The Role of Artificial Intelligence in Improving Criminal Justice System: Indian Perspective

Puneet Gawali

M.S. in Digital Forensics and Information Security, Institute of Forensic Science, Gujarat Forensic Sciences University. Address: Gandhinagar, Gujarat, India. E-mail: puneet.dfis1808@gfsu.edu.in

Reeta Sony

Assistant Professor, Centre for Studies in Science Policy, School of Social Sciences, Jawaharlal Nehru University, PhD. Address: New Mehrauli Road, JNU Ring Rd., New Delhi 110067, India. E-mail: reetasony@msil.jnu.ac.in

Abstract

The increasing cyber-attacks have created havoc in the criminal justice system. Understanding the purpose of crime and countering it is the crucial task for the law enforcement agencies. This research aims to present how Artificial Intelligence and Machine Learning along with Predictive Analysis using soft evidence can be used in sorting out the existing criminal record while making the use of metadata, and therefore predicting crime. Furthermore, it would surely help out the police and intelligence bodies to smartly investigate the cases by referring to the database and thus help the society in curbing the crime by quicker and more effective investigation processes. It would also assist the analyst in tracking the activities and associations of various criminal elements through their recent activities, by extracting the particular details from the documents or records. Prediction of the crime can be understood through this research. The present study reflects the accuracy level of threat from 28 states of India. By researching on this topic, it becomes evident that if proper data is fed to this model, the chances of prediction are higher and more accurate. The study also tried to find out the psychosocial perspectives of the crime and what would be the reason of individual indulges in such crime.

Keywords

Artificial Intelligence, law enforcement, criminal justice, prediction algorithm, accuracy, machine learning, motives, cyber attacks, information technology laws.

For citation: Puneet G., Sony R. (2020) The Role of Artificial Intelligence in Improving Criminal Justice: Indian Perspective // *Legal Issues in the Digital Era,* no 3, pp. 78–96.

DOI: 10.17323/2713-2749.2020.3.78.96

Introduction

In 1956, Artificial Intelligence (AI) was first introduced by its father, John McCarthy, in Dartmouth [McCarthy J., 2006: 12–14]. Digital transformation brings risks, as technology is the first layer [Dmitrik N., 2020: 54–78]. In recent years, technologies based on AI and Machine learning (ML) have progressively increased in their capability and accessibility, showing no sign of abating [Caldwell M., 2020: 1–13]. By understanding the AI law for the future, its advantages and disadvantages that can make AI advisable to humanity [Cui Y., 2020: 187–191]. AI research and its regulation aspire to balance innovation's social security against potential harms and obstructions [King T., Aggarwal N., Taddeo M., Floridi L., 2020: 89–120]. Development, adoption, and promotion of AI are the priorities of the Indian Government to make lives easier for society [Marda V., 2018: 1–19]. The preciseness and verisimilitude of the details about where the crimes occur, furthermore information on the depiction of crimes provided an approach to understanding such crimes in other countries [Furtado V., 2010: 4–17]. McGuire and Holt's further throws light on the impressive and much needed Routledge Handbook of Technology, Crime and Justice [McGuire M., Holt T., eds., 2017: 1–722] that has evidence of criminology's burgeoning of technological interest [Hayward K., Maas M., 2020: 1–25]. The most important lookout to implement this research would be to update judges to be specialist in the field of computer; such laws should be implemented wherein all the judges should be well trained to use this technology[1]. Using Artificial Intelligence which is the main emerging technology invented by John McCarthy and is beneficial as it perceives all the data as it is. In contrast, a human mind has to choose or make a selection from the different pieces of data before reasoning, leading to possible errors[2]. Information technologies and its applications has become more diverse and effective, such as COPLINK. As this COPLINK, is a licensed software that bridges the gap by conducting research as well as solving real world crimes by helping police officers as they serve the community in a sophisticated and understandable way [Chen H., 2003: 271–285]. This COPLINK project unites University of Arizona's Artificial Intelligence Lab with the Tucson Police Department's law enforcement, where crimes analyst, detectives, sergeants use this technology [Hauck R., 2002: 30–37]. In this paper, we have discussed on how the Artificial Intelligence could be used for the resolutions of criminal justice system, since it becomes difficult for the court of law to maintain the database of all the criminal activities, we have tried to sort this issue by feeding some criminal data to the model created by us and

