Crime Analysis and Prediction in 7 states of India using statistical software RStudio

Pankaj Kumar1, Smriti Tandon2*

1Associate Professor, Amity University, Kolkata
pkumar@kol.amity.edu,
2Associate Professor, Graphic Era Deemed to be University, Dehradun
smriti.tandon.mba@geu.ac.in
*Corresponding Author: smriti.tandon.mba@geu.ac.in

Abstract:

Crime analysis is a function of law enforcement. It comprises systematized analysis for understanding the movement of crime patterns. Crime is an increasing factor all over the world and the condition in India is the same. Common people feel insecure as they feel there is no safe place left in this world. It is also well known that crime increases with population and by looking at any year's population count of India it can be said how crime rate goes a step forward from the previous year. In this paper, data for the top 3 crimes in India namely Rape, Dowry death and Domestic violence is collected and pattern is analyzed using Holt's Linear Trend method (aka Double Exponential Smoothing). Through this paper, the respective crime patterns and future values are forecasted. The statistical computations and graphical presentation have been done with the help of the statistical software RStudio.

Keywords: Crime Analysis, Holt's Linear Trend method (aka Double Exponential Smoothing), RStudio

INTRODUCTION

Crime analysis can be done at various levels, including tactical, operational, and strategic. Crime report data as well as arrests report data and police calls are studied by crime analysts to identify developing patterns and trends as quickly as possible. The concept of crime has been defined from the social and legalstand point. According to Morwer, crimes are defined to be an anti-social act while Thorstein.Sellin says crime is nothing but violation of conduct norms of the normative group. Socio economic changes have been witnessed by India since independence. With increase in development and urbanization of big cities and towns there is also an increase in the rate of crimes. People feel very insecure in the presence of evil element in the society. According to Google, the major crimes of India are Rape, Dowry death, Domestic violence, Illegal drug trade, Arms trafficking etc. According to Crime Bureau of India, the cases of rape molestation and sexual assaults increased to about 44%, and 60% of them being girls under 18 years of age. This paper is limited to the study for the top 3 crimes in India namely Rape, Dowry death and Domestic violence due to lack of sufficient data for the rest of the crimes.

The National Crime Records Bureau states that in every twenty minutes one woman gets to be the victim of rape in India. A 23y/o rape victim was set on fire while she was on her way to the court for a hearing. And while there is an ongoing crisis due to the Covid 19 pandemic there is an exponential increase in marital rape, domestic violence, dowry death and sexual assault on both upper and lower class women. In India, violence acts against women is very common. Women have to face violence and insult not only from husbands but also from members of both the natal and the marital home. Girls and women in India do not usually get much privilege like the boys in terms of their position not only society but also in their families as well. The very idea of dowry is an important part in the domestic violence category.Laws implemented to fight these crimes :-

- Section 376A
- Section 376AB
- Section 376B
- Section 376C
- Section 376D
- Section 304B
- Section 113B
- Dowry Prohibition Act, 1961

The Protection of Women from Domestic Violence Act, 2005 Section 498A

Time series analysis is not a new phenomenon. It can be used in a multitude of business applications for forecasting a quantity into the future and explaining its historical patterns. Holt's Linear Trend Model is a time series forecasting procedure which is also known as Double Exponential Smoothing. It is a popular smoothing model for forecasting data with trend.

LITERATURE REVIEW

The remarkable rise in the information and communications technology has not only changed many of our beliefs, habits and traditions, but also has introduced a new wave of young billionaires. Holt's Linear Trend Method has been used to forecast stock market predictions [4], used for determining the major ambient air pollutants [5-6]. Michel and Makowski have compared statistical models for analyzing wheat yield time series using this process [7]. After using eight different forecasting models, their findings concluded that Holt's Linear Trend model shows better performance than others. CSRTI (Mysore) Scientists have used this method for time series forecasting in agricultural research and to forecast the export of tea from India up to the year 2020 [8-9]. According to a study conducted by Oni and Akanle[10] this method was proven to be the best one to describe the data having the lowest error in their study of comparing exponential smoothing models for forecasting Cassava production.

METHODOLOGY

The Methodology uses Holt's Linear Trend method (known as Double Exponential Smoothing) on data sets obtained from National Crime Records Bureau. The steps are:

1) DATA EXPLORATION
2) METHOD APPLIED
   2) ANALYSIS

These are detailed below:

1) DATA EXPLORATION

This analysis involves the yearly data for no. of Rapes committed, Dowry death and domestic violence from the year 2001-2019. The data has been collected for 7 states of India – Rajasthan, Uttar Pradesh, Madhya Pradesh, Maharashtra, Kerala, Assam and Haryana. The data has been collected from NCRB.gov.in.

