Crime Prediction and Analysis

*

Pratibha, Akanksha, Uprant, Suraina
(Final Year)
Department.of computer science and engineering
National Institute of Technology Hamirpur
Hamirpur,India

Dr. Lokesh Chouhan
(Assistant Professor)
Department.of computer science and engineering
National Institute of Technology Hamirpur
Hamirpur,India

***Abstract*—Crime is one of the dominant and alarming aspect of our society. Everyday huge number of crimes are committed, these frequent crimes have made the lives of common citizens restless. So, preventing the crime from occurring is a vital task. In the recent time, we have seen that artificial intelligence has shown its importance in almost all the field and crime prediction is one of them. But this requires keeping a track of almost all the crimes that has occurred and maintaining a proper database for the same which can be used for the future reference. The ability to predict the crime which can occur in future can help the law enforcement agencies in preventing the crime before it occurs. The capability to predict any crime on the basis of time, location and so on can help in providing useful information to law enforcement from strategical perspective. However, predicting the crime accurately is a challenging task because crimes are increasing at an alarming rate. Thus, the crime prediction and analysis methods are very important to detect the future crimes and reduce them. In Recent time, many researchers have conducted experiments to predict the crimes using various machine learning methods and particular inputs. For crime prediction, we used KNN problem, Decision trees and some other algorithms. For training, that algorithm is used which has better accuracy. The main purpose is to highlight the worth and effectiveness of machine learning in predicting violent crimes occurring in a particular region in such a way that it can be used by police to reduce crime rates in the society.**

***Index Terms*—Machine Learning, Crime Prediction, K-Nearest Neighbor, Decision trees.**

I. INTRODUCTION

Crime is increasing considerably day by day. Crime is among the main issues which is growing continuously in intensity and complexity[1]. Crime patterns are changing constantly because of which it is difficult to explain behaviours in crime patterns[2]. Crime is classified into various types like kidnapping, theft murder, rape etc. The law enforcement agencies collects the crime data information with the help of information technologies(IT). But occurrence of any crime is naturally unpredictable[3]. It is neither uniform nor random[4]. With rapid increase in crime number, analysis of crime is also required. Crime analysis basically consists of procedures and methods that aims at reducing crime risk. It is a practical approach to identify and analyse crime patterns. But, major challenge for law enforcement agencies is to analyse escalating number of crime data efficiently and accurately. So it becomes a difficult challenge for crime analysts to analyse such voluminous crime data without any computational support. A powerful system for predicting crimes is required in place of traditional crime analysis because traditional methods cannot be applied when crime data is high dimensional and complex queries are to be processed. Therefore a crime prediction and analysis tool were needed for identifying crime patterns effectively. This paper introduces some methodologies with the help of which it can be predicted that at what place and time which type of crime has a higher probability of occurrence. Classification helps in extracting features and predict future trends in crime data based on similarities. Methodologies used in this study are Extra Tree Classifier, MLP Classifier, K-Neighbour Classifier, Support Vector Machine (SVM), Decision Tree Classifier, XGB Classifier and Artificial Neural Network (ANN). The paper organisation is as follows. The introduction of the study of study is described in Section one. Section two consistes of the related works. Section III discusses the methodology for crime prediction methods. Section IV discusses its implementation. Section V consists of discussion. Section VI consists of Conclusion. Section VII discusses the future scope.

II. RELATED WORK

Many researches have been done which address this problem of reducing crime and many crime-predictions algorithms has been proposed.The prediction accuracy depends upon on type of data is used, type of attributes selected for prediction. In[5], mobile network activity was used to obtain human behavioural data which was taken from was used to predict the crime hotspot in London with an accuracy of about 70 percent when predicting that whether a specific area in London city will be a hotspot for crime or not. IN[6], data collected from various websites, newsletter was used prediction and classification of crime using Naive Bayes algorithm and decision trees and found that former performed better. In[7], a thorough study of various crime prediction method like Support Vector Machine(SVM), Artificial neural networks(ANN)