[1] Available at: https://builtin.com/artificial-intelligence (accessed: 25.11.2019)

[2] Available at: https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c (accessed: 25.11.2019)

therefore improving the way of investigation. Data plays a significant role in the criminal justice system, especially in predictive analysis[3] since the data itself reveals the information of the crime. It is crucial to think about the diverse and vast ethical dilemmas occurring in the criminal justice system, which involves making moral judgments and deciding about wrong and right. Data mining can be used in understanding and designing crime detection models [Nath S., 2006: 41–44].

Such ethics have been maintained since the model cannot be biased and gives accurate results. We understand that using predictive analysis is challenging in policing. Still, it should not mean that law enforcement agencies should not use analytics or intelligence for the improvement of investigation [Isaac W. 2017: 543]. With this research risk assessment and the investigation of the criminal justice system will become more sophisticated. The possible question raised would be who should be accountable for semi-automated decisions? [Završnik A., 2020: 567–583] since the accuracy is directly proportional to the data fed; therefore, the entire model depends on the specificity of the data. This tool can be useful for the lawyers as well those who are expert in technology and those who are not so technically advanced; they can make usage of this tool for predicting using different datasets [Alarie B., Niblett A., & Yoon A., 2018: 106–124].

1. Preparing the Model

The most crucial concept for approaching this topic would be the understanding of recidivism; through this model, we can keep a close watch on the behavior of various states and the crime committed. As shown in the Fig. (1) we can see how through certain steps our data is being processed in order to get the desired results. Different programming languages and environments enable ML research and development of its application. Python language has a tremendous growth within the scientific computing communities in the last decade, so in this case most recent ML and deep learning libraries are associated with Python based [Raschka S. et al, 2020: 193]. Python is used to prepare the model of predictive analysis and using the EDA (Exploratory Data Analysis) when a particular data becomes large or we need to understand some complex relationships in the variables. Through this paper we can perform the molding of such data for better investigation purpose.

First, the data is loaded in python and then we perform data cleaning and exploring the information in the variables. Pandas which provide data frames are imported using python, Matplotlib provides plotting support, and Numpy provides scientific computing within dimensional object support as seen in Fig. (2).

[3] How data plays a significant role. Available at: https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data (accessed: 09.04.2018)