Year	Rajasthan	UP	MP	Maharashtra	Kerala	Assam	Haryana
2001	1049	1958	2851	1302	562	817	398
2002	1051	1415	2891	1352	499	970	361
2003	1050	911	2738	1268	394	1095	353
2004	1038	1397	2875	1388	480	1171	386
2005	993	1217	2921	1545	478	1238	461
2006	1085	1314	2900	1500	601	1244	608
2007	1238	1648	3010	1451	512	1437	488
2008	1355	1871	2937	1558	568	1438	631
2009	1519	1759	2998	1483	568	1631	603
2010	1571	1563	3135	1599	634	1721	720
2011	1800	2042	3406	1701	1132	1700	733
2012	2049	1963	3425	1839	1019	1716	668
2013	3285	3050	4335	3063	1221	1937	971
2014	3759	3467	5076	3438	1347	1980	1174
2015	3644	3025	4391	4144	1256	1733	1070
2016	3656	4816	4882	4189	1661	1779	1189
2017	3305	4246	5562	1933	2035	2048	1104
2018	4335	3946	5433	2142	1972	1767	1296
2019	5997	3065	2485	2299	2044	1791	1480

Fig 1(a) : Data For Number Of Rapes Committed.

Year	Rajasthan	UP	MP	Maharashtra	Kerala	Assam	Haryana
2001	376	2211	609	308	27	59	285
2002	399	1893	674	303	17	70	256
2003	389	1322	648	368	33	60	222
2004	379	1708	751	314	31	74	251
2005	361	1564	739	341	21	99	212
2006	394	1798	764	387	25	105	255
2007	439	2076	742	436	27	100	269
2008	439	2237	805	390	31	103	302
2009	436	2232	858	341	20	170	281
2010	462	2217	892	393	22	175	284
2011	514	2322	811	339	15	121	255
2012	478	2244	743	329	32	140	258
2013	453	2335	776	320	21	170	263
2014	408	2469	733	279	28	188	293
2015	463	2335	664	268	8	216	243
2016	462	2473	629	248	25	157	260
2017	458	2711	635	239	12	193	254
2018	405	2521	547	200	17	176	216
2019	453	2424	550	196	8	164	248

Fig 1(a) : Data of Dowry Death.

Year	Rajasthan	UP	MP	Maharashtra	Kerala	Assam	Haryana
2001	5532	7365	2562	6090	2561	1248	1513
2002	5691	5679	3117	5353	2836	1694	1565
2003	5733	2626	2938	5452	2930	1808	1618
2004	6781	4950	3436	5646	3222	1945	2026
2005	4505	5997	2989	6233	3283	2206	2075
2006	7038	5204	2989	6738	3708	2548	2254
2007	7650	8170	3294	7356	3999	3000	2412
2008	8113	8312	3185	7829	4138	3478	2435
2009	10371	8566	3983	7681	4007	4398	2617
2010	11145	7978	3756	7434	4797	5410	2720
2011	12218	7121	3732	7136	5377	5246	2740
2012	13312	7661	3988	7415	5216	6407	3137
2013	15094	8781	4988	8542	4820	8636	3617
2014	15905	10471	6451	7696	4919	9626	3478
2015	14383	8660	5281	7640	3668	11225	3525
2016	13814	11166	6264	7215	3462	9321	3314
2017	11508	12895	6144	6711	3021	10598	3389
2018	12363	14361	4160	6882	2070	11261	4195
2019	18453	18617	5487	8561	3039	11943	4875

Fig 1(a) : Data of Domestic Violence.

2) METHOD APPLIED

The method used here is Holt's Linear Trend method which is also known as Double Exponential Smoothing. It is a widespread and prevalent data-driven method for forecasting series with a trend but no seasonality. The main idea in this method is to take the simple exponential smoothing set up but to add a trend component.

This method involves 3 equations:

i. $\hat{y}_{t+k} = a_t + kc_t$: forecast equation

ii. $a_t = \gamma y_t + (1 - \gamma)(a_{t-1} + c_{t-1})$: level equation

iii. $c_t = \delta(a_t - a_{t-1}) + (1 - \delta)c_{t-1}$: trend equation Here,

γ: smoothing constant for level. $0 < \gamma < 1$

δ: smoothing constant for trend. $0 < \delta < 1$

a_t: estimate of the level. c_t: estimate of the trend.

y_t: observation at time t.

