IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

Analyze Crime against Women in India State Wise using Data Science

Mr K S Chandrasekaran M.E[1]., V Kaviya[2], M Reshma[3], D S Shugeertha. [4]

Associate Professor, Dept. of Computer Science Engineering, Saranathan College of Engineering, Trichy, India[1]

UG Students, Dept. of Computer Science Engineering, Saranathan College of Engineering, Trichy, India[2 3 4]

ABSTRACT: The major goal of this research is to examine the crimes that occur in each area and forecast future crime patterns using crime data from the last fourteen years (2001-2014). The National Crime Records Bureau (NCRB) has disclosed the information, which is available on the Open Government Data Platform India (OGD). This information can be utilised to learn more about the various crime trends that are occurring across the country. Molestation, rapes, dowry, exploitation, kidnapping, relative cruelty, and other crimes are included in the dataset. To acquire crime trends, an EDA analytical data mining approach will be used, and linear regression and time series modelling will be used to predict and show crime rates. This article provides an in-depth examination of crimes against women.Criminal analysis is a systematic strategy to finding and analysing criminal patterns and trends. Crime data analysts can assist law enforcement personnel in speeding up the process of solving crimes as the use of digital systems grows. A system can examine previously unknown, usable information from unstructured data using the notion of data mining. Predictive policing entails identifying criminals using analytical and predictive tools, and it has been found to be fairly effective in doing so. Because of the rising crime rate over time, the system would have to deal with a massive amount of crime data stored in warehouses, which would be extremely difficult to evaluate manually, and criminals are getting more technologically advanced, so there will be additional challenges.

KEYWORDS: EDA Analysis, Time series Modelling, Linear regression

I. INTRODUCTION

According to the National Crime Records Bureau, crimes against women have increased dramatically in recent years, including dowry, kidnapping, insult to modesty, rape, assault, girl importation, and relative cruelty. The administration has made it a high priority to uphold law and order in order to curb the rising rate of crime against women. Crimes against women have always been a severe concern in India, prompting different legislative regulations and efforts to combat it. With each passing year, more crime reports are produced, resulting in a massive volume of data. Such information can be utilised to develop analyses and statistics that will assist governments and non-governmental organisations in implementing appropriate schemes and policies. Uttar Pradesh has the highest rate of cognisable crime in 2012.

The National Crime Records Bureau computed the rates as the number of incidences per 100,000 people. Each state's crime rate is distinct, so the measures to be implemented must be different as well. This study primarily focuses on the analysis of crime against women, which will aid in the generation of predictions and, as a result, may assist in the generation of useful information from current data. The first phase entails obtaining data from the government and processing it for further analysis and prediction.

A rising crime rate has become a major source of concern, threatening to stifle the development of effective government. Crimes are neither systematic nor random, thus they can't be studied. We cannot investigate the victims of crime, but we can investigate the location where the crime occurred or occurred. One of the big challenges is that crime is becoming more intense and complicated. In recent years, one of the social issues affecting the nature of life and economic development in a community has been crime. There are several sorts of crime, including property crime (theft, burglary, and robbery) and crime of aggression (homicides, assaults and rape). Law enforcement has been able to acquire detailed criminal data because to the availability of information technologies. With the rise in crime, crime analysis is required, which includes measures and procedures aimed at lowering the risk of crime. Both quantitative and qualitative methodologies can be used to investigate crime. In anticipating the future of criminal activity, qualitative tools such as scenario writing and environmental scanning are useful. In the meanwhile, a quantitative method is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

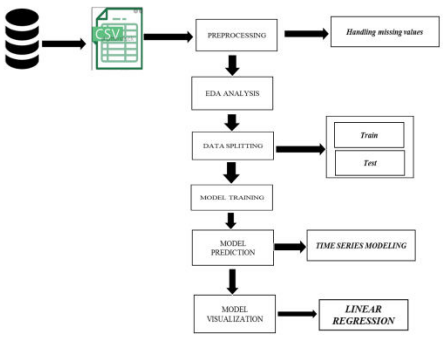
employed to forecast future crime rates. Furthermore, crime analysis is a useful tool for analysing and identifying criminal patterns. The task of uncovering and detecting crimes, as well as their links to criminals, is part of crime prevention.

