CRIME DATA ANALYSIS USING MACHINE LEARNING TECHNIQUES

Dr.L.RASIDHAR, Principal, Adarsh College of Engineering, Chebrolu

SATYA NARAYANA REDDY AATLA, Project Manager, Despro Mines and MedTech Pvt.Ltd, Nuzvid

NIKITHA MARUDHURI, Project Manager, Despro Mines and MedTech Pvt.Ltd, Nuzvid

Abstract: In this paper, criminal cases in India are increasing rapidly due to which number of cases pending are also piling up. This continuous increase in the criminal cases is proving to be difficult to be classified and to be solved. Recognizing the criminal activity patterns of a place is important in order to prevent it from happening. The crime solving agencies can do a better work if they have a good idea of the pattern of criminal activities that are happening in a particular area. This can be done by using machine learning by employing different algorithms to find the patterns of the criminal activities in a particular area. This paper uses crime data set and predicts the types of crimes in a particular area which helps in speeding up the classification of criminal cases and proceed accordingly.

This paper uses the data of past 18 years that is collected from various trusted sources. Data pre-processing is as important as final prediction, this paper used feature selection, removing null values and label encoding to clean and nourish the data. This research gives an efficient machine leaning model for predicting the next criminal case.

INTRODUCTION

At present, the criminal cases that are pending in India are rapidly increasing with the number of crimes committed are increasing. To solve a case based upon a particular data there should be a thorough investigation and analysis that is to be done internally [1]. With the amount of crime data that is present in India currently the analysis and decision making of these criminal cases is too difficult for the officials. Identifying this a major problem this paper concentrates on creating a solution for the decision making of crime that is committed. Machine Learning is the branch of science where computers decide without human intervention. In recent times Machine Learning is being used in various domains one of the examples of such cases is automated or self-driving cars. By Machine Learning algorithms there is a way where we can predict certain results based upon our inputs given and provide a solution to solving crime cases in India. The two common types of prediction techniques are classification and

regression. This crime data prediction is a domain where classification is applied. Classification is a supervised prediction technique and it has been used in various domains like forecasting stock, medicinal area, etc. [2].The main aim of this paper is to consider some algorithms which can be used to predict and analyse the crime data and improve the accuracy of those models by data processing in order to obtain better results. The purpose is to train the required model to predict the data using the training data set by validation of the test data set [3]. The models which are being used here are Logistic Regression, Decision Tree classification, Random Forest classification.

MOTIVATION

To solve a case based upon a particular data there should be a thorough investigation and analysis that is to be done internally [1]. With the amount of crime data that is present in India currently the analysis and decision making of these criminal cases is too difficult for the officials. Identifying this a major problem this paper concentrates on creating a solution for the decision making of crime that is committed. The vehicle starts driving on its own. An autonomous driving vehicle performs various actions to arrive at its destination, repeating the steps of recognition, judgment and control on its own.

PROBLEM STATEMENT

Many researchers have gone through this problem regarding the criminal cases being unsolved for a long period. They proposed different crime prediction algorithms. In all these models the accuracy will surely vary depending on the data set and the features or attributes we select during data pre-processing. In Crime prediction done on the Mississippi crime data set where models like linear regression and Decision stump model are used gave a result of 83%, 88% and 67% respectively [1]. Although these accuracies of the predictions may vary accordingly because it is discovered that many machine learning algorithms are implemented on data sets consisting of different places having distinctive features, so predictions are changing in all cases

Drawbacks

- By this methodology they had less accuracy in prediction

Proposed System:

The proposed system is made on the basis of the research work that is done by going through varioussuch documentations. Nearly all of the crimes are predicting based on the location and the types ofcrimes that are occurring in those areas.On surveying previous works, Linear Regression, Decision Tree and Random Forest tend to give goodaccuracy so these models are used in this paper to predict crimes. The dataset used in this paper is fromdata.world.com. The data set

contains different types of crimes that being committed in India accordingto the state and year respectively [4]. This paper takes types of crimes as input and gives the area inwhich crimes are committed as output. The data pre-processing involves data cleaning, feature selection, dropping null values, data scaling by normalizing and standardizing. After data preprocessingthe data is free of null values which m ay alter the accuracy of the model significantly andfeature selection is used to select only the required features that won't affect the accuracy of model.After data pre-processing the models chosen i.e., Logistic Regression, Decision Tree and RandomForest are trained by splitting the data into as train and test data. As the output required is a categoricalvalue classification models are used here. Python language is used for the data prediction.

