iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4369

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

A Linear Regression Model for the Prediction and Prevention of Cybercrimes in India

Dr. G. Nagusubramaniyan[1], Adithya Vikram Sakthivel[2],
[1]*Associate Professor, Department of Humanities and Social Sciences, National Institute of Technology, Tiruchirappalli*
[2]*Student, Department of Electronics and Communication Engineering, National Institute of Technology, Tiruchirappalli*

Abstract: Recent years reveal that there has been an unprecedented exponential growth in cybercrime globally. This paper intends to propose an accurate and effective multi-dimensional linear regression model to predict the number of individuals who will be arrested under cyber-crime in India based on data collected for each state in India's Federal Setup (excluding the Union Territories), by law enforcement agencies and governmental authorities. Further, this paper also explores certain parameters which contribute towards such criminal activities and suggests methods to combat the exponential growth of such crimes occurring in the country.

I. INTRODUCTION

Cyber-crimes can be broadly being labeled[1] as crimes against individuals, crimes against organizations, crimes against property, crimes against society and cybercrimes emanating from computer networks. These crimes may use various methods from hacking/cracking to spreading inaccurate and deceptive material online.

Cybercrimes can be defined as: "Offences that are committed against individuals or groups of individuals with a criminal motive to intentionally harm the reputation of the victim or cause physical or mental harm, or loss, to the victim directly or indirectly, using modern telecommunication networks such as Internet (networks including but not limited to Chat rooms, emails, notice boards and groups) and mobile phones (Bluetooth/SMS/MMS)"[2]. Debarati Halder and K. Jaishankar (2011) further defined cybercrime as a new phenomenon from the perspective of gender and defined 'cybercrime against women' as "Crimes targeted against women with a motive to intentionally harm the victim psychologically and physically, using modern telecommunication networks such as internet and mobile phones"[2]. In simpler terms, cybercrimes can be defined as an unlawful activity in which a computer is the object of the crime, or is used as a tool to commit an offence[3]. Any use of a computer as an appendage for further illegal ends, like doing fraud, child pornography and intellectual property, stealing identities or privacy violations[4]. Being a new field of criminology, some aspects of cybercrimes are only recently being discovered, and new methods are developed by computer-savvy criminals at a high frequency.

II. NEED FOR THE STUDY

Cybercrime in India has been growing rapidly. The Information Technology Act, 2000 passed by the Parliament of India in May 2000, had aimed to curb cybercrimes and to provide a legal framework for e-commerce transactions[5]. In 2001, India and United States had set up an India-US cyber security forum as part of their counter-terrorism dialogue. Cybercrime cases in India, registered under the IT Act, increased at a rate of 300 per cent between 2011 and 2014[6]. In 2015, there were 11,592 cases of cybercrime registered in India, making it a major threat which needs to be addressed in the public domain[7].

The research work carried out by S. Hinduja and J. W. Patchin (2008)[8], who studied cyberbullying in North America, noting the sharp rise in such activities over the last few years, and developed a logistic regression model for the identification and prevention of this antisocial activity, one could state that a similar approach could be developed for a general overview of cybercrime in India, as in most cases, cyberbullying can be classified as cybercrime, and it can broadly be assumed that the general nature of cybercrime is similar on a global scale. Also the Federal Bureau of Investigation's (FBI) show a drastic increase in the rate of cybercrimes committed globally with increase in time[9].

III. METHODOLOGY

This study covers twenty-eight states of India, covering all the directions and from the states mentioned, data have been collected from the data available with National Crime Records Bureau (NCRB), Indian Police Service[10] and the Ministry of Home Affairs, Government of India[11] as well as other agencies of the Government of India[12] for the time period covering 2012-2016. To analyze

and validate the data collected, a multidimensional linear regression model was formulated by using Matlab 2016. Further a technique to review findings in the literature involving in the specific statistical methods in integrating the findings from different empirical studies, known as meta-analysis approach, is adopted for verifying the data collected for the study. The p-value, along with the R^2 value, are calculated. Percentage analysis is utilized.