II. RELATED WORK

In[1] AuthorsRohit Patil1, Muzamil Kacchi2, Pranali Gavali3, Komal PimpariaCriminal analysis is a methodical approach for identifying and analysing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, system can analyse previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analysed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In this paper, the main focus is on the review of algorithms and techniques used for identify the criminals. In [2] authors Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav Crimes are a social irritation and cost our society deeply in several ways. Any research that can help in solving crimes quickly will pay for itself. About 10% of the criminals commit about 50% of the crimes. The system is trained by feeding previous year's 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. As per data of Indian statistics, which gives data of various crime of past 14 years (2001-2014) a regression model is created and the crime rate for the following years in various states can be predicted. The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviours in crime pattern are difficult to explain. This paper illustrates how social development may lead to crime prevention. We have used supervised, semi-supervised and unsupervised learning technique on the crime records for knowledge discovery and to help in increasing the predictive accuracy ofthe crime. In[3] authors,Nurul Hazwani Mohd Shamsuddin1 , Nor Azizah Ali2 , Razana Alwee In the recent past, crime analyses are required to reveal the complexities in the crime dataset. This process will help the parties that involve in law enforcement in arresting offenders and directing the crime prevention strategies. The ability to predict the future crimes based on the location, pattern and time can serve as a valuable source of knowledge for them either from strategic or tactical perspectives. Nevertheless, to predict future crime accurately with a better performance, it is a challenging task because of the increasing numbers of crime in present days. Therefore, crime prediction method is important to identify the future crime and reduces the numbers of crime. Currently, some researchers have been conducted a study to predict crime based on particular inputs. The performance of prediction models can be evaluated using a variety of different prediction methods such as support vector machine, multivariate time series and artificial neural network. However, there are still some limitations on their findings to provide an accurate prediction for the location of crimes. A large number of research papers on this topic have already been published previously. Thus, in this paper, we thoroughly review each of them and summarized the outcomes. Our objective is to identify current implementations of crime prediction method and the possibility to enhance it for future needs.

III. PROPOSED SYSTEM

SYSTEM ARCHITECTURE



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

A. *Data Selection*

- Initial the input data was collected from dataset repository.
- In our process, the crime in India dataset is used.
- The data selection is the process of analyzing the crime.
- This dataset contains complete information about various aspects of crimes happened in India from 2001.
- There are many factors that can be analyzed from this dataset.
- In our process, we have to take the crimes of women.
- In python, we have to read our input dataset with the help of panda's package.

B. *Data pre-processing:*

- Data pre-processing is the process of removing the unwanted data from the dataset.
- Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning.
- This step also includes cleaning the dataset by removing irrelevant or corrupted data that can affect the accuracy of the dataset, which makes it more efficient.Missing data removal
- Encoding Categorical data
- Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0.
- Missing and duplicate values were removed and data was cleaned of any abnormalities.
- Encoding Categorical data: That categorical data is defined as variables with a finite set of label values.
- That most machine learning algorithms require numerical input and output variables.

C. EDA Analysis:

- Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.
- EDA PROCESS: Exploratory Data Analysis or EDA is used to take insights from the data. Data Scientists and Analysts try to find different patterns, relations, and anomalies in the data using some statistical graphs and other visualization techniques. Following things are part of EDA: Get maximum insights from a data set.
- EDA or Exploratory analysis is a type of unsupervised learning, which is to maximize the analyst's insight into a data set and into the underlying structure of a data set.
- Steps to follow in EDA:

1. Understanding the dataset
2. Cleaning the dataset
3. Analysis the relationship between the variable.
   - Here we use different type of charts like bar chart, pie chart, lines charts to analysis them easily to understand them.