Advantages

We had a high accuracy in this model prediction methodology

LITERATURE SURVEY

1. McClendon, Lawrence, and NatarajanMeghanathan. "Using machine learning algorithms to analyze crime data." Machine Learning and Applications: An International Journal (MLAIJ) 2.1

Data mining and machine learning have become a vital part of crime detection and prevention. In this research, we use WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Un normalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. We implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the linear regression algorithm performed the best among the three selected algorithms. The scope of this project is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns.

2. AlkeshBharati, Dr Sarvanaguru RA. K," Crime Prediction and Analysis Using Machine Learning" in International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | September 2018.

The criminal cases in India are increasing rapidly due to which number of cases pending are also piling up. This continuous increase in the criminal cases is proving to be difficult to be classified and to be solved. Recognizing the criminal activity patterns of a place is important in order to prevent it from happening. The crime solving agencies can do a better work if they have a good idea of the

pattern of criminal activities that are happening in a particular area. This can be done by using machine learning by employing different algorithms to find the patterns of the criminal activities in a particular area. This paper uses crime data set and predicts the types of crimes in a particular area which helps in speeding up the classification of criminal cases and proceed accordingly. This paper uses the data of past 18 years that is collected from various trusted sources. Data pre-processing is as important as final prediction, this paper used feature selection, removing null values and label encoding to clean and nourish the data. This research gives an efficient machine leaning model for predicting the next criminal case. Various Machine learning models such as Logistic Regression, Decision Tree Classification, and Random Forest Classification were used to find the most efficient model to predict the type of crime at a particular location This paper discusses the about existing system which uses Knearest neighbour to predict next type of crime at a particular location, and also shows how the proposed system is better than the present existing system. This paper compares many machine learning models among themselves to find most efficient machine learning to tackle this problem. Keywords: criminal cases, Machine learning, Crime data, Algorithms, Data pre-processing, Logistic regression, Decision tree classification, Random forest classification. 1. Introduction At present, the criminal cases that are pending in India are rapidly increasing with the number of crimes committed are increasing. To solve a case based upon a particular data there should be a thorough investigation and analysis that is to be done internally [1]. With the amount of crime data that is present in India currently the analysis and decision making of these criminal cases is too difficult for the officials. Identifying this a major problem this paper concentrates on creating a solution for the decision making of crime that is committed. Machine Learning is the branch of science where computers decide without human intervention. In recent times Machine Learning is being used in various domains one of the examples of such cases is automated or self-driving cars. By Machine Learning algorithms there is a way where we can predict certain results based upon our inputs given and provide a solution to solving crime cases in India. The two common types of prediction techniques are classification and regression. This crime data prediction is a domain where classification is applied. Classification is a supervised prediction technique and it has been used in various domains like forecasting stock, medicinal area, etc. [2].The main aim of this paper is to consider some algorithms which can be used to predict and analyse the crime data and improve the accuracy of those models by data processing in order to obtain better results. The purpose is to train the required model to

predict the data using the training data set by validation of the test data set [3]. The models which are being used here are Logistic Regression, Decision Tree classification, Random Forest classification.

MODULE DESCRIPTION

1) **Data Acquisition:** Upload the URL data from the local host

2) **Data Preprocessing:** In this module, we will perform label encoding, convert the text data into token counts and quantify a word in documents, we generally compute a weight to each word which signifies the importance of the word in the document and corpus.

3) **Spliting:** In this module we will split the data into train and test data. x Train and y Train become data for the machine learning, capable to create a model.Once the model is created, input x Test and the output should be equal to y Test. The more closely the model output is to y Test: the more accurate the model is.

4) **Modelling:** in this module, we will apply the CNN-LSTM and CNN-BiLSTM on URL text and we will apply the ,machine learning algorithms on the features of URL.

5) **Compariosn**: Visualize the varies accuracy of modeling

About Dataset

Crime prediction:

The dataset used in our paper can be downloaded here:

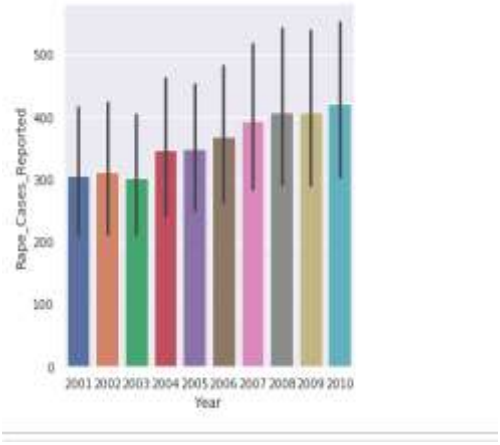
For crime prediction we take crime data in 2013.

