International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

Crime Predictive Model for Hotspot Mapping

[1]Mrs. P. Rajeswari, [2]S.Haneesh Chowdary, [3]L.Sunandhkumar, [4]T.Venkata Hanish

[1] *Assistant Professor, Computer Science and Engineering, RMD Engineering College*
[2,3,4]*Student, Computer Science and Engineering, RMD Engineering College*

Date of Submission: 20-03-2023 Date of Acceptance: 30-03-2023

ABSTRACT:
It is critical to recognise crime patterns in order to be better prepared to respond to criminal behaviour by analysing data scrapped from public source. The use of artificial intelligence (AI) and machine learning to identify criminality using sound or video or data.The Artificial Intelligence and Machine Learning usage in order to predict the happening of the crimes is still yet to found, which can also be of great use when it worked as inspected.We present and discuss our findings using several classifiers, as well as discuss future research directions.

I. INTRODUCTION :

Crime Datamining which is an Urdu word meaning protector. With respect to the safety and protection of the public the main query is, if anyone get to know about happening of a particular crime, then the public can be able to take the necessary safety action. So understanding the patterns of the happening crimes can
help with public safety and protection and can lead to targeted and sensitive practices by law enforcement authorities to mitigate crime, and more concerted efforts by citizens and authorities to create healthy environments.
Due to the huge increase in the development of the cities and towns. There is a huge increase in the happening of the crimes, and information regarding it can help with the prevention. So people who gets to know with these kind of things can take precautious actions and can even at least be careful. The crimes like more theft, robberies, accidents and many more. The common people have faced fear not only when they step out of the house and also taking necessary action to even safe guard their houses, so that no third person can find any chance to get or break into their place of living.

Much of the current work is focused in two major directions:

- Predicting surges and hotspots of crime, and
- Understanding patterns of criminal behaviour that could help in solving criminal investigations.

With the advent of the Big Data era and the availability of fast, efficientalgorithms for data analysis, understanding patterns in crime from data is an active andgrowingfield of research.
The algorithm takes the date, time and name of the region as the input. The result will also vary based on the selected timestamp. The crimes are classified into:

- Act 13 - Gambling
- Act279 – Accident
- Act302 – Murder
- Act323-Violence
- Act363 – Kidnapping
- Act379 - Robbery

The algorithms will be taking the input and identify the provided user location with geolocater and analyse the data which has been previously been recorded with respect to the latitude and longitude also with the taken timestamp, and check for the crime (among the taken 6 classes) that might have the high possibility of happening. If no chance of crime being predicted the results as safe zone. For these few algorithms like Decision Trees, Random Forests and K-Nearest Neighbours have been considered.

II. LITERATURE SURVEY :

[01] The EPSRC crime prevention and detection technologies programme(A.E.A Rose - 2005)

The function and role of the Engineering and Physical Sciences Research Council (EPSRC) is described. The EPSRC has launched a research

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

programme in Technologies for Crime Prevention and Detection. The aim is to encourage the engineers and physical scientists to develop the next generation of technologies for preventing and detecting crime. The Crime Technology programme is stimulating scientists and engineers to think about how their discipline and expertise can help combat crime and improve security.

[02] Identify Crime Detection Using Data Mining(Sharmistha Dutta, Ankit Kumar, Neetu Narayan - 2017)

Identity Crime is considered as crimes which involve masquerading one's identity and steal confidential information with respect to the concerned person's identity. This paper mainly deals with identity crime related to credit card application, which nowadays is quite prevalent and costly even. The existing non data-mining techniques for eliminating identity theft have some flaws and to combat them a new data- mining layer of defence has been proposed.

[03] Time, Place, and Modus Operandi: A Simple Apriori Algorithm Experiment for Crime Pattern Detection(Peng Chen, Justin Kurland - 2018)

Given the fast-paced nature of modern police work, the development and use of advanced data mining tools for crime analysis can play a critical factor in mitigating future harm and helping with crime prevention. This paper aims to solve the problem of identifying potential serial offending patterns using previously underutilised attributes from police-recorded crime data.