IV. OBJECTIVES

A. To evaluate the causes of cyber-crime based on this model.
B. To identify the parameters that can accurately be used to predict cyber-crime.
C. To develop an efficient linear regression model to accurately predict the number of people arrested under cyber-crime.

V. HYPOTHESES

A. There is no relationship between the number of police personal per state and the number of people detained under charges of cyber-crime.
B. There is no relationship between the overall literacy rate of the state and the number of people arrested under cyber-crime.
C. There is no relationship between the unemployment rate of the state and the number of people arrested under cyber-crime.

VI. LIMITATIONS

In the proposed linear regression model, data 0f twenty-eight states have been utilized, and the data of Union Territories of India have been excluded on benchmark bases.

VII. DATA ANALYSIS

Table 1: collected data with percentage analysis

S.No.	State	Number of Police Officers	Unemployment (per 1000 people)	Literacy Rate (Percentage)	Number of People Arrested Under Cyber Crime
1	Andhra Pradesh	89790 (7.09%)	39	67.4%	665 (11.77%)
2	Arunachal Pradesh	7155 (0.565%)	89	66.95%	2 (0.035%)
3	Assam	27867 (2.20%)	61	73.18%	351 (6.21%)
4	Bihar	55316 (4.37%)	60	63.82%	111 (1.96%)
5	Chhattisgarh	36774 (2.90%)	19	71.04%	105 (1.86%)
6	Goa	4722 (0.373%)	96	87.4%	4 (0.071%)
7	Gujarat	54955 (4.34%)	9	79.31%	174 (3.08%)
8	Haryana	36502 (2.88%)	47	76.64%	121 (2.14%)
9	Himachal Pradesh	8661 (0.684%)	106	83.78%	16 (0.283%)
10	Jammu & Kashmir	47051 (3.71%)	72	68.74%	4 (0.071%)
11	Jharkhand	41999 (3.31%)	77	67.63%	57 (1.01%)
12	Karnataka	64261 (5.07%)	125	75.6%	372 (6.58%)
13	Kerala	38569 (3.04%)	43	93.91%	283 (5.01%)
14	Madhya Pradesh	63612 (5.02%)	21	70.63%	386 (6.83%)
15	Maharashtra	156227 (12.33%)	57	82.91%	942 (16.67%)
16	Manipur	12199 (0.963%)	48	79.85%	3 (0.053%)
17	Meghalaya	5948 (0.469%)	30	75.48%	12 (0.212%)
18	Mizoram	3946 (0.304%)	85	91.58%	4 (0.071%)
19	Nagaland*	8604 (0.679%)	50	80.11%	0 (0%)
20	Orissa	30850 ((2.43%)	60	73.45%	17 (0.301%)
21	Punjab	57244 (4.52%)	71	76.68%	159 (2.81%)
22	Rajasthan	81008 (6.39%)	181	67.06%	248 (4.39%)
23	Sikkim	1905 (0.15%)	42	82.2%	2 (0.035%)
24	Tamil Nadu	101219 (7.99%)	197	80.33%	120 (2.12%)
25	Tripura	10441 (0.824%)		87.75%	1 (0.018%)
26	Uttar Pradesh**	140983 (11.13%)	74	69.72%	1223 (21.64%)
27	Uttarakhand	14144 (1.12%)	70	79.63%	57 (1%)
28	West Bengal	65186 (5.14%)	49	77.08%	212 (3.75%)
	Total	1267138	NA	NA	5651

*The state with the lowest number of individuals arrested under cyber crime**The state with the largest number of individuals arrested under cyber crime

It should be noted that the parenthesis is used to represent the values calculated in percentage analysis in table 1. From the collected data, it can be observed that the maximum number of arrests under cybercrime has occurred in Uttar Pradesh, while the lowest numbers has occurred in Nagaland. This can be explained by the overall high crime rates in Uttar Pradesh due to a low literacy rate and a relatively larger population. The exact opposite could be stated for Nagaland, an above average literacy rate, a small civilian population, and an overall low to nonexistent crime rate.

