Predictive Analysis of Crimes against Women in India

A Dissertation submitted in the partial fulfillment for the award of

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

by

Anjana Kumari

Roll no. 2k11/CSE/22

Under the Essential Guidance of

Dr. Kapil Sharma

Department of Computer Engineering

Delhi Technological University

New Delhi

2013-2014

DECLARATION

I, **Anjana Kumari**, student of M.Tech, Delhi Technological University, hereby declare that the thesis entitled "**Predictive Analysis of Crimes against Women in India"** which is being submitted to the **Delhi Technological University**, in partial fulfillment of the requirements for the award of degree of **Master of Technology in Computer Science & Engineering** is an authentic work carried out by me. The material contained in this thesis has not been submitted to any university or institution for the award of any degree.

Anjana Kumari
Department of Computer Engineering
Delhi Technological University,
Delhi.

CERTIFICATE

This is to certify that the thesis entitled "**Predictive Analysis of Crimes against Women in India"** submitted by **Anjana kumari (Roll Number: 2K11/CSE/22),** in partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science & Engineering, is an authentic work carried out by her under my guidance. The content embodied in this thesis has not been submitted by her earlier to any institution or organization for any degree or diploma to the best of my knowledge and belief.

Dr. Kapil Sharma
(Project Guide)
Associate Professor
Department of Computer Engineering
Delhi Technological University
Delhi-110042

ACKNOWLEDGEMENT

I take this opportunity to express my deepest gratitude and appreciation to all those who have helped me directly or indirectly towards the successful completion of this thesis.

Foremost, I would like to express my sincere gratitude to my guide **Dr. Kapil Sharma**, **Associate Professor, Department of Computer Engineering**, **Delhi Technological University, Delhi** whose benevolent guidance, constant support, encouragement and valuable suggestions throughout the course of my work helped me successfully complete this thesis. Without his continuous support and interest, this thesis would not have been the same as presented here.

Besides my guide, I would like to thank **Head of Department**(Computer Science), the entire teaching and non-teaching staff in the Department of Computer Science, DTU for all their help during my course of work.

I am very much grateful to **Dr. Manju Bala**, Assistant Professor, IP college for Women, Delhi University, for her kind supervision and guidance throughout the project work.

ANJANA KUMARI

M. Tech (CSE)

Department of computer science
Delhi technological university

TABLE OF CONTENTS

List of Figures

List of Tables

ABBREVIATIONS USED

1.	NCRB	-	National Crime Records Bureau
2.	CR	-	Crime rate
3.	WEKA	-	Waikato Environment for Knowledge Analysis
4.	ANOVA	-	Analysis of Variances
5.	NCW	-	National Commission of Women
6.	CII	-	Crime in India
7.	UN	-	United Nation
8.	IPC	-	Indian Penal Code
9.	SLL	-	Special & Local Laws
10.	GNU	-	GNU's Not Unix
11.	GP	-	Gaussian Process
12.	LR	-	Linear regression
13.	MLP	-	Multilayer Perceptron
14.	ME	-	Mean Error
15.	MAE	-	Mean Absolute Error
16.	MSE	-	Mean Square Error
17.	PE	-	Percentage Error
18.	MSB	-	Mean Square Between
19.	MPE	-	Mean Percentage Error
20.	MAPE	-	Mean Absolute Percentage Error
21.	SSQ	-	Sum of Squares

ABSTRACT

In this thesis work, predictive analysis of crimes against women in India has been performed for crimes domestic violence, rape and dowry system. Time Series Forecasting is done using Gaussian Method, Linear Regression Model and Multilayer Perceptron Model using WEKA tool. ANOVA is used to compare the variances of the three methods to identify the most effective algorithm for prediction. Data has been collected from National Crime Records Bureau, which compiles all crime related data from all over country for research and analysis purposes. Various acts related to crime against women and the amendments done to them are also compiled. The analysis shows us how effective the laws are in curbing theses crimes. After assessing the results we found out that Multilayer Perceptron algorithm used for prediction is giving the most effective result as evident from ANOVA p- value. Crime rate, defined as the number of crimes per one lakh population, is used for analysis of crimes. 'Crime rate' for crimes committed against women for any year has been calculated using only female population based on projected mid-year female population for that year. Projected population is calculated based on population Census of India 2001, as supplied by Demographic Division, Registrar General of India, Ministry of Home Affairs. It has been observed that the number of dowry deaths in year 2013 is on a decline where as rape and dowry prohibition cases show an increasing trend after 2010.

Chapter 1

INTRODUCTION

India is constantly progressing towards development. In a developing country like India, safety and security of women plays a key role in reflecting the image of our country before the world. The status of women in a country determines how secure a country is. However, in the recent years there has been an enormous rise in the crimes against women, particularly rapes and sexual offence cases. Here the crimes sexual offences/rapes, dowry harassment and deaths and domestic violence against women have been analyzed.

Crime against woman can be defined as an act committed in violation of law against any woman causing physical violence, mental stress or sexual abuse to a woman. First time the violence on basis of gender was recognized as violation of human rights in the World Human Rights Conference in Vienna in the year 1993. In the United Nations statement in 1993, violence against women was defined as "any act of violence on basis of gender whose consequences causes' physical hurt, mental stress, sexual or emotional suffering to a woman". It also includes coercion of such acts like bullying or withdrawal of autonomy etc (Violence against Women).

Here the main focus is on the predictive analysis of crimes against women. Forecasting of crime rate (CR) which would be registered after the enforcement of the Acts and amendments for crimes against women have been done so that the impact analysis of the amendments can be carried out. Then conclusions have been made and some suggestions are given based on prediction, what necessary steps can be taken to control such crimes in India. Crime Predictive analysis methods help police to work more sensibly. The objective of predictive analysis is to develop efficient planning and tactics that can aid in prevention of crime or make investigation pains very less (Perry, McInnis, Price, Smith, & Hollywood, 2013).

It is a naked truth that all crimes are not reported to the police due to a variety of reasons. Some of the reasons are social and some are administrative. For example, violence against women, children or elderly, sexual offences against women, teenage or minor often goes unreported, because of social reasons as criminals are their own

relatives or family members. Many times crimes are not reported to the police because of the fear of resentment or harassment likely to be faced by the victims or the complainants at the police stations. Many times police personals themselves do not register complaint due to some administrative reasons. Such factors can vary across the country. Thus, the reported number of crimes does not always represent the actual number of incidents of crimes.

Government of India set up the National Commission for Women under the National Commission for Women Act in January 1992, with a mandate to fight for women related issues and their security (National Commission of Women).

In this thesis work data has been collected from National Crime Records Bureau (NCRB) for three major crimes against women i.e. rape, dowry prohibition act and domestic violence; and carried out predictive analysis so that proper preventive measures can be taken to curb these crimes.

Organization of Thesis

This report has been organized in following format:

- Chapter 2 summarizes the work done by other researchers in the same field.
- Chapter 3 provides the list of amendments done to the laws related to the crimes under analysis i.e. Domestic Violence, Dowry System and Rape.
- Chapter 4 gives the detailed description of the tools and techniques used in the research and analysis.
- Chapter 5 summarizes the results.
- Chapter 7 gives conclusion and future work related to this study.
- Finally, the references for this work are listed.

Chapter 2

LITERATURE REVIEW

2.1 Women in India

The condition of women in India has gone through many transformations over the centuries. It has ranged from equal position as men in society in the historical times to the profound status in the medieval period. Today many reformers have come up to the support of women for equal rights. In modern India, women have held high positions in India including that of the President, Prime Minister, Governor, Leader of the Opposition and Chief Ministers of many states of India (Women in India). However, the condition of women in Indian society have not changed much; still women keep on suffering barbarity such as domestic violence, rape, sexual assault, acid attack, dowry deaths etc.

The Muslim invasion on India brought the 'purdah' system. In some rural areas of the country, Devadasis were exploited physically and sexually in the name of God. (Violence against Women).

Customs such as Sati, Jauhar, and Devadasi were common in society earlier. Now many communities have barred these customs and are outdated in modern India. However, these old age customs are still being followed in some rural areas of India. Let us see what the terms like sati, purdah, devdasi etc refer to in India.

- Sati

Sati is an ancient, almost obsolete custom. In this ritual the widow was burnt alive on her husband's funeral pyre. The act of performing immolation was believed to be voluntary on the widow's part, but then it became a practice to burn women forcefully. After the British control on Indian subcontinent, this custom started to mark its presence, as women were frequently raped or kidnapped by the British forces. Sati was eliminated by the British in 1829; still after independence around forty cases of sati were registered. The popular Roop Kanwar case in Rajasthan showed the way to the Commission of Sati (Prevention) Act in 1987 (THE COMMISSION OF SATI (PREVENTION) ACT, 1987, 2006).

- Jauhar

  Jauhar is an age long tradition of deliberate massacre by self immolation by women in the family of warriors who are defeated in wars. They do so in order to avoid detention and subsequent assault by the enemy. Generally the practice was followed by the wives of defeated Rajput Kings. Mainly such practice was prevalent during Muslim invasion period (Women in India).

- Purdah

  Purdah is a custom in which women cover up themselves so as to hide their body from others. It enforces limitations on the women, and is a symbol of the slavery of women. Indian women had to use purdah during the Islamic empire in Indian Subcontinent, as a result of fear of Muslim invaders (Women in India).Still it is being followed in many urban and rural parts of the country.

- Devadasis

  Devadasi was being practised in southern areas of India. In this women were told to marry a idol of the temple. In the name of God, priests and others used to assault women. By 1988 India declared it as a crime, and laws were made to stop it completely (Devdasi, 2007).

  During the British Empire, many reformers and social activists such as Ram Mohan Roy, Ishwar Chandra Vidyasagar, Dayanand Saraswati and Jyotirao Phule fought for women to improve the plight of women in India. Raja Rammohan Roy's efforts led to the eradication of Sati custom in the year 1829. Ishwar Chandra Vidyasagar's movement for the development in the state of widows led to the formation of Widow Remarriage Act of 1856. Many women reformers also supported and fought women rights (Women in India).

  The crime statistics and police records show that the number of incidence of crimes against women are rising. Earlier people did not report the crimes to police due to various reasons.

  In today’s world violence against women has taken various forms, such as physical assaults on marital/live-in partners, separation/divorce sexual assault, stalking, date rape, and coercive control (DeKeseredy, 2012). These and other highly violent behaviors that many women from different backgrounds experience behind closed

doors and elsewhere, is raising a serious concern about women's safety across the world. Lets discuss some of these violent behaviors:

- Acid throwing

  A survey by Thomas Reuters Foundation has said that India is the fourth most dangerous place in the world for women to live in. Women are the victims of this cruel form of brutality and mutilation irrespective of the caste or religion. It is a deliberate crime intended to kill or disfigure permanently (The world's most dangerous countries for women, 2011). In India, acid attacks on women are generally done to take revenge. Acid is inexpensive, easily accessible, and the fastest way to cause destruction to a woman's life by causing the damage of face structure etc. leaving the victim to face social boycott & in many cases drive them to commit suicide.

- Female infanticide and foeticide

  In the recent years in India, the male-female sex ratio has skewed noticeably in favor of males, the main reason for this is that females are often killed before reaching maturity. Many experts and studies suggest that the skewed male to female ratio can be credited to female infanticides and sex-selective abortions.

  Modern medical technologies have played a major role in providing good health to all. However, technology like ultrasound scan often reveals the sex of the baby, allowing a mother to decide whether to abort female foetuses and later have a male child. This negligence of the practice is usually considered the main reason for the drop in the ratio of male to female children being born. Indian government passed a law in 1994 to stop this malpractice. However, in practice this law is generally ignored, and level of abortion of female foetuses remain high and the sex ratio at birth keeps getting more skewed. Female infanticide (killing of girl infants) is still common in some rural areas of the country (Tandon & Sharma, 2006).

- Child marriage

  Child marriage can be defined as marriage of girls and boys before attaining adulthood. Child marriage has been a widespread practice in some states of India and still continues. Many times very young brides are often married to older men; as a result they become widow at a very young age. Earlier, child widows were fated to a life of great torture, shaved heads, living in loneliness, and being boycott by society. Although child marriage was forbidden in 1860, it is still prevalent (Child Marriage in

India). To eradicate the evil of child marriage, Child Marriage Restraint Act was passed in 1929, with an objective to protect female child who are not in a position to bear the stress and strains of married life (The Child Marriage Restraint Act, 1929).

2.2 Crime against Women

In this thesis work, three major crimes against women, their victimization and current scenario in the country has been analyzed.

2.2.1 Domestic Violence

Domestic violence against women generally covers cruelty by a person or his family towards a married woman. Domestic violence Act also includes the elderly and children besides women (Acharya, 2012). In 1983, domestic violence was recognized as a specific criminal offence by the introduction of Section 498-A into the Indian Penal Code. Following types of cruelties are dealt with by these laws:

- behavior that is likely to compel a woman to suicide,
- conduct which is likely to cause severe injury to the life, body parts or health of the woman,
- harassment with the intention of forcing the woman or her relatives to get property or money.

Forms of "cruelty" recognized by the Courts -

i. continual denial of food,
ii. Insisting on unwilling sexual conduct,
iii. always locking a woman inside the house,
iv. Stopping woman from meeting her children, thereby causing psychological torture,
v. Physical violence,
vi. hurtful, demoralizing and putting down the woman with the intention of causing mental torture,
vii. Intimidating woman and threatening of divorce unless dowry is given.

With an objective to protect women from domestic violence, the Domestic Violence Act 2005 came into existence, enforced in October 2006. The Domestic Violence Act - 2005, is an act to provide more effective protection of the rights of women guaranteed under the Constitution who is victim of violence of any kind occurring within the family and for matters connected therewith (The Gazette of India, 2005).

Earlier violence by husband for dowry demand was registered under section 498A of IPC which came into existence in 1983 (National Commission of Women). Today Domestic violence has become an inherent part of the society. The causative factors could be the desire to gain control over another family member, the desire to exploit someone for monetary benefits, the desire to be in a commanding position all the time showcasing one's superiority and so on. On various occasions, emotional problems and social influence also add to the violence. The victims of domestic violence can be an elderly person, children or women (Felson, 2009).

a. Children

Violence against children involves physical, psychological abuse, injury and sexual abuse. Children who survive abuse and violence often suffer long term physical and psychological damage that impairs their ability to learn and socialize. The behavioral and psychological consequences of growing up in a violent environment can be just as devastating for children who are not directly abused themselves. Stress disorder, depression, anger are some of the after effects of violence on children (The Facts on Children and Domestic Violence, 2008).

Working in someone's home also involves the risk of violence. Child domestic workers, who are generally girl child, are often abused by their employers including physical punishment, sexual harassment and humiliation. According to survey conducted by International centre for research on women, the girls who are married before 18 were twice as likely to report being beaten, sexually abused, slapped or threatened by their husbands than girls who married later (Grovert, 2008).

b. Women

Domestic violence is a pattern of behavior which involves violence or other abuse by one person against another in a domestic perspective, such as in marriage or cohabitation (Grovert, 2008). Usually women, being considered weaker are abused physically or sexually in a relationship.

Victimized women have a tendency to remain silent, worried and psychologically disturbed after the occurrence of the torture. A psychological set back and distress because of domestic violence affects women's efficiency in all forms of life (Patel, 2013). The suicide case of such ill-treated women is also a fatal outcome and the number of such cases is increasing. A working woman may quit from work place

because of the ill-treatment at home. Her health may get worse if she is not well physically and mentally (Patel, 2013).

c. Elderly

The elderly abuse is one of the most ill-fated incidents for the elderly class in their lives. In this phase of life old people like to spend a peaceful life without any stress but are prone to such kind of disgraceful behaviour by the family or society. Ironically elderly class themselves are involved in harming each other. Many of the elderly men keep on beating and harassing their wives throughout their lives (Mishra & Patel, january-june 2013).