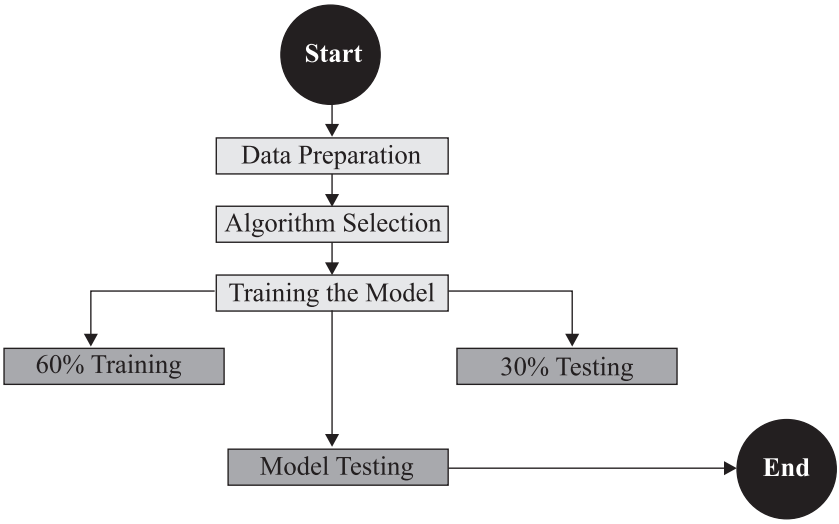

Fig. 1. Model Process Flowchart

```
In [2]:  import pandas as pd                          # pandas is a
         import numpy as np                           dataframe library
         import matplotlib.pyplot as plt              # numpy provides
         import seborn as seabornInstance             N-dim object support
         from sklearn.linear_model import             # matplotlib.pyplot
         LogisticRegression                           plots data
         from sklearn.model_selection import train_
         test_split
         from sklearn import metrics
         import os
         % matplotlib inline
```

Fig. 2. Importing Libraries

Secondly, standardization and visualization of data is very important to ensure that data fits the assumptions of the models. The Universal Rule of Law states that human rights, democracy and development depend on the level of progress the organizations and governments can achieve on the criminal justice front. The primary and crucial objectives of the criminal justice are controlling and preventing crime, maintaining law and order, protecting fundamental rights of victims along with the people in conflict with law, punishment and rehabilitation of those adjudged guilty of committing of crimes, and protection of life and property against crime and criminality in general. It is considered to be the primary obligation of the state under the constitution of India [Dhillon K., 2011: 27].

This paper would thus give an overview how every police station can update their data and predict the criminal behavior of the crime or any data available. Im-

porting various libraries and functions is the positive point of using python in this research paper since the data could be easily adjusted, it can be seen in Fig. 3 and 4.

Accurately predicting rare events is difficult, so the probability of having them in data is low, and the probability of training the algorithm is also low. Therefore, we only need a few percentages of the event to be able to train, to ensure that we have a reasonable chance to define how correctly a person or state is likely to develop the behavior or motive of committing a crime. Importing pandas will let us easily search the columns by name and see how many times this is true. Also, in the last column seen in the Fig.3 threat columns are mentioned which is categorically divided into binary 1s and 0s where 1s define that the attacks are increasing drastically whereas 0s define that the motives are mild. When a crime is predicted there will be questions arise regarding how an algorithm or code can be trustworthy[4]. This research would, therefore, throw light on this area where the data itself would be deciding everything, the more real the data the more effective the accuracy would be. Data mining and predictive analysis play an essential role in our life[5]. Now if we look into the data available very carefully, we can find whichever states having high unemployment rate (according to report by the Centre for Monitoring Indian Economy). It is noteworthy, that such states have high cybercrime rates which further denotes that in various states computer is used as a source to dupe money through various online frauds. The reason behind this is maintaining the anonymity and causing the harm because of vengeance or other motives. Cybercriminals mostly exploit the high-speed internet available at a lower cost to commit various criminal activities without being caught unless the states possess properly well-maintained cybersecurity labs to curb such crimes. The CMIE report further reveals that people belonging to age group 40 to 59 years have been successfully able to retain their jobs whereas people aged below 40 years were expelled out of their respective jobs which lead to social tension, desire of revenge, anger and other motives to launch such cyber-attacks[6].

The data shown in Fig. (3) presents the topmost cyber-crimes happened in various states of India until 2019. So far, which includes such crimes as bullying on social media and not full-fledged crimes wherein a lot of technical skills are required, this shows that certain age groups of people have launched such attacks to malign the image of the victim[7].