3) 7 STATES ANALYSIS

RStudio is used for analyzing datasets. RStudio is recommended professional data science solution IDE. It comprises of a set of integrated tools designed to be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and

managing workspace. The states are analysed for three different crimes one by one. The descriptive analysis is mentioned below:

1) RAJASTHAN
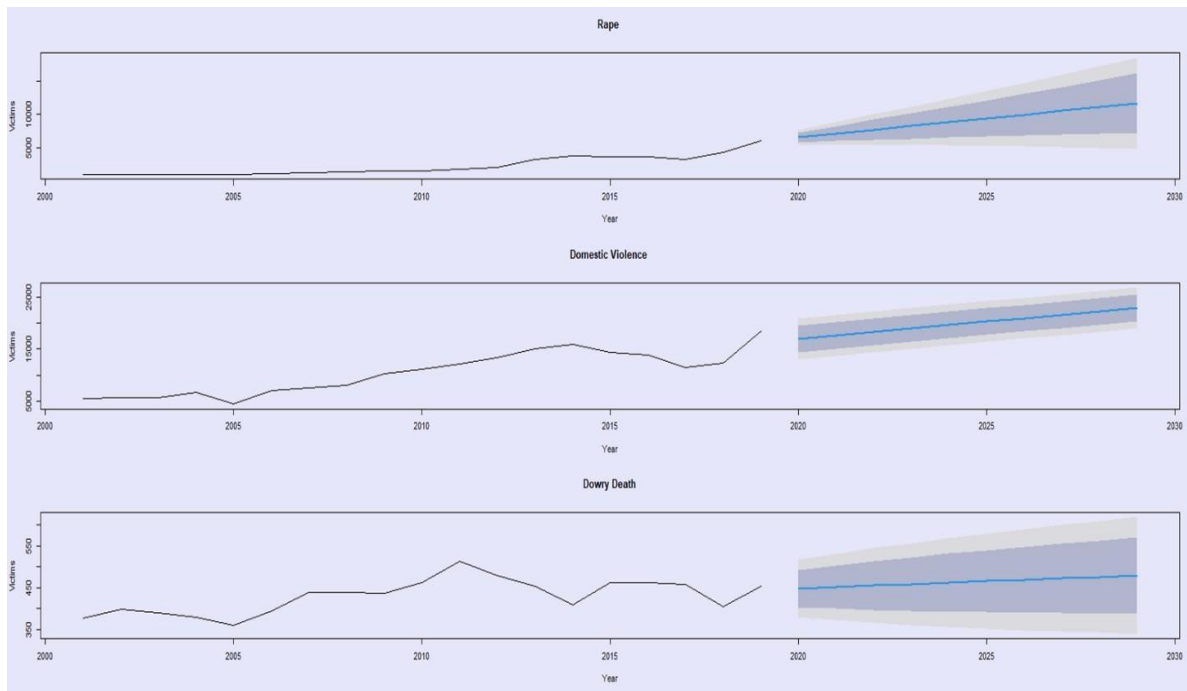
The number of rape cases in Rajasthan were more or less the same from 2001 to 2012 and after that a slight increasing trend is observed. from the graphical plot, the no. of rapes committed between 2020 to 2029, will increase upto 11649 by the end of 2029 from 5997 by the end of 2019. According to the 95% confidence interval there can be a hike of about 18479 inrape cases by 2024 or there can be a huge decline in the crime of about 4819.

In case of domestic violence an upward trend can be observed from the year 2005 to 2015 and after that a slight decline is observed till the year 2018. From the plot, the domestic violence numbers between 2020 to 2029, will increase upto 22892 by the end of 2029 from 18453 by the end of 2019. According to the 95% confidence interval there can be a hike of about 26783 in domestic violence by 2024 or there can be a decline in the crime of about 19000.

Now dowry death is a category that falls under domestic violence. It means that at the end of 2019 about 2% of domestic violence ended in the death of woman. Here the 95% ci is 339 and 618 ie By the end of 2029 the number of death due to dowry can increase upto 618 or can decline up to 339.

The next 10 years forecast(approx.. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	6562	17018	448
2021	7127	17670	451
2022	7693	18323	455
2023	8259	18976	458
2024	8823	19629	461
2025	9388	20281	465
2026	9954	20934	469
2027	10519	21587	472
2028	11084	22239	476
2029	11649	22892	479