Some of the olds are exiled from home by their children; some are exploited socially and many times beaten to death. A sense of insecurity and uncertainty dodges them all the time. They are cut off from society in some cases where son and daughter-in-law do not let them interact and move around freely in the society. The old people are not cared properly and their health problems are ignored. Due to the violence and mental suffering they go through, some of them leave home and stay in old age homes like Help Age India and other Senior Citizen Home (Mishra & Patel, january-june 2013).

2.2.2 Dowry System

The payment of a **dowry** gift at the time of marriage, often monetary, has a long history in many parts of the India. Taking and giving dowry has been prohibited in India since 1961 under Indian civil law. In this Act, "dowry" means any property or valuable gifts given or agreed to be given either directly or indirectly by one party or parents of either party to the other party in a marriage (Dowry system in India).

The Government of India passed the Dowry Prohibition Act in 1961 under which the dowry demands in wedding arrangements were made illegal. Still many take and give high priced gifts in weddings, and many cases of dowry-related violence against women, suicides and murders have been reported in India. In the 1980s, several such cases were reported, hence government decided to make some amendments in the existing law.

In 1985, the Dowry Prohibition Rules (maintenance of lists of presents to the bride and bridegroom) were framed. As per these rules, a signed list of gift items should be maintained, given at the time of the marriage to the bride and the bridegroom. The list should contain a short description of each present, its

approximate value, the name of person who has given the gift and relationship to the recipient. Although, these rules are rarely enforced. Many forms of crimes like harassment, domestic violence and even death can be caused if the demands are not met. To protect women against this intimidation, the Indian government had enacted the Dowry Prohibition Act and the Protection of Women from Domestic Violence Act and cruelty under Sec 498A of the Indian Penal Code in 1983 (Dowry Law in India).

Consequently, Sections 304B and 498A of the Indian Penal Code were enacted, making it easier for the wife to seek rights from harassment by the husband's family. Often men's rights groups disapprove of anti-dowry laws, who blame women and their families of misusing the laws (What Is Section 498A Of The Indian penal code?).

Dowry deaths are deaths of women who are murdered or driven to suicide by continuous harassment and suffering by husband and in-laws in an attempt to extract more dowries (Dowry Death). Dowry death is considered one of the heinous crimes against women, along with rape, eve teasing, violence and acid throwing.

Government of India has taken strict actions against such crimes as a result there is a decline in percentage of dowry death in 2013. It decreased by 1.8% during 2013 as compared to 2012. The victims of dowry system are married women.

a. Married Women

Dowry violence is usually carried out by the husband or the in-laws in a bid to extract a higher dowry from the bride's family. The dowry money given at the time of marriage may be a large amount, but with time greediness of husbands and in-laws increases. This greediness leads to physical, mental or sexual violence against the bride (Dowry system in India).

Even though giving and taking dowry has been forbidden in India since 1961, the law enforcement has been a challenge. An amendment to the law in 1986 mandated that any death or violence within the first seven years of marriage would be tried relating it to dowry. Generally most cases of dowry violence go unreported. According to reports of National Crime Records Bureau, every hour a woman dies in India because of the dowry demands of in-laws and husband. The incidents of cases registered under Dowry prohibition Act has been increased by 17.8 % as compared to 2012 (Crime in India, 1971-2013).

2.2.3 Rape and other Sexual Offences

An act is defined as rape if the victim does not react, has no sexual orientation to the criminal and tries to resist the attempt of attacker. Rape in India has been described by Radha Kumar as one of India's most common crimes against women and by the UN's human-rights chief as a "national crisis". New Delhi has the highest rate of rape-reports among Indian cities. Marital rape is still not counted as a criminal offence. Crime statistics show that rape cases in India have doubled between 1990 and 2008 (Rape in India). According to the National Crime Records Bureau 2013 annual report, 24,923 rape cases were reported across India in 2012 and Delhi reported the highest crime rate of 18.6 against the national average of 5.7 during the year 2013 (Crime in India, 1971-2013).

Sexual harassment

Eve teasing usually refers to sexual harassment of women in public places such as the streets, road side places, public transportation, parks etc. Eve teasing includes verbal assaults such as making passes or unwelcome sexual jokes; nonverbal assaults such as showing obscene gestures, whistling, staring and winking; and physical assaults such as pinching and rubbing against women in public places. In 1987, the Indecent Representation of Women (Prohibition) Act was passed to forbid indecent depiction of women through advertisements, publications, writings, paintings or in any other manner (Haldar & Jaishanker, 2012).

In 1997, in a landmark judgement, the Supreme Court of India took a tough stand against sexual harassment of women in the workplace. The Court also laid down detailed guidelines for prevention and redressal of grievances. The National Commission for Women subsequently elaborated these guidelines into a Code of Conduct for employers.

The victims of rape can be anybody; here women and children victims are discussed.

a. Children

Today one in every three rape victims is below 18 years. Children rape survivors can experience a wide range of emotions following a sexual assault including weakness, depression, nervousness, shame, and fear. It destroys the whole life of children as it is very difficult for them to come out of it. Often children do not report the incident to the parents because of fear. But parents should raise an alarm as soon as they are

aware of it. According to Crime in India report 2013, Incest rape (rape by blood relatives) cases have increased by 36.7% from 392 cases in 2012 to 536 cases in 2013 as compared to 35.2% increase in overall rape cases. It is to be noted that 52.2% of incest rape victims (286 out of 548 victims) were in age group of 10-18 years. (Crime in India, 1971-2013)

b. **Women**

Sexual harassment may lead to temporary or prolonged stress and/or depression depending on the women's psychological abilities to cope and the type of harassment, and the social support or lack thereof. Psychologists and social workers report that severe continual sexual harassment can have the same psychological effects as rape or sexual assault. Victims who do not submit to harassment may also experience various forms of revenge, including loneliness and aggravation. According to NCRB report, there were 33,764 victims of rape out of 33,707 reported rape cases in the country during the year 2013. 13.1% (4,427 out of 33,764) of the total victims of rape were girls under 14 years of age, while 26.3% (8,877 victims out of 33,764) were teenage girls (14-18 years) and 46.1% (15,556 victims) were women in the age-group 18-30 years. (Crime in India, 1971-2013)

To fight the crimes specifically against women various legal provisions and amendments have been done in existing law to handle such crimes effectively. There are two categories of law:

- Indian Penal Code (IPC) – It include law against rape, kidnapping, assault, insult to modesty of woman etc.
- Special and Local Laws (SLL) - Special & Local Laws are those laws which applies to a specific geographic area, or a particular class of persons identified by name in the legislation. Many laws are create specifically for women like Immoral Trafficking, Dowry Prohibition Law 1961, Commission of Sati Prevention Act 1987 (Crime in India, 1971-2013).

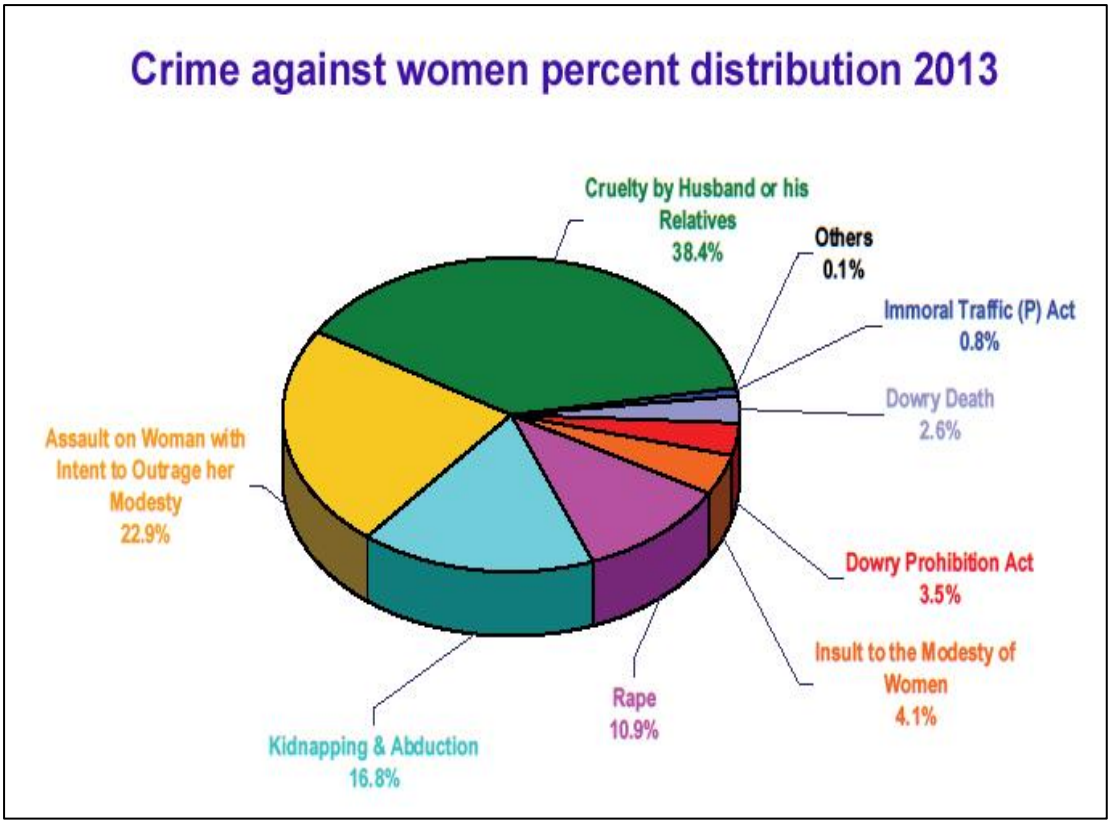

Figure 2.1 : Percentage distribution of crime against woman in 2013.
Source - Crime in India, NCRB

Figure 2.1 depicts the distribution of crime against women in 2013. A total of 3,09,546 cases of crimes against women were reported, out of which 38.4% of the total crimes were reported for domestic violence and 22.9% of the total crimes were reported for assault on woman with intent to outrage her modesty.

Chapter 3

AMENDMENTS IN LAW

In the previous chapter three crimes against women i.e. domestic violence, dowry violence and deaths and rape and sexual offences has been studied. The following amendments are done by Government of India to curb these crimes. After predictive analysis, the actual impact of these laws on curbing the crimes like domestic violence, dowry violence and deaths and rape and sexual offences can be found. Table 3.1 lists the laws and amendments made to them (Deshmukh, 2012) .

Table 3.1 :Amendments in Laws related to crimes against women

Case I(Rape and sexual offences)	S375 in 1860 IPC
	Addition of 164A in code of criminal procedure 1973
	Criminal Law (amendment)1983
	Sexual Offences Act 2003, came into force in 2004
	Criminal Law (Amendment) Act 2013
CaseII (Dowry Law)	Dowry prohibition Act 1961
	IPC498A(1983) protection from cruelty for soughting dowry
	IPC1986 (include dowry death)S304B
Case III (Domestic violence)	Domestic violence Act 2005 operational in 2006

Chapter 4

TOOLS AND TECHNIQUES USED

4.1 WEKA

Weka (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from a Java code as WEKA is written in Java. It was developed at the University, New Zealand. Weka is freely available under the GNU General Public License. Weka contains tools for data pre-processing, classification, clustering and forecasting. There are many advantages of using Weka for data mining tasks, visualization or forecasting (WEKA Manual). Some of them are:

- It is freely available under the GNU General Public License
- It is portable as it is fully implemented in the Java.
- It is a complete collection of data pre-processing, modellings and visualization techniques
- easy to use due to its graphical user interfaces

4.2 Data acquisition and processing

In this step, data is extracted from NCRB reports for the three crimes against woman under study i.e. domestic violence, dowry and rape. The data have been divided into two parts. One is the training set based on which prediction is carried out. Other is the test data, whose value is compared with the predicted values to draw inferences. For prediction following methods are used – Gaussian Process, Multilayer Perceptron and Linear Regression.

4.3 Methods used for Forecasting

WEKA has been used for time series forecasting. Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data. Time series forecasting is the process of using a model to generate predictions for future events based on known past events (Time Series Analysis and Forecasting with Weka).

4.3.1 Gaussian Processes

The Gaussian Processes (GP) is a renowned model in time series analysis. Gaussian process is a generalization of the Gaussian probability distribution. A probability distribution depicts random variables as scalars or vectors; a stochastic process directs the properties of functions. It can happen that the input values to a prediction problem can be uncertain. For example, for a discrete time series one can perform multi-step-ahead predictions by iterating one-step-ahead predictions. However, if the one-step-ahead predictions include uncertainty, then it is necessary to propagate this uncertainty forward to get the suitable multi-step-ahead predictions. One straightforward approach is to use sampling methods. It is possible to compute the mean and variance of the output analytically when using the SE (squared exponential) covariance function and Gaussian input noise (Rasmussen, 2006).

A Gaussian process is fully specified by its mean function and covariance function. The covariance function is an important function to construct GP. It is essential to determine several parameters of the covariance function. GP regards them as probabilistic variables. They are determined by setting up the prior distribution of them and evaluating the posterior distributions with the Bayes' theorem (Mori & Ohmi, 2005).

Mean function $m(x)$ and the covariance function $k(x,x')$ of a real process $f(x)$ is defined as

$$m(x) = E[f(x)]$$

$$k(x,x') = E[(f(x) - m(x))(f(x') - m(x'))]$$

And, Gaussian process as $f(x) \sim GP(m(x), k(x,x'))$

4.3.2 Multilayer Perceptron

A Multilayer Perceptron (MLP) is a feed forward artificial neural network model that maps different input data instances onto a set of suitable output. An MLP has multiple layers of nodes in a directed graph, with each layer entirely linked to the subsequent one. Each node in all the layers is a neuron associated with a nonlinear activation function except for the input nodes. MLP makes use of a supervised learning technique called back-propagation for training the network. MLP is a variation of the standard linear perceptron, which can make out data that is not linearly separable (Gupta, Mishra, & Pandey, 2014).

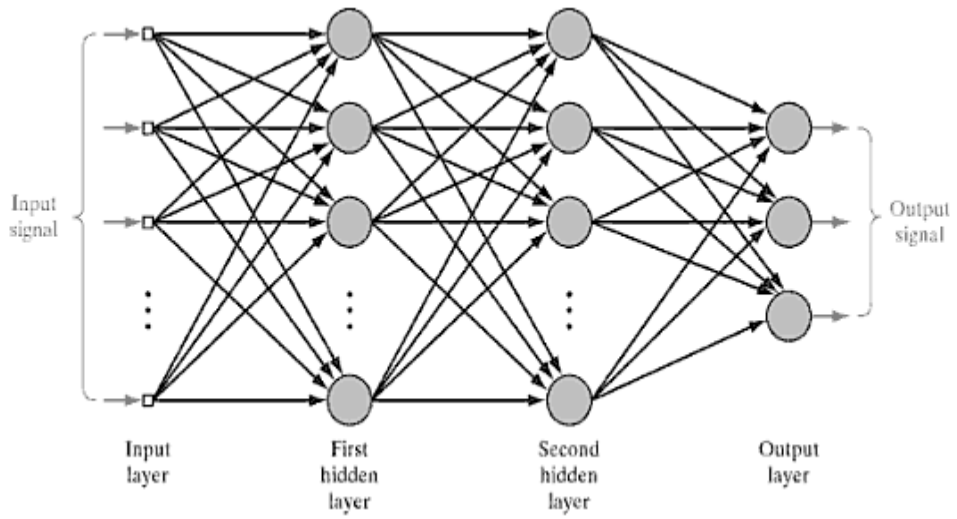

Figure 4.1:Architecture of Multilayer Perceptron

The Algorithm

The training of MLP proceeds in 2 phases:

- In the forward phase, the synaptic weights are fixed and the values in the input pattern are propagated through the network layer by layer until it reaches the output.
- In the backward phase, an error is generated by comparing the observed output of the network with the target response. The resulting error is propagated throughout the

network, layer by layer in the backward direction. In this phase consecutive adjustments are applied to the synaptic weights (MultiLayer Perceptron).