[04] Crime pattern detection, analysis & prediction(Sunil Yadav, Ajit Yadav, Rohit Vishwakarma, Nikhilesh Yadav - 2019)

Crimes are a social irritation and cost our in solving crimes quickly will pay for itself. About 10% of the criminals commit about 50% of the crimes [9]. The system is trained by feeding previous years record of crimes taken from legitimate online portal of India listing various crimes such as murder, kidnapping and abduction, dacoits, robbery, burglary, rape and other such crimes.society deeply in several ways. Any research that can help

III. IMPLEMENTATION :

In this Project, We use three different classification problems to solve, which we proceeded to attack with an assortment of classification algorithms. The following are the algorithms which we are using:

- Dataset & Preprocessing
- KNN(K- Nearest neighbors)
- Decision Tree
- Random Forests

Dataset & Pre-processing

In this Project, Dataset are scraped daily from website of data.gov.in or update the dataset from Kaggle Source which is publically available. There are some datasets in different Regional Languages and in order to perform machine learning this data cannot be used as it is. Hence the data needs to be processed.

Before implementing machine learning algorithms on the data, we went through a series of pre-processing steps with our classification task in mind. These included separating the data into set of data frames for Time & Location

The timestamp will be processed according to the year, date and time of occurrence of each crime. This was decomposed into five features: Year (2018), Month (1-12), Date (1-31), Hour (0- 23) and Minute (0-59).

Following these pre-processing steps, we ran some out-of-the box learning algorithms as a part of the initial exploratory steps. The new feature set consisted of 9 features, all of which were now numeric in nature.

Timestamp	Robbery	Gambling	Accident	Violence	Murder	Kidnapping	latitude	longitude
28-02-2021 21:00	1	0	0	0	0	0	22.73726	75.87599
28-02-2021 21:15	1	0	0	0	0	0	22.72099	75.87608
28-02-2022 10:15	0	0	1	0	0	0	22.73668	75.88317
28-02-2022 10:15	0	0	1	0	0	0	22.74653	75.88714

Fig 1. Dataset after Pre-processing

KNN (K-Nearest neighbour)

K - Nearest Neighbour is the Base classifier used in this project to handle the First Set of Classification Processes Made with 3 - 10 Base Neighbours. It works by finding the distances between a query and all the examples in the data, selecting the specified examples that are closest to the query, and then votes for the most frequent label. It classifies a new set of cases based on a similarity measure of data newly added. Damaged Data will be handled by Elbow Error, The Elbow Method is a widely used method which helps in determining the optimal value of K which takes the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

Error K Range as an input and an average array of data will be allocated according to N.

It's an instance-based Learning method & Elbow method though it is not parametric, which implies that it does not make any supposition on the primary data distribution. To put it in simple words, the model structure is decided by the data. It's pretty useful because in reality, most of the data does not follow the typical theoretical norms made. Hence, the K-Nearest-Neighbor Algorithm has been used in this project.