Table 2: Averages of the collected data

Average Number of Police Officers	Average Unemployment (per 1000 people)	Average Literacy Rate (Percentage)	Average Number of People Arrested Under Cyber Crime
45255	68	76.781%	202

A. Interpretation of Data

The data utilized in the constuction of the linear regression model are given in table 1. It should be noted that the distribution of the data is not Gaussian in nature as illustrated in images 1-4.

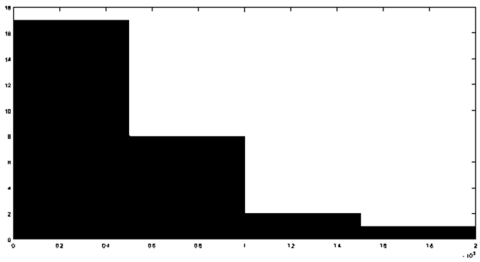


Figure 1: number of police officers per state

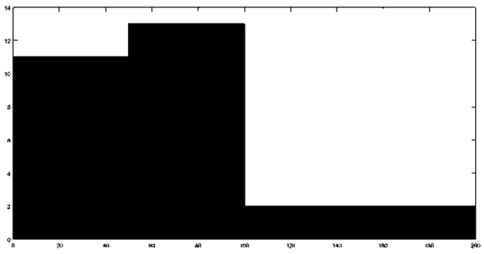


Figure 2: unemployment per 1000 people (per state)

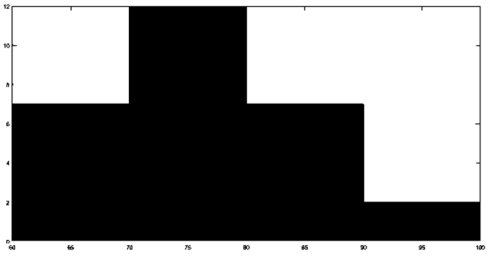


Figure 3: literacy rate (percentage) per tate

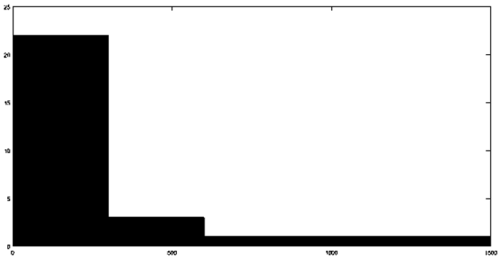


Figure 4: number of people arrested under cyber-crime (per state)

From figure 1, it can be noted that the distribution of the number of police officers employed by each state maintain a large right-hand skewness, this can be explained by the fact that the recruitment, training and equipping of law enforcement officers costs a comparatively large amount, an amount which most state governments might see as an unnecessary expense. It can be stated after observing figure 2, that a large amount of the distribution for unemployment is of the range of 60 to 100, as the national unemployment rate is of a similar value. The literacy rate, as illustrated by figure 3, maintains a distribution similar to Gaussian distribution. Despite there being a large number of cybercrimes reported in today's world, the number of cybercriminals arrested or identified is relatively low, as explained earlier, a single cybercriminal could commit several cyberattacks, this is illustrated in figure 4. The following Linear Regression Model has been constructed from the collected data in table 1, which we analyzed

$$f(X,Y,Z) = -1.2012 \times 10^{3} + 0.032(X) + 3.7746(Y) + 13.4309(Z) - 1.3705 \times 10^{-4}(XY) - 3.182 \times 10^{-4}(XZ) - 0.0405(YZ) + 1.6144 \times 10^{-6}(XYZ)$$

X = number of police officers per state

Y = unemployment rate per state

Z = literacy rate per state

$f(X, Y, Z)$ = number of people arrested under cybercrime per state

This linear regression model can alternatively be represented as shown in table 3, while table 4 gives some statistical measurements linked to the proposed linear regression model.

