*Global Journal on Application of Data Science and Internet of Things*
*Vol. 2, Issue 2 - 2018*

# A BIG DATA ANALYTICAL METHODOLOGY USING LINEAR REGRESSION FOR CLASSIFICATIONS IN CRIME DATA PATTERN EVALUATION

**PUSHPENDRA SINGH SHEKHAWAT[*], BRIJ KISHORE[*], PRATEEK DADHICH[**]**

## ABSTRACT

Big data analytics is a very familiar term in computer science world especially in data science work which has been associated with huge velocity or data and versatility in the dataset. In the dissertation a huge data set has been considered in order to get better accuracy in the result section .The Weka tool that has been used before in the existing work but We have applied Rapid Miner analytics for better pattern evaluations and meaningful result mining in huge volume of data. The tool rapid miner is having better feature and configuration to get better pattern evaluations in bulk data and provide better result in data science analytics. Our work primarily considered process cycle for result optimization. In our study we've implemented big data in the local domain. We have used a tool for implementing big data and for finding better outcomes. Most of the studies of big data were about the simple implementation and finding the result as mining. But in our work we transformed data for efficient result and more optimized process. For this kind of job we use various operators or filters for transforming and filtering the data.

**KEYWORDS***:* Weka, Big Data.

## INTRODUCTION

Internet is the fourth mode of communication after gesture, speech and written documents. Nowadays information plays vital role for organization and it is handled by computerized system. Internet is currently gaining popularity but unlike other modes of communication, laws ruling them are also at nascent stage. (Mitchel 2015).

Information services and systems are starting to have important effect on the country economy, trade and commerce. The stock exchange in the country is carried out all kinds of transaction and information transfer through Internet. All E-commerce solutions like Business to Business (B2B), Business to Customer (B2C) and Business to Government (B2G) relation, Database driven Shopping cart, Electronic-Payment System through credit card are few examples of application of Internet. crime such as hacking, online financial fraud and planting computer viruses, have potential of shaking economies. crime is growing at high rate, posing challenge for technology and criminal justice system. So before generation of Law one should understand the question like, what is crime?

[*]Department of Computer Science & Engineering, Apex Institute of Engineering & Technology, Jaipur.
[**]BTech Scholar, Poornima University, Jaipur. ***Correspondence E-mail Id:*** editor@eurekajournals.com

 **ISSN: 2581-4370**

## BACKGROUND

### OVERALL PROCESS OF THE CRIME PREVENTION

The important things here are to prevent the cyber crime by the composition of all the matter like network security, cyber crime detection, case filing, digital evidence collection, criminal prosecution and punishment, and awareness.

### NETWORK SECURITY

As we know the "prevention is better than cure" our first responsibility is to develop the secure system before we start actual use of the computerized system. So the software professional has to make the standards to prevent the cyber crime and take all the precautionary steps like firewall and data encryption techniques are primary concern with the any computerized system. Other than that the computer industries has to build certain policy and procedure to prevent the cyber crime. (Hansen, James 2014).

### CYBER CRIME DETECTION

Other prime concern for the crime removal is to detect the crime at earlier stage before it creates the havoc among the Internet users. So one should know when the computer is starting for the misuse by the attacker.

### CASE FILING

The case filing and lodging the cyber crime by victim is also important issue in the cyber crime, so instead of worry of anything, victim should come forward and complain about the cyber crime to the cyber cell.