[4] How code can be trustworthy. Available at: https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337 (accessed: 05.03.2018)

[5] What is Data Mining? Definition of Data Mining, Data Mining Meaning — The Economic Times (indiatimes.com). Available at: https://economictimes.indiatimes.com/definition/data-mining (accessed: 07.12.2020)

[6] The recent unemployment data. Available at: https://www.cmie.com/kommon/bin/sr.php?kall=warticle&dt=2020-01-21%2009:51:47&msec=203 (accessed: 21.01.2020)

[7] National Crime Records Bureau Empowering Indian Police with Information Technology Available at: https://ncrb.gov.in/en (accessed: 22.10.2020)

State_UT	Personal_ Revenge	Anger	Fraud	Extortion	Causing_ Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Developing own business	Spreading	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Risk
Andhara Pradesh	34	26	733	45	7	0	92	12	1	1	1	0	2	14	2	0	1	236	0
Arunachal Pradesh	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
Assam	239	46	389	153	234	0	113	9	4	3	0	0	0	0	0	0	0	832	1
Bihar	5	8	351	2	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
Chhattisga	0	1	23	2	25	0	21	4	0	1	0	0	1	0	0	0	0	61	0
Goa	0	0	11	0	12	0	4	0	0	0	0	0	0	0	0	0	0	2	0
Gujarat	17	32	401	24	154	16	23	0	0	17	0	0	1	0	0	3	0	14	1
Haryana	6	9	137	21	11	2	75	0	12	2	0	0	2	0	0	0	0	141	1
Himachal	4	1	18	1	3	1	15	4	0	0	0	0	0	0	0	0	0	22	0
Jammu &	2	0	20	7	7	3	10	3	1	2	3	0	0	1	0	0	0	14	0
Jharkhand	16	6	783	44	16	0	16	1	16	0	0	0	32	0	0	0	0	0	0
Karnataka	27	10	5441	97	49	1	85	22	1	3	3	0	5	5	1	1	0	88	1
Kerala	69	18	93	8	48	3	50	18	3	0	0	0	6	0	0	0	0	24	0
Madhya P	93	10	230	19	109	2	49	1	3	29	1	0	4	20	0	0	1	169	1
Maharash	99	129	1998	31	64	18	724	20	0	33	2	2	13	6	0	3	0	369	1
Manipur	0	0	14	3	0	0	9	0	1	1	0	0	0	0	0	0	0	1	0
Meghalaya	0	0	35	0	3	0	3	3	0	11	0	2	0	6	0	0	0	11	0
Mizoram	0	4	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
Nagaland	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Odisha	7	0	506	224	0	0	37	0	0	0	0	0	0	0	0	0	0	69	1

State_UT	Personal_ Revenge	Anger	Fraud	Extortion	Causing_ Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Developing own business	Spreading	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Risk
Punjab	14	7	48	15	19	4	85	2	0	3	0	2	2	0	0	0	0	38	0
Rajasthan	9	11	499	31	66	14	60	3	0	9	0	0	17	2	0	0	0	383	1
Sikkim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Tamil Nad	39	13	55	7	17	6	36	52	2	39	1	0	3	1	0	4	0	20	0
Telangana	19	3	732	51	3	2	77	14	0	3	0	0	0	0	0	0	0	301	1
Tripura	5	0	8	0	0	0	3	4	0	0	0	0	0	0	0	0	0	0	0
Uttar Prac	47	73	2351	199	343	191	343	45	0	59	9	0	75	614	0	0	0	1931	1
Uttarakha	3	41	46	16	12	22	13	0	0	0	0	0	0	0	0	5	0	13	0
West Ben	28	9	68	25	2	3	39	1	0	2	0	0	0	0	0	0	0	158	1
A & N Island	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Chandigar	0	0	19	3	0	7	0	0	0	0	0	0	0	0	0	0	0	1	0
D&N Have	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Daman &	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Delhi UT	11	4	36	11	4	1	35	0	0	0	0	0	35	2	0	0	0	50	0
Lakshadow	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0

Fig. 3. Raw Data (7)

In Fig. 4 we can see the features of data, a feature is something that's used to determine a result, and a column is a physical structure that stores the value of a feature or a result. In Fig. 12, using shape function the data is displayed in the format of rows and columns; here we have 36 rows and 13 columns; also, we check whether there are any null values present in the data sets shown in Fig. 12. Matplot library is used to create a function that cross plots feature so that we can see when they are correlated. Data is then inspected in order to eliminate any additional columns or rows to with no values that we no longer required. The duplicates including the same values are removed the same way. This is done to arrange our data since visual inspection may be error-prone and cannot deal with the critical issue of correlated columns. Thus, pandas help in understanding such null values and therefore identifying it in our data as we can see in Fig. 6, Is Null method will check each value on the data frames for null values. Similarly, Matplot library is used to create a function plots features so that we can see when the data is correlated: the color in yellow denotes the very positive correlation as seen in Fig. 11 and other color denotes that the data is not well correlated. In Fig. 11 we can see that column names on the horizontal and vertical axes is a matrix showing which column contains the data that are correlated with values.

```
In [3]:  os.getcwd()
         os.chdir ('C:/Users/Puneet/CRIME RECORD')
         os.getcwd()

Out[3]:  'C:\\Users\\Puneet\\CRIME RECORD'

In [20]: Cyber_data = pd.read_csv('Cyber new.csv') # read dataset
         Cyber_data.head()

Out[20]:
```

	State_UT	Personal_ Revenge	Anger	Fraud	Extortion	Causing_ Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Developing own business
0	Andhara Pradesh	34	26	733	45	7	0	92	12	1	1	1	0	
1	Arunachal Pradesh	0	0	2	0	0	0	0	0	0	0	0	0	
2	Assam	239	46	389	153	234	0	113	9	4	3	0	0	
3	Bihar	5	8	351	2	0	0	8	0	0	0	0	0	
4	Chhattisga	0	1	23	2	25	0	21	4	0	1	0	0	

Fig. 4. Selecting the data

As we can see in Fig. (4) and (5), data is fetched from the file path and utilized for the further data cleaning and correlating.

```
In [3]:  os.getcwd()
         os.chdir ('C:/Users/Puneet/CRIME RECORD')
         os.getcwd()

Out[3]:  'C:\\Users\\Puneet\\CRIME RECORD'

In [20]: Cyber_data = pd.read_csv('Cyber new.csv') # read dataset
         Cyber_data.head()

Out[20]:
```

Causing_ Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Developing own business	Spreading	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Risk
7	0	92	12	1	1	1	0	2	14	2	0	1	236	0
0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
234	0	113	9	4	3	0	0	0	0	0	0	0	832	1
0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
25	0	21	4	0	1	0	0	1	0	0	0	0	61	0

Fig. 5. Showing the data

2. Molding the Data

After cleaning the data of any extra columns or null values, we proceed to molding the data by inspecting if there are any issues. Algorithms are largely mathematical models which work best with numeric quantities and once the data molding is done, we can use this data for further training the algorithm as seen in Fig. 6 count, mean, std, etc. is calculated so that the data is molded accurately. Therefore, in machine learning, a lot of data manipulation is done for trial and error and predicting the best of the accuracy. When the data is manipulated it's very easy to change the meaning of the data what also helps in understanding if data has gone wrong anywhere. The entire model is created in *Jupyter Notebook*, therefore keeping track of all the changes and updates have been done automatically [Perkel J. et al, 2018: 145–147]. We also have the interactivity of the python interpreter using which we can make our data simpler for the prediction, as seen in Fig.6 and 7.

```
In [6]: Cyber_data.describe()
Out [6]:
```

	Personal_ Revenge	Anger	Fraud	Extortion	Causing_ Disrepute	Prank	Sexual Exploita-tion	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale pur-chase illegal drugs
count	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000
mean	22.055556	12.805556	418.083333	29.166667	33.666667	8.222222	56.388889	6.055556	1.222222	6.055556	0.583333	0.166667
std	44.940560	25.484807	1007.891615	54.345193	72.200119	31.825516	129.880886	12.094732	3.330475	13.259917	1.645340	0.560612
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	6.750000	0.000000	0.000000	0.000000	2.750000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	5.000000	3.500000	41.000000	7.500000	3.500000	0.000000	15.500000	0.500000	0.000000	0.000000	0.000000	0.000000
75%	21.000000	10.250000	392.000000	26.500000	20.500000	3.000000	52.500000	4.000000	1.000000	3.000000	0.000000	0.000000
max	239.000000	129.000000	5441.000000	224.000000	343.000000	191.000000	724.000000	52.000000	16.000000	59.000000	9.000000	2.000000

```
In [7]: cyber_data.isnull().any()
Out [7]:  State_UT                          False
          Personal_Revenge                  False
          Anger                             False
          Fraud                             False
          Extortion                         False
          Causing_Disrepute                 False
          Prank                             False
          Sexual Exploitation               False
          Political Motives                 False
          Terrorist Activities              False
          Inciting Hate against Country     False
          Disrupt Public Service            False
          Sale purchase illegal drugs       False
          Developing own business           False
          Spreading                         False
          Psycho or Pervert                 False
          Steal Information                 False
          Abetment to Suicide               False
          Others                            False
          Risk                              False
```

Fig. 6. Null values are checked

```
In [6]: Cyber_data.describe()
Out [6]:
```

Prank	Sexual Exploi-tation	Political Motives	Terrorist Activities	Incit-ing Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Develop-ing own business	Spreading	Psycho or Pervert	Steal Infor-mation	Abetment to Suicide	Others	Risk
36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000	36.000000
8.222222	56.388889	6.055556	1.222222	6.055556	0.583333	0.166667	5.500000	18.638889	0.111111	0.444444	0.055556	137.666667	0.305556
31.825516	129.880886	12.094732	3.330475	13.259917	1.645340	0.560612	14.484474	102.147275	0.398410	1.229273	0.232311	349.023454	0.467177
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	2.750000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
0.000000	15.500000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	14.000000	0.000000
3.000000	52.500000	4.000000	1.000000	3.000000	0.000000	0.000000	2.250000	1.000000	0.000000	0.000000	0.000000	101.250000	1.000000
191.000000	724.000000	52.000000	16.000000	59.000000	9.000000	2.000000	75.000000	614.000000	2.000000	5.000000	1.000000	1931.000000	1.000000

```
In [7]: cyber_data.isnull().any()
Out [7]: State_UT                          False
         Personal_Revenge                 False
         Anger                            False
         Fraud                            False
         Extortion                        False
         Causing_Disrepute                False
         Prank                            False
         Sexual Exploitation              False
         Political Motives                False
         Terrorist Activities             False
         Inciting Hate against Country    False
         Disrupt Public Service           False
         Sale purchase illegal drugs      False
         Developing own business          False
         Spreading                        False
         Psycho or Pervert                False
         Steal Information                False
         Abetment to Suicide              False
         Others                           False
         Risk                             False
```

Fig. 7. Null Values shown in risk column

3. Testing Model's Accuracy

In this section we will discuss the role of the Machine Learning algorithm. An algorithm can be defined as an engine that drives the entire process. For our prediction, we will use data containing examples of the results and try to predict the future using the scikit learn and the algorithm's logic the data is analyzed. This analysis evaluates the data concerning a mathematical model and logic associated the algorithm, and the algorithm then uses the results of this analysis to adjust internal parameters to produce a model that has been trained to best fit the features and give the best results. The best result is defined by evaluating a function specific to a particular algorithm. Therefore, the fit parameters are stored and hence the model is now trained. Further, we use this model to predict on the real data. We use the Scikit learn package in python to predict on the real data. The parameters of the trained model along with the python code is used to predict whether the state is in threat of cyber-attack or no. Selecting an appropriate algorithm from scikit learning was the toughest part which we faced while researching on this paper.

