



Analytical Study of Some Selected Classification Algorithms and Crime Prediction

Sapna Singh Kshatri, Bhawana Narain

Abstract: To prevent the crime these days police exercises particularly in the case of investigation, emphasis on Artificial Intelligence, data mining and Machine learning aspect. To prevent future crimes it is necessary to understand the crime behavior from the earlier crime records. The more numbers of recorded crime data approaches the research to analysis the data for prediction and prevention of crime. Most of the researcher use to cluster the data for further classification to predict the definitive of crime. To prevent the crime these tools are xeric applicable to predict the most sensitize zone in the city. Thesis concentrated on the methods to predict the crime and on his hidden arrays in the existing past records. The objectives of the thesis is to predict the certain possibility of crime by applying data ruining approach through WEKA are applied to confirm criminality y proclamation. three algorithms are referred from different groups of methods: SMO Zero R and J 48 decision trees. Over 10000 records from Indian police department are collected to predict the frequency of crime in overall and its behavior in which Naive algorithm shows the reliable prediction represent against crime frequency. This paper compares the three diverse order classification to be specific, SMO, Zero R and J 48 Decision Tree for anticipating Crime Category' for various states in India. The outcomes from the examination demonstrated that, Decision Tree calculation out performed calculation and sieved 41.44% and 73.33%% Accuracy in anticipating Crime Category for various conditions of India. order classification to be specific, SMO, Zero R and J 48 Decision Tree for anticipating Crime Category' for various states in India. The outcomes from the examination demonstrated that, Decision Tree calculation out performed calculation and sieved 41.44% and 73.33%% Accuracy in anticipating Crime Category for various conditions of India.

Keywords: crime prediction, J48, SMO, Zero-R, Indian crime, WEKA

I. INTRODUCTION

Crime is a disgusting critique which happens worldwide in both developed and under developing nations. To secure occupants, the legislature of the state for the most part keeps up a police center. The fundamental mission of the police is in general to secure the peoples or individuals, their proper and influence them to obey the rules and regulations of the constitution. More often, everyone needs to have a need of

security in where he lives, and expect that it is the administration commitment to keep up request and guarantee his security and safety. Public generally collaborate or support police with activities intercessions, in any case, the job of the police isn't just capturing culprit, however the y additionally search the offence scenes, gather proof and data about offences and guilt y parties. The police additionally need to acts preventively in hazard zones undermined with an offender. Keep up request isn't a simple objective it requires the inclusion of numerous disciplines. One of this exercises is additionally guiltiness research and examination. The police normally gather information about offences, guilt y parties and environment where a wrongdoing happened. These information are essential for catching criminals be that as it may, can likewise be valuable for guiltiness investigation and afterward for culpability prevention. To break down wrong extents, information and history of guiltiness, the hazard territories can be distinguished, and the methodology to build security in these zones can be created. A considerable lot of elements can impact the guiltiness, and a significant number of them can be identified by information of the earth and the fundamental driver of specific sort of wrongdoing. The urban areas are developing, and with an expanding accentuation on work effectiveness, new ways to deal with keep up wellbeing are created. Additionally, the enthusiasm for information handling and AI caused the prescient investigation turns out to be part of the police exercises. The expectation of future wrongdoing does not intend to heavenly definite spot where a particular individual perpetrates a wrongdoing, anyway utilizing information investigation, scientific and factual strategies to distinguish the most powerless spots to perpetrate a wrongdoing and make activity to ensure it. It likewise does not mean that IPCs totally supplant the cop's basic leadership. Today, many Police utilize a few devices for basic leadership from basic mapping methods to complex programming dependent on calculations. They additionally recorded first accomplishment with guiltiness decrease using these devices. It additionally enabled them to do their work productively and the sky is the limit from there financially. Be that as it may, how the prescient strategies work and how they effective? in this postulation, the tricky of guiltiness expectation is investigated from the theoretical and furthermore from the down to earth perspective. The fundamental targets are to present the prescient investigation of guiltiness and potential techniques, depict the procedure of investigation, and principally these discoveries set in motion with genuine information. wrongdoing is an offense against the overall public that is frequently accused and deserving of the law (Tayal,

Revised Manuscript Received on July 20, 2020.

* Correspondence Author

Sapna Singh Kshatri*, Computer Science and Application, MATS University, Raipur, India

Dr. Bhawana Narain, Computer Science and Application, MATS University, Raipur India .