D. DATA SPLITTING:

- During the machine learning process, data are needed so that learning can take place.
- In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works.
- In our process, we considered 70% of the dataset to be the training data and the remaining 30% to be the testing data.
- Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes.
- One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

E. Model Training/Prediction:

- This is the stage where the ML algorithm is trained by feeding datasets.
- This is the stage where the learning takes place. Consistent training can significantly improve the prediction rate of the ML model.
- The weights of the model must be initialized randomly. This way the algorithm will learn to adjust the weights accordingly.
- We train and test the dataset and make them to give the correct output or prediction for the crime.
- Prediction Model building step involves working on different statistical method, training those models with the help of test and then use the same to predict values out of it to get some the useful information which can help the end user in some manner, let it be predicting crimes or detect a safe zone.
- As for prediction we use two types of models

1. Time Series Modelling
2. Linear regression
    - In order to forecast the crime rate for future years, linear regression technique is being used. This technique consists of a dependent and an independent variable. The linear regression line has an equation of the form y = mX + c, where m is the slope of the line, c is the coefficient of the line, X is the independent variable and y is the dependent variable. Here, the independent variable (X) is Year and the dependent variable (Y) will be the rate of specific crime from the dataset. The core idea is obtaining a line that best fits the data which can be used to predict any new feature value. The best fit line is the one for which total prediction error (all data points) are as small as possible. This best fitting line is called the regression line.

F. Model Visualization:

- Data visualization is essential for representing insights from data in a graphical manner.
- With the large amount of data in dataset, one of the greatest challenge is to easily communicate the hidden patterns and findings in an easy and understandable manner.
- To visualize the data, there are many visualization techniques available. Some of techniques that we are utilizing for the project are line chart, bar chart, Pie chart, scatterplot.
- This helps us in better understanding and accurate interpretations of useful insights from the raw dataset with the help of data analysis and data visualization.

G. RESULT GENERATION:

The essential step in any machine learning model is to evaluate the accuracy of the model. The Mean Squared Error, Mean absolute error, Root Mean Squared Error.

1. The Mean absolute error represents the average of the absolute difference between the actual and predicted values in the dataset. It measures the average of the residuals in the dataset.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}|$$

