Pravin K, 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques

Pravin K, Dr. Naga Sundaram
Department of Computer Applications School of Computer Sciences,
Vels Instititute of Science and Advanced Studies (VISTAS),Pallavaram,Chennai

Abstract- Crime forecasting is one of the most wanted possible forecasts, as it could lead to fewer crimes and fewer police forces to secure threatened areas. However, predicting when and where crime will happen is challenging. Even modern predictive policing methods don't provide a reasonable or accurate approximation to forecast crimes. It has a long history of experiments and tests to reduce and to prevent crime. Deep Learning is a relatively new machine learning topic, which achieved state-of-the-art performance in many tasks and slowly but surely changes the machine learning area. For a few tasks, it is even better than the human himself.

Keywords: crime rate, number of crimes, regression algorithm, machine learning, Python, data classification, pattern identification, prediction, and visualization.

Keywords: Machine Learning, Crime predicitn.

I. INTRODUCTION

This success leads to the questions "Is deep learning able to forecast crime as ac- curate or even better as a human?" and "How to forecast crime using deep learning techniques?".

Therefore, this study applied different machine learning algorithms, namely, the logistic regression, support vector machine (SVM), k-nearest neighbors (KNN)time series analysis by long-short memory ,and RCNN (Regression Convolution Neural Network) Overall,these results provide early identification of crime, hot spots with higher crime rate, and future trends with improved predictive accuracy than with other methods and are useful for directing police practice and strategies.it compares a similardeep learning forecast architecture to the two proposed ones and shows how good orbad the two designs can handle unseen data. In the end, it concludes concerning the results and illustrates what is possible to enhance the methods.

II. LITERATURE SURVEY

Crime Prediction And Analysis Using Machine Learning By Alkesh Bharati And Dr SarvanagurU RA. K A have proposed a system that focus to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithm, using python as core we can predict the type of crime which will occur in a particular area. The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The K-Nearest Neighbor (KNN) classification and other algorithm will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in the country. This work helps the law enforcement agencies to predict and detect crimes in Chicago with improved accuracy and thus reduces the crime rate "CRIME

PREDICTION USING K- NEARESTNEIGHBOURING ALGORITHM" Many experts have worked on crime rate prediction and analysis using different methods like k-means clustering, KNN, naive bayes, Fuzzy c algorithm etc. Among all the methods Naive Bayes Algorithm was found to be one of the best methods for predicting the crimes which may happen in future. So, we proposed a system in which Bernoulli NB with linear regression is used. Bernoulli Naive Bayes is a variant of Naive Bayes. Naive Bayes classifier C.P. Chaithanya, N. Manohar, Ajay Bazil Issac, Crime Rate Predicition

It describes Text detection is the method of locating areas in a picture wherever, text is present. Text detection and classification in natural pictures is very important for several computer vision applications like optical character recognition, distinguish between human and machine inputs and spam removal. Currently the challenge in text identifying is to detect the text in natural pictures due to many factors like, low- quality image, unclear words, typical font, image having a lot of color stroke than the background color, blurred pictures due to some natural problems like rain, sunny, snow, etc. The main aim of this work is to identify and classify the text in natural pictures. Here system

III. PROPOSED SYSTEM

Here it provides the result as crime rate in some specific location along with the red zones and highly occurring crimes in that area. There is an additional feature which removes the fake crime reported by machine learning.

This work helps the law enforcement agencies to predict and detect crimes in improved accuracy and thus reduces the crime rat

• Creating a website that helps the police department to analyse and predict the crime rate of a particular area using Machine Learning technique.

• Help people travelling to different place to understand the crime trends of a particular area.

• Enable the common people to file a complaint through online portal as well as track their complaint.

IV. EXISTING SYSTEM

Some type of news such as various bad events from natural phenomenal or climate are unpredictable. When the unexpected events happen, there are also fake news that are broadcasted that creates confusion due to the nature of the events

Very few people know the real fact of the event while the most people believe the forwarded news from their credible friends or relatives. These are difficult to detect whether to believe or not when they receive the news information

Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data

PREDICTION AND FORECASTING:

Crime prediction and forecasting approaches have transformed dramatically in recent years since the introduction of

commercial software packages. Crime prediction refers to the accuracy of reported crimes in the past, whereas forecasting direct towards the future crime trends. However, a quick overview of criminal activities has been achieved by investigation authorities through the available software packages, whereas for deep analysis, only learning approaches may ensure the optimum solution. Therefore, different machine learning techniques can be used to predict crime patterns and thus may assist in further necessary actions based on historical data.