- **Weight Training Calculation in Backward phase**

  Let the input pattern be E. Let the target and observed response for node 'i' be $t_i(E)$ and $o_i(E)$ respectively. Let w_{ij} to specify weight between node i and node j.

1. The Error Term for output unit k is calculated as:

$$\delta_{O_k} = o_k(E)(1 - o_k(E))(t_k(E) - o_k(E))$$

2. The Error Term for hidden unit k is given by:

$$\delta_{H_k} = h_k(E)(1 - h_k(E)) \sum_{i \epsilon\ outputs} w_{ki}\delta_{O_i}$$

3. For each weight w_{ij} between input node i and hidden node j, calculate

$$\Delta_{ij} = \eta \delta_{H_j} x_i$$

   where, x_i is the input to the network to the input node i for input pattern E and η is learning rate.

4. For each weight w_{ij} between hidden node i and output node j, calculate:

$$\Delta_{ij} = \eta \delta_{o_j} h_i(E)$$

   where, $h_i(E)$ is the output from hidden node i for E.

5. Finally, add on each Δ_{ij} on to w_{ij}

$$w_{ij} = w_{ij} + \Delta_{ij}$$

6. In this way, the error is propagated back through the MLP.

4.3.3 Linear Regression

In simple linear regression (LR), scores are predicted on one variable from the scores on a second variable. The variable being predicted is referred to as Y and is called

the criterion variable. The variable on which prediction is based is referred to as X and is called the predictor variable. In simple linear regression, there is only one predictor variable. In LR the prediction of Y when plotted as a function of X forms a straight line. Linear regression consists of finding the best-fitting straight line through the points and the best-fitting line is called a regression line (Lane). The vertical lines from the points to the regression line symbolize the errors of prediction. Linear regression refers to a model in which the conditional mean of *y* given the value of *X* is an affine function of X. Linear regression focuses on the conditional probability distribution of *Y* given *X*, instead of on the joint probability distribution of *Y* and *X*, which is the field of multivariate analysis (Linear Regression)

Linear regression was the first type of regression analysis to be studied severely, and to be widely used in practical applications. This is because models which are linearly dependent on their unknown parameters are easier to fit than models which are non-linearly associated to their parameters and because the statistical properties of the resulting estimators are easier to decide (Linear Regression). Most applications of Linear Regression fall into one of the following two broad categories:

- If the goal is prediction, linear regression can be used to fit a predictive model to an observed data set of y and X values. After development of such model, if an additional test value of X is given without its associated value of y, the fitted model can be used to make a prediction of the value of y.
- Given a variable y and a set of variables $X_1,.....,X_p$ that may be related to y, now linear regression analysis can be applied to measure the strength of the association between y and the X_j , to decide which X_j may have no relationship with y at all, and to make out which subsets of the X_j have redundant information about y.

Given a data set $\{y_i, x_{i1},.....,x_{ip}\}n_i = 1$ of n statistical units, a linear regression model presumes that the relationship between the dependent variable y_i and the p-vector of regressors x_i is linear. This relationship is modeled through error variable ε_i — an unnoticed random variable that adds noise to the linear relationship between the dependent variable and regressors. Therefore the model takes the form

$$y_i = \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i = x_i^T \beta + \varepsilon_i \quad , \qquad \text{i=1,…..,n,}$$

where T denotes the transpose, so that $x_i^T \beta$ is the inner product between vectors x_i and β. Often these n equations are stacked together and written in vector form as $y = X\beta + \varepsilon$

$$\text{where, } \quad y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1^T \\ x_2^T \\ . \\ . \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & . & . & x_{1p} \\ x_{21} & . & . & x_{2p} \\ . & . & . & . \\ . & . & . & . \\ x_{n1} & . & . & x_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

Some generally used terminologies are:

- y_i is called the regressed, exogenous variable, or dependent variable. The decision as to which variable in a data set is modelled as the dependent variable and which are modelled as the independent variables may be based on a presumption that the value of one of the variables is influenced by the other variables.
- $x_{i1}, x_{i2}, ..., x_{in}$ are called regressors, endogenous variables, or independent variables. The matrix X is sometimes called the design matrix.
- Usually a constant is included as one of the regressors. For example we can take x_{i1} = 1 for i = 1, ..., n. The analogous element of β is called the intercept.
- Sometimes one of the regressors can be a non-linear function of another regressor or data, as in a polynomial regression and segmented regression. The model remains linear provided that it is linear in the parameter vector β.
- The regressors x_{ij} may be viewed either as random variables, which is merely observe, or they can be considered as predetermined fixed values which can be selected.
- β is a p-dimensional parameter vector. Its elements are called regression coefficients. Statistical estimation and inference in linear regression focuses on β. The elements of

this parameter vector are interpreted as the partial derivatives of the dependent variable with respect to the various independent variables.

- ε_i is called the error term, disturbance, or noise. This variable considers all other factors which influence the dependent variable y_i other than the regressors x_i (Kuzhda, 2012).

Assumptions in Linear Regression:

- Weak exogeneity - This essentially means that the predictor variables x can be treated as fixed values, rather than random variables.
- Linearity - This means that the mean of the response variable is a linear combination of the parameters (regression coefficients) and the predictor variables.
- Constant variance - This means that different response variables have the same variance in their errors, despite the values of the predictor variables.
- Independence of errors - This assumes that the errors of the response variables are uncorrelated with each other.

4.4 Evaluation Criteria for the Methods Used in Forecasting

1) Mean Error (ME) : A quick way of computing forecast errors is the mean error (ME) which is a simple average of all the errors of forecast for a time series data. This involves the summing of all the individual forecast errors and dividing by the number of forecast. The formula for calculating mean absolute deviation is given as:

$$ME = \frac{\sum_{i=1}^{N} e_i}{N}$$

Disadvantage of this measure is that if forecasts are both positive errors and negative errors the actual values, ME will include some cancellation effects that may potentially misrepresent the actual magnitude of the forecast error.

2) Mean Absolute Error (MAE): The mean absolute error (MAE) is the mean or average of the absolute values of the errors. The formula for calculating mean absolute deviation is given as:

$$MAE = \frac{\sum_{i=1}^{N} | e_i |}{N}$$

Relative to the mean error (ME), the mean absolute error (MAE) is commonly used because by taking the absolute values of the errors, it avoids the issues with the cancelling effects of the positive and negative values. N denotes the number of forecasts. If this value is 0 (zero), the forecast is perfect. As compared to the mean squared error value, unique or large error values will affect the MAE less than the MSE value.

3) Mean Square Error (MSE): Another popular way of computing forecast errors is the mean square error (MSE) which is computed by squaring each error and then taking a simple average of all the squared errors of forecast. This involves the summing of all the individual squared forecast errors and dividing by the number of forecast. The formula for calculating mean square error is given as:

$$MSE = \frac{\sum_{i=1}^{N} e_i^2}{N}$$

The MSE is preferred by some because it also avoids the problem of the cancelling effects of positive and negative values of forecast errors.

4) Percentage error (PE): All the above measures depend on the actual error value. It may seem rational to rather express the lack of fit in terms of the *relative* variation of the one-step-ahead forecasts from the observed values, that is, relative to the magnitude of the observed values. In other words, the absolute errors may be not so much of interest as are the relative errors in the forecasts. Percentage error value is used to assess the relative error, various indices have been proposed. The percentage error value is computed as:

$$PE_t = 100 * (X_t - F_t) / X_t$$

where X_t is the observed value at time t and F_t is the forecast (smoothed values) (Test of Hypothesis - People of Statistics, 2001).

5) Mean Percentage Error (MPE): The mean percent error (MPE) is the ratio of the error to the actual value being forecast multiplied by 100. The formula for calculating mean percent error is given as:

$$MPE = \frac{\sum_{i=1}^{N}\left(\frac{e_i}{Y_i}.100\right)}{N}$$

6) Mean Absolute Percentage Error (MAPE): Similar to the mean percent error (MPE), the mean absolute percent error (MAPE) is the average of the absolute values of the percentage of the forecast errors. The formula for calculating mean absolute percent error is given as:

$$MAPE = \frac{\sum_{i=1}^{N}\left(|\frac{e_i}{Y_i}|.100\right)}{N}$$

A mean percentage error near 0 (zero) can be produced by large positive and negative percentage errors that cancel each other out. Therefore, a better measure of relative fit is the mean absolute percentage error (Test of Hypothesis - People of Statistics, 2001).

4.5 ANOVA –analysis of variance

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. An ANOVA performed on a design in which there is only one factor is called a one-way ANOVA. If experimentation has two factors, then it is called a two-way ANOVA (Lane, Analysis of Variance).

Analysis of variance is a method for testing differences among means by evaluating variance. The ANOVA test is based on two approx values of the population variance ($\sigma 2$). One value is called the mean square error (MSE) and is based on differences of scores within the groups. MSE estimates $\sigma 2$ in spite of whether the null hypothesis is true. The second value is called the mean square between (MSB) which is based on differences among the sample means. MSB estimates $\sigma 2$ only if the population means are equal. If they are not equal, then MSB estimates a number larger than $\sigma 2$. Therefore, if the MSB is much larger than the MSE, then the population means are not likely to be equal. On the other hand, if the MSB is about the same as MSE, then the

data are consistent with the hypothesis that the population means are equal (Lane, Analysis of Variance).

Before proceeding with the calculation of MSE and MSB, let's see the assumptions made by ANOVA:

1. The populations have the same variance. This is called the assumption of homogeneity of variance.
2. The populations are normally distributed.
3. Each value is sampled independently from each other.

The ratio of MSB is to MSE is called F ratio, named after statistician R. Fisher. One of the important characteristics of ANOVA is that it partitions the variation into its various sources. In ANOVA, the term sum of squares (SSQ) is used to indicate variation. The total sum of variation is defined as the sum of squared differences between each score and the mean of all subjects. The mean of all subjects is called the Grand Mean and is denoted as GM. The total sum of squares is defined as

$$SSQ_{total} = \sum (X - GM)^2$$

which means to subtract the grand mean from each score, square the difference, and then sum up the squared values.

Chapter 5

RESULTS

This chapter describes the results of predictive analysis performed on datasets of Domestic violence, Dowry and Rape. Statistical methods like Gaussian Process, Multilayer Perceptron and Linear Regression are employed to forecast the crime rate for these crimes against woman. The performance analysis values and ANOVA test results are also compiled for each method.

5.1 Impact of Domestic Violence Act 2005 enforced in 2006.

Input to the system is data from year 1991 to 2006, based on input the developed model predicts the number of cases per lakh of population that must be registered in coming years. Using this analysis one can analyse the impact of this law on society.

a. Prediction using Gaussian Process

Here Gaussian Process has been used for forecasting. Table 5.1 gives the values of actual number of cases that are registered and the predicted number of cases that would be registered from the year 2007 to 2014.

Table 5.1: Prediction of cases registered under Domestic Violence Act using Gaussian Process

Year	Actual No. of cases registered for cruelty by husband and in-laws	Predicted no. of cases that would be registered under Domestic Violence Act
1991	15949	-
1992	19750	-
1993	22064	-
1994	25946	-
1995	28579	-
1996	35246	-
1997	36592	-
1998	41375	-
1999	43823	-
2000	45778	-
2001	49170	-

2002	49237	-
2003	50703	-
2004	58121	-
2005	58319	-
2006	63128	-
2007*	75930	51037.5274
2008*	81344	50188.0673
2009*	89546	48686.3901
2010*	94041	50488.9029
2011*	99135	52872.8918
2012*	106527	58975.087
2013*	118866	61534.1861
2014*	-	63505.3991

a. Evaluation on Training data :

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.2 tabulates these values calculated by WEKA for prediction.

Table 5.2: Evaluation data for prediction using Gaussian Process

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	11	10	9	8	7	6
Mean absolute error	7132.5669	6945.5779	6366.0363	5870.9779	6357.2564	9764.3533
Root mean squared error	8443.796	8303.0442	7708.5532	7316.2495	7807.4188	10060.6853
Mean absolute percentage error	15.5878	14.3061	12.1308	10.5537	11.114	17.5985
Direction accuracy	60	55.5556	50	57.1429	83.3333	100
Mean squared error	71297690.43	68940543.6588	59421792.9944	53527506.8551	60955788.3374	101217389.4272

b. Future forecast visualization

Figure 5.1 shows the actual and predicted number of cases for domestic violence, in the future forecast visualisation in WEKA. Domestic violence data has been compiled by NCRB after the year 1991. Hence the training dataset consists of data from year 1991 to 2006, so as to find out the impact of Domestic violence Act 2005 which was operational in 2006. By observing both the actual data and predicted data, the impact of the Act can be found.

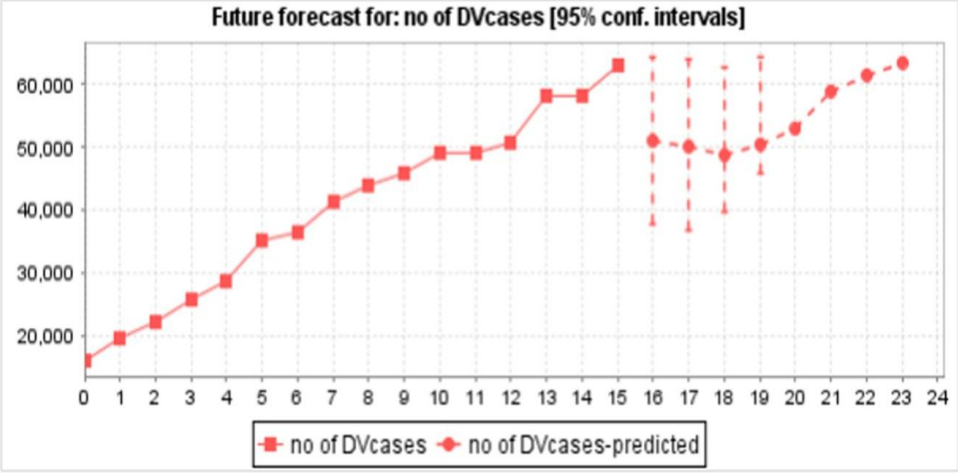

Figure 5.1: Actual and predicted values for Gaussian Process

c. Training prediction for target visualization

The comparison of predicted number of cases per lakh of population and actual number of cases per lakh of population are depicted in Figure 5.2.

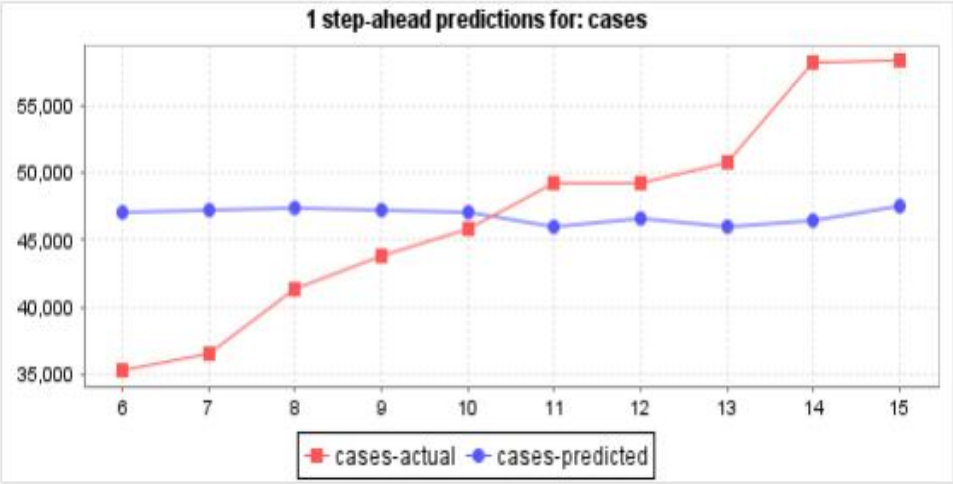

Figure 5.2 : Comparison of predicted and actual data for Gaussian Process

d. ANOVA Analysis data for Gaussian Prediction.