2) UTTAR PRADESH
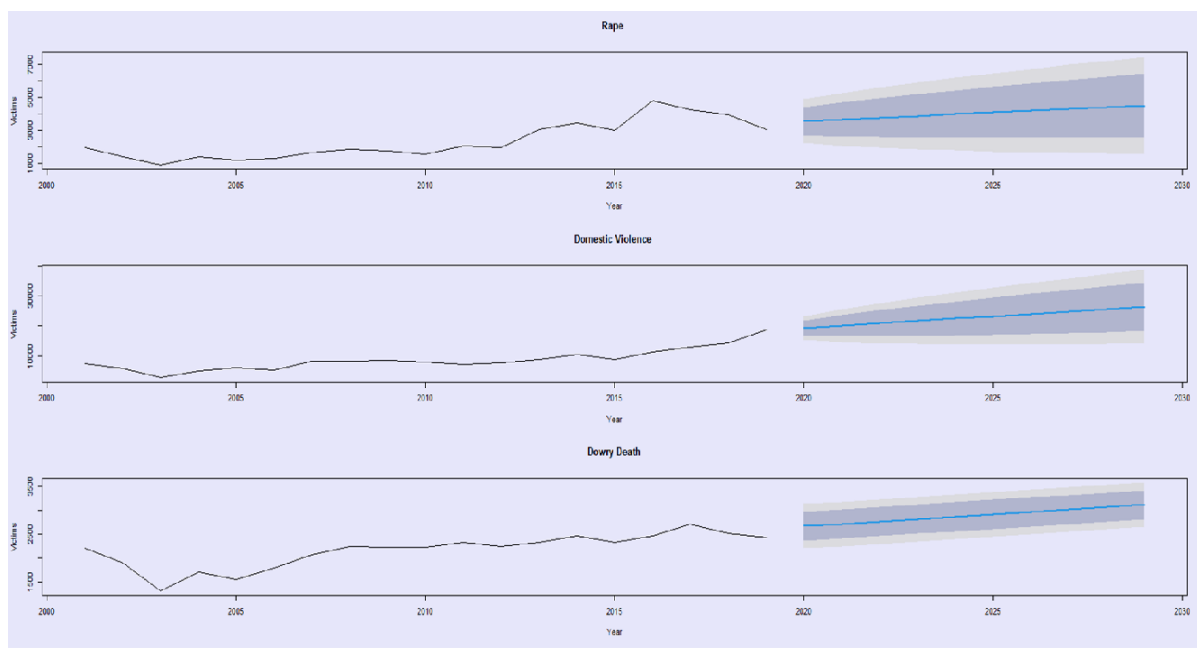
In case of Uttar Pradesh after a sudden drop in rape cases in the year 2003 the number of the crime has been more or less the same from 2004-2012. After that an increase in rape cases is observed till the year 2016. Again there is a drop in cases from 2016-2019. From the forecasted region an increasing trend is observed. The number of rapes committed will reach 4525 by the year 2029 from 3065 which was observed at the end of the year 2019. Here the 95% confidence interval is (1577, 7472) which means by 2029 the number of rape cases may be as high as 7472 or as low as 1577.

The number of domestic violence victims are more or less the same in case of Uttar Pradesh till the year 2015 and after that a slight increase is observed. The domestic violence numbers will increase upto 26458 by the end of 2029 from 18617 by the end of 2019. Here the confidence interval is (14081, 38835) which means the crime numbers can reach up to 38835 by the year 2029 or may fall up to 14081.

The plot shows a significant drop in cases in the year 2003 but since then an increasing trend is observed. At the end of 2019 about 13.02% domestic violence cases ended in dowry death. This percentage will decrease to 11.75% by the end of the year 2029. The 95% confidence interval says that the number of cases may be as high as 3574 or as low as 2647 by the year 2029.

The next 10 years forecast(approx.. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	3559	19248	2662
2021	3667	20049	2712
2022	3774	20850	2762
2023	3881	21651	2812
2024	3988	22453	2861
2025	4096	23254	2911
2026	4203	24055	2961
2027	4310	24856	3011
2028	4417	25657	3061
2029	4525	26458	3110

3) MADHYA PRADESH
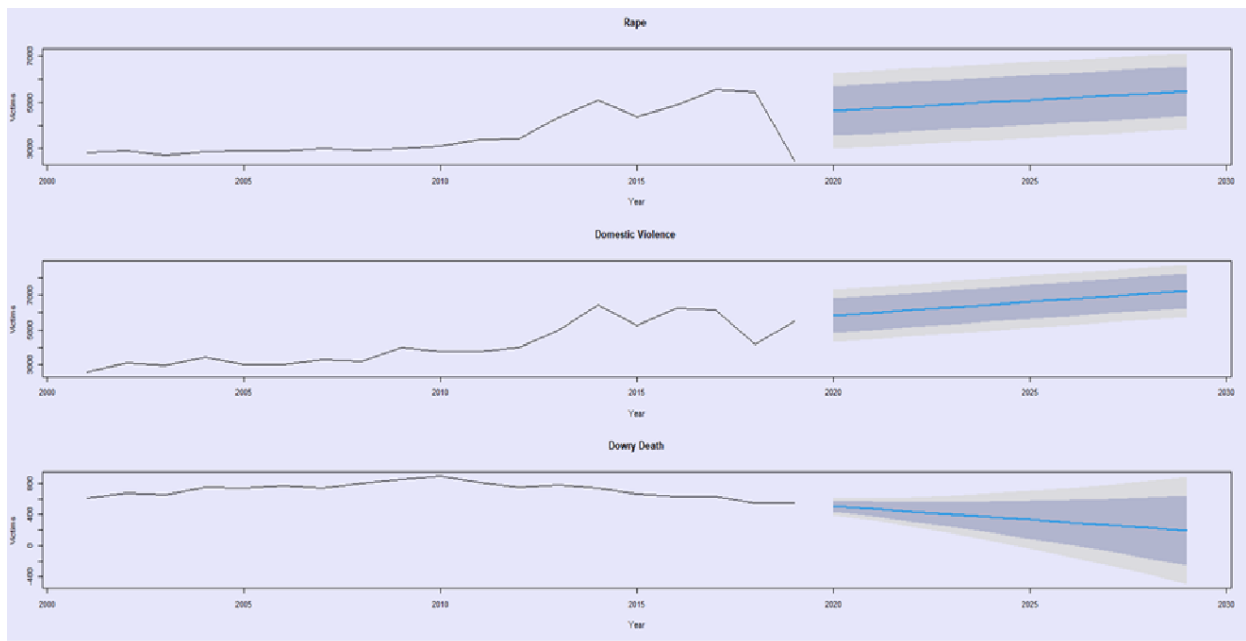
A significant drop in number of cases is observed in the year 2018 in rape cases in Madhya Pradesh. The number of cases are almost same over the years from 2001-2012 and after that the case numbers are seen to be getting high. According to the forecasted region there is a rise of 45.35% in the cases from the year 2019-2029. Here the 95% confidence interval is (6267,7111)

After being more or less the same in number of domestic violence cases from the years 2001 to 2011 a significant increase is seen. The number of cases drop at the year 2018. According to the forecast by the year 2029 the number of cases will increase upto 7243 from 5487 in the year 2019. The 95% confidence interval indicates that there may be a rise in cases of 8743 or a drop in cases by 5744 by the end of 2029.

The number of death due to dowry is more or less high in case of Madhya Pradesh. But from the year 2010 a decreasing trend is observed. At the end of 2019 about 10.02% domestic violence cases ended in dowry death. This rate will decrease a significant amount-2.65% by the end of the year 2029. From the 95% confidence interval the number of cases canbe as high as 881 or there may be no cases at all.

The next 10 years forecast(approx.. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	4636	5819	503
2021	4730	5978	468
2022	4823	6136	434
2023	4917	6294	399
2024	5011	6452	365
2025	5105	6610	330
2026	5198	6769	296
2027	5292	6927	262
2028	5386	7085	227
2029	5480	7243	192

4) MAHARASHTRA
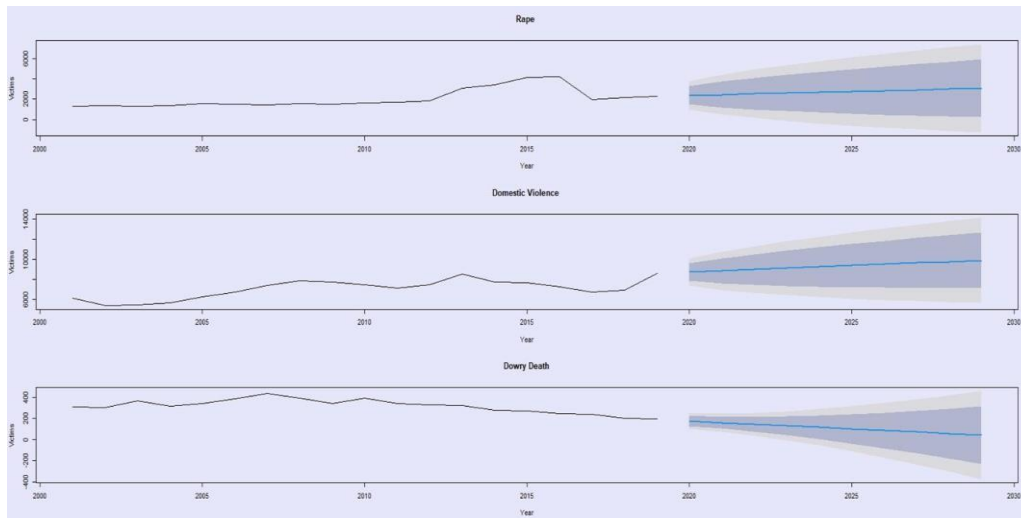
In case of Maharashtra the number of rape cases are at a constant level from 2001-2012 after which there is a little increase from the year 2012-2016. According to the forecasted region the number in cases goes to 3064 in the year 2029 from 2299 in the year 2019. The 95% confidence interval is (-1279, 7408).

In Maharashtra the numbers are more or less the same in domestic violence cases. According to the forecast by the year 2029 the number of cases will increase upto 9883 from 8651 in the year 2019. The 95% confidence interval indicates that there may be a rise in cases of 14128 or a drop in cases by 5693 by the end of 2029.

The number of death due to dowry is more or less high in case of Maharashtra. But from the year 2010 a decreasing trend is observed. At the end of 2019 about 2.3% domestic violence cases ended in dowry death. This rate will decrease a significant amount- 0.42% by the end of the year 2029. From the 95% confidence interval the number of cases can be 459 or there may be no cases at all.

The next 10 years forecast (approx. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	2375	8693	174
2021	2452	8825	160
2022	2529	8958	145
2023	2605	9090	130
2024	2682	9222	116
2025	2758	9354	101
2026	2835	9487	86
2027	2911	9619	72
2028	2988	9751	57
2029	3064	9884	42

5) KERALA
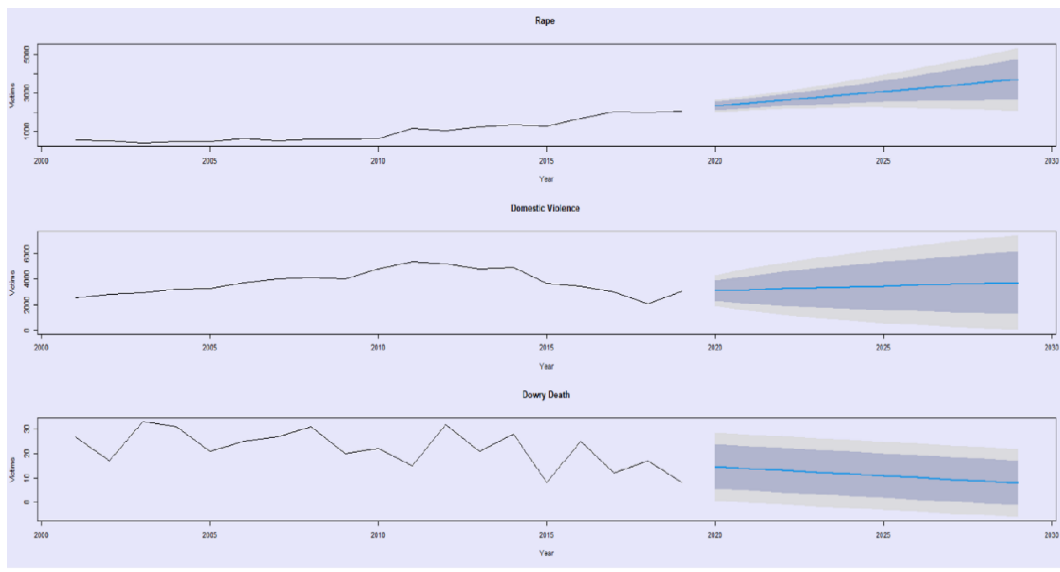
From the plot it can be seen that the number of rape case is increasing in the upcoming 10 years. We see that, by the end of 2019, the crime number is 2044 which will be increasing by the end of 2029 and becomes around 3710. The 95% C.I. is (38.55, 7455.80). Hence, we can say that, by 2024, the crime number can go up as high as approximately 7446 or as low as 39(approx.). So, looking at the fluctuations, it can be lesser than the crime rate observed by the end of 2019. But in general, the forecasts show a more or less increasing rate over the time period.
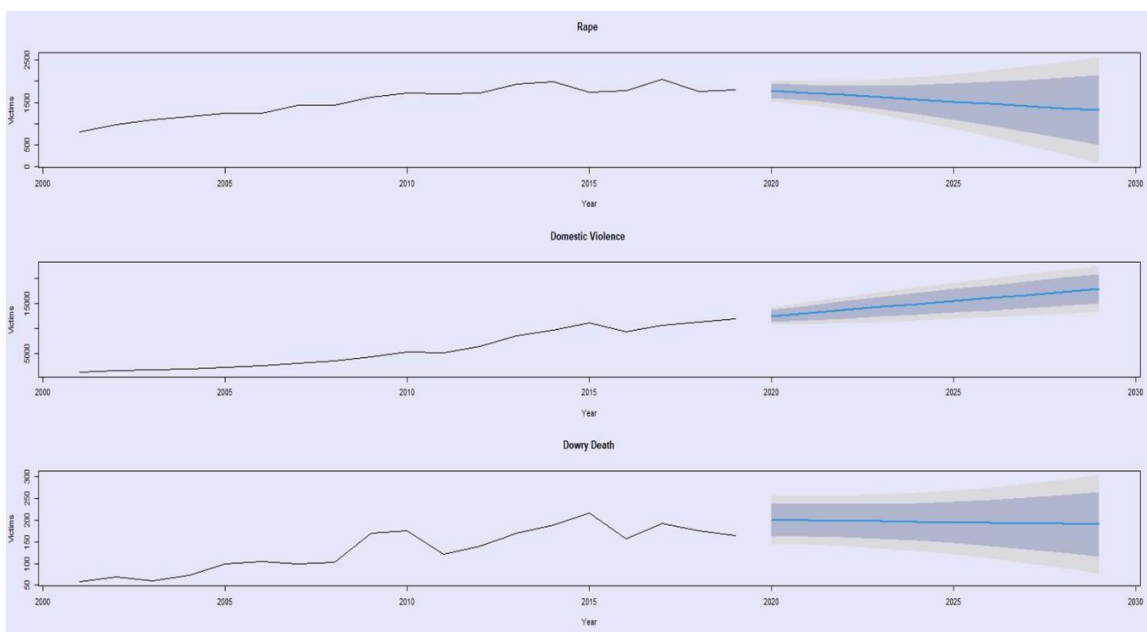
In Kerala the numbers are more or less the same in domestic violence cases, with a decrease from the year 2014 to the year 2018. According to the forecast by the year 2029 the number of cases will increase upto 3747 from 3039 in the year 2019. The 95% confidence interval indicates that there may be a rise in cases of 7456(approx.) or a significant drop in cases by 39(approx.) by the end of 2029.

An unusual trend is observed throughout the years in case of dowry death in Kerala. At the end of 2019 about 0.26% domestic violence cases ended in dowry death. By the end of 2029 the number of dowry death cases will be 8(approx..) which is same at the year 2019. From the 95% confidence interval the number of cases can be 22(approx..) or there may be no cases atall.

The next 10 years forecast (approx. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	2323	3110	15
2021	2477	3181	14
2022	2631	3251	13
2023	2785	3322	12
2024	2939	3393	12
2025	3094	3464	11
2026	3248	3535	10
2027	3402	3605	9
2028	3556	3676	9
2029	3710	3747	8

6) ASSAM

In Assam, the number of rape committed shows a decreasing nature in the next 10 years. In the beginning of 2001, the crime number was as high as 1473 which decreased by the end of 2029 to 1315(approx..). The 95% confidence interval is (67.49, 2561.85).
This means that the number of the crime is as low as 67(approx..) to as high as 2562(approx..).

In Assam the numbers are more or less the same in domestic violence cases till 2010, after which an increasing trend is observed. According to the forecast by the year 2029 the number of cases will increase upto 17927 from 11943 in the year 2019. The 95% confidence interval indicates that there may be a rise in cases of 22465(approx.) or a significant drop in casesby 13389(approx.) by the end of 2029.

Again, in case of Assam an unusual trend is observed throughout the years. At the end of 2019 about 1.37% domestic violence cases ended in dowry death. By the end of 2029 the number of dowry death cases will be 190(approx..). From the 95% confidence interval the number of cases can be as high as 303(approx..) or as low as 77(approx.)

The next 10 years forecast(approx.. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	1776	12522	200
2021	1725	13123	199
2022	1674	13723	198
2023	1622	14324	197
2024	1571	14924	196
2025	1520	15525	195
2026	1469	16125	193
2027	1417	16726	192
2028	1366	17327	191
2029	1315	17927	190