Therefore, this study is divided into two sections: crime prediction.crime forecasting.

Eight different machine learning algorithms are implemented to achieve highly accurate predictions in both the Chicago and Los Angeles datasets. The machine-learning algorithms implemented in this study were namely logistic regression, decision tree, random forest, MLP, Naïve Bayes, SVM, XGBoost, and KNN to get the crime prediction accuracy. Detailed information about these machine learning algorithmsmodels architecture is given in the supplementary information

(SI) and an experimental flow chart is given in Fig. 2. The prediction results further identify areas with high crime

density, all crime types and the crime rate over the past years.

Additionally, the statistical model ARIMA for time series analyses was applied to foresee future crime trends andanalytics. Crime forecasting based on time series data was also implemented in a later part of this study. A time-series analysis involves forecasting based on a sequence of events or data points that forms a series with respect to time . Research groups around the globe have recently used different approaches, including unsupervised models such as the bilinear model, the threshold autoregressive (tar) model, the autoregressive conditional heteroscedastic (ARCH) and deep learning approaches, to identify future trends Real-time crimeorecastingis always critical; especially in unknown circumstances; when and where the next crime will happen remains difficult to predict accurately. Therefore, we used an ARIMA model for future forecasting and calculated the RMSE to aggregate the magnitudes of the errors and crime predictions. The details of the ARIMA model are discussed in the SI. The forecasting results illustrate future crime trends by highlighting the crime hot spots, top five crimes and overall crime rates until 2024.

V. PREDICTION AND FORECASTING

Crime prediction and forecasting approaches have transformed dramatically in recent years since the introduction of commercial software packages. Crime prediction refers tothe accuracy of reported crimes in the past, whereas forecasting direct towards the future crime trends. However, a quick overview of criminal activities has been achieved by investigation authorities through the available software packages, whereas for deep analysis, only learning approaches may ensure the optimum solution. Therefore, different machine learning techniques can be used to predict crime patterns and thus may assist in further necessary actions based on historical data. Therefore, this study is divided into two sections: i) crime prediction and ii) crime forecasting.Eight different machine learning algorithms are implemented to achieve highly accurate predictions in both the Chicago and Los Angeles datasets. The machine-learning algorithms implemented in this study were namely logistic regression, decision tree, random forest, MLP, Naïve Bayes, SVM,
XGBoost, and KNN to get the crime prediction accuracy. Detailed information about these machine learning algorithms models architecture is given in the supplementary information(SI) and an experimental flow chart is given in Fig. 2. The prediction results further identify areas with high crime density, all crime types and the crime rate over the past years.
Additionally, the statistical model ARIMA for time series analyses was applied to foresee future crime trends and analytics. Crime forecasting based on time series data was also implemented in a later part of this study. A time-series analysis involves forecasting based on a sequence of events or data points that forms a series with respect to time. Research groups around the globe have recently used different approaches, including unsupervised models such as the

VI. RESULTS

The results and discussion part is divided into four sections based on methodology as shown in Fig. 1; predictive accuracy, time series analysis through LSTM, exploratory data analysis, and forecasting with an ARIMA model. The experimental results are also shown and discussed in each section. First, the predictive accuracy is discussed based on different algorithms.
In the second part, time series analysis was performedthorough LSTM to measure the performance of the model. Thereafter, crime particulars are thoroughly discussed in the exploratory data analysis section, and finally, crime forecasting and future crime trends are shown through the
ARIMA model. Different Python libraries were applied including Keras with Tensor Flow, Sk Learn, Pandas, Numpy, Seaburn, Scipy, and many others to generate the results.

A. PREDICTIVE ACCURACY

This study used different parameters to assess the performance of multiple algorithms, which better reflect the real dataset application. Eight different algorithms were applied to the Chicago and Los Angeles datasets to investigate the detailed

predictive accuracy of the trained models.. To the best of our knowledge, these algorithms have not been implemented together for Chicago and Los Angeles datasets. Consequently, the main reason to choose these cities is population density, which reported higher crime rates in the past with big data. The implemented algorithms have different methodologies to refine the data that involves supervised, unsupervised and reinforcement learning approaches. Additionally, Random Forest and XGBoost were also implemented which prompts an ensemble learning approach.

