dataset, one for training and other for testing. But now we have only one. So lets split this in two with a ratio of 80:20. We will also divide the dataframe into feature column and label column.

Here we imported train_test_split function of sklearn. Then use it to split the dataset. Also, *test_size = 0.2*, it makes the split with 80% as train dataset and 20% as test dataset.

Once the model is trained, we need to Test the model. For that we will pass *test_x* to the predict method.

5)ANALYZE AND PREDICTION:

In the actual dataset, we chose only 3 features :
STATE:State in India
DISTRICT: District in the state of India.
Year: 2001-2018

6) PREDICTION :

1. Total number of murder rate
2. Total number of rape rate
3. Total number of theft rate
4. Total number of total crime rate

7)ACCURACY ON TEST SET:

We got a accuracy of 95.1%,97.1%, 98.1%, 96.5%, on test set.

8)SAVING THE TRAINED MODEL:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle .

Make sure you have pickle installed in your environment.

Next, let's import the module and dump the model into .pkl file

V. CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with Good Accuracy. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding Indian crimes datasets that can help in capturing the factors that can help in keeping society safe.

VI. REFERENCES

[1] U. Thongsatapornwatana, ``A survey of data mining techniques for analyzingcrime patterns," in Proc. 2nd Asian Conf. Defence Technol. (ACDT),Jan. 2016, pp. 123128.

[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. 360381, Apr. 2011.

[3] M. Cahill and G. Mulligan, ``Using geographically weighted regressionto explore local crime patterns," Social Sci. Comput. Rev., vol. 25, no. 2,pp. 174193, May 2007.

[4] A. Almehmadi, Z. Joudaki, and R. Jalali, ``Language usage on Twitterpredicts crime rates," in Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), 2017,pp. 307310.

[5] H. Berestycki and J.-P. Nadal, ``Self-organised critical hot spots of criminalactivity," Eur. J. Appl. Math., vol. 21, nos. 45, pp. 371399, Oct. 2010.

[6] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, ``Bayesian networkmodeling of offender behavior for criminal proling," in Proc. 44thIEEE Conf. Decis. Control, Eur. Control Conf. (CDC-ECC), Dec. 2005,pp. 27022709.

[7] W. Gorr and R. Harries, ``Introduction to crime forecasting," Int. J. Fore-casting, vol. 19, no. 4, pp. 551555, Oct. 2003.

[8] W. H. Li, L.Wen, and Y. B. Chen, ``Application of improved GA-BP neuralnetwork model in property crime prediction," Geomatics Inf. Sci. WuhanUniv., vol. 42, no. 8, pp. 11101116, 2017.

[9] R. Haining, ``Mapping and analysing crime data: Lessons from researchand practice," Int. J. Geogr. Inf. Sci., vol. 16, no. 5, pp. 203507, 2002.

[10] S. Chainey, L. Tompson, and S. Uhlig, ``The utility of hotspot mapping forpredicting spatial patterns of crime," Secur. J., vol. 21, nos. 12, pp. 428,Feb. 2008.

[11] S. Chainey and J. Ratcliffe, ``GIS and crime mapping," Soc. Sci. Comput.Rev., vol. 25, no. 2, pp. 279282, 2005.

[12] L. Lin,W. J. Liu, andW.W. Liao, ``Comparison of random forest algorithmand space-time kernel density mapping for crime hotspot prediction,"Prog. Geogr., vol. 37, no. 6, pp. 761771, 2018.

[13] C. L. X. Liu, S. H. Zhou, and C. Jiang, ``Spatial heterogeneity of microspatialfactors' effects on street robberies: A case study

of DP Peninsula,"Geograph. Res., vol. 36, no. 12, pp. 24922504, 2017.

[14] M. I. Jordan and T. M. Mitchell, ``Machine learning: Trends, perspectives,and prospects," Science, vol. 349, no. 6245, pp. 255260, Jul. 2015.

[15] X. Zhao and J. Tang, ``Modeling temporal-spatial correlations for crimeprediction," in Proc. Int. Conf. Inf. Knowl. Manag. Proc., vol. F1318, 2017,pp. 497506.

[16] A. Babakura, M. N. Sulaiman, and M. A. Yusuf, ``Improved methodof classication algorithms for crime prediction," in Proc. Int. Symp.Biometrics Secur. Technol. (ISBAST), 2015, pp. 250255.

[17] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, ``Mixed spatial-temporal characteristicsbased crime hot spots prediction," in Proc. IEEE 20th Int.Conf. Comput. Supported Cooperat. Work Design (CSCWD), May 2016,pp. 97101.

[18] A. R. Dandekar and M. S. Nimbarte, ``Verication of family relationfrom parents and child facial images," in Proc. Int. Conf. Power, Autom.Commun. (INPAC), 2014, pp. 157162.

[19] G. R. Nitta, B. Y. Rao, T. Sravani, N. Ramakrishiah, and M. BalaAnand,``LASSO-based feature selection and Naïve Bayes classier for crimeprediction and its type," Serv. Oriented Comput. Appl., vol. 13, no. 3,pp. 187197, 2019.

[20] H. Tyralis and G. Papacharalampous, ``Variable selection in time seriesforecasting using random forests," Algorithms, vol. 10, no. 4, p. 114,Oct. 2017.