was done and concluded that there doesnot exist particular method which can solve different crime datasets problems. IN[8], various supervised learning techniques, unsupervised learning technique[9] on the crime records were done which address the connections between crime and crime pattern for the purpose of knowledge discovery which will help in increasing predictive accuracy of crime. In [10], different approach for predicting like Data mining technique, Deep learning technique, Crime cast technique, Sentimental analysis technique were discussed and it was found that every method have some cons and pros. Every method gives better result for a particular instance. Clustering approaches were used for detection of crime and classification method were used for the prediction of crime, [11]. The K-Means clustering was implemented and their performance is evaluated on the basis of accuracy. On comparing the performance of different clustering algorithm DBSCAN gave result with highest accuracy and KNN classification algorithm is used for crime prediction. Hence, this system helps law enforcement agencies for accurate and improved crime analysis.In [12], a comparison of classification algorithms, Naïve Bayes and decision tree was performed with an data mining software, WEKA. The datasets for this study was obtained from USCensus 1990. In [13], the pattern of road accidents in Ethiopia were studied after taking into consideration various factors like the driver, car, road conditions etc.Different classification algorithms used were K-Nearest Neighbour, Decision tree and Naive Bayes on a dataset containing around 18000 datapoints. The prediction accuracy for all three methods was between 79 to 81 percent.

III. METHODOLOGY

We used Predictive modeling for making predictions since it has the method which is able to build a model and has the capability to make predictions. This method consists of different algorithms of ML that can study properties from the data used for training which is used for producing predictions. It is split in two major classes one is Regression and other is classification of patterns. Regression models are based upon analysis of the relationship that are present between trends and variable in order to make predictions about the continuous variables. Whereas, the job of classification is to assign a particular class labels to a data value as output of the prediction. Division of pattern classification is in two ways i.e., Supervised and Unsupervised learning. It is already known in supervised learning that which class labels are to be used for building classification models. In unsupervised learning, these class labels are not known. Here, we worked with supervised learning.

A. Data collection and Pre-processing

Data collection is a process which information is gathered from many sources which is later used to develop the machine learning models. The data should be stored in a way that makes sense for problem.

Data pre-processing basically involves methods to remove the infinite or null values from data which might affect the performance of the model. In this step the data set is converted into the format which can be fed into machine learning models after removing the null values.

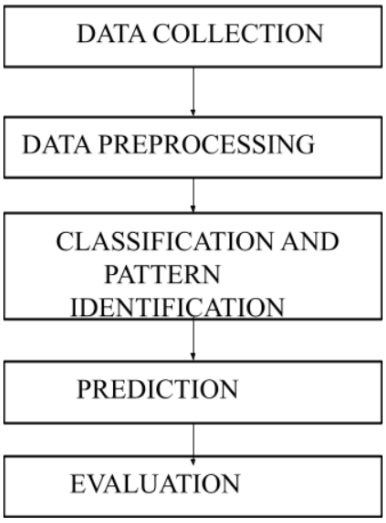

Fig. 1. Example of a figure caption.

B. Model selection

1) Support Vector Machine: Support Vector Machine performs well for regression, time prediction series and classification problems. Support vector machine performance can be measured against RNN. Thus, SVM had been applied in predicting hot-spots of crime [16] and predicting diseases like diabetic and pre-diabetic. Since it can make prototype of nonlinear relations in a coherent way. It performs well for anticipation of time series. For a predetermined degree of crime and data set it has to select a subset using K-clustering algorithm of crime data set and will determine a label for each data point in the set that is selected. Point where the crime rate is below given rate are called hotspots and where it is above given rate are called coldspots.

2) K-nearest Neighbour: Used for finding correlation between the test set and train set. If the given test set is close to the train set then it is assigned the class label of train set. The major limitation that emerges is when training set has less number of data points. To enhance it diverse techniques like K-NN algorithm has been used. This technique belongs to supervised learning domain. It finds its applications in data mining, intrusion detection and pattern recognition. In this the result is a membership of class. An object is categorized by neighbour's mass votes, where the object is being allocated to the most familiar of its k-nearest neighbours.