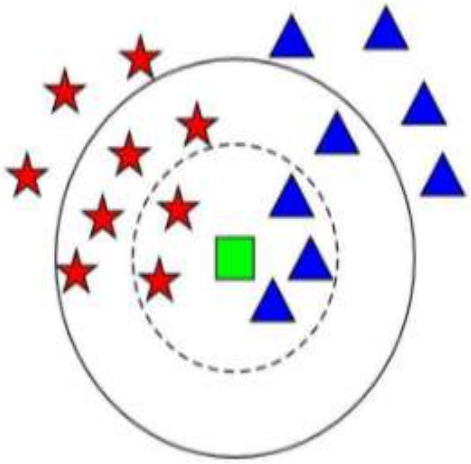

Fig 2.Principle diagram of KNN

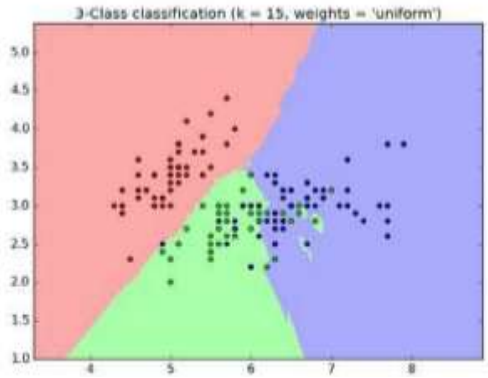

Fig 3.Graphical Representation of KNN

After analyzing the graph, the range in which the error rate was minimum came out to be 1-15. Furthermore, we checked all its values in range 1- 15, but from the values of k ranging 1-13, the accuracy remained constant. Hence, we selected k=3 that belonged from the range 1-1.

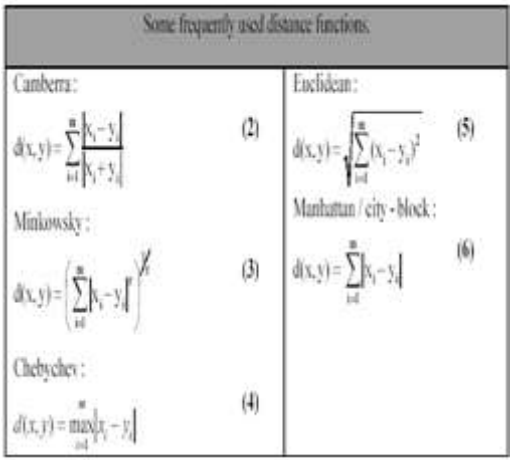

Fig 4. Functions

To calculate Distance values, we required test, train and predicted values of both x and y. So by performing Some quadratic scores using **Figure 4** functionswhich figures the average magnitude of the error. It is the square root of the square differences calculated between expectation and actual observation. An increase in k-value results in increased root mean square error. Hence the value of k was picked from range 1-15 because that is the only range with minimum error. The previous work had included extra factors which did not seem necessary in our case. The Base Classifier system was trained to learn using some particular inputs and apply the algorithms.

Decision Tree

Decision Tree is the Second Classification Process that handles the K Predicted Value and applying the Set of Rules to it, Decision Tree Training Data will be generated by going through process which will be represented in tree structure, Similarly Decision Forest Tree Classifier Creates a High Score nodes which is useful for estimators to deal even in random states, Using sklearn library it is helpful to stores the dataset in memory instances dealing with multiple classifiers handles the flow of prediction more accurate. The result of the decision tree is a tree-shaped structure that describes a series of decisions.

K Training samples take the form of sample data that will be used to build a tree that has been substantiated. Set of Rules Develop the Algorithms result by grouping several training sample data that will result in a decision tree based on the facts on the training data

