Insightful Intelligence: Anticipating Crime Through Machine Learning

Dr. G.Revathy[1], Ms.M.Nivethini[2], Dr.A.Gomathi[3]& Mr. M. Tamil Thendral[4]

1- Assistant Professor, Department of CSE, Srinivasa Ramanujan Centre, SASTRA Deemed University, Kumbakonam. revathyjayabaskar@gmail.com. 0000-0002-0691-1687
2- Assistant Professor, Department of Artificial Intelligence and Data Science, Builders Engineering College, Nathakadaiyur, Tiruppur Dt. Email mn.aid@builderscollege.edu.in
3- Associate Professor, Department of IT, Knowledge Institute of Technology (Autonomous),Salem. Email agit@kiot.ac.in. 0009-0005-6851-8572
4- Assistant Professor , Department of CSE , Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology , Chennai. Email: tamil.mtt@gmail.com, 0000- 0003- 0052- 6339
.

How to cite this article: G.Revathy, M.Nivethini, A.Gomathi M. Tamil Thendral (2024) Insightful Intelligence: Anticipating Crime Through Machine Learning. *Library Progress International,* 44(3), 18168-18173.

ABSTRACT

India is seeing a rapid increase in the number of criminal cases; thus, there are more cases awaiting judgment. It is becoming more difficult to classify and settle criminal cases because of their exponential growth. Understanding a location's trends of criminal activity is crucial to preventing it. A good grasp of the patterns of criminal activity in a particular area will help the organizations tasked with solving crimes perform more effectively. Identifying trends in criminal activity can be accomplished through the application of machine learning and the use of different algorithms. This study makes use of a database of crime data to predict the kinds of crimes that will happen in a certain location.

KEYWORDS ; Crime Database, Machine learning, Crime classification.

INTRODUCTION

The analysis predicts exactness by contrasting the outcomes of various administered algorithms. The comparison would yield the most precise algorithm, which would then forecast the outcome. Across the world, the amount of crime is rising daily. Because it is either skilled or unexpected, an offence cannot be anticipated. Crime is an issue that is economical and unpleasant to life. Depending on the sort of country and society, the specifics of how crime is committed undergo revolutionary change. Prior studies on crime prediction have shown that conditions, employment status, and factors like education and deficiency all have an impact on crime rates. More than 2,000,000 records worth of data were gathered between 2018 and 2022.

Finding patterns and contributing factors to high rates of violent crime in a given area or community may be the goal of using machine learning algorithms to predict violent crime rates. deep learning models can find correlations and forecast future crime rates by examining data on past crime rates, socioeconomic indicators & other pertinent variables. The ultimate objective of this analysis might be to assist law enforcement organisations and decision-makers in identifying regions that are vulnerable to violent crime and in developing preventative strategies. Deep learning algorithms may also be used to create more precise and effective interventions, including social services, community policing programmes, and crime prevention programmes.

LITERATURE SURVEY

Any culture that deals with crime faces challenges since it is unpredictable, can occur anywhere at any time, and is hard to forecast. Three popular prediction classification algorithms—Random Forest, Gradient Boosting Decision Tree, and Naive Bayes—are examined and contrasted before a crime prediction model is submitted. Predictions for many categories are generated by the algorithm by looking at the top 10 offenses, or 97% of the episodes. In order to find patterns and provide an explanation for trends in crime, exploratory data analysis, or EDA, is performed using a crime dataset. Specifically, the accuracy rates of Gradient Boosting Decision Tree, Random Forest, and Naive Bayes are 63.45%, 63.43%, and 65.82%, per source[1].

Python, supervised machine learning approach (SMLT) and machine learning-classification approach Forecasting the correctness of the result. Comparing the actual data with our predicted values for both years, the accuracy of the recommended techniques falls between 85% and 90%. An accuracy of no more than 90% is achieved. It is unable to look up and compare the Recall, Precision, and Confusion matrix to our earlier findings. It is impossible to assess the qualities' relevance using the well-known machine learning technique[2].

Using police records from 2012 to 2014, we employ machine learning techniques to predict corruption offenses in Italian communities. Our analysis accurately predicts more than 70% (but not quite 80%) of the towns that will have corruption incidents, or an increase in corruption offenses. We show how the 2012 anti-corruption law in Italy may be strengthened to better combat white-collar crime through the use of algorithmic forecasts[3].

This study aims to explore multiple approaches for estimating crime rates in different urban areas. In this paper, we looked at three different kinds of prediction models: logistic regression, linear regression, and gradient boosting. Feature selection approaches were used to identify the predictive variables for these models[7]. We were able to increase forecast accuracy while preventing the model from becoming overfit by employing this technique. Crime statistics from Saint Petersburg were used to assess the generated models. Gradient boosting is the most efficient method for estimating crime rates in a certain metropolitan area, according to our analysis of the results of model predictions[4].