Tha dataset consists ten column and 823 records.

SAMPLE RESULTS

Import required package and extracting the dataset

Data Processing



Year wise

CONCLUSION

It is clear that basic details of criminal activities in a neighbourhood contain indicators that will be employed by machine learning agents to classify a criminal activity given a location and date. The training agent suffers from imbalanced categories of the dataset, it had been ready to overcome the problem by oversampling and under-sampling the dataset. This paper presents a crime data prediction by taking the types of crimes as input and giving are in which these crimes are committed as output using Colab notebook having python as a core language and python provide inbuilt libraries such as Pandas and Numpy through which the work will be completed faster and Scikit provides all the processes of how to use different libraries providing by the python. Results of prediction are different for different algorithms and the accuracy of Random Forest Classifier found to be good with the accuracy of 95.122%.

REFERENCES:

1. McClendon, Lawrence, and NatarajanMeghanathan. "Using machine learning algorithms to analyze crime data." Machine Learning and Applications: An International Journal (MLAIJ) 2.1 (2015): 1-12.

2. AlkeshBharati, Dr Sarvanaguru RA. K," Crime Prediction and Analysis Using Machine Learning" in International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09| September 2018

3. McClendon, L., &Meghanathan, N. (2015). Using machine learning algorithms to analyze crime data. Machine Learning and Applications: An International Journal (MLAIJ), 2(1), 1-12.

4. Chen, Hsinchun, et al. "Crime data mining: a general framework and some examples." computer 37.4 (2004): 50-56.

5. Chen, H., Chungcdvsfdf, W., Xu, J. J., Wangsac, G., Qin, Y., &Chauascas, M. (2004). Crime data mining: a general

framework and some examples. computer, 37(4), 50-56.

6. Chen, Hsinchun, Wingyan Chung, Jennifer JieXu, Gang Wang, Yi Qin, and Michael Chau. "Crime data mining: a general framework and some examples." computer 37, no. 4 (2004): 50-56.

7. Chen, H., Chungsasda, W., Xu, J.J., Wang, G., Qin, Y. and Chau, M., 2004. Crime data mining: a general framework and some examples. computer, 37(4), pp.50-56.

8. Sathyadevan, Shiju. "Crime analysis and prediction using data mining." 2014 First International Conference on Networks & Soft Computing (ICNSC2014). IEEE, 2014.

9. Sathyadevan, S. (2014, August). Crime analysis and prediction using data mining.In 2014 First International Conference on Networks & Soft Computing (ICNSC2014) (pp. 406-412).IEEE.

10. Sathyadevan, Shiju. "Crime analysis and prediction using data mining."In 2014 First International Conference on Networks & Soft Computing (ICNSC2014), pp. 406-412.IEEE, 2014.

11. Sathyadevan, S., 2014, August. Crime analysis and prediction using data mining.In 2014 First International Conference on Networks & Soft Computing (ICNSC2014) (pp. 406-412).IEEE.

12. Sathyadevan S. Crime analysis and prediction using data mining. In2014 First International Conference on Networks & Soft Computing (ICNSC2014) 2014 Aug 19 (pp. 406-412).IEEE.

13. Nath, ShyamVaran. "Crime pattern detection using data mining."2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops.IEEE, 2006.

14. Nath, S. V. (2006, December). Crime pattern detection using data mining.In 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops (pp. 41-44).IEEE.

15. Nath, ShyamVaran. "Crime pattern detection using data mining."In 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, pp.41-44.IEEE, 2006.

16. Nath, S.V., 2006, December. Crime pattern detection using data mining.In 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops (pp. 41-44).IEEE.

17. Nath SV. Crime pattern detection using data mining. In2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops 2006 Dec 18 (pp. 41-44). IEEE.

18. Wang, Tong, et al. "Learning to detect patterns of crime." Joint European conference on machine learning and knowledge discovery in databases.Springer, Berlin, Heidelberg, 2013.