Table 3: Coefficients of the linear egression model

Parameter	Coefficient
x-axis cutoff	-1201.2
Number of Police Officers	0.032
Unemployment	3.7746
Literacy Rate	13.4309
(Number of Police Officers)*(Unemployment)	-0.00013705
(Number of Police Officers)*(Literacy Rate)	-0.0003182
(Unemployment)*(Literacy Rate)	-0.0405
(Number of Police Officers)*(Unemployment)*(Literacy Rate)	0.0000016144

Table 4: measurements linked to the proposed linear regression model

Measure	Value
R^2 *statistics* of the overall model	0.7636
$p - value$ of the overall model	4.2342×10^{-5}

VIII. RESULTS

From the p-value (approximately 0.00004), which is lesser than 0.05 (the maximum accepted p-value), we can conclude the following:

A. The null hypothesis that there is no relationship between the number of police personal per state and the number of people detained under charges of cyber-crime has been proven false.
B. The null hypothesis that there is no relationship between the overall literacy rate of the state and the number of people arrested under cyber-crime has been proven false.
C. The null hypothesis that there is no relationship between the unemployment rate of the state and the number of people arrested under cyber-crime has been proven false.

This shows a clear-cut relationship between the predictor values and the target values. Also from the R^2 statistic, we can see that the model explains more than 76% of the values observed in the datasets, making the model quite accurate. All this was achieved despite the data utilized lacking a normal (Gaussian) distribution.

IX. FINDINGS

In relation to the proposed model, a larger number of police officers could indicate that more resources and personal are assigned to prevention of cyber-crimes. Also, a larger literacy rate indicates a larger tech savvy population, hence more sophisticated criminals. A large unemployment rate could indicate that more people are willing to go down a path of crime for survival. A combination of a high literacy rate and a large number of people who are unemployed could result in a larger number of people taking on the profession of a cyber-criminal as a large population of educated people need a method to overcome the aftereffects of unemployment. Hence, it can be stated with confidence, that cyber-crime could be reduced drastically if most of the population is employed, or conversely, the unemployment rate is kept to a minimum. It can also be stated from the model that a larger number of officers being employed by a State's police department, the more likely it is for cyber-criminals to be captured.

X. FUTURE RESEARCH

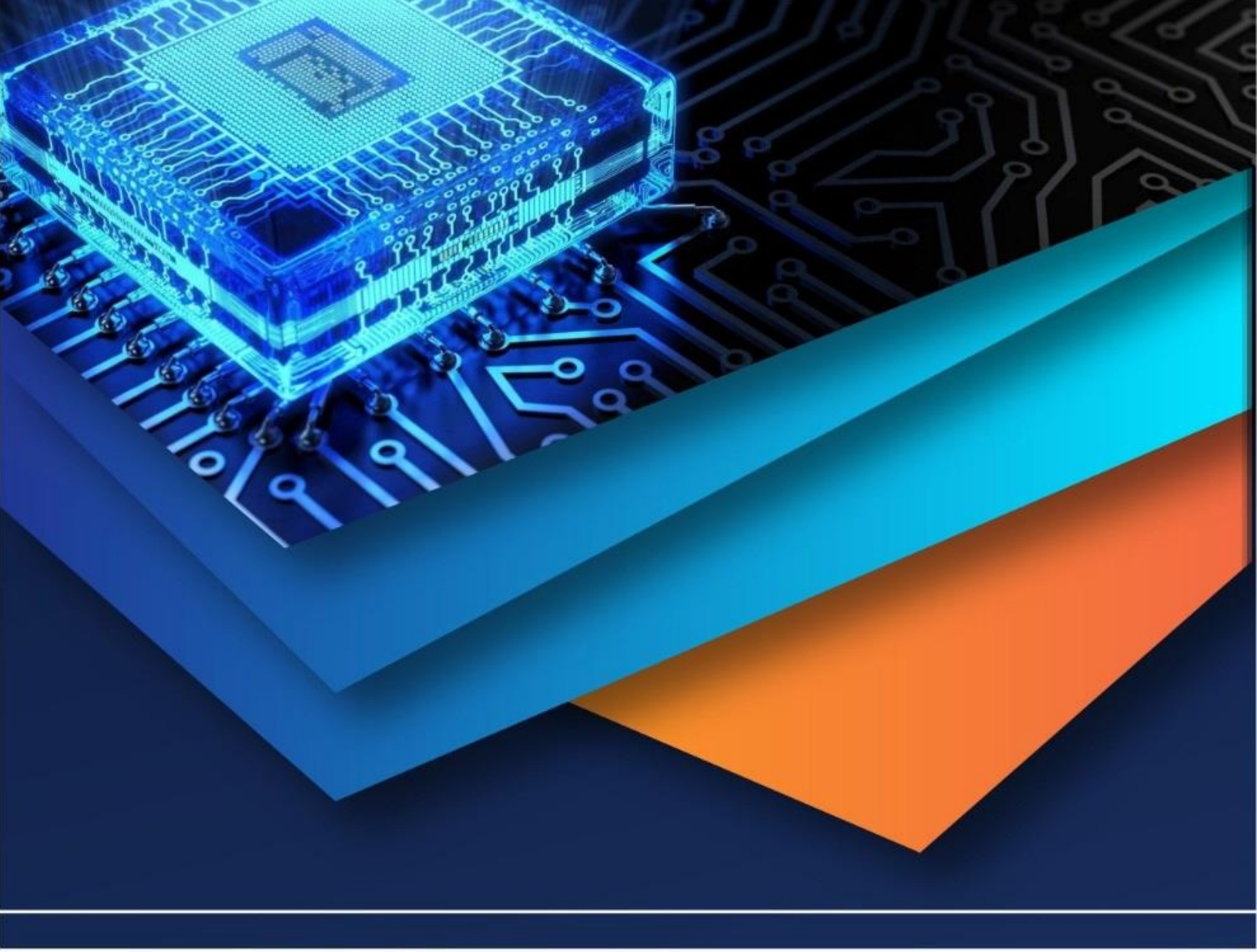
This model could be expanded upon in the near future with more parameters being considered, explaining the outlier points that exist in the present linear regression model, improving the accuracy of the model. It can also be stated with confidence that similar models could be constructed based on this one to identify and prevent similar criminal activities.