Table 5.3 summarizes the ANOVA output for the actual and predicted data. The p-value for this model is .0000124.

Table 5.3: ANOVA Summary for Gaussian Prediction

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	6.07E+09	1	6.07E+09	50.459	1.24E-05	4.747225
Within Groups	1.44E+09	12	1.2E+08			
Total	7.52E+09	13				

5.1.2 Prediction using Linear Regression Model

Here Linear Regression Model has been used for forecasting. Table 5.4 gives the values of actual number of cases registered in police stations and the predicted number of cases that would be registered from the year 2007 to 2014.

Table 5.4: Prediction of cases registered under Domestic Violence Act using Linear Regression

Year	Actual No. of cases registered for cruelty by husband and in-laws	Predicted no. of cases that would be registered under Domestic Violence Act
1991	15949	-
1992	19750	-
1993	22064	-
1994	25946	-
1995	28579	-
1996	35246	-
1997	36592	-
1998	41375	-
1999	43823	-
2000	45778	-
2001	49170	-
2002	49237	-
2003	50703	-
2004	58121	-
2005	58319	-

2006	63128	-
2007*	75930	65010.2985
2008*	81344	67993.9094
2009*	89546	69986.4865
2010*	94041	73180.2991
2011*	99135	75605.7516
2012*	106527	78518.3236
2013*	118866	81246.8903
2014*	-	84140.3441

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.5 tabulates these values calculated by WEKA for prediction For Linear Regression Model.

Table 5.5: Evaluation data for prediction using Linear Regression

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	11	10	9	8	7	6
Mean absolute error	1153.3673	1245.3946	1270.1896	1314.7283	1391.8184	1547.279
Root mean squared error	1359.2253	1423.6606	1461.4726	1516.0782	1594.0589	1682.8975
Mean absolute percentage error	2.3303	2.4972	2.4643	2.4962	2.6001	2.8971
Direction accuracy	100	100	100	100	100	100
Mean squared error	1847493.3605	2026809.3725	2135902.1746	2298492.9708	2541023.8703	2832144.0444

b. Future forecast visualization

Figure 5.3 shows the actual and predicted number of cases for domestic violence using linear regression, in the future forecast visualisation in WEKA. Domestic violence data has been compiled by NCRB after the year 1991. Hence the training dataset consists of data from year 1991 to 2006, so as to find out the impact of Domestic

violence Act 2005 which was operational in 2006. By observing both the actual data and predicted data, the impact of the Act can be found.

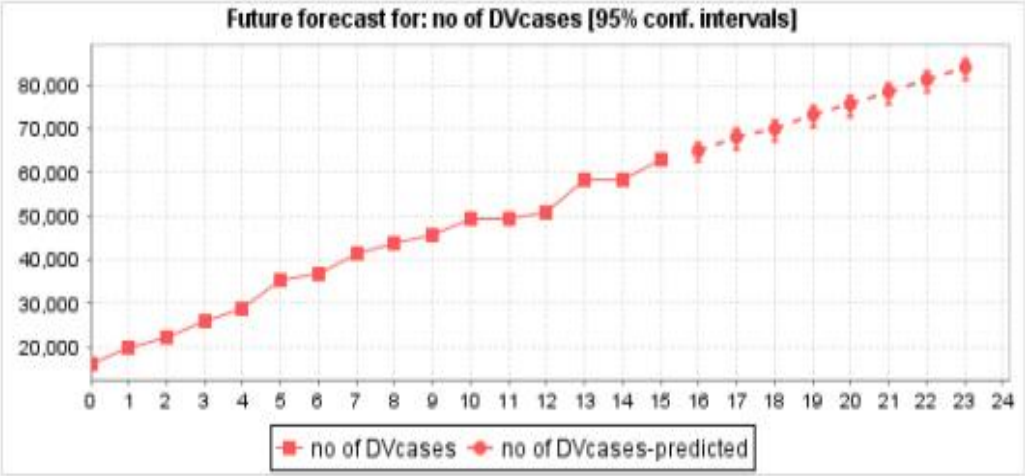

Figure 5.3: Actual and predicted values for Linear Regression

c. Training Predictions for targets visualization

Figure 5.4 shows the comparison of predicted and actual data for domestic violence dataset using linear regression.

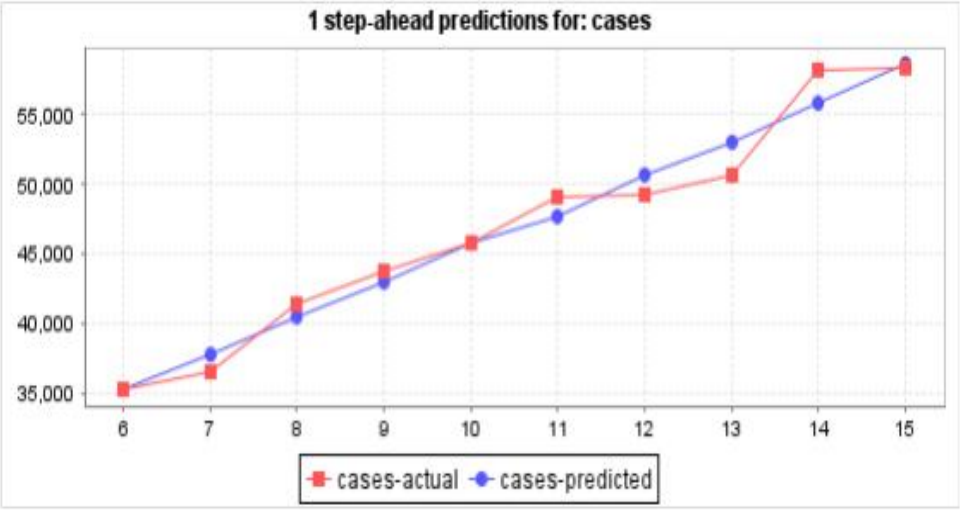

Figure 5.4 : Comparison of predicted and actual data for Linear Regression

d. ANOVA Analysis data for Linear Regression

Table 5.6 summarizes the ANOVA output for the actual and predicted data using Linear Regression. The p- value for this model is .003188.

Table 5.6:ANOVA Summary for Linear Regression Prediction

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	1.69E+09	1	1.69E+09	13.49296	0.003188	4.747225
Within Groups	1.5E+09	12	1.25E+08			
Total	3.19E+09	13				

5.1.3 Prediction using Multilayer Perceptron

Here Multilayer Perceptron Model has been used for forecasting. Table 5.7 gives the values of actual number of cases registered across India and the predicted number of cases that would be registered from the year 2007 to 2014.

Table 5.7: Prediction of cases registered under Domestic Violence Act using Multilayer Perceptron

year	Actual No. of cases registered for cruelty by husband and in-laws	Predicted no. of cases that would be registered under Domestic Violence Act
1991	15949	-
1992	19750	-
1993	22064	-
1994	25946	-
1995	28579	-
1996	35246	-
1997	36592	-
1998	41375	-
1999	43823	-
2000	45778	-
2001	49170	-
2002	49237	-
2003	50703	-
2004	58121	-
2005	58319	-
2006	63128	-
2007*	75930	63307.2146

2008*	81344	65931.8796
2009*	89546	67293.0197
2010*	94041	69587.7213
2011*	99135	71508.8873
2012*	106527	73461.7487
2013*	118866	75147.0675
2014*	-	76656.6057

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.8 tabulates these values calculated by WEKA for prediction for Multilayer Perceptron Model.

Table 5.8: Evaluation data for prediction using MultiLayer Perceptron

Target	1-steps-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	11	10	9	8	7	6
Mean absolute error	1153.5837	1168.7409	1187.4546	1313.2584	1457.5513	1582.6743
Root mean squared error	1394.3286	1463.9342	1431.1503	1551.085	1679.5453	1729.131
Mean absolute percentage error	2.4015	2.3589	2.2524	2.466	2.6994	2.9279
Direction accuracy	100	100	100	100	100	100
Mean squared error	1944152.1212	2143103.4342	2048191.2438	2405864.8093	2820872.3699	2989893.8659

b. Future forecast visualization

Figure 5.5 shows the actual and predicted number of cases for domestic violence using Multilayer Perceptron, in the future forecast visualisation in WEKA. Domestic violence data has been compiled by NCRB from the year 1991. Hence the training dataset consists of data from year 1991 to 2006, so as to find out the impact of Domestic violence Act 2005 which was operational in 2006. By observing both the actual data and predicted data, the impact of the Act can be found.

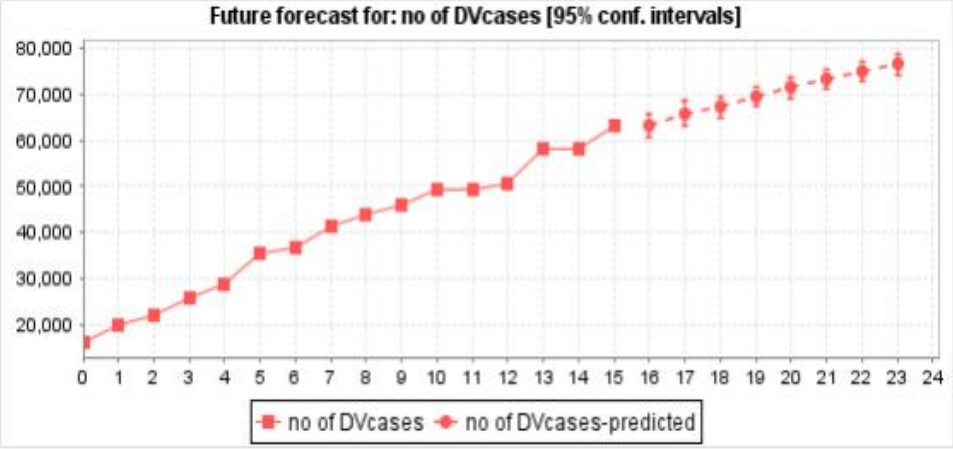

Figure 5.5: actual and predicted values for MultiLayerPerceptron

c. Train Prediction for Target Visualization

Figure 5.6 shows the comparison of predicted and actual data for domestic violence dataset using linear regression. The 1 – step ahead graph has been generated by WEKA.

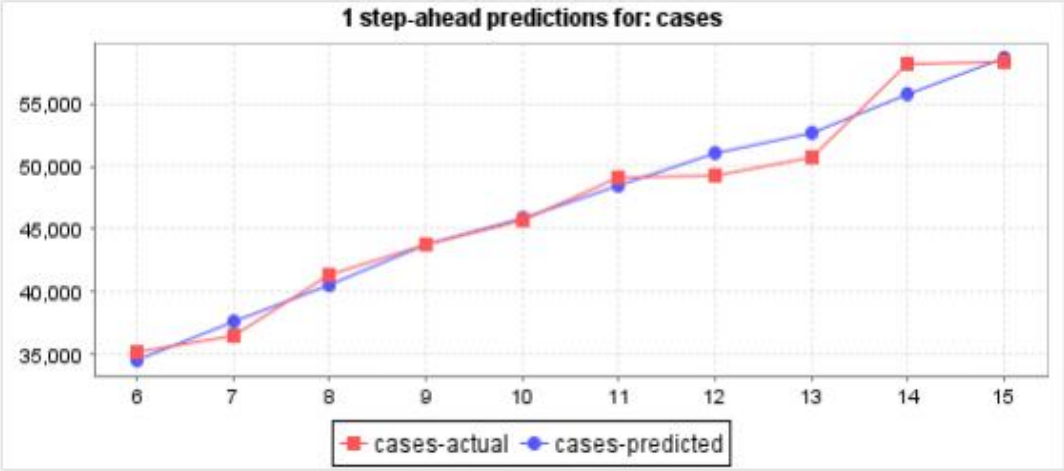

Figure 5.6 : Comparison of predicted and actual data for MultiLayerPerceptron

d. ANOVA Analysis data for Multilayer Perceptron

Table 5.9 summarizes the ANOVA output for the actual and predicted values using Multi Layer Perceptron. The p- value for this model is .000835.

Table 5.9: ANOVA Summary for MultiLayerPerceptron Prediction

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	2.29E+09	1	2.29E+09	19.53958	0.000835	4.747225
Within Groups	1.41E+09	12	1.17E+08			
Total	3.7E+09	13				

According to ANOVA results of all the three models used for prediction of domestic violence data, the p value of Gaussian Process is the smallest. Hence that is the most effective model for prediction.

5.2 Impact of Dowry Prohibition Act

Here the input to the system is data from year 1988 to 2000, since when the NCRB started compiling data of Dowry. Based on the input data, the number of cases that would be registered in coming years till 2013 are predicted. Using this analysis the impact of dowry prohibition law on society can be analyzed.

5.2.1 Prediction using Gaussian Process

Here, Gaussian Process has been used for forecasting. Table 5.10 gives the values of actual number of cases registered across the country and the predicted number of cases that would be registered after 2000 under Dowry Prohibition Act.

Table 5.10: Prediction of number of cases to be registered under Dowry Prohibition Act using Gaussian process

Year	Actual number of cases registered under Dowry Act	Number of cases predicted
1988	2084	-
1989	1918	-
1990	2155	-
1991	1841	-
1992	2102	-
1993	2679	-
1994	2709	-
1995	2814	-
1996	2647	-
1997	2685	-
1998	3578	-
1999	3064	-
2000	2876	-
2001*	3222	2870.1612
2002*	2816	2714.3372
2003*	2684	2734.3754
2004*	3592	2928.0078

2005*	3204	3030.5118
2006*	4504	3115.8266
2007*	5623	3185.8025
2008*	5555	3185.5313
2009*	5650	3173.8636
2010*	5182	3185.5332
2011*	6619	3207.739
2012*	9038	3249.0692
2013*	10709	3305.4541

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.11 tabulates these values calculated by WEKA for prediction for Gaussian Process.

Table 5.11: Evaluation data for prediction using Gaussian process

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	223.4939	265.6172	278.6617	321.2149	357.0779	471.2089
Root Mean Squared Error	333.4	365.189	394.66	432.418	483.4	573.24
Mean absolute percentage error	7.17	8.5	8.78	10.05	10.755	14.04
Direction Accuracy	42.85	16.66	0	0	0	0
Mean Squared Error	111155.8905	133363.159	155757.3324	186985.3785	233685.0737	328609.5248

b. Future forecast visualization

Figure 5.7 shows the actual and predicted number of cases for dowry using Gaussian Process, in the future forecast visualisation in WEKA. Dowry data has been compiled by NCRB from the year 1988. Hence the training dataset consists of data from year 1988 to 2000, so as to find out the impact of Dowry Prohibition Law. By observing both the actual data and predicted data, the impact of the Act can be found.

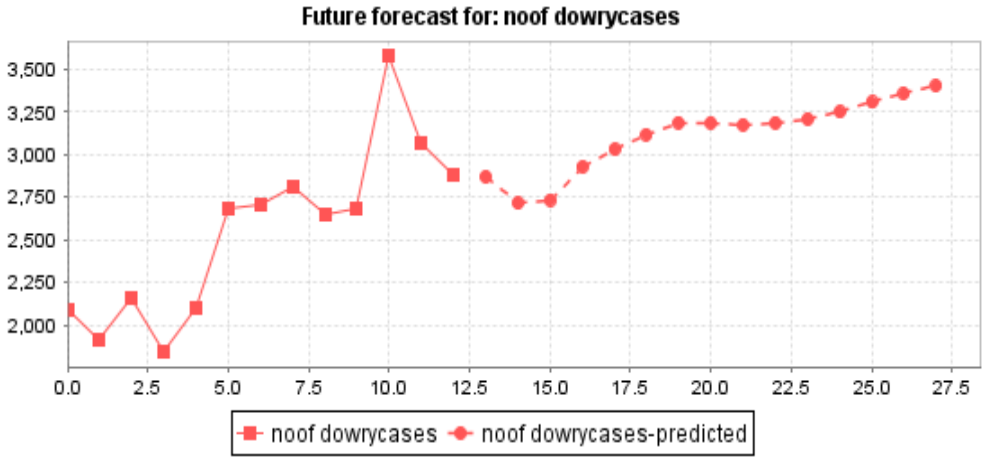

Figure 5.7: Actual and predicted values for dowry using Gaussian Process

c. Train Prediction for Target Visualisation

Figure 5.8 shows the comparison of predicted and actual data for dowry dataset using Gaussian Process.

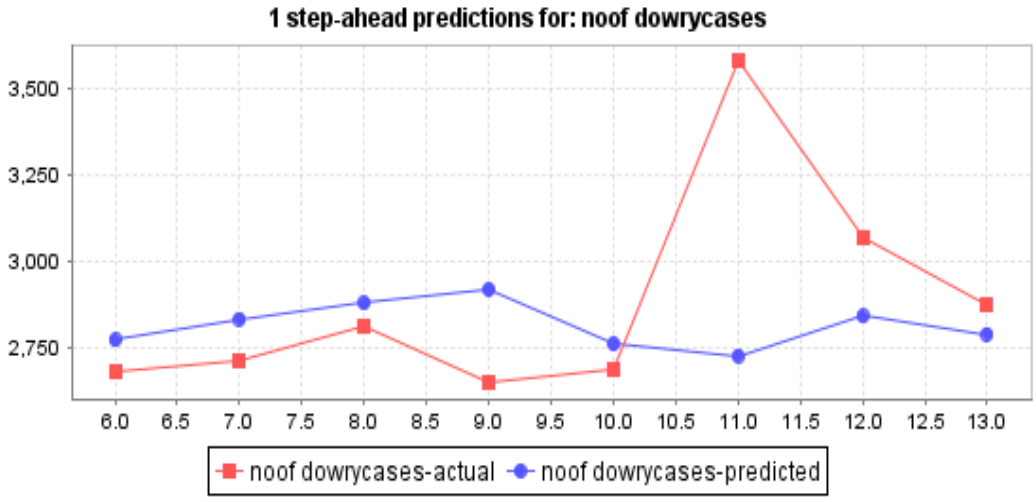

Figure 5.8: Comparison of predicted and actual data for dowry data using Gaussian Process

d. ANOVA Analysis data for dowry using Gaussian Process

Table 5.12 summarizes the ANOVA output for the actual and predicted values using Gaussian Process. The p- value for this model is .003392.

Table 5.12: ANOVA Summary

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	31266231	1	31266231	10.57003	0.003392	4.259677
Within Groups	70992194	24	2958008			
Total	1.02E+08	25				

b. Prediction using Linear Regression

Here we are using Linear Regression Model for forecasting. Table 5.13 gives the values of actual number of cases registered in all over India since 1988 and the predicted number of cases that would be registered after year 2000 under dowry prohibition Act.

Table 5.13:Prediction of cases registered under Dowry Prohibition Act using Linear Regression

Year	Actual number of dowry cases Registered	Predicted number of Dowry cases to be registered
1988	2084	-
1989	1918	-
1990	2155	-
1991	1841	-
1992	2102	-
1993	2679	-
1994	2709	-
1995	2814	-
1996	2647	-
1997	2685	-
1998	3578	-
1999	3064	-
2000	2876	-
2001*	3222	2870.1612
2002*	2816	2714.3372
2003*	2684	2734.3754
2004*	3592	2928.0078
2005*	3204	3030.5118
2006*	4504	3115.8266
2007*	5623	3185.8025

2008*	5555	3185.5313
2009*	5650	3173.8636
2010*	5182	3185.5332
2011*	6619	3207.739
2012*	9038	3249.0692
2013*	10709	3305.4541

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.14 tabulates these values calculated by WEKA for prediction using Linear Regression.

Table 5.14 :Evaluation Data for Linear Regression Model

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	227.766	226.2535	237.488	254.118	287.1	318.2436
Root Mean Squared Error	277.71	282.89	298.599	319.734	352.215	391.02
Mean absolute percentage error	7.64	7.46	7.735	8.18	9.07	9.7
Direction Accuracy	57.14	50	40	50	33.33	0
mean Squared Error	77127.06	80028.91	89161.72	1022230.19	124056.01	152903.153

b. Future forecast visualization

Figure 5.9 illustrates the actual and predicted number of cases for dowry using Gaussian Process, in the future forecast visualization in WEKA. Dowry data has been

compiled by NCRB from the year 1988. Hence the training dataset consists of data from year 1988 to 2000, so as to find out the impact of Dowry Prohibition Law. By observing both the actual data and predicted data, the impact of the Act can be found.

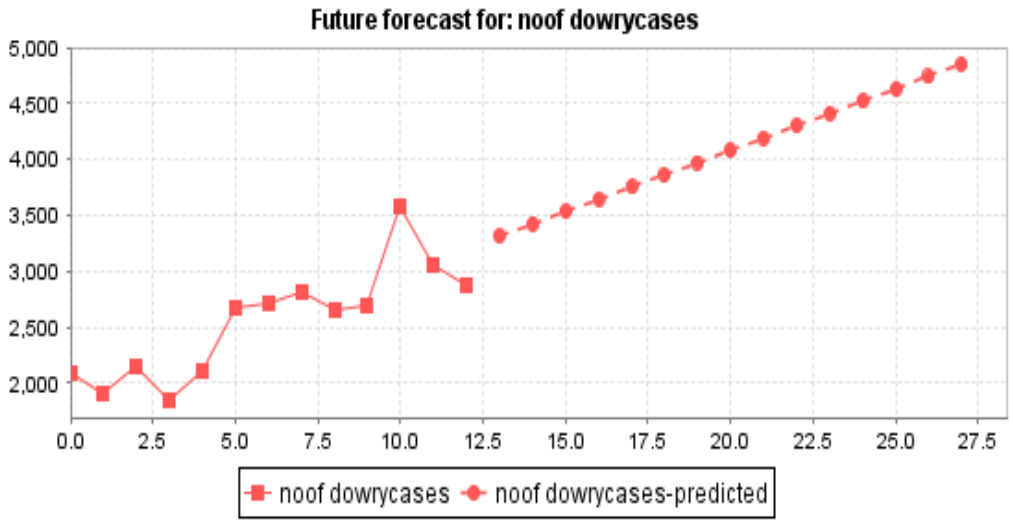

Figure 5.9: Actual and predicted values for dowry using Linear Regression

c. Train Prediction for Target Visualization

Figure 5.10 shows the comparison of predicted and actual data for dowry dataset using Gaussian Process.

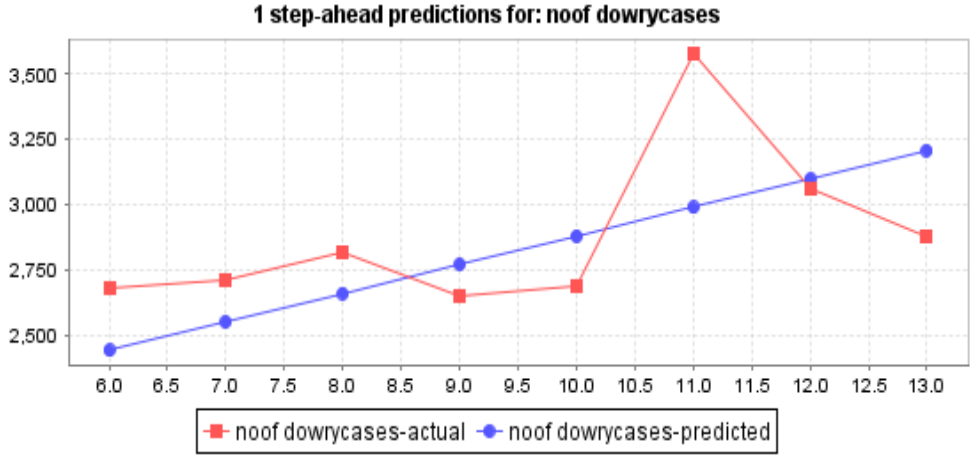

Figure 5.10: Comparison of Actual and Predicted data for Linear Regression Model

d. ANOVA Analysis data for dowry using Linear Regression

Table 5.15 summarizes the ANOVA output for the actual and predicted values using Linear Regression Model for the dowry dataset. The p- value for this model is 0.003392.

Table 5.15:ANOVA Summary

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	31266231	1	31266231	10.57003	0.003392	4.259677
Within Groups	70992194	24	2958008			
Total	1.02E+08	25				

c. Prediction using Multilayer Perceptron

Here, Multilayer Perceptron Model has been used for forecasting. Table 5.16 gives the values of actual number of cases registered all over the country since 1988 and the predicted number of cases that would be registered after year 2000 under dowry prohibition Act.

Table 5.16: Prediction of cases registered under Dowry Prohibition Act using MultiLayerPerceptron

Year	Actual number of dowry cases Registered	Predicted number of Dowry cases to be registered
1988	2084	-
1989	1918	-
1990	2155	-
1991	1841	-
1992	2102	-
1993	2679	-
1994	2709	-
1995	2814	-
1996	2647	-
1997	2685	-
1998	3578	-
1999	3064	-
2000	2876	-
2001*	3222	4357.282
2002*	2816	2429.307
2003*	2684	2012.521
2004*	3592	4516.851
2005*	3204	2315.985

2006*	4504	2077.245
2007*	5623	2918.826
2008*	5555	2265.929
2009*	5650	2095.804
2010*	5182	3292.131
2011*	6619	4241.358
2012*	9038	3448.383
2013*	10709	4189.598
2014*	-	4374.543

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.17 tabulates these values calculated by WEKA for prediction using Multilayer Perceptron.

Table 5.17 :Evaluation Data for MultiLayerPerceptron

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	0.1154	0.1228	0.1589	0.2636	0.342	0.0442
Root Mean Squared Error	0.1326	0.147	0.178	0.3174	0.3559	0.0553
Mean absolute percentage error	0.0042	0.0043	0.0055	0.0089	0.0111	0.0013
Direction Accuracy	100	100	100	100	100	100
mean Squared Error	0.0176	0.0216	0.0319	0.1008	0.1266	0.0031

b. Future forecast visualization

Figure 5.11 shows the actual and predicted number of cases for dowry using Multilayer Perceptron, in the future forecast visualization in WEKA. Dowry data has been compiled by NCRB from the year 1988. Hence the training dataset consists of data from year 1988 to 2000, so as to find out the impact of Dowry Prohibition Law. By observing both the actual data and predicted data, the impact of the Act can be found.

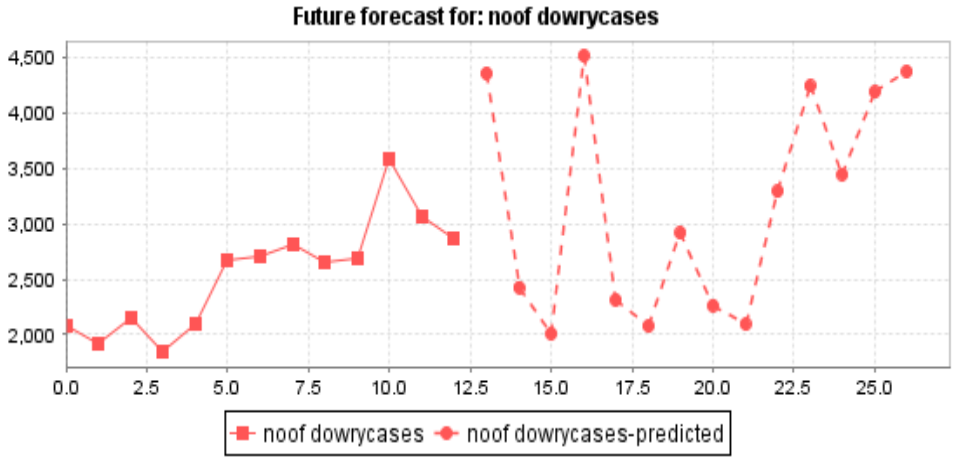

Figure 5.11: Actual and predicted values for dowry using MultiLayerPerceptron

c. Train Prediction for Target Visualisation

Figure 5.12 shows the comparison of predicted and actual data for dowry dataset using MultiLayer Perceptron.

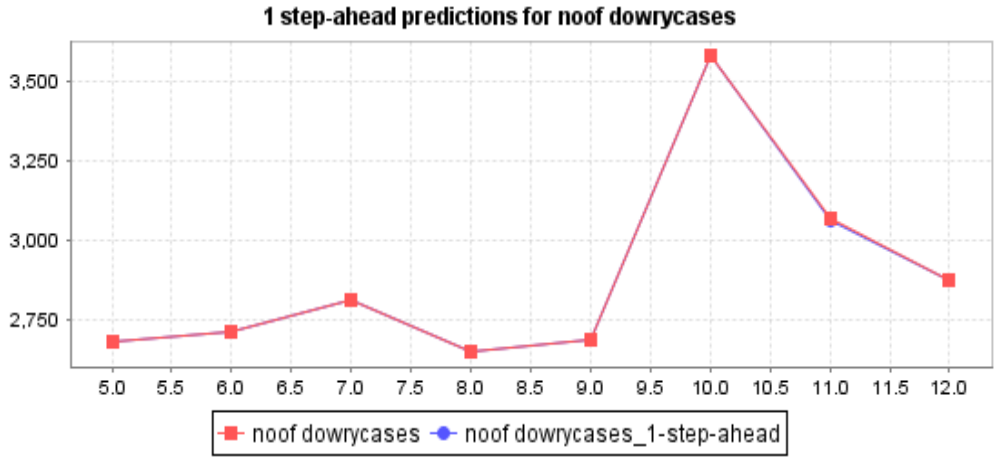

Figure 5.12: : Comparison of Actual and Predicted data for Multilayer Perceptron

d. ANOVA Analysis data for dowry using Multilayer Perceptron

Table 5.18 summarizes the ANOVA output for the actual and predicted values using Multilayer Perceptron Model for dowry dataset. The p- value for this model is .006211.

Table 5.18: ANOVA Summary

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	30665990	1	30665990	8.997689	0.006211	4.259677
Within Groups	81796976	24	3408207			
Total	1.12E+08	25				

5.3 Impact of Amendment in Criminal Law 1983

Here the input to the system is data from year 1971 to 1983, since when the NCRB started compiling data of Rape. Based on the input data the number of cases that would be registered in coming years till 2013 are predicted. Using this analysis, the impact of Amendment to the rape law can be analyzed.

5.4.1 Prediction using Gaussian Process

Here, Gaussian Process has been used for forecasting. Table 5.19 gives the values of actual number of cases registered all over India since 1971 and the predicted number of cases that would be registered after amendment in criminal law in 1983.