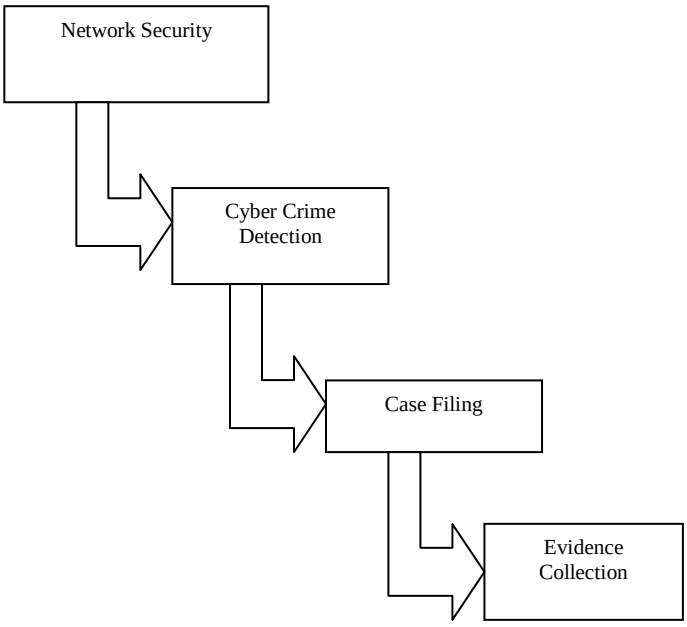


**Figure 1.cyber crime to cyber cell**

## PROBLEM FORMULATION AND PROPOSED SOLUTION

### PREVIOUS WORK

In previous works, we analyzed that the researchers have took the dataset, and apply the algorithms on them. They had generated results and analyzed the result. Some of the studies were about the big data implementation on any tool, some of the work was about introduction of big data, handling of big data etc. They introduced some models for big data.

But the actual implementation was missing in some studies. Some of the studies were about

 **ISSN: 2581-4370**

the implementation of algorithm on big data. In which they found the result for decision making.

## COMPARISON OF WORK

The work previously done was about the Big data domains and their implementation on Hadoop or other tools. But we implemented the big data analysis on the tool which was not dedicatedly designed for Big data such as Hadoop by optimizing and transforming the data without losing any important aspects of data or attribute.

In our work we'll analyze the data with its statistical information. Then we transformed the data for modelling and for better results. So that we can have a clear scenario about the historical data. We found some techniques to handle the big data with care. This clear scenario about the data gives us the clear objective to find and then we can find the better outcome from the huge amount of data. Because of huge amount of data, the result formulation is very tough task, but because of machine learning techniques and big data handling techniques we can achieve desired output

## PROPOSED ALGORITHM

To do so, we applied the classification on the dataset. In which we used the decision tree for classify the data accordingly.

### STEPS

1. Extract data from Training data (Big Data).
2. Build Predictive model which make decisions on the basis of data. It may include ETL process.
3. Take decisions from the predictive model and from incorrectly classified data.
4. Now apply operators to filter data, reconstruct dataset and apply removal of un usable datasets or results.
5. Repeat step 4 until you get the Decisions on the big data.

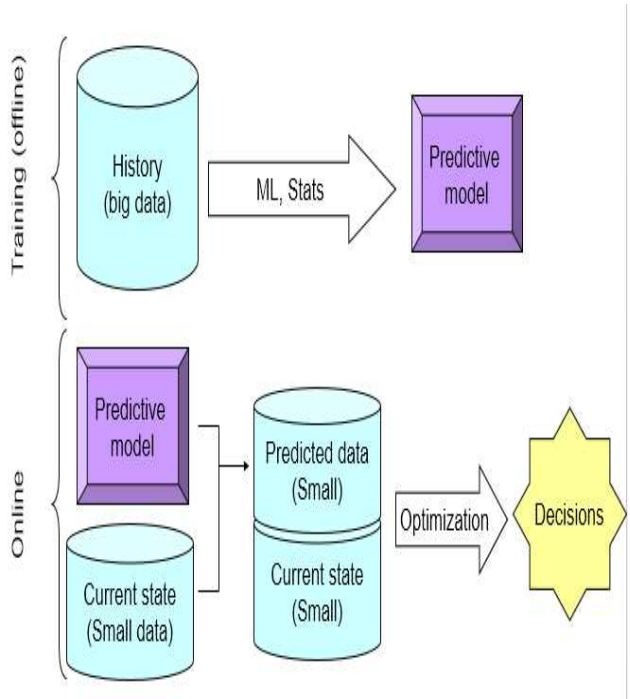


**Figure 2.Decision Tree**

## RESULT

Actualizing new learning methods in Rapid Miner is simple. It is only important to make two Java classes, of which one plays out the learning on the preparation dataset, i.e. the assessing of the model parameters. Alternate class must spare these parameters and have the capacity to apply

 **ISSN: 2581-4370**

the model to new information i.e. make expectations

## ATTRIBUTE ANALYSIS

Now we can analyze the particular attribute in details. The figure given below describes the attribute in details. We described the attribute in graphical format. It displays the statistical information about the attribute values such as missing, infinite, valid and category count, percentage etc.

## CRIME CLASSIFICATION

It contains the crime classes with their respective data values.

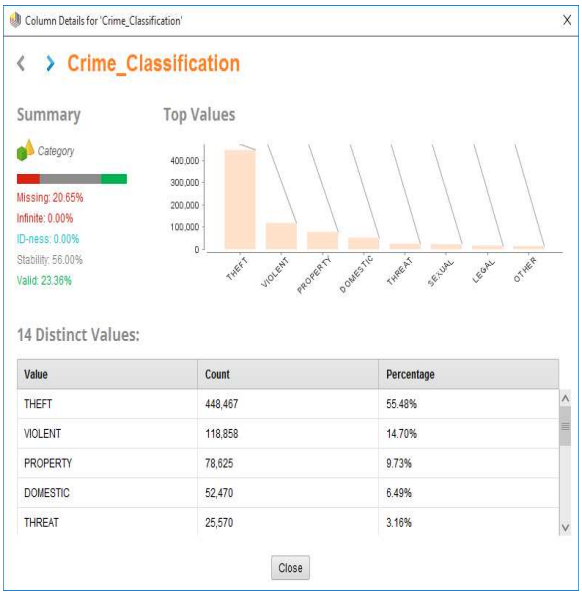


**Figure 3.Crime Classification attribute analysis**

## AREA

This field contains the area id in numeric form.

We in the result section we got the minimum, maximum, average, std. Deviation, values.

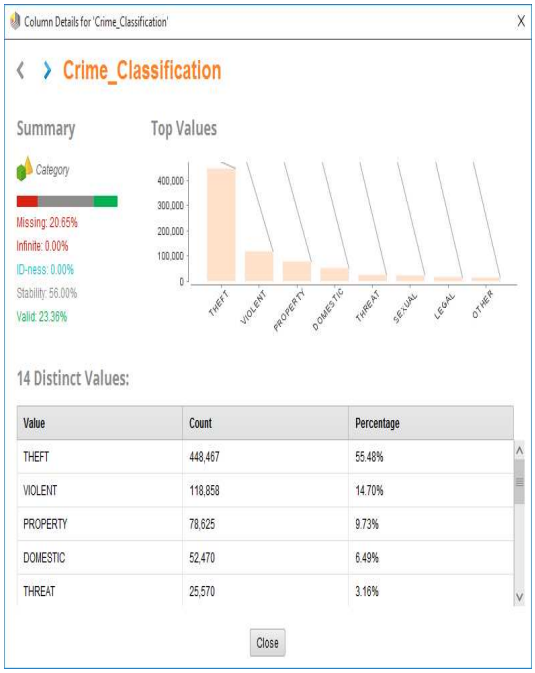


**Figure 4.Area attribute analysis**

 **ISSN: 2581-4370**

**AREA_NAME**

This field contains the Area name. It displays the statistical information about the attribute values such as missing, infinite, valid and category count, percentage etc.