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

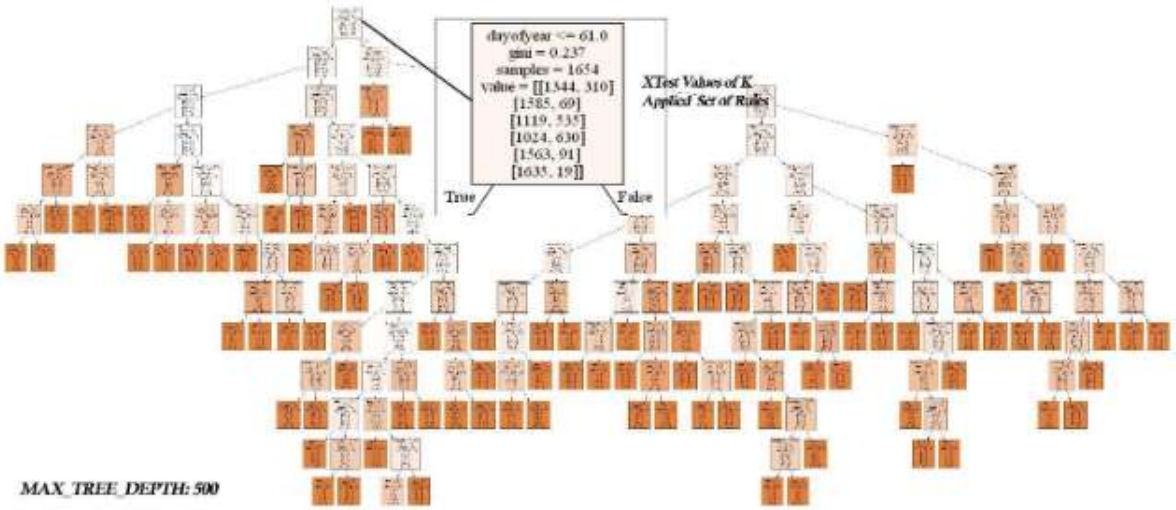

Fig 5. Decision Tree

In **Figure 5**, The Generated Tree applied with Set of Crime Patterns thus the accuracy or precision and recall have the highest values compared to other Crime pattern algorithms

Random forest

Random Forest is the Estimator used to finalize the decision made by decision tree. Random forest algorithm there creates data instances based upon decision tree and then gets the prediction from each of them. The Process of Random Forest functions with the k-th tree, a random vector Θ will be generated, independent of the past random vectors Θ1 , … , Θk−1 , but with the same distribution; then the training set and Θk are used to generate the tree to obtain the classifier h(x, Θk),

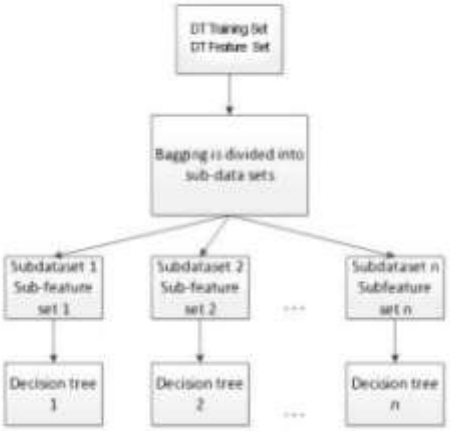

Fig 6. Random Forest 1

The nature and dimension of Θ depends on its use in the tree structure. After generating a large number of trees, they vote for the most popular class. Randomly select m of the Tree n characteristic variables. Split the decision tree to the variable with the largest tree associated with the DT variable.

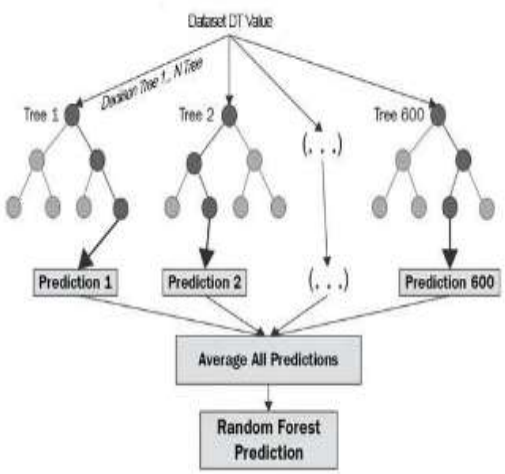

Fig 7. Random Forest 2

After constructing the decision tree and the N Estimation Process has been decided, the random forest will classify the number of tests according to the category with the most votes obtained from the B decision tree.

Data Visualization :

Visualization of Test Data prediction, this also displays every one of Crime data predictions,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

these patterns are according to each hour of day. Here you can see that the spike at nighttime is quite regular across each day of the week for city, but residential show flatter peaks during the weekends.

This is what it Expect to see, as the majority of people work during the weekdays. But it took specific domain knowledge of crime pattern knowledge on how to best aggregate the data. It needs to know what to chart to begin With to be able to make any sense of any complicated dataset That visualized of every Crimes

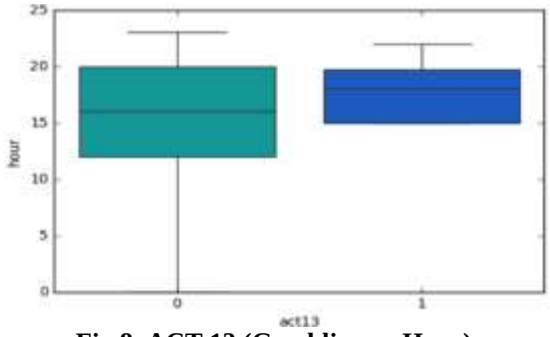

Fig 8. ACT 13 (Gambling vs Hour)