3) Decision trees: Decision trees are one of the most popular and powerful tool for classification and prediction. It has a structure like a tree, where all of the intermediate node represents a test on a peculiarity and the end product of test is denoted by every branch, and label of class are held by every

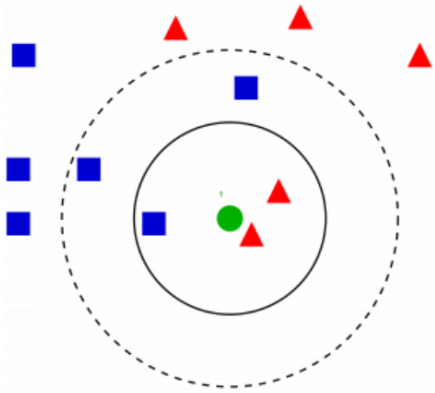

Fig. 2. Example of a KNN

leaf node. The target variable is generally categorical. Decision trees are used either for calculating the probability that a given record belongs to each category or to classify records (which is done by assigning records to the most similar class).

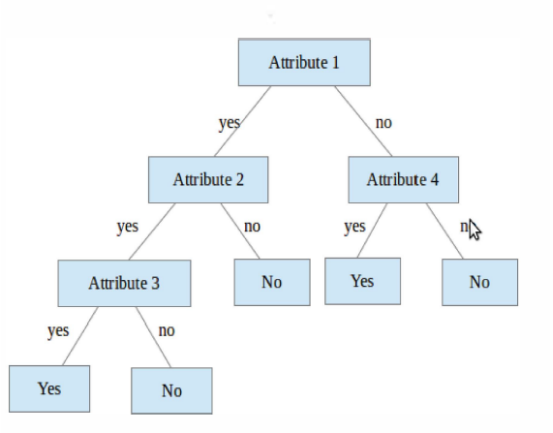

Fig. 3. Example of a decision tree

4) Extra Tree classifier: It is a type of technique which aggregates the results of many de-correlated decision trees collected in a forest to give its classification result. It is quite similar to a random forest classifier. It only differs in the structure of decision trees in the forest. Each of the decision tree in the extra tree forest has been constructed from the original training set .At each test node, each of the decision tree is given a random set of k-features from the universal feature set. From this random set each decision tree has to select the best feature to split the data based on some mathematical criteria. This leads to the creation of multiple de-correlated decision trees.

5) Artificial Neural Network: Artificial Neural network is developed after taking inspiration from biological neuron and try simulate decision making process of human brain. It comprises of huge number of constituents which work cooperatively to process and resolve problems. It is based on prediction by analysing trends in an already existing large amount of historical data. It has more general and flexible functions forms and can effectively deal with than traditional statistical methods. These were used to estimate the relation between inputs and outputs by adjusting the weights in every iteration. ANN can realize and study patterns for obtaining knowledge. It displays a link between an input neuron and an output neuron. Neurons have some specified weights. Output is calculated by multiplying the input with the specific neuron weight and then comparing it with the threshold value. If its above given threshold then it is contemplated as the output.

C. Training and testing

In this step, after validating the assumptions of the algorithm that we have chosen. Model is trained on the basis of given training Sample. After training, the performance of the model is checked on the basis of error and accuracy. At last, the trained model is tested with some unseen data and the model performance is checked on the basis of various performance parameters depending on the problem.

IV. IMPLEMENTATION

A. Data collection

The crime data set of San Francisco city is taken from kaggle.com in csv format which was derived from incidents derived from SFPD Crime Incident Reporting system. the attributes of the data set are Date,category of crime, description, DayofWeek, Police Department District, Address, Latitude and longitude and contained 884k data points.