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Analytical Study of Some Selected Classification Algorithms and Crime Prediction

(2015)) It has been seen that hoodlums carry out violations at wherever and in any structure. One of the notable isolations on the planet is the dread assault on World Trade Center on September 11, 2001 (Hahn, 1992)

A. Literature Review

we examine the utilization of the zone under the receiver operating characteristic (ROC) bend (AUC) as execution measure for Artificial intelligence classification. As a contextual investigation we assess six AI calculations (k-Nearest Neighbors, C4.5, Multiscale Classifier, Quadratic Discriminant Function, Multi-layer Perceptron, and a Perceptron) on six "genuine world" clinical diagnostics informational collections (Bradley., Volume 30, Issue 7, July 1997,).

This paper talks about the circle of expert social work with detainees and by method of outline presents the experience of the Voluntary Action for Rehabilitation and Development (VARHAD) in social work mediation with detainees (Rajagiri, December 2011).

This was cross sectional investigation comprise two gatherings: sentenced crooks gathering and under preliminary lawbreakers gathering. Apparatuses Socio-Demographic: This is a semi organized Performa. It contains data about socio segment factors, for example training, gender, age, religion, conjugal status, habitation and control of the subjects (Nanjundeswaraswamy, 2016).

The women's activist hypothesis of female wrongdoing is by all accounts generally reasonable for breaking down the female culpability. In India ladies are significantly increasingly defenseless as man centric society is considerably more grounded in India in contrast with western nations (Banerjee, 2019).

This paper talks about the circle of expert social work with detainees and by method of delineation presents the experience of the Voluntary Action for Rehabilitation and Development. A significant part of the criminal equity framework is the jail association. Detainment facilities in India are represented by the Prisoners Act of 1894, amended occasionally. According to the measurements of 2008 (NCRB 2008) (Ronald, 2011)

The proposed wrongdoing investigation instrument takes out the rescanning of the database for each new inquiry too need of aptitude clients. It gives the client an intuitive and quick approach to do procedure of distinguishing proof of wrongdoing problem areas and wrongdoing zones just as information examination among different region of intrigue. An ODBC information source is required before running this module to make a connection from CCIS database. (Kuruville, (2012)).

II. METHODOLOGY

Subsequent subgroups in this area include: 2.1 classification of protocols, 2.2 proposed method. In this segment the information that was utilized in our investigations is portrayed in detail. The philosophy is likewise portrayed, including a short layout of the calculations utilized in our analyses, just as the assessment measurements utilized.

A. Data collection

In this progression the dataset is gathered from the police office about wrongdoing. Figures for individual states from

2001 to 2005 are considered. The dataset includes of 35 attributes where the 36th trait is the class variable. There are 2100 occurrences in the dataset. The dataset includes 7 Association areas and 28 states with the comparing locales.

The dataset comprises of 7 Association domains and 28 states with the corresponding regions. The bad crime in every locale is recorded from 2001 to 2005. In the dataset various kinds of wrongdoings (qualities) are viewed as like homicide, assault, hijacking, thievery, dacoity, theft, cheating, share passing, fire related crime, and so forth.

Table 1 types of crime

S.No.	Crime Type	No. of Records
1	MURDER	469878
2	ATTEMPT TO MURDER	429292
3	CULPABLE HOMICIDE NOT AMOUNTING TO MURDER	51343
4	RAPE	277215
5	CUSTODIAL RAPE	38705
6	OTHER RAPE	273016
7	KIDNAPPING & ABDUCTION	422325
8	KIDNAPPING AND ABDUCTION OF WOMEN AND GIRLS	315980
9	KIDNAPPING AND ABDUCTION OF OTHERS	145009
10	DACOITY	66300
11	PREPARATION AND ASSEMBLY FOR DACOITY	71227
12	ROBBERY	284504
13	BURGLARY	1300152
14	THEFT	3898568
15	AUTO THEFT	1437577
16	OTHER THEFT	2437773
17	RIOTS	878237
18	CRIMINAL BREACH OF TRUST	223262
19	CHEATING	879326
20	COUNTERFEITING	29149
21	ARSON	127378
22	HURT/GREIVIOUS HURT	3709380
23	DOWRY DEATHS	137375

With the end goal of our examinations we center around the Indian Mainland and information from January 2001 to Walk 2005 (60 months). Table 1 depicts the properties/highlights of the dataset. On the entirety of our information the "Revealed by" and "Falls inside" qualities have the worth "Hampshire Constabulary"; the documentation makes reference to that despite the fact that these characteristics are right now the equivalent, the "Falls inside" property will change sooner rather than later.