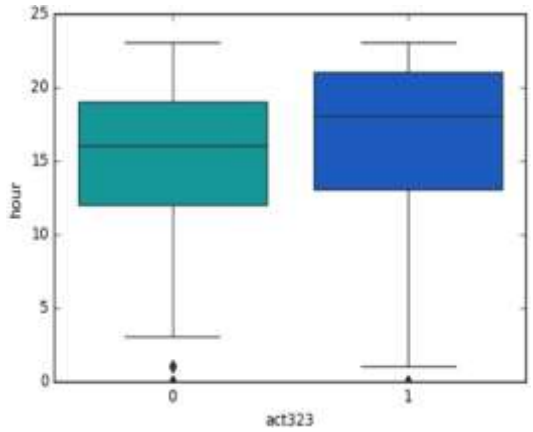

Fig 9. ACT323 (Violence vs Hour)

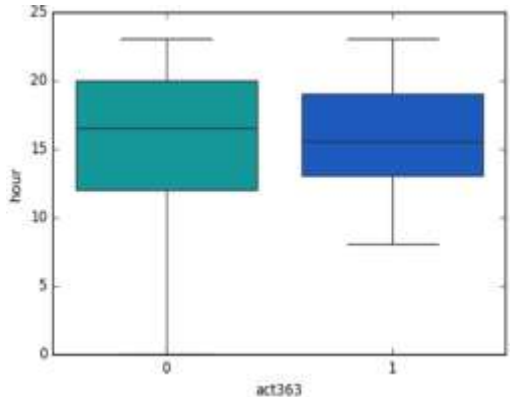

Fig 10. ACT 363 (Kidnapping vs Hour)

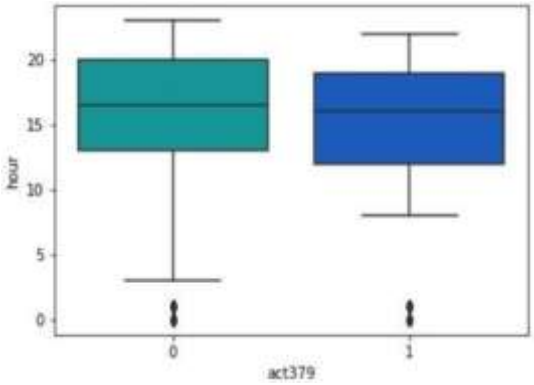

Fig 11. ACT 379 (Robbery vs Hour)

IV. SYSTEM TESTING

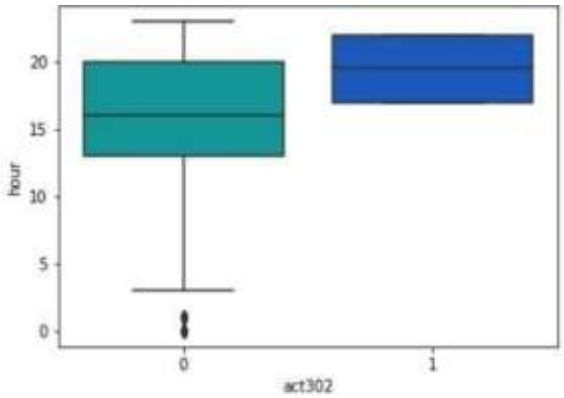

Fig 12. ACT 302 (Murder vs Hour)