[21] K. K. Kandaswamy, K.-C. Chou, T. Martinetz, S. Möller, P. N. Suganthan,S. Sridharan, and G. Pugalenthi, ``AFP-pred: A random forest approach forpredicting antifreeze proteins from sequence-derived properties," J. Theor.Biol., vol. 270, no. 1, pp. 5662, Feb. 2011.

[22] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, andJ. P. Rigol-Sanchez, ``An assessment of the effectiveness of a random forestclassier for land-cover classication," ISPRS J. Photogramm. RemoteSens., vol. 67, pp. 93104, Jan. 2012.

[23] L. Lin, J. Jiakai, S. Guangwen, L. Weiwei, Y. Hongjie1, and L. Wenjuan,``Hotspot prediction of public property crime based on spatial differentiationof crime and built environment," J. Geo-Inf. Sci., vol. 21, no. 11,pp. 16551668, 2019.

[24] Z. Jun and H. Wenbo, ``Recent advances in Bayesian machine learning,"J. Comput. Res. Develop., vol. 52, no. 1, pp. 1626, 2015.

[25] J. T. Huang, J. Li, and Y. Gong, ``An analysis of convolutional neuralnetworks for speech recognition," in Proc. IEEE Int. Conf. Acoust., SpeechSignal Process. (ICASSP), South Brisbane, QLD, Australia, Apr. 2015,pp. 49894993.

[26] Z. Feiyan, J. Linpeng, and D. Jun, ``Review of convolutional neural network,"Chin. J. Comput., vol. 40, no. 6, pp. 12291251, 2017.

[27] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X.Wang, ``A CFCC-LSTMmodel for sea surface temperature prediction," IEEE Geosci. Remote Sens.Lett., vol. 15, no. 2, pp. 207211, Feb. 2018.

[28] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, ``PredictingAlzheimer's disease using LSTM," IEEE Access, vol. 7, pp. 8089380901,2019.

[29] L. Mou, P. Zhao, and Y. Chen, ``Short-term trafc ow prediction: A longshort-term memory model enhanced by temporal information," in Proc.19th COTA Int. Conf. Transp. Prof. CICTP Transp. China-Connect.World,2019, pp. 24112422.

[30] L. E. Cohen and M. Felson, ``Social change and crime rate trends:Aroutineactivity approach," Amer. Sociol. Rev., vol. 44, no. 4, p. 588, Aug. 1979.

[31] G. Gudjonsson, ``The reasoning criminal. Rational choice perspectives onoffending," Behav. Res. Therapy, vol. 26, no. 3, pp. 246287, 1988.

[32] P. Brantingham and P. Brantingham, ``Criminality of placeCrime generatorsand crime attractors," Eur. J. Crim. Policy Res., vol. 3, no. 3, pp. 526,1995.

[33] Enhancing Urban Safety and Security. Global Report on Human Settlements 2007, UN-Habitat, Nairobi, Kenya, 2007.

[34] G. Owusu, C. Wrigley-Asante, M. Oteng-Ababio, and A. Y. Owusu,``Crime prevention through environmental design (CPTED) and builtenvironmentalmanifestations in Accra and Kumasi, Ghana," Crime Prevention Community Saf., vol. 17, no. 4, pp. 249269, Nov. 2015.

[35] Y. Wenhao and A. Tinghua, ``The visualization and analysis of POI featuresunder network space supported by kernel density estimation," ActaGeodaetica et CartographicaSinica, vol. 44, no. 1, pp. 8290, 2015.

[36] G. Song, L. Xiao, S. Zhou, D. Long, S. Zhou, and K. Liu, ``Impact ofresidents' routine activities on the spatial-temporal pattern of theft fromperson," Acta Geography Sinica, vol. 72, no. 2, pp. 356367, 2017.

[37] L. Lin, D. Fang-Ye, X. Lu-Zi, S. Guang-Wen, and J. C. L. Kai, ``The densityof various road typesand larceny rate: An empirical analysis of ZG city,"Hum. Geeography, vol. 32, no. 6, pp. 3239, 2017.

[38] C. Xu, L. Liu, and S. H. Zhou, ``The comparison of predictive accuracy ofcrime hotspot density maps with the consideration of the near similarity:A case study of robberies at DP Peninsula," ScientiaGeographicaSinica,vol. 36, no. 1, pp. 5562,2016.

[39] G. Rosser, T. Davies, K. J. Bowers, S. D. Johnson, and T. Cheng, ``Predictivecrime mapping: Arbitrary grids or street networks," J. QuantumCriminol., vol. 33, no. 3, pp. 569594, 2017.

[40] D. Grifth, Multivariate Statistical Analysis for Geographers.Upper Saddle River, NJ, USA: Prentice-Hall, 1997.

[41] A. Rummens,W. Hardyns, and L. Pauwels, ``The use of predictive analysisin spatiotemporal crime forecasting: Building and testing a model in anurban context," Appl. Geography, vol. 86, pp. 255261, Sep. 2017.

[42] S. Favarin, ``This must be the place (to commit a crime). Testing the lawof crime concentration in Milan, Italy," Eur. J. Criminol., vol. 15, no. 6,pp. 702729, Nov. 2018.

[43] M. Felson and E. Poulsen, ``Simple indicators of crime by time of day,"Int. J. Forecasting, vol. 19, no. 4, pp. 595601, Oct. 2003.

[44] A. Sagovsky and S. D. Johnson, ``When does victimisation occur?" Australia. New Zealand J. Criminol., vol. 40, no. 1, pp. 216, 2007.