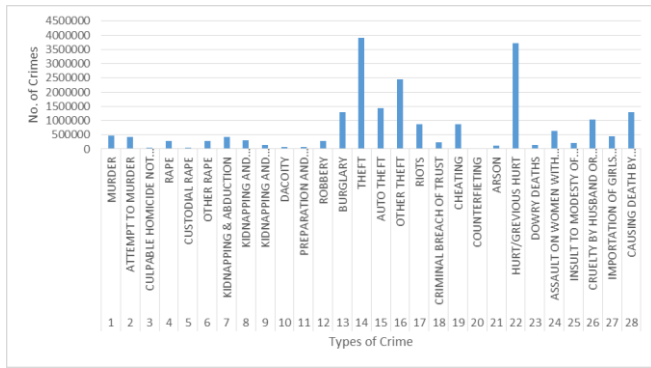


Figure II-1 graph representation of crimes

The wrongdoing type is one of the 16 classifications utilized by the police, which are recorded in Table 2 in diving request of recurrence on the Hampshire Constabulary information utilized in our exploration. The "Last result class" has choices, for example, under scrutiny; incapable to indict suspect; examination complete – no speculate distinguished; wrongdoer given admonition; guilty party fined, and so forth. The "Unique situation" characteristic is a literary depiction of the setting of wrongdoing; on as of late distributed information, this is constantly vacant.

B. Proposed approach

In this area the work process of the proposed model is portrayed. The initial phase in the model is Data pre-processing (DP) which incorporates filling the missing qualities, information cleaning and change of information. These days, a colossal measure of information being gathered and put away in databases on wrongdoing wherever over the globe. Important data and information covered up in such databases however without programmed strategies for separating this data it is for all intents and purposes incomprehensible. So to dig for those numerous algorithms were made to separate information from huge arrangements of information (Shiju Sathyadeven, 2014.), (Ubon Thansatpornwatana, 2016, ISBN: 978-1-5090-2258-8/16). Information mining order comprises of foreseeing certain result dependent on given info classification is a methodical way to deal with construct arrangement models from input information.

Research has shown that crime is expanding at a disorienting rate, so it is important to control it. So to reduce the percentage of crime, we have to consider the crime percentage of different places of a nation. This re- search has inaccurate ID detailing the various types of violations in all states and association domain crime percentages. Since uneducated and managed learning strategies are used, it helps to improve the filtration of vastly inaccurate databases. In this way, the crime percentage can be reduced in time and diligence by following the approach presented by the master. . Subsequently, this field talks about insights in relation to the experiment and the results that the WEKA instrument is using (Dr. Sudhir B. Jagtap, 2013).

1) Selection of Indian states

In this approach the dataset consists of 28 states and 7 union territories with the corresponding districts. The Indian states and union territories chosen for the clustering are Andhra Pradesh, Bihar, Gujarat, Karnataka, Kerala, Madhya Pradesh, Rajasthan, Tamil Nadu, Uttar Pradesh and Delhi UT. The average IPC crimes for each of the states (based on the

districts) during 2001 – 2005 are analyses with a line graph. The results are generated using the attributes “States/UT”, versus attribute “average_IPC_crime” (Fig 2).

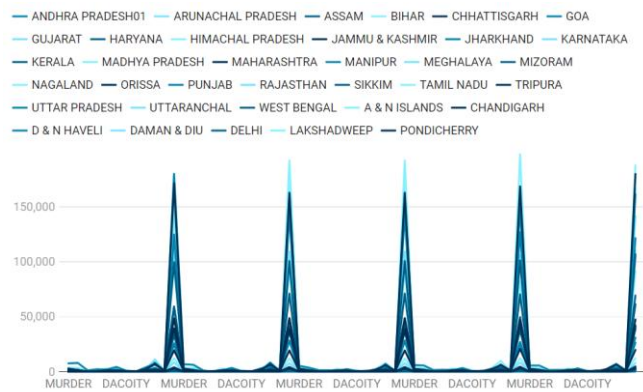


Figure II-2 violent crime in india

2) Crime analysis using WEKA

WEKA is a valuable open source free instrument. It is utilized for real word datasets. WEKA device helps in confirming the information mining algorithms (B. M. Patil, 2009). The after effects of k-implies are confirmed with WEKA. The Random Forest algorithms and the Neural Networks (characterization) chose are applied on the dataset and the accuracy is confirmed. The quantity of accurately ordered examples helps in realizing what number of cases are characterized effectively and the quantity of inaccurately grouped cases can be decreased by re-moving the properties which influence the exactness. The Random Forest is applied on the dataset and the test alternative chose is Cross-approval (Folds 5). The quantity of accurately arranged instances are 44 and wrongly classified instances are 16. The classifier checks a precision of 73.333 %.