Crimes are common in undeveloped countries like India. We must always be aware of our surroundings since cities are progressively getting more urbanized. In order to avoid the unlucky, we will attempt to track crime rates using the KNN prediction algorithm. It will attempt to forecast the sort of crime, as well as when, where, and how it will occur. This data will illustrate crime tendencies in a certain place, which may be valuable in criminal investigations. It will also show us which locations have the greatest crime rates. In this study, the k-nearest neighbour machine learning technique will be applied[5].

For criminal justice choices, retributive and utilitarian motives are at odds. This is due in part to the retributive goal's rejection of prediction, whereas all utilitarian goals need it. In the context of this discussion, we examine studies on violence prediction and find that, due to their low accuracy, such forecasts are only marginally beneficial for public policy development or individual decision-making. One is determining the severity of crime and strategies to reduce it. Second, we propose the idea of social stakes and argue that it should also be evaluated. Finally, we offer a paradigm that might help to reduce the tension between retributive and utilitarian approaches[6]s.

METHODOLOGY

In this system, we proposed lasso, multiple linear regression, the KSN model, and the null model. Lasso is effective when numerous factors may be influencing the result, but only a small number of them are actually crucial. When working by huge classified data that has a great number of potential predictors, Lasso can be quite helpful. In order to create a prediction, KNN traces the K data facts that are contiguous to the input data point and uses their tags (for classification) or values (for regression). When there are distinct patterns in the data that can be quickly identified visually or by using basic metrics like distance, KNN can be useful. A statistical technique known as multiple linear regression predicts the linear connection between numerous predictor variables. A null model is a straightforward baseline model that makes no assumptions about the relationship between the predictors and the outcome. It offers a benchmark for assessing the performance of more complicated models, which makes it helpful for comparisons.

3.1.1. BLOCK DIAGRAM

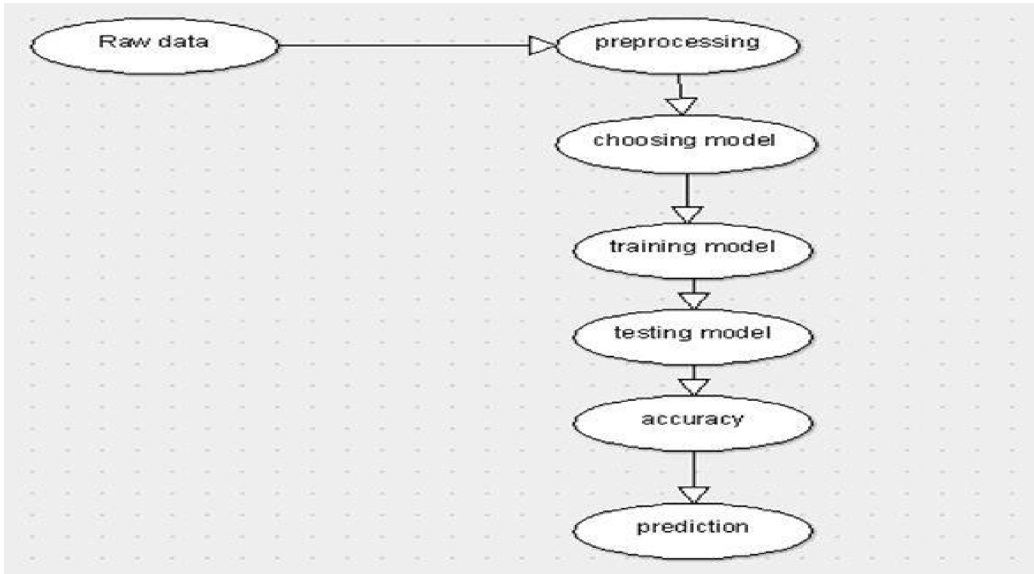


Figure 1 Architecture Diagram

DATASET

For testing algorithms that select or learn weights for qualities, a large number of variables are supplied. Clearly unrelated components to crime were eliminated, but; features were chosen if there was a strong correlation between them and crime (N=122) in addition to the variable that was to be predicted (Per Capita Violent Crimes). The neighborhood, law enforcement, and percentage of policemen assigned to drug units are among the factors in the dataset. Other neighborhood-related variables include the percentage of the population classified as urban and the median family income. The per capita violent crime variable was determined by adding up the total number of offenses that were classed as violent crimes and the population of the United States. It seems that rape counting has stirred up some debate in certain places. The values of per capita violent crime were inaccurate due to the absence of rape statistics. These cities are not represented in the dataset. Many of the localities that were excluded were in the Midwest of the United States.