B. Data Pre-processing

In this step all the null values are remove. The categorical attributes are converted into numeric using Label Encoder which can be understood by the classification models. There exists some samples which are considered to be outliners, those samples have been removed after checking location of each point and if not in San Francisco range then it was removed. There also exist two features Descript and Resolution are considered redundant as they does not exist in testing values so they were removed.

	Dates	Category	Descript	DayOfWeek	PdDistrict	Resolution	Address	X	
0	2015-05-13 23:53:00	WARRANTS	WARRANT ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.425892	3
1	2015-05-13 23:53:00	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	OAK ST / LAGUNA ST	-122.425892	3
2	2015-05-13 23:33:00	OTHER OFFENSES	TRAFFIC VIOLATION ARREST	Wednesday	NORTHERN	ARREST, BOOKED	VANNESS AV / GREENWICH ST	-122.424363	3

Fig. 4. Data before pre-processing

The data was imbalanced as it contained crime with different frequencies. So, to solve from this problem, two techniques were used one was to Replicate data having small frequency(lower than 1000) and second is to put more weights on the data having smaller frequency.

There are features from data that are not numbers, so they are converted into numbers so that we can train the models on them by using Label Encoding and One-hot Encoding. One-hot Encoding might produce very high number of dimensions due to lot of data labels in very feature but even then this is better because problem with the Label Encoding technique is that it assume higher the categorical value, better is the category which results in more errors.

	X	Y	Dates_int	Address_int	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
0	-122.425892	37.774599	7140	13297	0	0	0	0	0	0	1
1	-122.425892	37.774599	7140	13297	0	0	0	0	0	0	1
2	-122.424363	37.800414	7140	14853	0	0	0	0	0	0	1
3	-122.426995	37.800873	7140	3086	0	0	0	0	0	0	1
4	-122.438738	37.771541	7140	1255	0	0	0	0	0	0	1

Fig. 5. Data after pre-processing

C. Training models

For training the data splitted in the ratio of 80 percent for training and 20 percent for testing using sklearn library. As a result we train size of around 114000 data points and test size of around 28000 data points.

Here basically five models as mentioned in the methodology

```
model_KNN = KNeighborsClassifier(n_jobs=-1, weights='distance')
model_tree = DecisionTreeClassifier(class_weight='balanced')
model_extraTrees = ExtraTreesClassifier(n_jobs=-1, class_weight='balanced'
)
model_NN = MLPClassifier(learning_rate='invscaling', shuffle=True)
model_SVC = SVC(probability=True, class_weight='balanced') # One-to-One
```

Fig. 6. test time of different models

were used for traning purpose. After trying different combinations of parameters for each of the models different models were trained and their fscore, loggloss etc were calculated. For KNN the parameters used were njobs and weights and their values were set to minus one and balanced respectively. Similarly for Decision tree the parameter used was class weight and whose value was set to balance and similarly the parameter tuning was done for other models to get best output from each case possible. For MLP after trying different parameter tuning it was found that it worked best for 100 hidden layers and adam solver. Here as evident from the graph that SVC is taking lot of training time hence this cannot be considered. Where as if we look for MLP it has good value of fscore along with less logloss and optimal training time as compared to others. So, it is best to use neural network for this purpose.

D. Model evaluation and Metrics

For evaluating classification models that were implemented for the purpose of classification and prediction. It was found that call, accuracy, f1-score. Precision is a measure which identifies positive cases from all the predicted cases.

$$precision = \frac{TruePositive(TP)}{(Truepositive(TP) + FalsePositive(FP))}$$

Next is recall it measure which correctly identifies positive cases from all the actual positive cases.

$$recall = \frac{TruePositive}{(Truepositive + FalseNegative(FN)}$$

Accuracy is one of the most commonly used metric which measure all the correctly identified value without caring about the wrongly identified values. So, instead of using accuracy the measure that is used to check the performance is F-beta score.

$$accuracy = \frac{TP + TN}{(TP + FP + TN + FN)}$$

F1-Score is the harmonic mean of Recall and precision which gives a better measure of incorrectly classified cases than that of Accuracy Metric.