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

Where,

$\hat{y}$ – *predicted value of y*

$\bar{y}$ – *mean value of y*

2. Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

3. Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

IV. PERFORMANCE OUTPUT

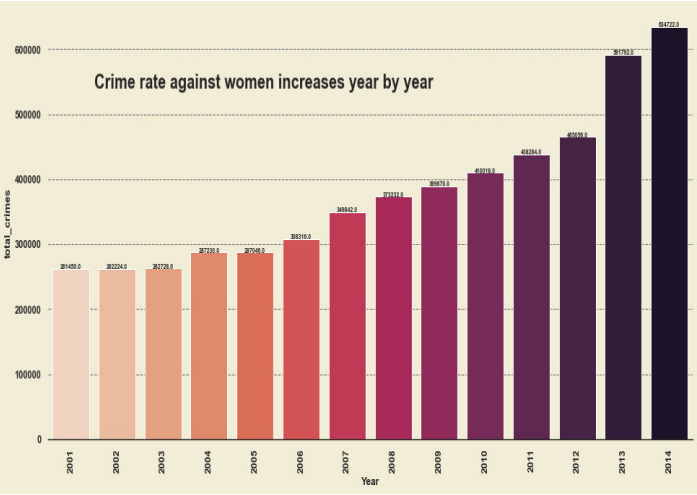


Fig 1:crime rate increments year by year

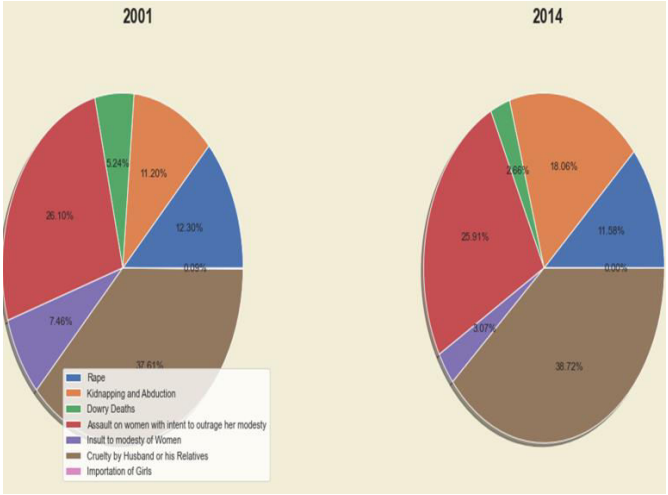


Fig 2:Comparison Between 2001 and 2014

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

TOP FIVE STATES WITH HIGHEST NUMBER OF EACH CRIME:

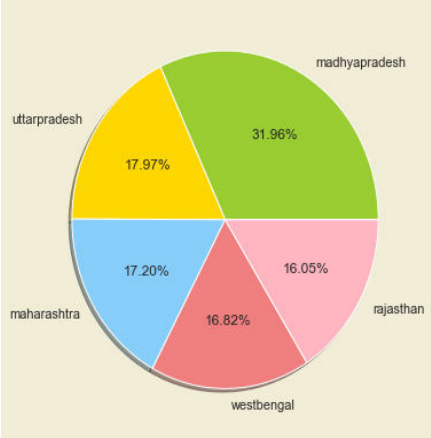


Fig 3: Top 5 states against rape

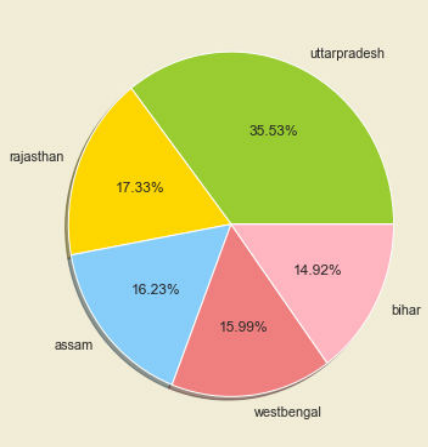


Fig 4 :Top 5 states against kidnapping and abduction

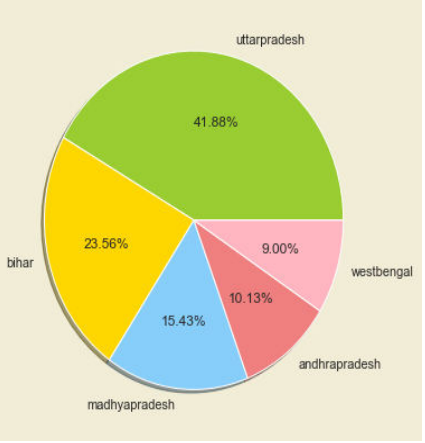


Fig 5 : Top 5 states against dowry Death

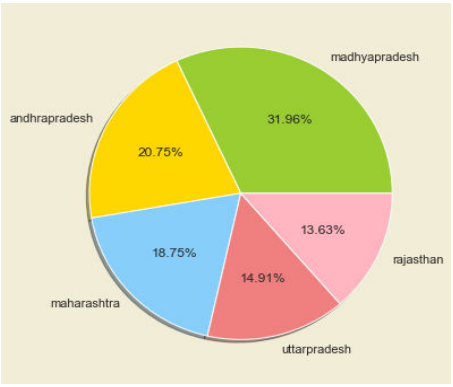


Fig 6 : Top 5 States against Assault on women

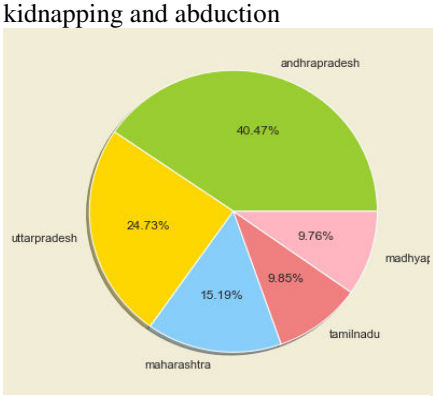


Fig 7:Top 5 states against Insult on women

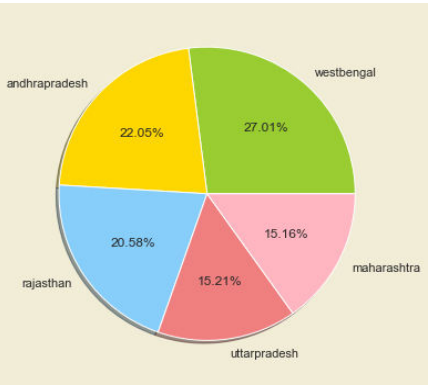


Fig 8:Cruetly by husband or his relative

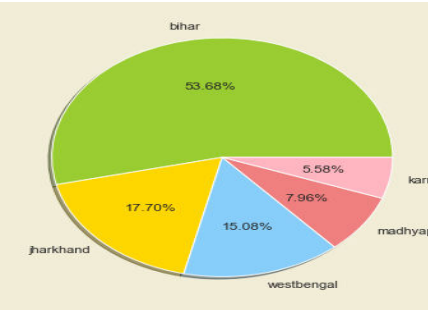


Fig 9: Importation of girls

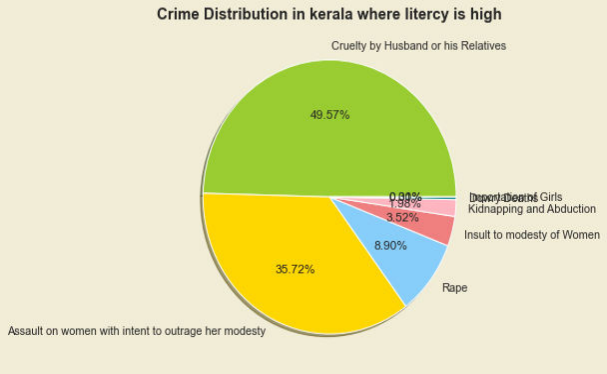
International Journal of Innovative Research in Computer and Communication Engineering

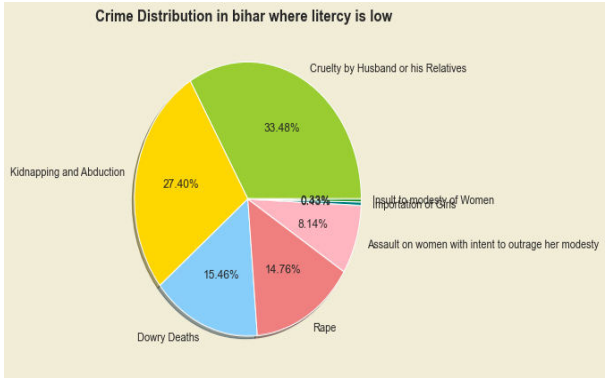
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