State	county	community	communityName	fold	Population	householdsize	racepctblack	racePctwhite	racePctAsia	racePctHisp	agePct12t21	agePct12t29	agePct16t24
8	?	?.1	Lakewoodcity	1	0.19	0.33	0.02	0.9	0.12	0.17	0.34	0.47	0.29
53	?	?	Tukwilacity	1	0	0.16	0.12	0.74	0.45	0.07	0.26	0.59	0.35
24	?	?	Aberdeentown	1	0	0.42	0.49	0.56	0.17	0.04	0.39	0.47	0.28
34	5	81440	Willingborotownsh	1	0.04	0.77	1	0.08	0.12	0.1	0.51	0.5	0.34
42	95	6096	Bethlehemtownshi	1	0.01	0.55	0.02	0.95	0.09	0.05	0.38	0.38	0.23
6	?	?	SouthPasadenacity	1	0.02	0.28	0.06	0.54	1	0.25	0.31	0.48	0.27
44	7	41500	Lincolntown	1	0.01	0.39	0	0.98	0.06	0.02	0.3	0.37	0.23
6	?	?	Selmacity	1	0.01	0.74	0.03	0.46	0.2	1	0.52	0.55	0.36
21	?	?	Hendersoncity	1	0.03	0.34	0.2	0.84	0.02	0	0.38	0.45	0.28
29	?	?	Claytoncity	1	0.01	0.4	0.06	0.87	0.3	0.03	0.9	0.82	0.8
6	?	?	DalyCitycity	1	0.13	0.71	0.15	0.07	1	0.41	0.4	0.52	0.35
36	?	?	RockvilleCentrevill	1	0.02	0.46	0.08	0.91	0.07	0.1	0.34	0.36	0.22
25	21	44105	Needhamtown	1	0.03	0.47	0.01	0.96	0.13	0.02	0.29	0.32	0.2
55	87	30075	GrandChutetown	1	0.01	0.44	0	0.98	0.04	0.01	0.35	0.53	0.32
6	?	?	DanaPointcity	1	0.04	0.36	0.01	0.85	0.14	0.26	0.32	0.46	0.3
19	187	91370	FortDodgecity	1	0.03	0.34	0.06	0.93	0.03	0.03	0.39	0.41	0.28
36	1	1000	Albanycity	1	0.15	0.31	0.4	0.63	0.14	0.06	0.58	0.72	0.65
34	27	17650	Denvilletownship	1	0.01	0.53	0.01	0.94	0.2	0.03	0.34	0.39	0.27
18	?	?	Valparaisocity	1	0.02	0.47	0.01	0.97	0.07	0.02	0.7	0.67	0.63
42	129	66376	Rostravertownship	1	0	0.41	0.05	0.96	0.01	0.01	0.37	0.37	0.24
6	?	?	Modestocity	1	0.25	0.54	0.05	0.71	0.48	0.3	0.42	0.48	0.28

Figure 2 Sample Dataset

FEATURE EXTRACTION

When processing needs must be reduced without losing important or pertinent data, the feature extraction method can be helpful. Reduced amounts of duplicate data for a particular study can also be achieved by feature extraction. In addition, machine learning accelerates the learning and generalization processes by reducing the amount of data and the computer's work in generating variable combinations (features).

Figure 3 Feature Extraction Output

RESULTS AND DISCUSSIONS

The highest crime rate is been targeted as output in Figure 4

```
In [39]: ith the highest correlation with the target
         tures = crime.corr().loc['ViolentCrimesPerPop'].apply(np.abs).sort_values(ascending=False).index[13:24]
         tures = list(top_corr_features)
         tures

Out[39]: ['NumUnderPov',
          'PctHousLess3BR',
          'medFamInc',
          'PctNotHSGrad',
          'PctHousOccup',
          'PctVacMore6Mos',
          'Population',
          'numbUrban',
          'medIncome',
          'HousVacant',
          'PctLess9thGrade']
```

Figure 4 High crime rate output

The highest violent crime target is been achieved in Figure 5

```
In [61]: with the highest correlation with the target
         atures = crime.corr().loc['ViolentCrimesPerPop'].apply(np.abs).sort_values(ascending=False).index[1:12]
         atures = list(top_corr_features)
         atures

Out[61]: ['PctKids2Par',
          'PctFam2Par',
          'NumImmig',
          'racePctWhite',
          'PctTeen2Par',
          'PctYoungKids2Par',
          'racepctblack',
          'FemalePctDiv',
          'pctWInvInc',
          'pctWPubAsst',
          'NumIlleg']
```

Figure 5 Highest Violent crime rate target

The mean square error for the models are plotted and given in Figure 7 . The model predicted values are given in Figure 6.