XI. CONCLUSION

It can be concluded with confidence that the proposed linear regression model can be used to predict a generalized number of individuals who would be arrested under cybercrime, and can also be utilized to identify issues that contribute to cyber-criminal activities, giving law enforcement agencies and governments a rough idea on how to curb its growth.

REFERENCE

[1] Cybercrime classification, [Online], Available: http://shodhganga.inflibnet.ac.in/ bitstream /10603/7829/12/12_chapter%203.pdf [29 September 2013]
[2] Halder, D., & Jaishankar, K. (2011) Cyber crime and the Victimization of Women: Laws, Rights, and Regulations. Hershey, PA, USA: IGI Global. ISBN 978-1-60960-830-
[3] Barkha, Rama Mohan, U. (2011) Cyber Law and Crimes, IT Act 2000 & Computer Crime analysis. (3rd ed.), ISBN: 978-93-81113-23-3
[4] Mayank, R. and Agarwal, P., 2016. Cybercrimes: An Indian perspective. International Journal of Engineering Science, 3863
[5] B.M.Gandhi. Indian Penal Code. India: Eastern Book Company. p. 41. ISBN 9788170128922
[6] Cybercrime in India up 300% in 3 years: Study". Economic Times. Bennett, Coleman & Co. Ltd. Retrieved 8 May 2017
[7] "11,592 cases of cyber crime registered in India in 2015: NCRB". LineMint. HT Media Ltd. Retrieved 8 May 20
[8] Hinduja, S. and Patchin, J.W., 2008. Cyberbullying: An exploratory analysis of factors related to offending and victimization. Deviant behavior, 29(2), pp.129-156
[9] Federal Bureau of Investigation's 2010 statistics, [Online], Available: www.fbi.gov [16 July 2013]
[10] Crime in India 2014, Chapter 18 (PDF), National Crime Records Bureau
[11] "Census 2011, Chapter 6 (State of Literacy), p.14" (PDF). Government of Indi
[12] "Report on Fifth Annual Employment-Unemployment Survey (2015-16)" (PDF). Ministry of Labour and Employment. p. 120. Retrieved 24 November 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY