



# Crime Analysis Using Supervised Learning

PRATIK KUMAR DUTTA   SABYASACHI CHAKRABORTY   SAHIL GANDHE

Under The Guidance of  
Prof. Jayakumar S

VELLORE INSTITUTE OF TECHNOLOGY UNIVERSITY,VELLORE

## ABSTRACT

***This paper describes that Crime analysis is performed using an ensemble of Naive Bayesian Classification, Random Forest and Support Vector machine algorithms. In this technique, team has calculated and given the result of the accuracy for the provided terms and data. System is able to get prediction automatically with this approach and approximate outcome for a big division of Crime analyzed data and expressions by comparing and deploying efficient algorithms that are significantly better and thereby achieving results in minimal span of time.***

## INTRODUCTION

Crimes are the increasing danger to the mankind. There are numerous crimes that happens normal interim of time. Maybe it is expanding and spreading at a quick and huge rate. Crimes occur from little town, town to enormous urban areas. Violations are of various kinds – burglary, murder, assault, ambush, battery, false detainment, hijacking, manslaughter. Since Crimes are expanding there is a need to illuminate the cases in a lot quicker way. The Crime exercises have been expanded at a quicker rate and it is the obligation of police division to control and decrease the crimes exercises. Crimes forecast also, criminal recognizable proofs are the serious issues to the police office as there are colossal measures of Crimes occurrence information that exist. There is a need of innovation through which the case explaining could be quicker.

## BACKGROUND

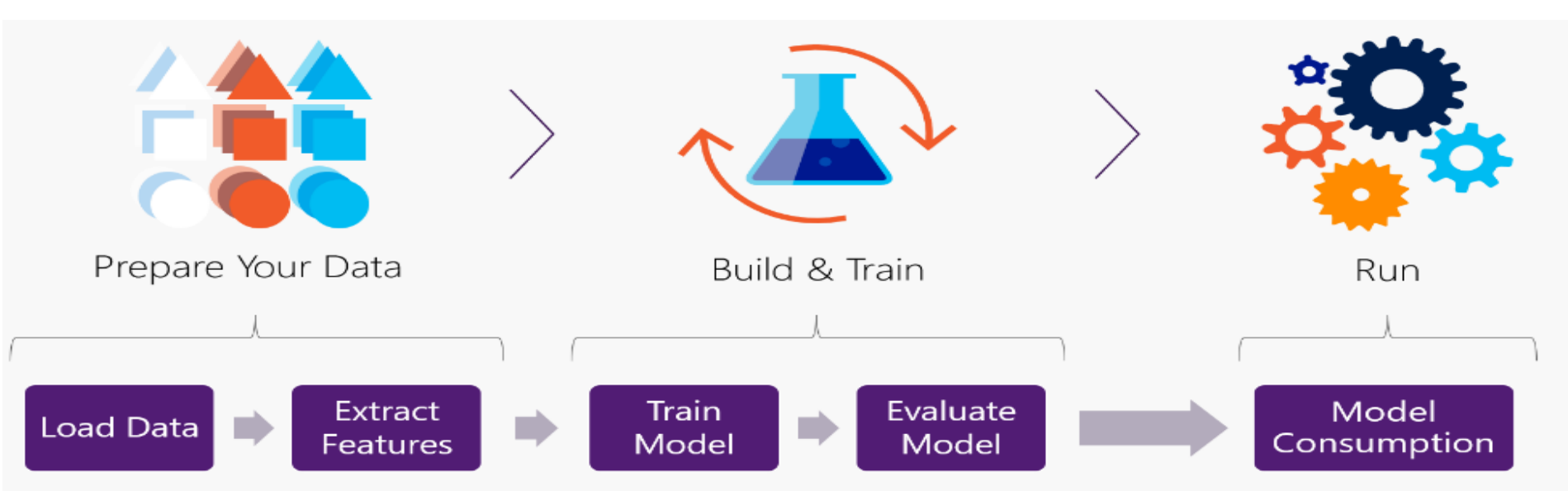
Procedure and techniques of crime analysis is the process of recognizing and sorting conclusions communicated in a piece of content, particularly with the end goal to decide if the item is positive, negative or unbiased along with comparing to get better accuracy after the pre processing. Different application like emotion detection and Crime identification, Age prediction, Selection and Multilayer Perception can be done using: 1. Naive Bayesian Classifier, 2. Random Forest, 3. Decision Tree Algorithms and 4. Support vector machine.