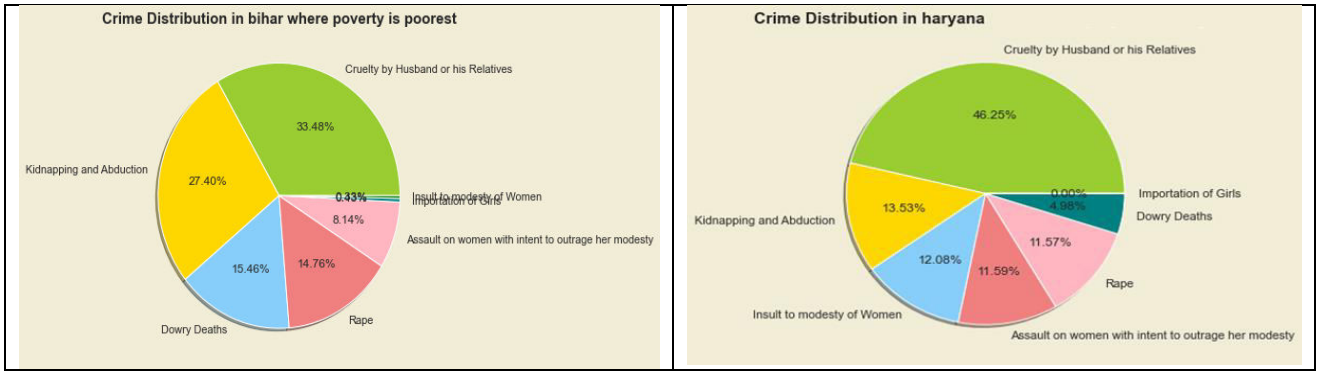
| DOI: 10.15680/IJIRCCE.2022.1006087 |

THE IMPACT OF LITERACY AGAINST WOMEN





THE IMPACT OF POVERTY AGAINST WOMEN



PREDICTED VALUES

Year	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
2012	2682	2806	1008	9632	7428	26778	0
2013	3270	3190	984	13860	9404	30168	0
2014	1922	1442	430	9094	5298	12724	0

	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls
2012	1991	4710	1343	8773	6660	33448	0
2013	2516	6256	1686	11445	9398	34498	0
2014	3444	6356	1782	14488	14366	37007	0

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

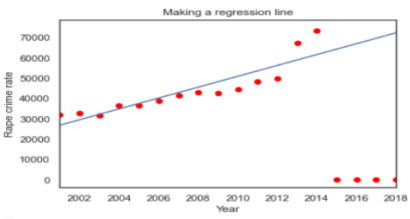
|| Volume 10, Issue 6, June 2022 ||

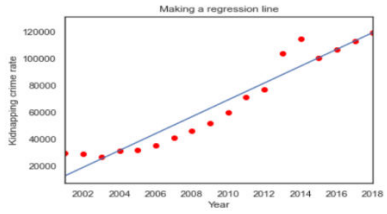
| DOI: 10.15680/IJIRCCE.2022.1006087 |

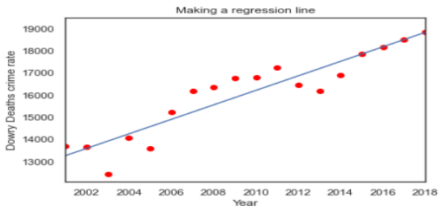
FORECASTING OF CRIMES

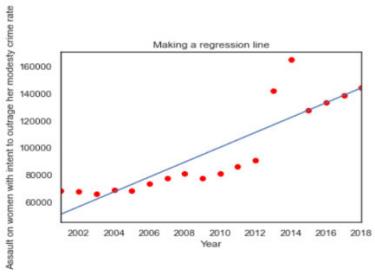
	Rape	Kidnapping and Abduction	Dowry Deaths	Assault on women with intent to outrage her modesty	Insult to modesty of Women	Cruelty by Husband or his Relatives	Importation of Girls	Total_crimes
2015	2276	3205	1041	12918	6691	19008	0	45139
2016	1592	2481	1056	6695	6705	11688	0	30219
2017	2710	4133	1652	9728	7961	24241	0	50427
2018	2494	3234	1400	5792	8037	26519	0	47479
2019	3706	4052	1423	12148	9304	37618	0	68255
2020	2854	2220	676	9812	7889	28612	0	52066
2021	3154	2875	690	15652	7890	27357	0	57621
2022	1669	1570	405	10072	6117	10985	0	30818