C. Classification

classificationis the demonstration of searching for a model that portrays a class mark so that such a model can be utilized to anticipate an obscure class name (H. I. Bulbul and Ö. Unsal, 2011). Consequently, classification is normally used to predict an obscure class names. For example, a classification model can be utilized to group bank credits as either sheltered or dangerous. Arrangement applies a few techniques like choice tree, Bayesian strategy and rule acceptance in building its models. Grouping process includes two stages. The initial step is the learning stage which includes building the models while the subsequent stage includes utilizing the model to anticipate the class names.

III. DECISION TREE


A decision tree is an supervised learning calculation that works for both discrete and constant factors. (Myles) It is a notable classification strategy that appears as tree structure and it is normally comprised of:

- 1) Testing node which holds the information for testing the condition
- 2) Start node is the parent and typically top most hub.
- 3) Terminal node (leaf node): is the anticipated class mark
- 4) Branches: speaks to consequences of a test made on an attributes.



Analytical Study of Some Selected Classification Algorithms and Crime Prediction

Figure 3.1 is an example decision tree that predicts the buying enthusiasm of a client in PC. Rectangular shapes are utilized for testing hubs while oval shapes are utilized for result nodes. It is generally binary while others are non parallel.

 Weka Classifier Tree Visualizer: 20:29:20 - trees.J48 (Book_temp)

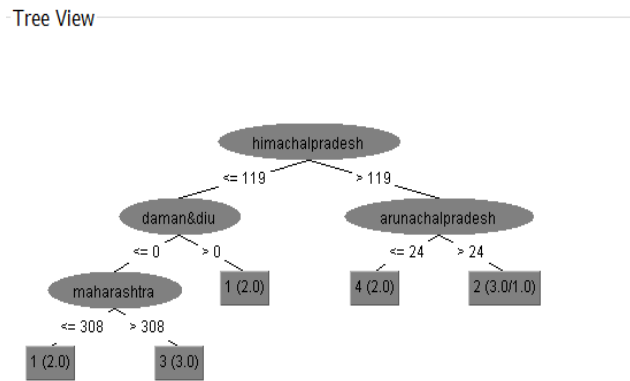


Figure III-1 building a decision tree

A. Building Decision Tree J48

decision tree can be assembled utilizing various techniques, the principal strategy created was ID3 (Interactive, Dichotomies) which later transformed into C4.5 classifier. J48 classifier is an improved adaptation of C4.5 choice tree classifier and has become a well known choice tree classifier. Grouping and Regression Trees (CART) was later evolved to deal with binary trees. In choice examination, a choice tree can be used to ostensibly and explicitly address decisions and dynamic. In data mining, a decision tree depicts data (yet the ensuing game plan tree can be a commitment for dynamic). To make a decision tree, the C4.5 (Quinlan, 1986.), J48 and SMO are basic methods of decision tree classification.

Algorithm of decision tree

Boundaries

Dataset's fields

Set of Attributes

Determination Technique for the Attribute

Result

Tree Classifier

System

1) A node is Created (call it)

2) Check if all records is in one gathering and compose nodes
The J48 calculation incorporates a few features as demonstrated as follows:

- It is available as an open source in WEKA interface in Java. Which is easily download in any operating framework its an assortment of representation instruments and algorithms. for information investigation and prescient modeling (G. Holmes, 1994).
- Computation directly encourages . for data analysis and predictive modeling
- The calculation uses unreserved and consistent properties.
- Enumeration gives a strategy, known as arithmetic, that manages missing properties. This method helps to resolve the issue of missing value, which is a sign in, cant highlight in view of fixing missing assets dependent on accessible information.
- Similarly the calculation gives the process of pruning a tree, which helps to create smaller trees and eliminate information.

- In addition, the calculation gives a sub-tree substitution process that reduces the degree of cutaneous blunder in view of suppressing the subtype with a leaf.

```

| | | asislands <= 11: 2 (10.0)
| | | asislands > 11: 1 (3.0)
daman&diu > 8
| jharkhand <= 2524: 4 (5.0)
| jharkhand > 2524: 5 (5.0)

Number of Leaves : 9
Size of the tree : 17

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      44      73.3333 %
Incorrectly Classified Instances    16      26.6667 %
Kappa statistic                    0.6503
Mean absolute error                 0.1177
Root mean squared error             0.3314
Relative absolute error             32.3381 %
Root relative squared error         84.7728 %
Total Number of Instances          60

=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0.750    0.125    0.750     0.750   0.750     0.625 0.752    0.427    1
0.600    0.080    0.600     0.600   0.600     0.520 0.752    0.427    2
0.967    0.009    0.765     0.967   0.813     0.747 0.977    0.649    3
0.500    0.040    0.714     0.500   0.588     0.534 0.730    0.440    4
1.000    0.018    0.833     1.000   0.909     0.905 0.987    0.783    5
Weighted Avg. 0.733    0.085    0.730   0.733   0.727     0.646 0.815    0.587

=== Confusion Matrix ===

 a b c d e <-- classified as
15 1 3 1 0 | a = 1
 3 6 0 1 0 | b = 2
 1 1 13 0 0 | c = 3
 1 2 1 5 1 | d = 4
 0 0 0 0 5 | e = 5
    
```