Table 5.19 :Prediction of rape data using Gaussian Process

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-
1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-
1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	4925.5629
1985	7289	4554.4219
1986	7952	4368.4147
1987	8559	4228.1826
1988	9099	4417.99
1989	9752	4952.5477
1990	10068	5317.8514
1991	10410	5614.6425
1992	11112	5777.3633

1993	12218	5718.5824
1994	13208	5556.806
1995	13754	5425.944
1996	14846	5350.1932
1997	15330	5388.4976
1998	15151	5544.9999
1999	15468	5753.5533
2000	16496	5956.8652
2001	16075	6115.251
2002	16373	6198.6435
2003	15847	6215.8211
2004	18233	6201.1828
2005	18359	6189.5763
2006	19348	6210.2061
2007	20737	6277.4673
2008	21467	6384.1231
2009	21397	6510.2126
2010	22172	6633.1086
2011	24206	6734.4884
2012	24923	6807.019
2013	33707	6855.6009

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.20 tabulates these values calculated by WEKA for prediction using Gaussian Process.

Table 5.20 : Evaluation Data of prediction using gaussian process

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	651.205	699.4686	696.7274	723.693	869.6479	1297.54
Root Mean Squared Error	703.419	771.4665	788.216	830.7653	972.9587	1321.2549
Mean absolute percentage error	13.63	14.09	13.407	13.65	15.45	22.935
Direction Accuracy	71.428	50	40	0	0	100
mean Squared Error	494798.9	595160.52	621285.105	690170.9	946648.719	1745714.6

b. Future forecast visualization

Figure 5.13 shows the actual and predicted number of cases for rape using Gaussian process, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 1983, so as to find out the impact of amendment in criminal law in 1983. By observing both the actual data and predicted data, the impact of the amendment can be found and analyzed.

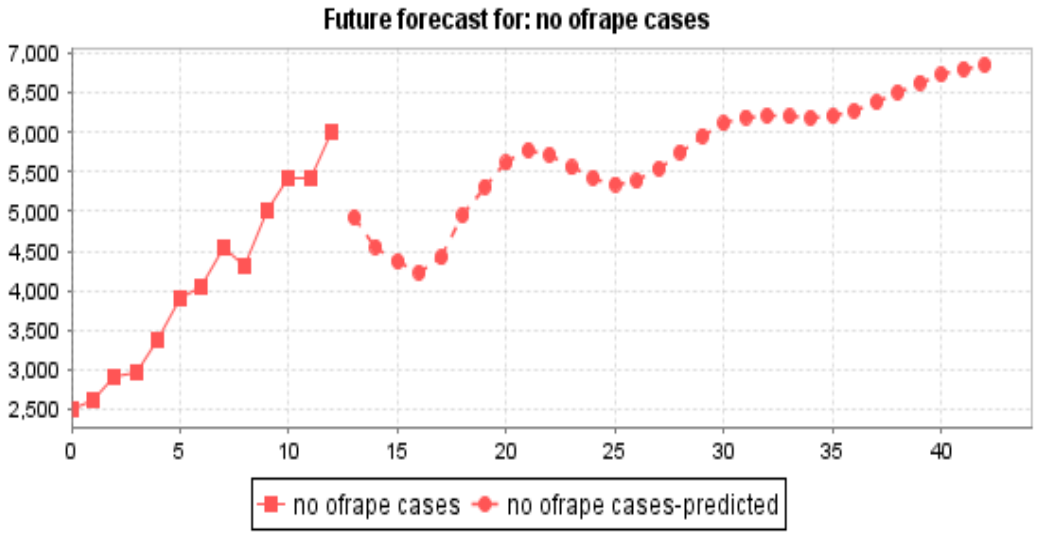

Figure 5.13:Actual and Predicted number of cases using Gaussian Process

c. Train Prediction for Target Visualisation

Figure 5.14 shows the comparison of predicted and actual data for rape dataset using Gaussian Process.

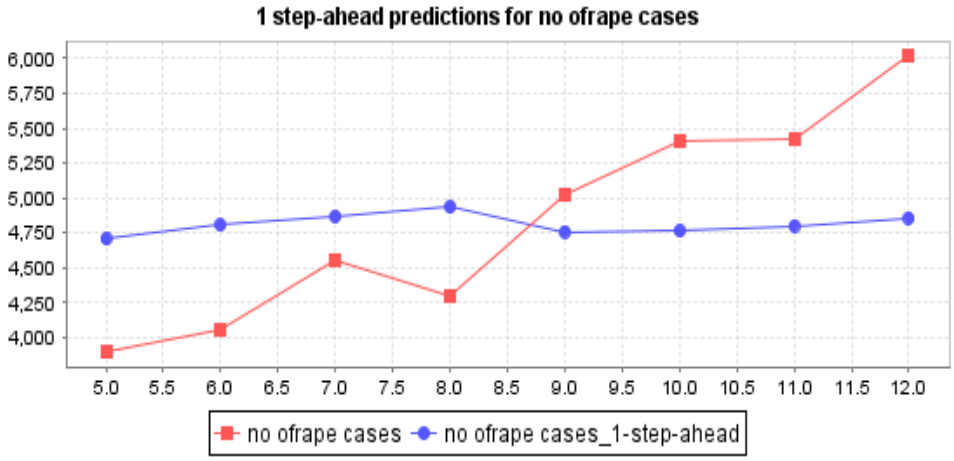

Figure 5.14: comparison of actual and predicted values using Gaussian Process

d. ANOVA Analysis data for dowry using Gaussian Processes

Table 5.21 summarizes the ANOVA output for the actual and predicted values for rape data using Gaussian Processes Model. The p- value for this model is very less i.e. $2.61*10^{-12}$.

Table 5.21: comparison of actual and predicted values using Gaussian Process

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.48E+09	1	1.48E+09	77.83341	2.61E-12	4.006873
Within Groups	1.1E+09	58	19029981			
Total	2.58E+09	59				

5.4.2 Prediction using Linear Regression

Here, Linear Regression Model has been used for forecasting. Table 5.22 gives the values of actual number of cases registered and the predicted number of cases that would be registered after amendment in criminal law in 1983.

Table 5.22 : Actual and Predicted number of cases using Linear Regression Model

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-
1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-

1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	6155
1985	7289	6451.473
1986	7952	6747.945
1987	8559	7044.418
1988	9099	7340.89
1989	9752	7637.363
1990	10068	7933.835
1991	10410	8230.308
1992	11112	8526.78
1993	12218	8823.253
1994	13208	9119.725
1995	13754	9416.198
1996	14846	9712.67
1997	15330	10009.14
1998	15151	10305.62
1999	15468	10602.09
2000	16496	10898.56
2001	16075	11195.03
2002	16373	11491.51
2003	15847	11787.98
2004	18233	12084.45
2005	18359	12380.92
2006	19348	12677.4
2007	20737	12973.87
2008	21467	13270.34
2009	21397	13566.81
2010	22172	13863.29

2011	24206	14159.76
2012	24923	14456.23
2013	33707	14752.7

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.23 tabulates these values calculated by WEKA for prediction using Linear Regression Model.

Table 5.23: Evaluation data for prediction using Linear Regression

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	147.3475	152.714	174.551	173.094	123.2088	146.315
Root Mean Squared Error	177.45	185.108	199.744	203.136	129.86	146.6968
Mean absolute percentage error	3.111	3.152	3.589	3.509	2.219	2.602
Direction Accuracy	85.714	83.333	80	100	100	100
mean Squared Error	31488.61	34265.32	39897.775	41264.524	16866.004	21519.958

b. Future forecast visualization

Figure 5.15 shows the actual and predicted number of cases for rape using Linear Regression, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 1983, so as to find out the impact of amendment in criminal law in 1983. By observing both the actual data and predicted data we can find out the impact of the amendment.

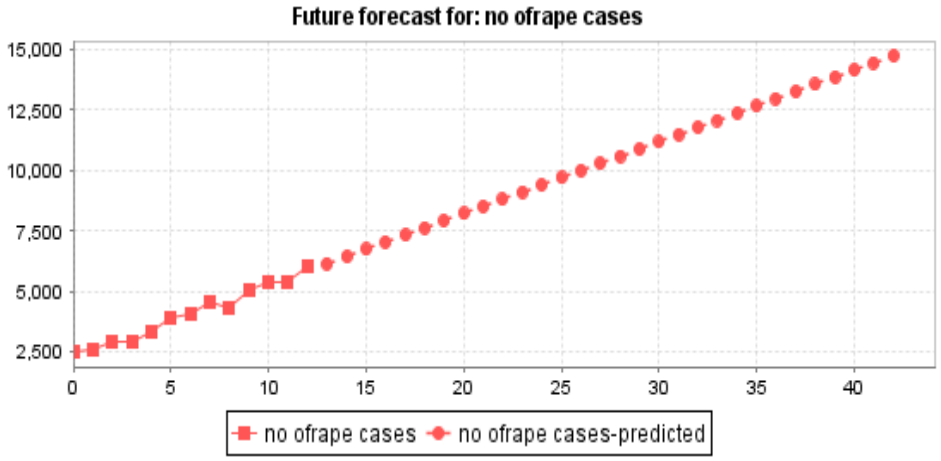

Figure 5.15:Actual and predicted data using Linear Regression

c. Train Prediction for Target Visualisation

Figure 5.16 shows the comparison of predicted and actual data for rape dataset using Linear Regression.

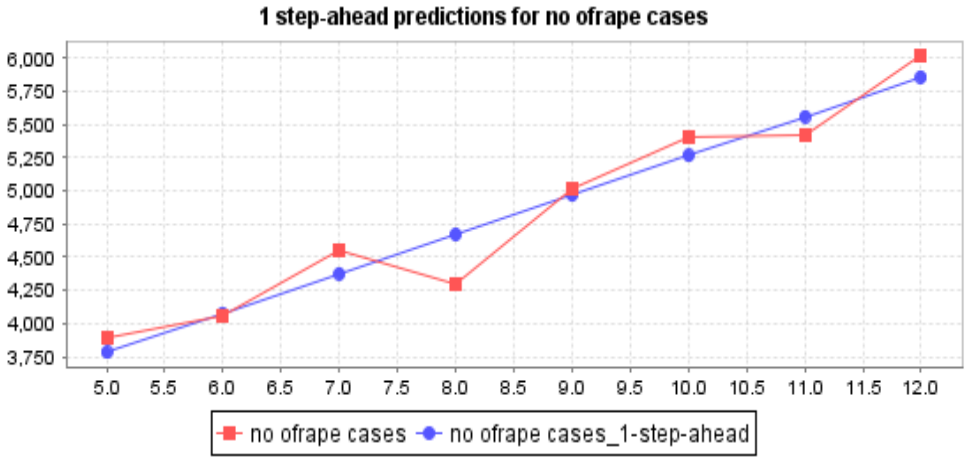

Figure 5.16 : Comparison of actual and predicted cases for Linear Regression

d. ANOVA Analysis data for rape using Linear Regression

Table 5.24 summarizes the ANOVA output for the actual and predicted values using Linear Regression Model. The p- value for this model is $6.7*10^{-5}$.

Table 5.24 ANOVA Summary

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	4.09E+08	1	4.09E+08	18.4628	6.7E-05	4.006873
Within Groups	1.29E+09	58	22160562			
Total	1.69E+09	59				

5.4.3 Prediction using Multilayer Perceptron

Here we are using Multilayer Perceptron Model for forecasting. Table 5.25 gives the values of actual number of cases registered across India since 1971 and the predicted number of cases that would be registered after amendment in criminal law in 1983.

Table 5.25: Actual and predicted data using MultiLayerPerceptron

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-
1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-
1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	5759.402
1985	7289	6226.762
1986	7952	6312.657
1987	8559	6286.536
1988	9099	6477.262
1989	9752	6394.988
1990	10068	6524.317
1991	10410	6557.095
1992	11112	6559.232
1993	12218	6622.616
1994	13208	6617.13
1995	13754	6658.326
1996	14846	6679.772
1997	15330	6693.355

1998	15151	6718.711
1999	15468	6730.397
2000	16496	6748.726
2001	16075	6763.389
2002	16373	6775.855
2003	15847	6789.183
2004	18233	6799.953
2005	18359	6810.709
2006	19348	6820.41
2007	20737	6829.219
2008	21467	6837.521
2009	21397	6845.054
2010	22172	6852.086
2011	24206	6858.596
2012	24923	6864.64
2013	33707	6870.294

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.26 tabulates these values calculated by WEKA for prediction using Multilayer Perceptron Model.

Table 5.26 : Evaluation Data for prediction using Multilayer Perceptron

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	8	7	6	5	4	3
Mean Absolute Error	54.88	80.8688	95.701	63.137	38.173	80.96
Root Mean Squared Error	87.758	115.117	127.601	90.689	50.976	100.667
Mean absolute percentage error	1.177	1.753	2.063	1.358	0.73	1.401

Direction Accuracy	85.714	83.333	80	100	100	100
mean Squared Error	7701.467	13252.016	16282.074	8224.6298	2598.626	10133.915

b. Future forecast visualization

Figure 5.17 shows the actual and predicted number of cases for rape using Multilayer Perceptron, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 1983, so as to find out the impact of amendment in criminal law in 1983. By observing both the actual data and predicted data we can find out the impact of the amendment

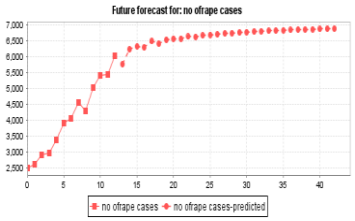

Figure 5.17: Evaluation Data for prediction using MultiLayer Perceptron

c. Train Prediction for Target Visualisation

Figure 5.18 shows the comparison of predicted and actual data for rape dataset using Multilayer Perceptron.

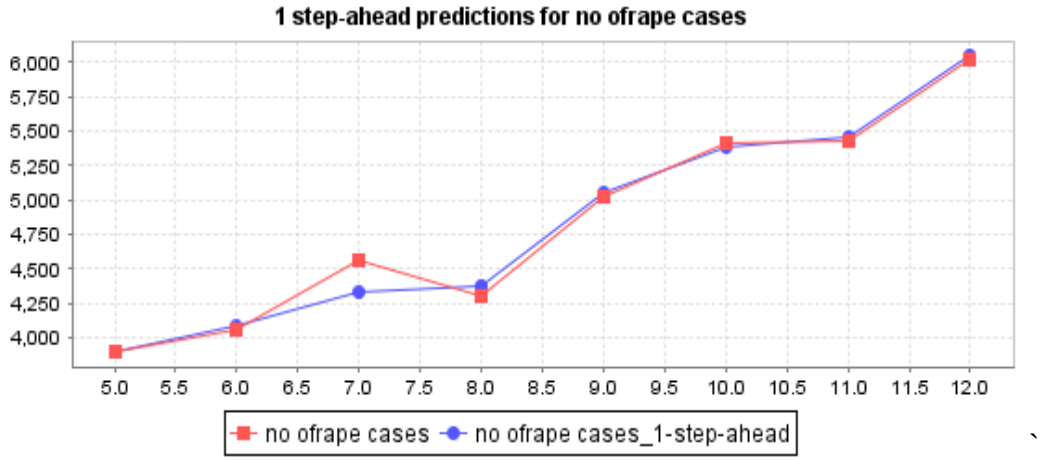

Figure 5.18: Comparison of actual and predicted data for Gaussian Processes

d. ANOVA Analysis data for dowry using Multilayer Perceptron

Table 5.27 summarizes the ANOVA output for the actual and predicted values using Multilayer Perceptron Model. The p- value for this model is 4.63E-11.