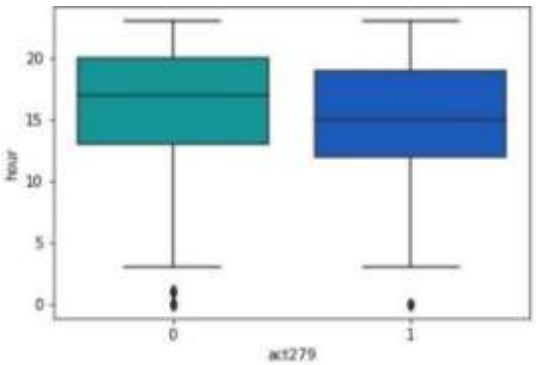

Fig 13. ACT 279 (Accident vs Hour)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 3 March 2023, pp: 1608-1614 www.ijaem.net ISSN: 2395-5252

Test Case ID	Test Name	Test Description	Steps	Executed result	Actual result	Test Case Statement
1	Check for correct entered numeric values and date and time	The entered values are in correct format.	Enter details and click submit	If format is correct details are sent to kernel successfully.	As expected.	Pass
2	Check for correct entered time	The entered values are in correct format.	Enter details and click submit	If format is correct details are sent to kernel successfully.	As expected.	Pass
3	Check for correct entered location	The entered values are in correct format.	Enter details and click submit	If format is correct details are sent to kernel successfully.	As expected.	Pass
4	Predicted result	Output displayed	Output displayed	If format is correct details are sent to kernel successfully.	As expected.	Pass
5	Analysis Button	Data visualization is displayed	Click analysis	If format is correct details are sent to kernel successfully.	As expected.	Pass

V. EXPECTED OUTCOME :

The idea behind this project is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement. This kind of data analysis was technologically impossible a few decades ago, but the hope is that recent developments in machine learning are up to the task.

The use of artificial intelligence (AI) and machine learning to identify criminality using sound or video or data. The Artificial Intelligence and Machine Learning usage in order to predict the happening of the crimes is still yet to found, which can also be of great use when it worked as inspected. The biggest challenge will probably be "proving" to politicians that it works. When a system is designed to stop something from happening, it is difficult to prove the negative.

Companies that are directly involved in providing governments with AI tools to monitor areas or predict crime will likely benefit from a positive feedback loop. Improvements in crime prevention technology will likely spur increased total spending on this technology.

Possible avenues through which to extend this work include time-series modelling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of

crime. It would also be interesting to explore relationships between surges in different categories of crimes – for example, it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

VI. CONCLUSION :

The initial problem of classifying 6 different crime categories was a challenging multi-class classification problem, and there was not enough predictability in our initial data-set to obtain very high accuracy on it. We found that a more meaningful approach was to collapse the crime categories into fewer, larger groups, in order to find structure in the data. We got high accuracy and precision on Prediction. However, the Violent/Non-violent crime classification did not yield remarkable results with the same classifiers – this was a significantly harder classification problem. Thus, collapsing crime categories is not

an obvious task and requires careful choice and consideration.

Possible avenues through which to extend this work include time-series modelling of the data to understand temporal correlations in it, which can then be used to predict surges in different categories of crime. It would also be interesting to explore relationships between surges in different categories of crimes – for example, it could be the case that two or more classes of crimes surge and sink together, which would be an interesting relationship to uncover. Other areas to work on include implementing a more accurate multi-class classifier, and exploring better ways to visualize our results.

REFERENCES :

[1]. Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex.2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction.

[2]. Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei.2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)

[3]. Kianmehr, Keivan and Alhajj, Reda. 2008. Effectiveness of support vector machine for crime hot-spots prediction, pages 433-458, Applied Artificial Intelligence, volume 22, number 5.

[4]. Toole, Jameson L and Eagle, Nathan and Plotkin, Joshua B. 2011 (TIST), volume 2, number 4, pages 38, ACM Transactions on Intelligent Systems and Technology

[5]. Wang, Tong and Rudin, Cynthia and Wagner, Daniel and Sevieri, Rich. 2013. pages 515- 530, Machine Learning and Knowledge Discovery in Databases

[6]. Friedman, Jerome H. ”Stochastic gradient boosting.” Computational Statistics and Data Analysis 38.4 (2002): 367-378.sts

[7]. Leo Breiman, Random Forests, Machine Learning, 2001,Volume 45, Number 1, Page 5.