Figure III-2 screen shot of result of J48

B. SMO algorithms

SMO decays the advancement issue and takes care of the littlest conceivable improvement issue at each progression. In standard SVM, it works by overhauling two Lagrangian multipliers without a moment's delay with the others fixed (J. Platt, 1998)

Choose the SVM algorithm:

SMO alludes to the particular effective improvement algorithms utilized inside the SVM usage, which represents Sequential Minimal Optimization.

```

Time taken to build model: 0.33 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      26      43.3333 %
Incorrectly Classified Instances    34      56.6667 %
Kappa statistic                    0.1791
Mean absolute error                 0.2807
Root mean squared error             0.3751
Relative absolute error             91.5217 %
Root relative squared error         95.9411 %
Total Number of Instances          60

=== Detailed Accuracy By Class ===

TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
1        0.85     0.37       1       0.541     0.575    1
0        0       0          0       0         0.57     2
0        0       0          0       0         0.567    3
0.1      0       1          0.1     0.182    0.55     4
1        0       1          1       1         1        5
Weighted Avg. 0.433    0.283    0.373    0.433    0.294    0.603

=== Confusion Matrix ===

 a b c d e <-- classified as
20 0 0 0 0 | a = 1
10 0 0 0 0 | b = 2
15 0 0 0 0 | c = 3
 9 0 0 1 0 | d = 4
 0 0 0 0 5 | e = 5
    
```

Figure III-3 Screen shot of result of SMO classifier

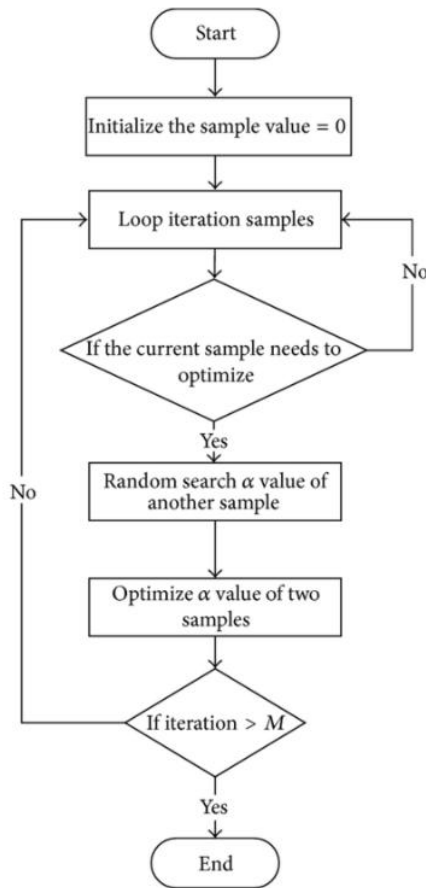


Figure III-4 SMO algorithm

The C boundary, called the multifaceted nature boundary in Weka controls how adaptable the procedure for adhering to a meaningful boundary to isolate the classes can be. An estimation of 0 permits no infringement of the edge, though the default is 1.

A key boundary in SVM is the kind of Kernel to utilize. The least difficult portion is a Linear part that isolates information with a straight line or hyper plane. The default in Weka is a Polynomial Kernel that will isolate the classes utilizing a bended or wiggly line, the higher the polynomial, the more wiggly (the type esteem).

A famous and ground-breaking part is the RBF Kernel or Radial Basis Function Kernel that is equipped for learning shut polygons and complex shapes to isolate the classes.

C. Zero R classifier

Zero R classifier is a straightforward classifier procedure that works with mode for the desire for apparent data and mean for the gauge of numeric data. It is regularly insinuated as larger part class system (Desai, 2012). In WEKA Zero-R is a fundamental classifier. Zero-R is a frivolous classifier, anyway it gives a lower bound on the show of a given dataset which should be inside and out improved by extra complex classifiers. As such it is a reasonable test on how well the class can be foreseen without thinking about different qualities. It might be used as a Lower Bound on Performance. Any learning figuring in WEKA is gotten from the hypothetical WEKA classifiers (Inamdar, 2011).