7)HARYANA
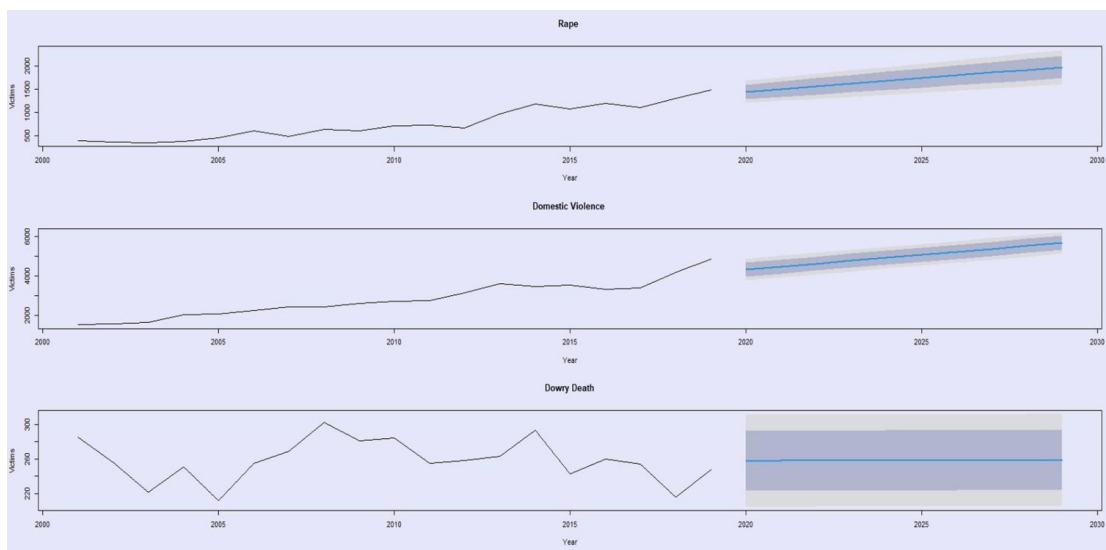
After being more or less the same from 2001-2011 the number of rape case show an increasing trend in Haryana. From the forecasted region also, an increasing trend is observed. The number of rapes committed will reach 1969 by the year 2029 from 1480 which was observed at the end of the year 2019. Here the 95% confidence interval is (1608.66, 2329.88) which means by 2029 the number of rape cases may be as high as 2330(approx..) or as low as 1609(approx..).

In case of domestic violence an increasing trend is observed in Haryana. The number this crime committed is also increasing in the upcoming 10yrs. According to the forecast by the year 2029 the number of cases will increase upto 5680 from 4875 in the year 2019. The 95% confidence interval indicates that there may be a rise in cases of 6227(approx.) or a drop in cases by 5113(approx.) by the end of 2029.

An unusual trend is observed throughout the years in case of dowry death in Haryana. At the end of 2019 about 5.08% domestic violence cases ended in dowry death. By the end of 2029 the number of dowry death cases will be 259(approx..). Here the 95% confidence interval is (205.74, 311.5)

The next 10 years forecast (approx.. values) :

Year	RAPE	Domestic Violence	Dowry Death
2020	1776	12522	200
2021	1725	13123	199
2022	1674	13723	198
2023	1622	14324	197
2024	1571	14924	196
2025	1520	15525	195
2026	1469	16125	193
2027	1417	16726	192
2028	1366	17327	191
2029	1315	17927	190

CONCLUSION & FUTURE SCOPE

The number of crimes reported is observed to be increasing over the years, Rape cases and Domestic Violence is increasing in 6 states namely Rajasthan, Madhya Pradesh, Uttar Pradesh, Maharashtra, Kerala and Haryana) and decreasing in 1 state i.e. Assam. However, Dowry death is decreasing in 3 states including Madhya Pradesh, Maharashtra, Kerala and increasing in 4 states. UP would be the most unsafe place in the upcoming years. Due to COVID-19, year 2020 datasets are not explored. The crimes in consideration are gender biased. However, Men also experience some sort of torture or violence at the hands of their wives or their intimate partners in their lifetime. Crimes likes these generally do not get reported and are not included in survey.

REFERENCES

[1]   . National Crimes Records Bureau (NCRB)[2]. www.statista.com
[3]. www.data.gov.in

[4].SR Sandipan Biswas, Indranil Sarkar, Prasenjit Das, Rajesh Bose: Examining the Effects of Pandemics on Stock Market Trends through Sentiment Analysis, Journal of Xidian University, Vol 14 (6), 2020.

[5]   Zakaria, N. and Noor, N.: Imputation methods for filling missing data in urbanair pollution data for malaysia, Urbanism. Architecture. Constructions Vol 9: 159–166, 2018.

[6]   Zhao, N., Liu, Y., Vanos, J. and Cao: Day-of-week and seasonal patterns ofpm2.5 concentrations over the united states: Time-series analyses using the prophetprocedure, Atmospheric Environment 192: 116 – 127, 2018.

[7]   Lucie Michel,David Makowski :Comparison of Statistical Models for Analyzing Wheat Yield Time Series, PLOS Sustainability and Transformation, October 29, 2013.

[8]   Oni, O.V. and Akanle, Y.O: Comparison of Exponential Smoothing Models for Forecasting Cassava Production. International Journal of Scientific Research in Mathematical and Statistical Sciences, 5(3), 65-68, 2018.

[9]   V Oni1, Y.O. Akanle, Ibrahim-Tiamiyu,: Inclusion of Birth and Death Rate in the Modelling of Lassa Fever in Nigeria , International Journal of Scientific Research in Mathematical and Statistical Sciences Volume-5, Issue-4, pp.22-32, August 2018.