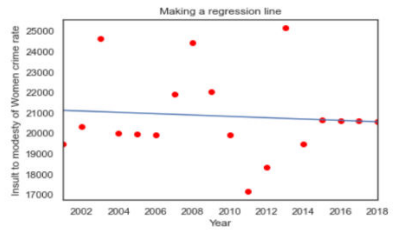
LINEAR REGRESSION

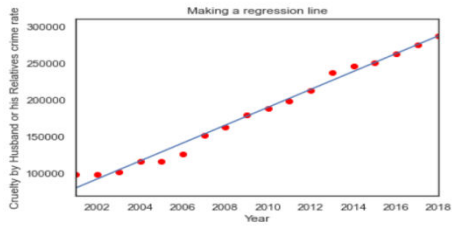


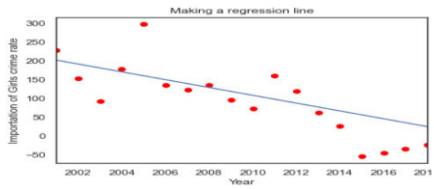












Rape:

```
Mean Absolute Error: 3244.349140321215
Mean Squared Error: 14873060.569540225
Root Mean Squared Error: 3856.5607177302713
```

Kidnapping and abduction:

```
Mean Absolute Error: 11333.858846365785
Mean Squared Error: 132727520.4209841
Root Mean Squared Error: 11520.743049863759
```

Dowry deaths:

```
Mean Absolute Error: 988.9805879090369
Mean Squared Error: 978809.4809039752
Root Mean Squared Error: 989.3480079850442
```

Assault on women with intent to outrage her modesty:

```
Mean Absolute Error: 12004.746533554047
Mean Squared Error: 176436158.74626437
Root Mean Squared Error: 13282.927340999211
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006087 |

Insult to modesty of women:

```
Mean Absolute Error: 1103.5047143649426
Mean Squared Error: 1234611.364526479
Root Mean Squared Error: 1111.1306694203338
```

Importation of girls:

```
Mean Absolute Error: 58.04631170271872
Mean Squared Error: 7785.493597516889
Root Mean Squared Error: 88.23544411129174
```

IV. CONCLUSION AND FUTURE WORK

India is one of the worst countries for women in the world. And in recent time it is getting worse when crimes against women are happening in the name of their religion. According to scroll.in, 26 crimes against women are reported every hour or one complaint every two minutes.Andhra Pradesh had registered the most crimes against women from 2001 to 2014. Uttar Pradesh and West Bengal were the second and third states with the most crimes against women. Andhra Pradesh also reported the highest number of cases of insult to the modesty of Women. Madhya Pradesh reported the most number of Rape cases and Assault on her modest cases. Madhya Pradesh was also the fifth highest state with the most number of registered crime cases against women.Uttar Pradesh reported the most number of Kidnapping and Abduction and Dowry Deaths. Seriously, still dowry! The world is in the 21st century, but UP is still in the 18th century.

REFERENCES

1. Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225-230). IEEE.

2. Suhong Kim; Param Joshi; Parminder Singh Kalsi; Pooya Taheri. Crime Analysis through Machine Learning, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).

3. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.

4.Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE.

5.Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri Fraser. Crime Analysis Through Machine Learning, November 2018, International College, Simon Fraser University.

6.Alkesh Bharati, Dr Sarvanaguru RA.K, Crime Prediction and Analysis Using Machine Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | Sep 2018.

7.Arpita Nagpal, Aman Jatain, Deepti Gaur. Review based on Data Clustering Algorithms, Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013).

8.Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi. Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, vol 4, No.8, 2015.

9. Lawrence McClendon and Natarajan Meghanathan*. USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA, Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015.

10.Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details