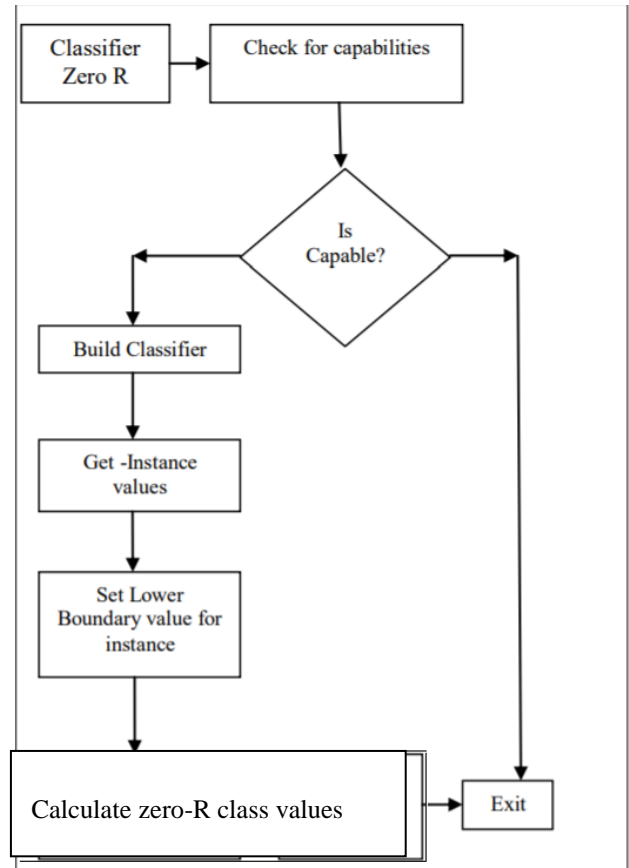


Figure III-5 Pre-process of Zero-R classifier

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      20          33.3333 %
Incorrectly Classified Instances    40          66.6667 %
Kappa statistic                    0
Mean absolute error                 0.3073
Root mean squared error             0.3917
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          60

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      1      1      0.333      1      0.5      0.5      1
      0      0      0      0      0      0.5      2
      0      0      0      0      0      0.389    3
      0      0      0      0      0      0.5      4
      0      0      0      0      0      0.227    5
Weighted Avg.  0.333  0.333  0.111  0.333  0.167  0.449