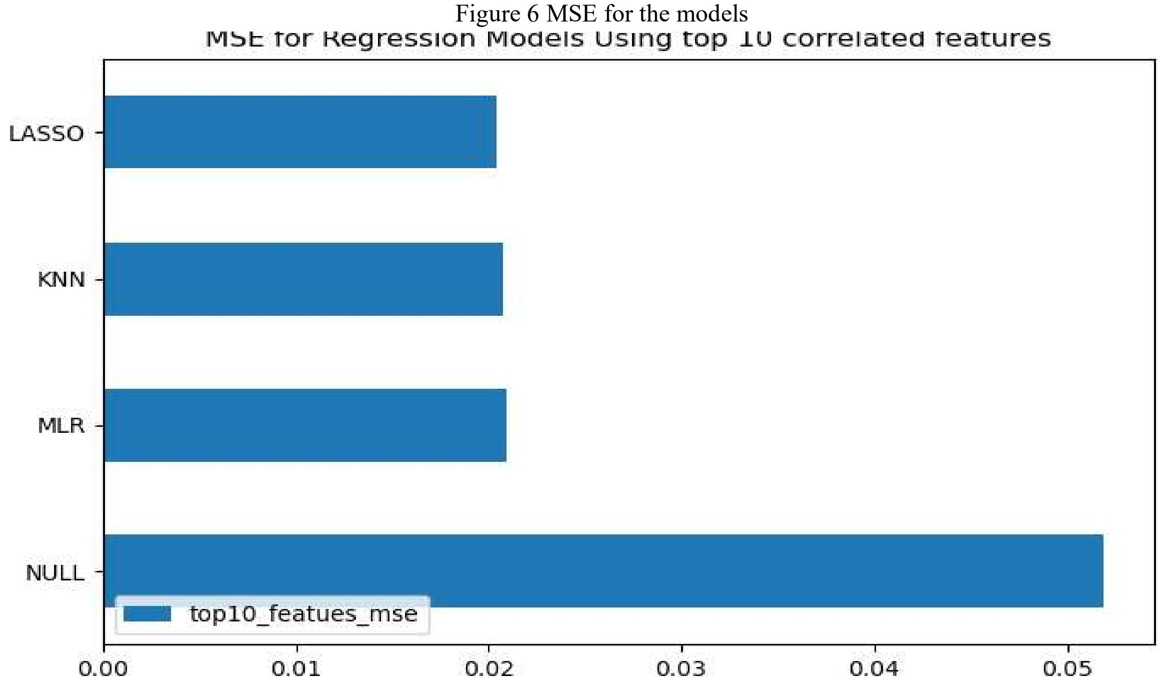

In [77]: models

Out[77]:

	NULL	MLR	KNN	LASSO
top10_featues_mse	0.051887	0.020959	0.020793	0.020379
full_model_mse	0.072208	NaN	NaN	NaN

Figure 6 Model predicted results

Figure 6 MSE for the models



CONCLUSION

Based on the given results, it can be inferred that the NULL method is more efficient than other Machine learning algorithms. We may continue our work with various sorts of cross validation approaches to improve accuracy. Future study plans include developing visual pictures and location maps to provide effective prediction of the anticipated criminal occurrence, as well as an opportunity to improve police patrolling system control.

1. REFERENCES

1. Aravindan, S., et al. "GUI based prediction of crime rate using machine learning approach." Int J Comput Sci Mob Comput 9.3 (2020): 221-229.
2. Khan, Muzammil, Azmat Ali, and Yasser Alharbi. "Predicting and preventing crime: a crime prediction model using san francisco crime data by classification techniques." Complexity 2022 (2022).
3. De Blasio, Guido, Alessio D'Ignazio, and Marco Letta. "Predicting corruption crimes with machine learning. A study for the Italian municipalities." Preprint at http://www. diss. uniroma1. it/sites/default/files/allegati/DiSSE_deBlasioetal_wp16_2020. pdf (2020).

4. Ingilevich, Varvara, and Sergey Ivanov. "Crime rate prediction in the urban environment using social factors." Procedia Computer Science 136 (2018): 472-478.
5. Kumar, Akash, et al. "Crime prediction using K-nearest neighboring algorithm." 2020 International conference on emerging trends in information *technology and engineering (IC-ETITE)*. IEEE, 2020.
6. Gottfredson, Don M., and Stephen D. Gottfredson. "Stakes and risks in the prediction of violent criminal behavior." *Violence and victims* 3.4 (1988): 247-262.
7. Chandrakala, S., Deepak, K., & Revathy, G. (2023). Anomaly detection in surveillance videos: a thematic taxonomy of deep models, review and performance analysis. *Artificial Intelligence Review*, *56*(4), 3319-3368.