$$F1 - score = \frac{(2 * (recall * precision))}{(recall + precision)}$$

TABLE I
PERFORMANCE OF VARIOUS MODELS DURING TRAINING

Algorithm	Performance metrices	
	F-score	***logloss***
Decisiontress	0.99	0.005
KNN	0.99	0.004
Extratress	0.99f	0.005
ANN	0.20	0.66
SVM	0.39	3.5

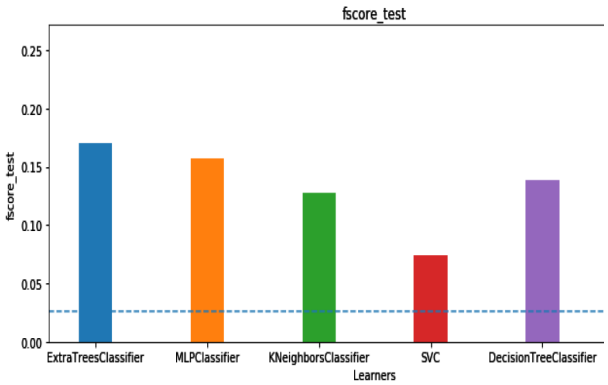

Fig. 7. Fscore for different models

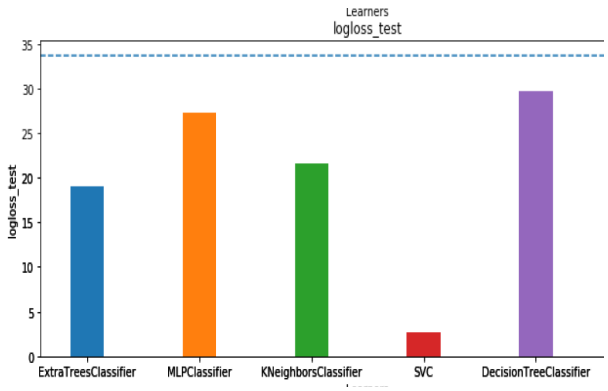

Fig. 8. logloss for different models

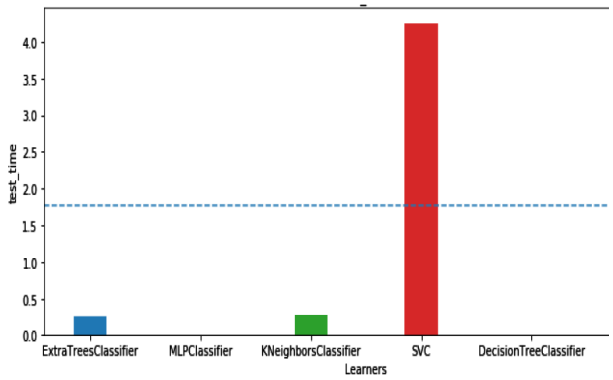

Fig. 9. test time of different models

E. Discussion

Crime solving is very difficult work which requires experience and intelligence of human along with Artificial intelligence approaches which assist them in problems of crime detection, as mentioned in [14]. Other than that predicting future patterns of crimes basically involves the changes in the rate of crime in next year and applied methods of prediction to help discover those changes in upcoming years.As proposed in [15], prediction of crime with the help of artificial neural network is usually better in accuracy and evaluate the target function much faster. Same is seen with the case of San Francisco data set in which neural networks performed better for testing data set than the other algorithm. As proposed by [16], the problem lies in finding out techniques which can analyse efficiently the growing data set of crime. The accuracy for predicting crime is basically depends upon on the crime data set used. If used training data set is very large, then model will be trained with very good accuracy while if the data set used for training purpose is having less size, then small degree of training is attained. Also, the prediction accuracy also dependent on dimension of training data set. This will give more right results if the model is highly trained and will not give good results if the model is not trained properly.