=== Confusion Matrix ===

 a b c d e  <-- classified as
20 0 0 0 0 | a = 1
10 0 0 0 0 | b = 2
15 0 0 0 0 | c = 3
10 0 0 0 0 | d = 4
 5 0 0 0 0 | e = 5
```

Figure III-6screen shot of Zero R result

D. Evolution matrix

The boundaries considered while evaluating the picked classifiers are:

- 1) Exactness This shows the degree of successfully organized cases in each portrayal model
- 2) Kappa: Measures the association between described cases and authentic classes. It generally speaking lies between [0, 1]. The estimation of 1 techniques impeccable relationship while 0 strategies subjective guessing.

Analytical Study of Some Selected Classification Algorithms and Crime Prediction

- 3) TP Rate Is the estimations that shows adequately portrayed models..
- 4) FP Rate: Is the report of cases mistakenly named as right events.
- 5) Review: Measures the degree of each and every significant datum that was returned by the classifier. A high survey suggests the model returns most of the critical data.
- 6) Accuracy- Measures the exactness of the relevant data recuperated. High precision suggests the model returns more appropriate data than unnecessary data.
- 7) Time: Time taken to play out the gathering

Table 2 SMO classifier result

Evaluati on Metrics	Time	Accu racy	TP Rate	FP Rate	Kap pa	Prec ision	Re call
2001	0.21 Secs	25%	0.250	0.322	0.65	.111	0.250
2003	0.09 Secs	41.66 %	0.417	0.292	0.1515	.205	0.417
2005	0.33 Secs	43.33 %	0.433	0.283	0.179	0.373	0.433

Table 3 Classifier result of J48

Evaluati on Metrics	Time	Accu racy	TP Rate	FP Rate	Kap pa	Prec ision	Re call
2001	0.01 Secs	8.33 %	0.083	0.353	-.2571	0.05	0.083
2003	0.02 Secs	58.33 %	0.583	0.123	0.462	0.598	0.583
2005	0.01 Secs	73.33 %	0.733	0.085	0.650	0.73	0.733

Table 4 Zero-R classifier result

Evaluati on Metrics	Time	Accu racy	TP Rate	FP Rate	Kap pa	Prec ision	Re call
2001	0.00 Secs	33.33 %	0.33	0.33	0.00	0.11	0.33
2003	0.00 Secs	33.33 %	0.33	0.33	0.00	0.11	0.33
2005	0.00 Secs	33.33 %	0.33	0.33	0.00	0.11	0.33

As we can see table numbers 2,3 and 4 and we can compare between all algorithms, the table shows us the results of the years 2001,2003 and 2005. In which J 48 is more accurate than other algorithms based on the time and accuracy of the address.

IV. RESULT AND DISCUSSION

Researched three algorithms of machine learning for crime production. J48,SMO which is a tree based classifier and Zero R which is the rule based classifier. All algorithms have performed very well in their field. All algorithms have performed very well in their field. But as is the nature of our

data, we found that the J48's performance was better than the rest of the algorithms and more efficient in less time.

Table 5 result of proposed model

Evaluation Metrics	J48	SMO	Zero R
Time	0.01 Secs	0.09 Secs	0.00 sec
Accuracy	94.44%	56.78%	33.33%
TP Rate	0.944	0.568	0.333
FP Rate	0.022	0.568	0.333
Kappa	0.921	0	0
Precision	0.945	0.322	0.11
Recall	0.944	0.568	0.333

As can be seen Accuracy rate and TP rate are much higher than the rest of the classifier. Same TN rate and time are very low. Therefore, the selection of J48 model will be best for our crime prediction.

Firstly we have to load the our proposed saved model to the right click menu on the "Result list" panel: All models were saved, but the selection was made of random forest. After comparing with the rest of the model, the accuracy rate of Random Forest is higher than the rest. And our aim is to create a high accuracy model so that prediction is more accurate. Now The proposed model will be loaded into the explorer .After that we can now use the loaded model to make predictions for new data.

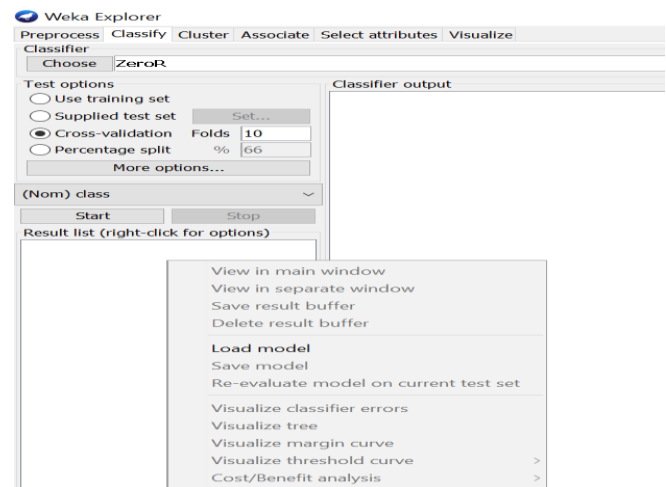


Figure IV-1 WEKA Load Model From File

The model will currently be stacked into the pilgrim .We would now be able to utilize the stacked model to make forecasts for new information. This data has been taken from 2007 to 2008.. In which the class is not defined.

V. CRIME PREDICTION USING J48

Three methodologies are proposed to remove criminal relations from the occurrence outlines and make wrongdoing forecast. After applying the model, I achieved an accuracy of 73.33 % , meaning that 44 instances out of 60 were correctly.

I also obtained the values of several performance measures for numeric prediction, presented in Table 6. From the confusion matrix result for the crime data established that five instances of the class were assigned to the different class.

Table 6 crime prediction result of J48

Performance measures	Result (J48)
Kappa statistic	.8356
MAE, mean absolute error	.0165
RMSE, root mean square error	.1277
RAE, relative absolute error	7.785
RRSE root relative squared error (%)	37.342
Correct classified instances out of 60	50

VI. CONCLUSION