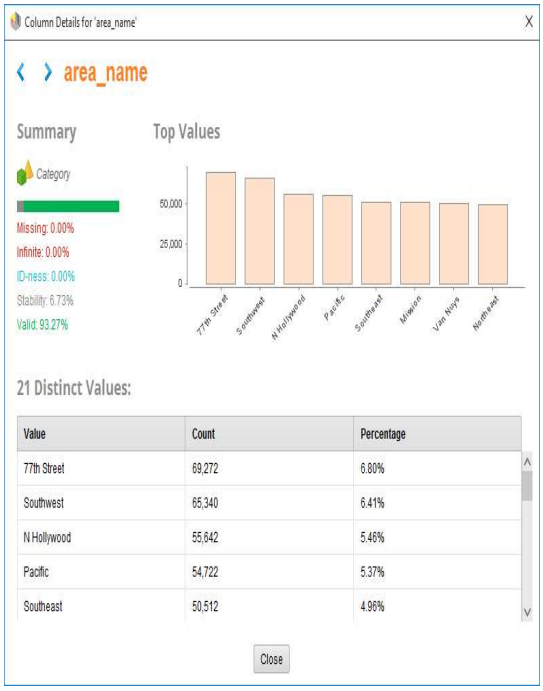


**Figure 4.Area_name attribute analysis**

## CONCLUSION AND FUTURE WORK

In the vast area of Big data, the analysis and outcome prediction is very tough task. In our work we analyzed the data with its statistical information. Compare them with correlation tests, then build the final predictive model using regression and then generated the result as desired. Because the data amount is very huge, we also removed the unused attributes through filters

## REFERENCES

[1]. Neesha Jothia, Nur'Aini Abdul Rashidb, Wahidah Husainc, a,-Relavant data dredging in Crime-A Review, Procedia Computer Science 72 ( 2015 ) 306-313.

[2]. Abhishek Kumar, Pramod Singh Rathore,-An Approach for Classification using Simple CART Algorithm in Weka, IEEE sponsored 3rd International Conference on Electroics and Communication System (ICECS 2016) 978-1-4673-7832-1/16/$31.00©2016 IEEE.

[3]. Pramod Singh Rathore,-An Contemplated Approach for Criminality Data using Mining Algorithm, International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248 Volume: 4 Issue: 2 236-240.

[4]. Pramod Singh Rathore,-Analysis Of Crime Data Using Data Mining Algorithm, international journal of engineering sciences & research Technology (2018) ISSN: 2277-9655.

[5]. Blessing Ojemea, Audrey Mboghob,-Selecting Learning Algorithms for Simultaneous Identification of Depression and Comorbid Disorders Procedia Computer Science 96 ( 2016 ) 1294-1303.

[6]. Lakshmi.B.Na, Dr.Indumathi.T.Sb, Dr.Nandini Ravic,-A study on C.5 Decision Tree Classification Algorithm for Risk Predictions during Crime, Procedia Technology 24 ( 2016 ) 1542-1549.

[7]. Btissam Zerhari1, Ayoub Ait Lahcen1,2, Salma Mouline1, Big Data Clustering:

 **ISSN: 2581-4370**

Algorithms and Challenges, International Conference on Very large Databases, pp - 487-499, 1994.

[8]. R.Agrawal, R.Srikant,-Mining Sequential Patterns, The 11th International conference on Data Engineering, pp-3-14, 1995.

[9]. J-S.R.Jang, C-T.Sun,-Neuro-Fuzzy and Soft Computing, ISBN-978-81-203-2243-1, PHI, 2011.

[10]. DU, P. and NAKAO, A., "Ddos defense as a network service," pp. 894–897, 2010.

[11]. DURCEKOVA, V., SCHWARTZ, L., and SHAHMEHRI, N., "Sophisticated denial of service attacks aimed at application layer," in ELEKTRO, 2012, pp. 55–60, IEEE, 2012.

[12]. Chan, S.C., Chu, Y.J., Zhang, Z.G.: A New Variable Regularized Transform Domain NLMS Adaptive Filtering Algorithm. IEEE Transactions on Audio, Speech, and Language Processing 21 (4), 868–878 (2013) Cross Ref Google Scholar.

[13]. Bhargava 2018 clustered, Clustered Comparative Analysis of Security Sensor Discrimination Data, Bhargava, Neeraj and Jain, Aakanksha and Kumar, Abhishek and Rathore, Pramod Singh and Bansal, Anuja}, Annals of Computer Science Poland, 2018.

[14]. Bhargava 2018 detection, Detection of Malicious Executables Using Rule Based Classification Algorithms, Bhargava, Neeraj and Jain, Aakanksha and Kumar, Abhishek and Le, Dac-Nhuong, 2018.

[15]. Bhargava 2017 prediction, Prediction of arthritis using classification and regression tree algorithm, Bhargava, Neeraj and Purohit, Renuka and Sharma, Sakshi and Kumar, Abhishek, Communication and Electronics Systems (ICCES), 2017 2nd International Conference on, 606--610, 2017, IEEE.

 **ISSN: 2581-4370**