*A. Naive Bayesian classifier*  
We can say Naive Bayesian classifiers defines that this classifiers are considering the characterization job from a Statistical perspective. The beginning stage is that the likelihood of a class is given by the back likelihood or probability given a preparation report. Here alludes to the majority of the content in the whole preparing set. Here is given by , where is the aspect (word) of record .  
 $P(A|B) = P(B|A) P(A)P(B)$  ----- [1]  
*P= probability of two objects A= first object*  
*B= Second object*

*B. Random Forest*  
Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction  
Using Gini Split / Gini Index  
 $G = 1 - \sum (P_i)^2$  ----- [2]  
*P=probability of Positive Occurrence*  
*G= Gini index*

*C. Support Vector Machine*  
Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges  
 $hw, b(x) = g(w^T x + b)$  ----- [3]  
*w, b= two attributes*  
Can be generalized to more than two values  
b. MODULES • Data Collection  
• Data Pre-processing  
• Algorithm extraction:  
• Features of algorithm extraction  
• Compare the result based on accuracy  
• Analyze the performance

- a) Naive Bayesian prediction
- b) Random Forest prediction
- c) Support Vector Machine (SVM) prediction



## RESULTS

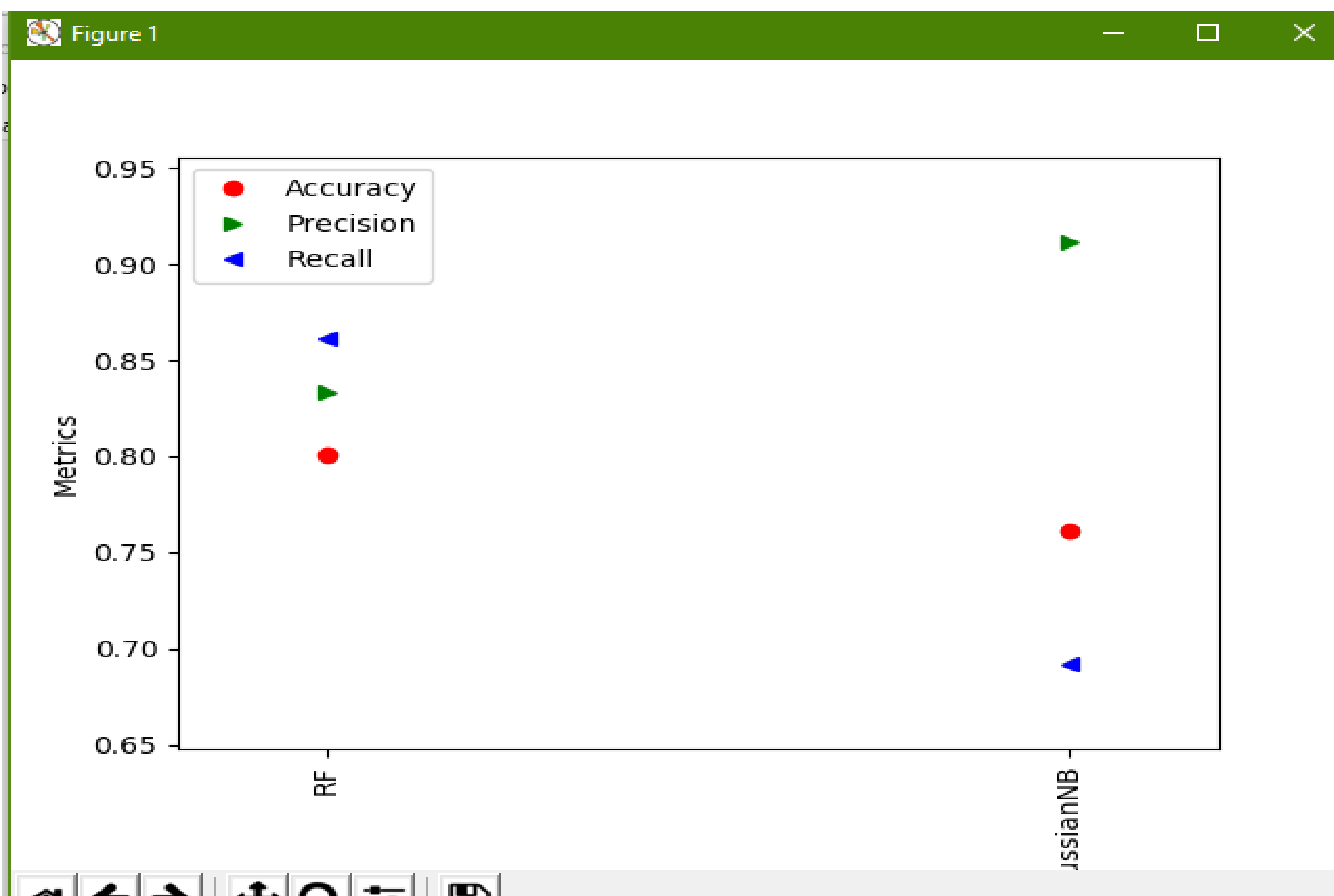
Accuracy for RandomForestClassifier is:- 0.8012537688442212  
Precision for RandomForestClassifier is 0.8333878362200029  
Recall for RandomForestClassifier is 0.8615999999999999