Different methods have been executed so as to control the phishing assaults. Different tools and software are there to determine such sites. This paper took a gander at the utilization of information digging for distinguishing wrongdoing designs utilizing grouping rule mining strategies. The model will currently be stacked into the pilgrim .We would now be able to utilize the stacked model to make forecasts for new information. The created framework likewise adequately anticipated the rundown of potential suspects by essentially examining information recovered from the wrongdoing scene with previously existing information in the database with 80% exactness. The aftereffect of this work told the best way to manage crime expectation utilizing choice tree methods. Information mining had been applied in different regions of security, wrongdoing and criminal recognition. In this investigation, empowering results were gotten, a test information was utilized for testing and preparing classifiers due to time limitation.

REFERENCE

- Shiju Sathyadeven, D. M. (2014.). Crime Analysis and prediction using data mining. *IEEE*.
- B. M. Patil, D. T. (2009). "Predicting Burn Patient Survivability Using Decision Tree In WEKA Environment," . *IEEE International Advance Computing Conference*, (pp. pp. 1353-1356, doi: 10.1109/IADCC.2009.4809213.). Patiala.
- Banerjee, S. a. (2019). Private Versus Public: On Possibilities of Private Being Inclusive and Public Being exclusive . *An International Refereed Journal of English Language and Literature*.
- Bradley., A. P. (Volume 30, Issue 7, July 1997.). Cooperative Research Centre for Sensor Signal and Information Processing., *Department of Electrical and Computer Engineering*, (pp. Pp 1145-1159).
- Desai, A. a. (2012). "Analysis of machine learning algorithms using WEKA". ." *International Journal of Computer Applications* 975 (2012): 8887.
- Dr. Sudhir B. Jagtap, D. K. (2013). Census Data Mining and Data Analysis using WEKA. *International Conference in "Emerging Trends in Science, Technology and Management-*, (pp. 35-40). Singapore.
- G. Holmes, A. D. (1994). "WEKA: a machine learning workbench,". *Intelligent Information Systems Conference, Brisbane, Queensland, Australia*, pp. pp. 357-361, doi: 10.1109/ANZIIS.1994.396.
- H. I. Bulbul and Ö. Unsal, . (2011). "Comparison of Classification Techniques used in Machine Learning as Applied on Vocational Guidance Data,". *10th International Conference on Machine Learning and Applications and Workshops*, (pp. pp. 298-301, d). Honolulu.

- Hahn, B. H. (1992). *Risk Society: Toward a New Modernity*. Sage, Ltd. Becker.
- In, I. J. (October – December, 2015). *The International Journal of Indian Psychology*, ISSN 2348-5396 (e) | ISSN: 2349-3429 (p) Volume 3, Issue 1, No.5,.
- Inamdar, S. A. (2011). "Preprocessor agent approach to knowledge discovery using Zero-R algorithm." . *International Journal 2 (2011)* .:
- J. Platt. (1998). "Sequential minimal optimization: A fast algorithm for training support vector machines,". *Tech. Rep. MSR-TR-98-14*.
- Kuruvilla, A. ". (2012).). "Voice of the voiceless-innovative nutrition-health strategies to boost the morale of women prisoners.
- Myles, A. J. (n.d.). "An introduction to decision tree modeling." . *A Journal of the Chemometrics Society* 18.6 (2004):, (pp. 275-285.).
- Nanjundeswaraswamy, T. S. (2016). "Quality of work life components: A literature review." . *The International Journal of Indian Psychology* , 12-36.
- Quinlan, J. R. (1986.). *Induction of Decision Trees*. Machine Learning. (pp. 81-106.). Kluwer Academic Publishers, .
- Rajagiri, Y. R. (December 2011). School of Social Work, Tata Institute of Social Sciences, Deonar, . *Journal of Social Development Volume*.
- Ronald, Y. (2011). "Social Work Intervention with Prisoners: The case of Varhad in Maharashtra. " *Rajagiri Journal of Social Development* 3, 1&2 (2011): 105-114.
- Tayal, D. K. ((2015)). "Crime detection and criminal identification in India using data mining techniques." . *AI & society* 30.1 : , pp. 117-127.
- Ubun Thansatopornwatana, . (2016, ISBN: 978-1-5090-2258-8/16). *A Survey of Data Mining Techniques for Analyzing Crime Patterns*.

AUTHORS PROFILE



Mrs Sapna Singh kshatri is a research scholar of P.hd.(computer science and application)MSIT Department of MATS University, Raipur(C.G).She has completed her M.C.A. from Rungta college of Eng & Technology.This is her Third paper on the basis of her P.hd dissertation which is on Crime prediction.



Dr. Bhawana Narain is Associate Professor in MSIT Department of MATS University Raipur(C.G). she has done P.hd (computer science), M.phil(computer science).Her work experience is 15 years and area of interest is computer network (ad hoc, mesh) and digital image processing. She has published 41 papers in national and international journals and conference. Two books are published, she has worked as Co-pi in two minor projects. She has been working as state students co-coordinator of computer society of INDIA and awarded as best teacher and research by national organizer.