F. Conclusion

Crime prediction is one the current trends in the society. Crime prediction intends to reduce crime occurrences. It does this by predicting which type of crime may occur in future. Here, analysis of crime and prediction are performed with the help pf various approaches some of which are KNN, Artificial Neural network, Decision trees, Extra trees and Support vector machine. It was found that the best algorithm for predicting training data is Decision Tree and KNN but they did not performed well for testing data. On the other hand, for MLP it takes more time for training but provided good results for both training and testing data set. Every method have some cons and pros. Every method gives better result for a particular instance but here Artificial neural network performed best.

G. Future Scope

This work in future can be extended to have better classification algorithms which can detect criminals more accurately. We can also increase privacy and some other security measures to protect data set that we are using. Along with this, this work can be further extended to predict who will commit a crime and this can be done using Face recognition. The system will detect if there is any suspicious change in the behavior or usual movements. For example if a person is moving back and forth in same region over and over might indicate that he is a pickpocket and it will also track person over time.

ACKNOWLEDGMENT

We have immense pleasure in expressing our sincerest and deepest sense of gratitude towards our guide Dr. Lokesh Chouhan for the assistance, valuable guidance and co-operation in carrying out this project successfully. We also take this opportunity to thank Head of the Department Dr. Kamlesh Dutta for providing the required facilities in completing this project. We are greatly thankful to our parents, friends, and faculty members for their motivation, guidance and help whenever needed.

REFERENCES

[1] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, “Once upon a crime: towards crime prediction from demographics and mobile data”,IEEE,Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 427-434

[2] Ubon Thansatapornwatana, A Survey of Data Mining Techniques for-Analyzing Crime Patterns Second Asian Conference on Defense Technology ACDT, IEEE, 2016, ISBN: 978-1-5090-2258-8/16.

[3] J. L. LeBeau, ”The Methods and Measure of Centrography and the spatial Dynamics of Rape”Journal of Quantitative Criminology, Vol.3, No.2, pp.125-141, 1987

[4] J. L. LeBeau, ”The Methods and Measure of Centrography and the spatial Dynamics of Rape” Journal of Quantitative Criminology, Vol.3, No.2, pp.125-141, 1987.

[5] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, Alex Pentland.”Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data”, arXiv:1409.2983v1 [cs.CY] ,2014.

[6] Shiju Sathyadevan,Devan M. S.,Surya S Gangadharan, First,”Crime Analysis and Prediction Using Data Mining” International Conference on Networks Soft Computing (ICNSC), 2014

[7] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav,International Conference on Electronics, Communication and Aerospace Technology,ICECA 2017

[8] Amanpreet Singh,Narina Thakur,Aakanksha Sharma,"A review of supervised machine learning algorithms",IEEE,3rd International Conference on Computing for Sustainable Global Development,2016

[9] Bin Li,Yajuan Guo,Yi Wu,Jinming Chen,Yubo Yuan,Xiaoyi ZhangAn unsupervised learning algorithm for the classification of the protection device in the fault diagnosis system,IEEE,China International Conference on Electricity Distribution (CICED),2014

[10] Varshitha D N Vidyashree K P,Aishwarya P Janya T S,K R Dhananjay Gupta Sahana R,"Paper on Different Approaches for Crime Prediction system",International Journal of Engineering Research Technology (IJERT),NCETEIT - 2017 Conference Proceedings

[11] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.

[12] Malathi. A,Dr. S. Santhosh Baboo,"An Enhanced Algorithm to Predict a Future Crime using Data Mining",International Journal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011

[13] T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," Proc. of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.

[14] K.B.S. Al-Janabi, "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-Means Mining Algorithm," in Journal of Kufa for Mathematics and Computer, Vol. 1, No. 3, 2011, pp. 8-24.

[15] A. Malathi, S.S. Baboo, "An Enhanced Algorithm to Predict a Future Crime using Data Mining," in International Journal of Computer Applications, Vol. 21, 2011, pp. 1-6.

[16] C.H. Yu, "Crime Forecasting Using Data Mining Techniques," in International Conference on Data Mining Workshop, 2011, pp. 779-786.