Table 5.27: ANOVA Summary

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	1.22E+09	1	1.22E+09	65.16398	4.63E-11	4.006873
Within Groups	1.09E+09	58	18785274			
Total	2.31E+09	59				

5.4 Impact of Sexual Offences Act 2003

Here the input to the system is data from year 1971 to 2003, since when the NCRB started compiling data of Rape. Based on input we predict the number of cases that must be registered in coming years till 2013. Using this analysis we can determine the impact of Amendment to the rape law.

5.4.1 Prediction using Gaussian Process

Here we are using Gaussian Process for forecasting. Table 5.28 gives the values of actual number of cases registered and the predicted number of cases that would be registered after the Sexual Offences Act in 2003.

Table 5.28 : Prediction of data using Gaussian Process

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-
1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-
1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	-
1985	7289	-
1986	7952	-
1987	8559	-
1988	9099	-
1989	9752	-
1990	10068	-
1991	10410	-
1992	11112	-
1993	12218	-
1994	13208	-
1995	13754	-
1996	14846	-
1997	15330	-
1998	15151	-
1999	15468	-
2000	16496	-
2001	16075	-
2002	16373	-

2003	15847	-
2004	18233	12964.03
2005	18359	11347.36
2006	19348	9993.54
2007	20737	8743.649
2008	21467	8018.609
2009	21397	7984.688
2010	22172	8087.396
2011	24206	8372.335
2012	24923	8762.325
2013	33707	9090.621

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.29 tabulates these values calculated by WEKA for prediction using Gaussian Process.

Table 5.29: Evaluation Data for prediction using Gaussian Process

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	28	27	26	25	24	23
Mean Absolute Error	2739.3596	3836.4474	4746.869	5456.34	5703.175	5322.3687
Root Mean Squared Error	3116.0589	4365.9143	5420.1542	6242.7826	6522.625	6051.7249
Mean absolute percentage error	41.581	55.477	65.3317	71.99	70.477	61.835
Direction Accuracy	66.666	65.3846	24	12.5	13.0435	13.63
mean Squared Error	9709823.172	19061207.31	29378071.22	38972334.59	42544636.4	36623374.66

b. Future forecast visualization

Figure 5.19 shows the actual and predicted number of cases for rape using Gaussian process, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 1983, so as to find out the impact of amendment in criminal law in 1983. By observing both the actual data and predicted data we can find out the impact of the amendment.

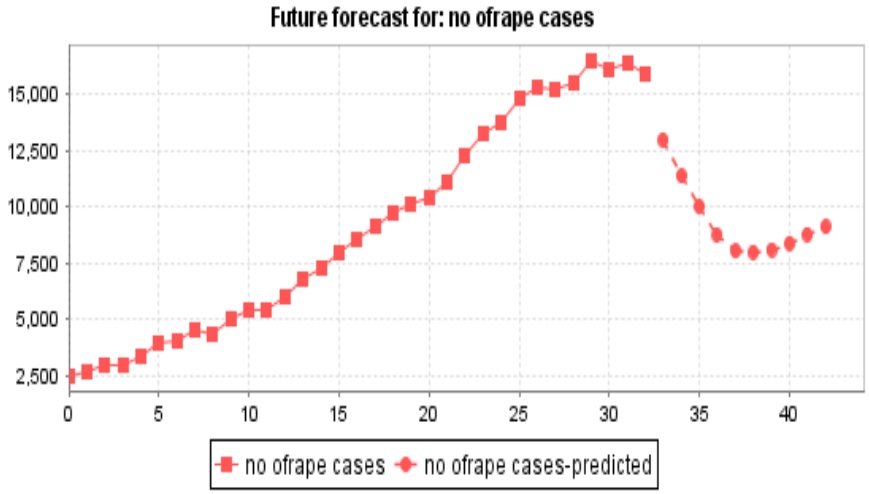

Figure 5.19: Actual and Predicted data for sexual Offences using Gaussian Process

c. Train Prediction for Target Visualisation

Figure 5.20 shows the comparison of predicted and actual data for rape dataset using Gaussian Process.

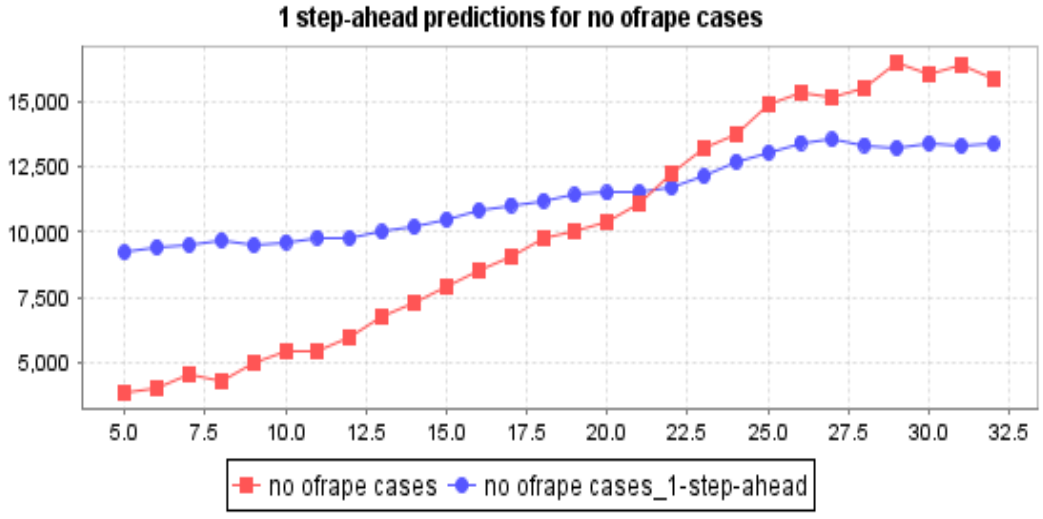

Figure 5.20: Comparison of actual and predicted data

d. ANOVA Analysis data for rape using Gaussian Process.

Table 5.30 summarizes the ANOVA output for the actual and predicted values for rape data using Gaussian Process Model. The p- value for this model is 8.67E-08.

Table 5.30: ANOVA Summary

Source of Variation	SS	Df	MS	F	P-value	F crit
Between Groups	8.6E+08	1	8.6E+08	73.90307	8.67E-08	4.413873
Within Groups	2.1E+08	18	11643197			
Total	1.07E+09	19				

5.4.2 Prediction using Linear Regression

Here we are using Linear Regression Model for forecasting. Table 5.31 gives the values of actual number of cases registered and the predicted number of cases that would be registered after amendment in sexual offences Act 2003.

Table 5.31 : Actual and predicted data using Linear Regression

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-
1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-
1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	-
1985	7289	-
1986	7952	-
1987	8559	-
1988	9099	-
1989	9752	-

1990	10068	-
1991	10410	-
1992	11112	-
1993	12218	-
1994	13208	-
1995	13754	-
1996	14846	-
1997	15330	-
1998	15151	-
1999	15468	-
2000	16496	-
2001	16075	-
2002	16373	-
2003	15847	-
2004	18233	17502.84
2005	18359	18250.37
2006	19348	18720.63
2007	20737	19358.16
2008	21467	19937.49
2009	21397	20476.1
2010	22172	21023.67
2011	24206	21567.41
2012	24923	22106.43
2013	33707	22645.66

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.32 tabulates these values calculated by WEKA for prediction using Linear Regression Model.

Table 5.32 : Evaluation Data for prediction using Linear Regression

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	28	27	26	25	24	23
Mean Absolute Error	419.5962	478.64	495.128	518.743	530.267	547.552
Root Mean Squared Error	526.659	592.583	610.842	631.559	644.295	656.936
Mean absolute percentage error	4.71	5.09	5.09	5.128	5.029	5.132
Direction Accuracy	85.18	84.61	84	87.5	86.95	86.36
mean Squared Error	277369.98	351155.25	373128.03	398866.95	415116.65	431565.439

b. Future forecast visualization

Figure 5.21 shows the actual and predicted number of cases for rape using Linear Regression, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 2003, so as to find out the impact of Sexual Offences Law 2003. By observing both the actual data and predicted data we can determine the impact of the amendment.

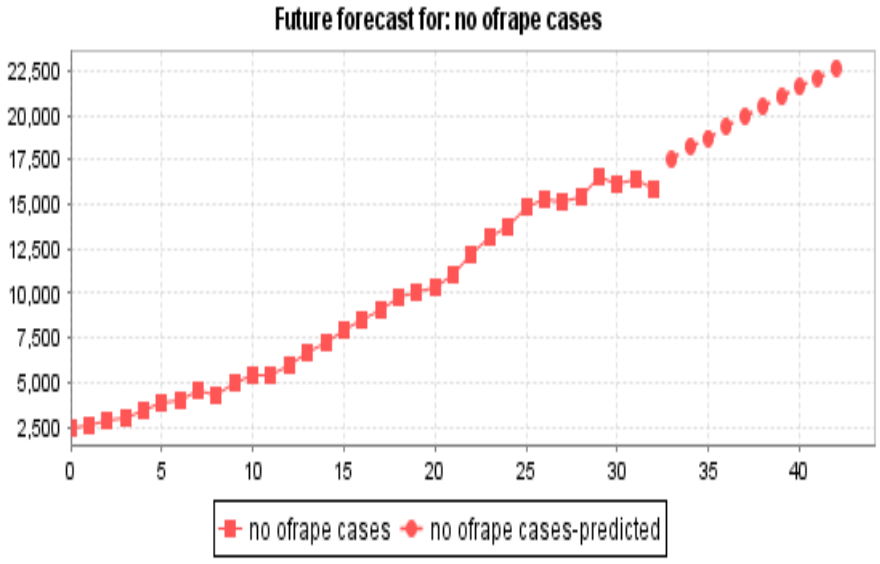

Figure 5.21: Actual and Predicted data using Linear Regression

c. Train Prediction for Target Visualisation

Figure 5.22 shows the comparison of predicted and actual data for rape dataset using Linear Regression.

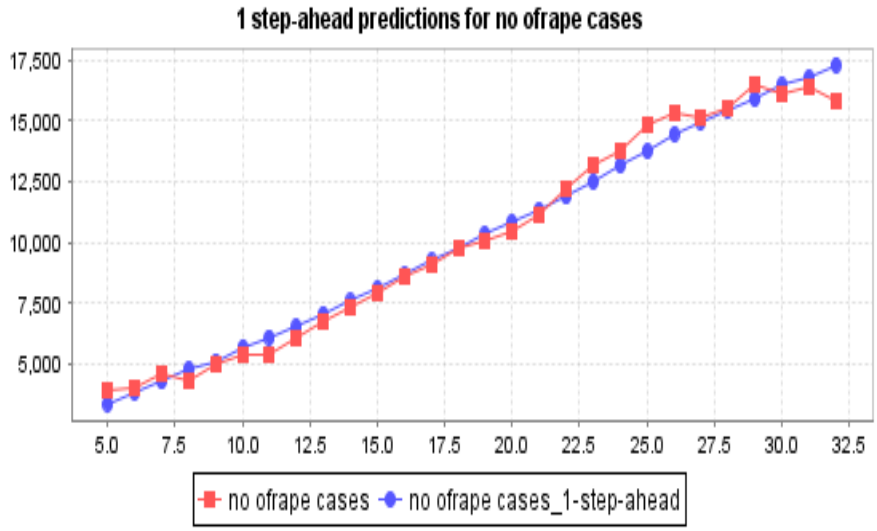

Figure 5.22 : comparison of actual and predicted data for Linear Regression Model

d. ANOVA Analysis data for rape data using Linear Regression Model

Table 5.33 summarises the ANOVA output for the actual and predicted values using Linear Regression Model. The p- value for this model is 0.151447.

Table 5.33:ANOVA Summary

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	26358631	1	26358631	2.244213	0.151447	4.413873
Within Groups	2.11E+08	18	11745153			
Total	2.38E+08	19				

5.4.3 Prediction using Multilayer Perceptron

Here we are using Multilayer Perceptron Model for forecasting. Table 5.34 gives the values of actual number of cases registered in India since 1971 till 2013 and the predicted number of cases that would be registered after amendment in Sexual Offences Act in 2003.

Table 5.34 : Actual and Predicted data using Multilayer Perceptron

Year	Actual number of rape cases Registered	Predicted number of rape - to be registered
1971	2487	-
1972	2605	-

1973	2919	-
1974	2962	-
1975	3376	-
1976	3893	-
1977	4058	-
1978	4558	-
1979	4300	-
1980	5023	-
1981	5409	-
1982	5427	-
1983	6019	-
1984	6740	-
1985	7289	-
1986	7952	-
1987	8559	-
1988	9099	-
1989	9752	-
1990	10068	-
1991	10410	-
1992	11112	-
1993	12218	-
1994	13208	-
1995	13754	-
1996	14846	-
1997	15330	-
1998	15151	-
1999	15468	-
2000	16496	-
2001	16075	-
2002	16373	-
2003	15847	-
2004	18233	17170.37
2005	18359	17433.98
2006	19348	17662.27

2007	20737	17782.01
2008	21467	17958.16
2009	21397	18052.79
2010	22172	18166.03
2011	24206	18268.06
2012	24923	18364.19
2013	33707	18448.44

a. Evaluation on training data

Training Data is evaluated by calculating Mean absolute error, Root mean squared error, Mean absolute percentage error, Direction accuracy and Mean squared error for each prediction. Table 5.35 tabulates these values calculated by WEKA for prediction using Multilayer Perceptron Model.

Table 5.35 :Evaluation Data for prediction using Multilayer Perceptron

Target	1-step-ahead	2-steps-ahead	3-steps-ahead	4-steps-ahead	5-steps-ahead	6-steps-ahead
N	28	27	26	25	24	23
Mean Absolute Error	278.238	313.4656	324.037	334.687	328.432	348.669
Root Mean Squared Error	358.226	392.781	403.735	410.906	407.106	419.936
Mean absolute percentage error	2.926	3.249	3.235	3.294	2.966	3.127
Direction Accuracy	85.185	84.615	84	87.5	86.95	86.36
Mean Squared Error	128326.259	154277.07	163002.684	168844.374	165735.79	176346.8

b. Future forecast visualization

Figure 5.23 shows the actual and predicted number of cases for rape using Multilayer Perceptron, in the future forecast visualisation in WEKA. Rape data has been compiled by NCRB from the year 1971. Hence the training dataset consists of data from year 1971 to 2003, so as to find out the impact of Sexual Offences Law 2003. By observing both the actual data and predicted data we can find out the impact of the amendment.

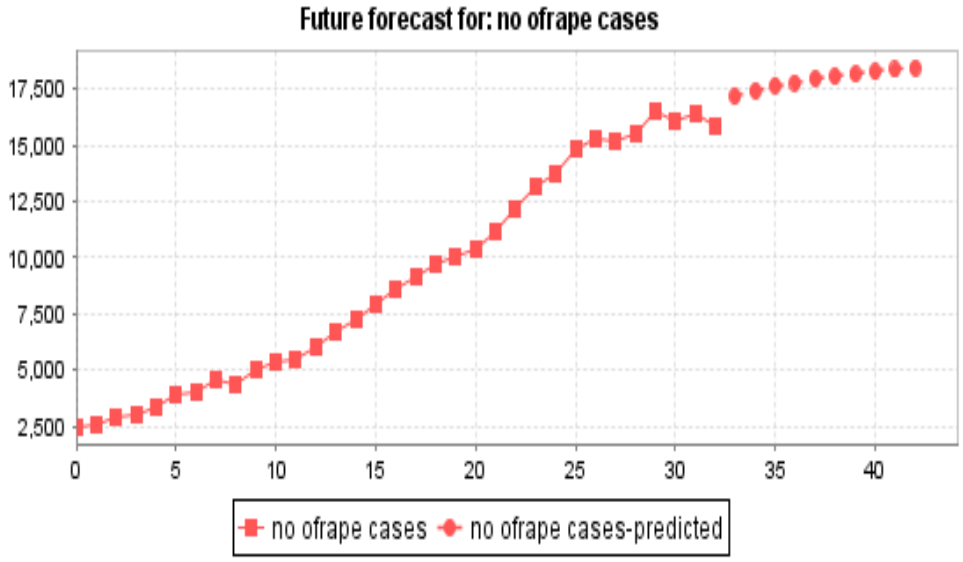

Figure 5.23: Actual and predicted number of cases for rape using Multilayer Perceptron

c. Train Prediction for Target Visualisation

Figure 5.24 shows the comparison of predicted and actual data for rape dataset using Multilayer Perceptron.