Decision Tree layout the significant decisions, while SVM and Naïve Bayes are used for better classification and KNN for advance regression. To handle dependent variables Logistic regression is implemented along with MLP which refers to the network of multiple layers of the perceptron. Since all these mathematical expressions help to seek improved accuracy to the best of their proficiency, with other performance metrics such as precision, recall and F1-score, as listed in Table 1. The accuracy estimates the proportion of instances that are correctly classified to obtain the optimum threshold for crime prediction. XGBoost performs better than other algorithms with 94% and 88% accuracy on both the Chicago and Los Angeles datasets, as multiple innovative algorithms work behind XGBoost. The Naïve Bayes, MLP (with hidden layer sizes of 24, 28, 30, and 34), and SVM algorithms also achieve a better performance on the Chicago dataset than on the Los Angeles dataset with maximum accuracy. The decision tree algorithm achieves an accuracy of approximately 66% (Chicago) and 60% (Los Angeles).

The MLP (87 and 84%) and KNN (88 and 89%) algorithms also approach the maximum accuracy on both datasets. The logistic regression model determines the statistical relationship between variables to achieve optimal results; here, it depicts consistent performance with 90% accuracy on the Chicago dataset and achieves below average results on the Los Angeles dataset. All these reported accuracy results are higher as compared with the literature.

VII.DISSCUSSION

Criminality is a phenomenon that occurs seemingly random and multiple research efforts have been made to develop rigorous and independent assessments. However, this study highlights the practical perspective of criminology by introducing predictive analysis through possible methods based on real-time data. Therefore, implementation of different machine learning algorithms were examined including LSTM and ARIMA modeling. First, the performance of different machine learning algorithms namely logistic regression, SVM, Naïve Bayes, KNN, decision tree, MLP, random forest and XGBoost were examined on datasets of Chicago and Los Angeles. The efficiency of prediction accuracy achieved by different algorithms is comparatively better than those reported earlier and suggests better performance. The performance of machine learning algorithms is more consistent for the Chicago dataset as compared with the Los Angeles dataset; where XGBoost achieves improved efficiency for prediction accuracy (around 94% and 88%) followed by KNN (around 88% and 89%) on both crime datasets.

VIII.CONCLUSION

Crimes are serious threats to human society, safety, and sustainable development and are thus meant to be controlled. Investigation authorities often demand computational predictions and predictive systems that improve crime analytics to further enhance the safety and security of cities and help to prevent crimes. Herein, we achieved an improved predictive accuracy for crimes by implementing different machine learning algorithms on Chicago and Los Angeles crime datasets. Among the different algorithms, XGBoost achieves the maximum accuracy on Chicago datasets and KNN achieves the maximum accuracy on Los Angeles. Data preprocessing was followed by splitting the dataset into training and testing sets, and later the performance parameters were examined. This study further applied a deep learning architecture for time series analysis through LSTM, by which the Chicago crime count had intense variations

compared with Los Angeles, as shown by the RMSE and MAE. Also, the exploratory data analysis exhibited extensive visualizations regarding crime particulars, including crime rates in different periods from daily to yearly trends, crime types, and highintensity areas based on historical patterns. Moreover, the implementation of an ARIMA model to predict the five-year trends regarding the crime rate and hot spots having high crime density suggest moderate variations for Chicago and a decline for Los Angles. For future work, this study will be expanded by using satellite imagery data, and the implementation of different learning techniques with corresponding visual data for different crime datasets.

APPENDIX

The machine-learning algorithms implemented in this study (Logistic Regression, SVM, Naïve Bayes, KNN, Decision Tree, MLP, Random Forest, XGBoost) LSTM and ARIMA models are detailed in SI.

ACKNOWLEDGMENT

Wajiha Safat acknowledges the financial support for M.S. study from COMSATS University, Islamabad. She especially thank Dr. Abdul Ghaffar (Institute of Metal Research, Chinese Academy of Sciences, Shenyang) for fruitful discussions.

COMPLIANCE WITH ETHICAL STANDARDS : Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] G. Mohler, Marked point process hotspot maps for homicide and gun crime prediction in Chicago, Int. J. Forecast. 2014, 30, 491–497.

[2] A. Iriberri, G. Leroy, Natural language processing and eGovernment: Extracting reusable crime report information, In: IEEE International Conference on Information Reuse and Integration, Las Vegas, IL, USA, August 13-15, 2007 pp. 221–226.

[3] V. Pinheiro, V. Furtado, T. Pequeno, D. Nogueira, Natural language processing based on semantic inferentialism for extracting crime information from text, In: IEEE International Conference on Intelligence and Security Informatics, Vancouver, BC, Canada, May 23-26, 2010, pp. 19–24.