Accuracy for gaussian Naive Bayes : 0.761608040201005  
Recall for gaussian: 0.692  
Precision for gaussian: 0.9117998148278733

Accuracy for SVM is: 0.8072713567839195  
Precision for SVM is: 0.8549456169037868  
Recall for SVM is 0.8448

## CONCLUSION

In this paper, Crime Analysis is performed by collecting the Crime data and individual pacifiers such as Naïve Bayesian Classifier, Random Forest, and Support Vector Machine. All classifiers however in the future, an ensemble of naïve Bayesian Classifier, Support Vector Machine and Random Forest is to be deployed for better accuracy.



## REFERENCES

[1] Agrawal, R., Imielinski, T. and Swami, A. (1993) "Mining Association Rules between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD Conference, pp. 207-216.  
Agrawal, R. and Srikant, R. (1994) "Fast Algorithms for Mining Association Rules", Proceedings of 1994 International Conference on Very Large Data Bases, pp. 487-499.  
Bailey, T. C. and Gatrell, A. C. (1995) Interactive Spatial Analysis. Wiley: New York.  
Berry, M.J.A. and Linoff, G. (1997) Data Mining Techniques --- for Marketing, Sales and Customer Support. Wiley: New York.  
Besag, J. E. and Newell, J. (1991) "The Detection of Clusters in Rare Diseases", Journal of the Royal Statistical Society A, vol. 154, pp. 143-155.  
Eldershaw, C. and Hegland, M. (1997) "Cluster Analysis using Triangulation", in Noye, B. J, Teubner, M. D. and Gill, A. W. (eds.), Computational Techniques and Applications: CTAC97. World Scientific: Singapore, pp. 201-208.  
[2] G.C. Oatley, J. Zeleznikow, B.W. Ewart, "Matching and Predicting Crimes," In Applications and Innovations in Intelligent Systems XII in Proceedings of AI2004, The Twenty-fourth SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence. Ann Macintosh, Richard Ellis and Tony Allen Ed. London: Springer, , pp. 19-32, 2004.  
R. William Adderley, "The use of data mining techniques in crime trend analysis and offender profiling," Ph.D. thesis, University of Wolverhampton, Wolverhampton, England , 2007.  
Y. Xiang, M. Chau, H. Atabakhsh, H. Chen, "Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list," Decision support systems, Elsevier Science Publishers, vol. 41 no.1, pp: 69-83, Nov. 2005  
[3] Keyvanpour, Mohammad Reza, Mostafa Javideh, and Mohammad Reza Ebrahimi. "Detecting and investigating crime by means of data mining: a general crime matching framework." Procedia Computer Science 3 (2011): 872-880.  
[4] Al-Janabi, Kadhim B. Swadi. "A proposed framework for analyzing crime data set using decision tree and simple k-means mining algorithms." Journal of Kufa for Mathematics and Computer 1.3 (2011): 8-24.  
Al-Janabi, K.B.S., 2011. A proposed framework for analyzing crime data set using decision tree and simple k-means mining algorithms. Journal of Kufa for Mathematics and Computer, 1(3), pp.8-24.  
[5] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Hong Atabakhsh, "Crime Data Mining: An Overview and Case Studies", AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003  
[6] Alves, L.G., Ribeiro, H.V. and Rodrigues, F.A., 2018. Crime prediction through urban metrics and statistical learning. Physica A: Statistical Mechanics and its Applications, 505, pp.435-443.  
Alves, Luiz GA, Haroldo V. Ribeiro, and Francisco A. Rodrigues. "Crime prediction through urban metrics and statistical learning." Physica A: Statistical Mechanics and its Applications 505 (2018): 435-443.  
[7] Sathyadevan, Shiju, and Surya Gangadharan. "Crime analysis and prediction using data mining." 2014 First International Conference on Networks & Soft Computing (ICNSC2014). IEEE, 2014.  
Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In 2014 First International Conference on Networks & Soft Computing (ICNSC2014) (pp. 406-412). IEEE.