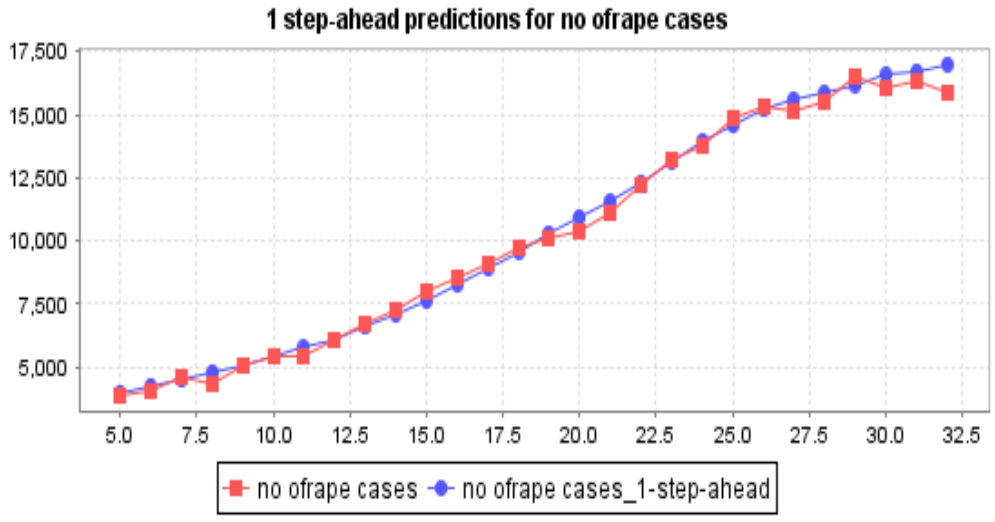

Figure 5.24 : Comparison of actual and predicted data for rape dataset using Multilayer Perceptron

d. ANOVA Analysis data for dowry using Multilayer Perceptron

Table 5.36 summarizes the ANOVA output for the actual and predicted values using Multilayer Perceptron Model. The p- value for this model is 0.005647.

Table 5.36 : ANOVA Summary

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.02E+08	1	1.02E+08	9.866837	0.005647	4.413873
Within Groups	1.87E+08	18	10372634			
Total	2.89E+08	19				

5.5 Result Analysis of Models used for Prediction

We have carried out predictive analysis for four different cases. Based on ANOVA p value we have also determined if that model can be effectively used for prediction for that case or not. For Domestic Violence case we have determined that Gaussian Process model is appropriate whereas for Dowry Prohibition we must use Linear Regression mode. For Rape and sexual offences case we must use Multilayer Perceptron Model for prediction analysis.

Chapter 6

CONCLUSIONS AND FUTURE WORK

After analyzing the results of predictive analysis of the crimes Domestic Violence, Dowry and Rape, following conclusions have been drawn.

- The predicted values show that there is a big difference between actual number of cases registered per lakh of female population and the predicted number of cases per lakh of female population for all the crimes. For domestic violence actual number of cases registered per lakh of female population in 2013 is 20.06 whereas the predicted value is 6.5 cases per lakh of female population. For Dowry cases the actual crime rate for 2013 is 1.81 whereas predicted value for this is nearly 0.5. For dowry deaths crime rate for actual and predicted is nearly same. But for rape cases, situation is alarming after 2009 as the actual crime rate for rape in 2013 is 5.69 whereas the predicted crime rate is 1.76. Therefore we can conclude that there is not much influence of the laws, acts and policies on society.
- Much more amendments to the existing laws and new laws are required to be made to ensure women safety and they should cover all the crimes which are being committed against women but due to lack of appropriate policies or loopholes in the laws criminals are set free.
- One of the reasons of difference between actual number of cases registered per lakh of population and predicted number of cases to be registered per lakh of population can be that women are more aware about the laws now than before and ensure that their complaints get registered. Earlier people were afraid of going to police or were not aware about the laws and policies.
- It seems that all the laws and policies are not being enforced strictly to curb the crimes against women.

SUGGESTIONS

- More and more women should be educated about their rights, laws and policies made for their safety and security.

- A National level campaign should be launched by Government to make awareness among women about their safety and security.
- Every girl should be nurtured in a way that they could get a regular source of income so that they are self reliable and not dependent on others.
- Men and boys should be advised or counseled to respect women and share household chores.
- Women should also get share in her ancestral property.
- Government should take steps to enforce women safety related law and act strictly

FUTURE WORK

This work can be further extended by finding out the scenario of other crimes against women and plight of women victims in all other states of India and what steps other states have taken to combat such crimes. It can further be extended to analyse other types of crimes prevalent in society and what steps are being taken to curb those crimes. Also the dataset taken here is of across India, one can do the analysis on state wise data.

Hierarchical clustering and other predictive algorithms may be applied to determine the hotspots of the crime so that police and other resources required for controlling the crime in that area can be sent on time.

Chapter 7

REFERENCES

1. (n.d.). Retrieved from National Commission of Women, India: http://ncw.nic.in/
2. Acharya, B. d. (2012, March). Domestic violence. *Uthan Thoughts towards development, Domestic violence*, pp. 5-11.
3. *CHAPTER-8 : TEST OF HYPOTHESIS - People of Statistics.* (2001). Retrieved from http://gauss.stat.su.se/gu/e/slides/Time%20Series/Forecast.pdf
4. *Child Marriage in India*. (n.d.). Retrieved from Centre for Social Research: http://www.csrindia.org/index.php/child-marriage-in-india
5. (1971-2013). *Crime in India.* National Crime Records Bureau.
6. DeKeseredy, W. S. (2012). Violence against Women. *Oxford Bibliographies*.
7. Deshmukh, D. B. (2012). Amendments to the laws relating to rape. *The Journal of Mimer Medical College*.
8. Devadoss, A. V., & Ligori, T. A. (2013). Forecasting of Stock Prices Using Multi Layer Perceptron. *International Journal of Computing Algorithm*, 440-449.
9. (2007). Devdasi. encyclopedia britannica.
10. *Dowry Death*. (n.d.). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Dowry_death
11. *Dowry Law in India*. (n.d.). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Dowry_law_in_India
12. *Dowry system in India*. (n.d.). Retrieved from Wikipedia: http://en.wikipedia.org/wiki/Dowry_system_in_India
13. Felson, R. B. (2009). Violence, Crime, and Violent Crime. *IJCV: Vol. 3 (1)*, 23 – 39.
14. Goel, A. (2004). Violence and Protective Measures for Women Development and Empowerment. In *Violence and Protective Measures for Women Development and Empowerment* (pp. 3-4). New Delhi: Deep & Deep Publications.
15. Groff, E. R., & Vigne, N. G. (2002). Forecasting the Future of Predictive Crime Mapping. *National Criminal Justice Reference Service (NCJRS)*, 29-57.

16. Grovert, A. J. (2008, July). *Domestic Violence Against Women: A Literature Review.* Retrieved from http://commons.pacificu.edu/spp/92

17. Gupta, M., Chandra, B., & Gupta, M. P. (2006). Crime Data Mining for Indian Police Information System. *Journal of Crime Vol.2 Issue. 6*, 43-54.

18. Gupta, P., Mishra, S., & Pandey, S. K. (2014). Time Series Data Mining in Rainfall Forecasting Using Artificial Neural Network. *International Journal of Scientific Engineering and Technology*, 1060-1065.

19. Haldar, D., & Jaishanker, K. (2012). *Cyber crime and the Victimization of Women: Laws, Rights, and Regulations.* IGI Global.

20. Kuzhda, T. (2012). RETAIL SALES FORECASTIHG WITH APPLICATIOH THE MULTIPLE REGRESSION. *Socio-Economic Problems and the State*, 91-101.

21. Lane, D. M. (n.d.). Analysis of Variance.

22. Lane, D. M. (n.d.). *Introduction to Linear Regression*. Retrieved from http://onlinestatbook.com/2/regression/intro.html

23. *Linear Regression*. (n.d.). Retrieved from WIKIPEDIA- The free encyclopedia: http://en.wikipedia.org/wiki/Linear_regression

24. Mishra, J. A., & Patel, A. B. (january-june 2013). Crimes against the Elderly in India: A Content Analysis on factors causing Fear of crime. *International Journal of Criminal Justice Sciences*, 13-23.

25. Mori, H., & Ohmi, M. (2005). Probabilistic Short-term Load Forecasting with Gaussian Processes. *IEEE transactions*.

26. *MultiLayer Perceptron*. (n.d.). Retrieved from WIKIPEDIA- The free encyclopedia: http://en.wikipedia.org/wiki/Multilayer_perceptron

27. Patel, P. R. (2013). Domestic Violence: A Social Evil. *Galaxy: International Multidisciplinary Research Journal*.

28. Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.* National Criminal Justice Reference Service (NCJRS).

29. Rani, A., & S., R. (2014). Crime Trend Analysis And Prediction Using Mahanolobis Distance And Dynamic Time Warping Technique. *Researchjournali's Journal of Computer Science Vol. 1| No. 3*.

30. Ranjan, A. K. (2013, January 30). *Crime Against Women In India.* Retrieved from Countercurrents.org: http://www.countercurrents.org/ranjan300113.htm

31. *Rape in India*. (n.d.). Retrieved from wikipedia: http://en.wikipedia.org/wiki/Rape_in_India

32. Rasmussen, C. E. (2006). *Gaussian Processes in Machine Learning.* The MIT Press.

33. Tandon, S. L., & Sharma, R. (2006). Female Foeticide and Infanticide in India : An Analysis of Crimes against Girl Children. *International Journal of Criminal Justice Sciences*.

34. *The Child Marriage Restraint Act, 1929*. (n.d.). Retrieved from http://wcd.nic.in/cmr1929.htm

35. *THE COMMISSION OF SATI (PREVENTION) ACT, 1987.* (2006, December). Retrieved from http://www.wcd.nic.in/commissionofsatiprevention.htm

36. *The Facts on Children and Domestic Violence.* (2008). Retrieved from Futures Without Violence: http://www.lessonsfromliterature.org/docs/Children-and-Domestic-Violence.pdf

37. *The Gazette of India.* (2005). Retrieved from http://wcd.nic.in/wdvact.pdf: http://wcd.nic.in/wdvact.pdf

38. *The world's most dangerous countries for women*. (2011). Retrieved from TrustLaw: www.trust.org/trustlaw

39. *Time Series Analysis and Forecasting with Weka.* (n.d.). Retrieved from http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka#TimeSeriesAnalysisandForecastingwithWeka-1Introduction

40. *Violence against Women*. (n.d.). Retrieved from WIKIPEDIA - The free encyclopedia: http://en.wikipedia.org/wiki/Violence_against_women

41. *WEKA Manual.* (n.d.). Retrieved from Machine Learning Group at the University of Waikato: http://www.cs.waikato.ac.nz/ml/weka/

42. *What Is Section 498A Of The Indian penal code?* (n.d.). Retrieved from https://ipc498a.files.wordpress.com/2007/11/what-is-section-498a-of-the-indian-penal-code.pdf

43. *Women in India*. (n.d.). Retrieved from WIKIPEDIA- The free encyclopedia: http://en.wikipedia.org/wiki/Women_in_India

Chapter 8

APPENDIX-I

DATASETS USED

1. Domestic Violence

This dataset has data from year 1991 to 2013. It represents the yearly number of cases registered all over India for dowry domestic violence.

Table 8.0.1: Dataset for Domestic violence

Year	Number of Cases Registered	Rate
1991	15949	3
1992	19750	3
1993	22064	3.1
1994	25946	3.1
1995	28579	3.1
1996	35246	3.8
1997	36592	3.8
1998	41375	4.3
1999	43823	4.4
2000	45778	4.6
2001	49170	4.8
2002	49237	4.7
2003	50703	4.7
2004	58121	5.4
2005	58319	5.3
2006	63128	5.6
2007	75930	6.7
2008	81344	7.1
2009	89546	7.7
2010	94041	7.9
2011	99135	8.2
2012	106527	18.2
2013	118866	20.06

2. Dowry Prohibition

This dataset has data from year 1988 to 2013. It represents the yearly number of cases registered all over India under dowry prohibition act.

Table 8.2 : Dowry Prohibition dataset

Year	Number of cases Registered	Rate
1988	2084	0.3
1989	1918	0.2
1990	2155	0.3
1991	1841	0.2
1992	2102	0.2
1993	2679	0.3
1994	2709	0.3
1995	2814	0.3
1996	2647	0.3
1997	2685	0.3
1998	3578	0.4
1999	3064	0.3
2000	2876	0.3
2001	3222	0.3
2002	2816	0.3
2003	2684	0.3
2004	3592	0.3
2005	3204	0.3
2006	4504	0.4
2007	5623	0.5
2008	5555	0.5
2009	5650	0.5
2010	5182	0.4
2011	6619	0.5
2012	9038	1.54
2013	10709	1.81

3. Dowry Deaths

This dataset has data from year 1991 to 2013. It represents the yearly number of cases registered all over India for dowry death.

Table 8.3 : Dataset for Dowry Deaths

Year	No of Cases Registered	Crime Rate
1991	5157	0.7

1992	4962	0.6
1993	5817	0.7
1994	4935	0.6
1995	4648	0.5
1996	5513	0.6
1997	6006	0.6
1998	6975	0.7
1999	6699	0.7
2000	6995	0.7
2001	6851	0.7
2002	6822	0.6
2003	6208	0.6
2004	7026	0.6
2005	6787	0.6
2006	7618	0.6
2007	8093	0.7
2008	8172	0.7
2009	8383	0.7
2010	8391	0.7
2011	8618	0.7
2012	8233	1.41
2013	8083	1.36

4. Rape

This dataset has data from year 1971 to 2013. It represents the yearly number of cases registered all over India for rape and sexual offences.

Table 8.4 : Dataset for Rape

Year	Number of Cases Registered	Crime Rate
1971	2487	0.5
1972	2605	0.5
1973	2919	0.51
1974	2962	0.5
1975	3376	0.6
1976	3893	0.6
1977	4058	0.6
1978	4558	0.7
1979	4300	0.7
1980	5023	0.8
1981	5409	0.8
1982	5427	0.8
1983	6019	0.8
1984	6740	0.9
1985	7289	1

1986	7952	1
1987	8559	1.1
1988	9099	1.1
1989	9752	1.2
1990	10068	1.2
1991	10410	1.2
1992	11112	1.3
1993	12218	1.4
1994	13208	1.5
1995	13754	1.5
1996	14846	1.6
1997	15330	1.6
1998	15151	1.6
1999	15468	1.6
2000	16496	1.6
2001	16075	1.6
2002	16373	1.6
2003	15847	1.5
2004	18233	1.7
2005	18359	1.7
2006	19348	1.7
2007	20737	1.8
2008	21467	1.9
2009	21397	1.8
2010	22172	1.9
2011	24206	2
2012	24923	4.26
2013	33707	5.69