Prediction means supervised learning so eliminating all other algorithms was my main goal, furthermore, prediction can be divided into two more categories regression and classification, where regression means a continuous set of values. Predicting binary outcome whether the threat is there or not; we further eliminated all the algorithms that do not support classification in general and especially binary classification. Naïve Bayes, Logistic Regression and Decision Tree are algorithms which support classic machine learning algorithms and also provide excellent help in understanding more complex algorithms.

```
In [8]:   plt.figure(figsize=(15,10))
          plt.tight_layout()
          seabornInstance.distplot(cyber_data['Risk'])

Out [8]:  <matplotlib.axes._subplots.AxesSubplot at 0x1c303d03808>
```

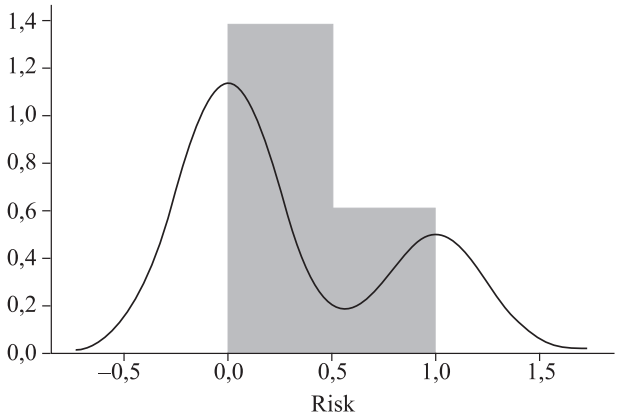

Fig. 8. Graph Denoting the Risk

```
In [13]: df.corr()
Out [13]:
```

	Online Banking Frauds	Cyber Blackmailing/ Threatening Sec 506, 503, 384	Fake News on Social Media Sec 505	Cyber Terrorism sec 66F	Tampering Computer Source	Identity Theft sec 66C	Computer related offence sec 66	Ransom-ware	Offences other than Ransom-ware	Cyber Stalk-ing/Bullying of Women/ Children Sec 354D IPC
Online Banking Frauds	1.000000	0.297261	0.210440	–0.081656	0.297991	–0.029883	0.278153	0.276576	0.322202	0.799312
Cyber Blackmailing/ Threatening Sec 506, 503, 384	0.297261	1.000000	0.472718	0.368497	0.154649	–0.049758	0.156595	0.128745	0.154715	0.300828
Fake News on Social Media Sec 505	0.210440	0.472718	1.000000	0.589458	0.252916	–0.041483	0.231064	0.235354	0.270136	0.205406
Cyber Terrorism sec 66F	-0.081656	0.368497	0.589458	1.000000	0.028105	–0.042616	0.061458	0.034039	0.025907	–0.047158
Tampering Computer Source	0.297991	0.154649	0.252916	0.028105	1.000000	0.065687	0.991130	0.992945	0.937691	0.028343
Identity Theft sec 66C	-0.029883	-0.049758	-0.041483	-0.042616	0.065687	1.000000	0.014784	0.003531	0.074809	0.000353
Computer related of-fence sec 66	0.278153	0.156595	0.231064	0.061458	0.991130	0.014784	1.000000	0.998242	0.955151	0.010488
Ransomware	0.276576	0.128745	0.235354	0.034039	0.992945	0.003531	0.998242	1.000000	0.949917	0.006252
Offences other than Ransomware	0.322202	0.154715	0.270136	0.025907	0.937691	0.074809	0.955151	0.949917	1.000000	0.060779
Cyber Stalking/Bully-ing of Women/Chil-dren Sec 354D IPC	0.799312	0.300828	0.205406	–0.047158	0.028343	0.000353	0.010488	0.006252	0.060779	1.000000

Fig. 9. Correlation Performed

Logistic regression algorithm has a dubious name since in statistics a regression often implies continuous values but logistic regression returns a binary result. The algorithm measures the relationship of each feature and compares them based on their impact on the result. The result and value are then mapped against a curve seen in Fig. (8), which is equivalent to threat or no threat.

```
def     plot_corr(df, size=11):
        """
        Function plots a graphical correlation matrix for each pair of columns
        in the dataframe

        Input:
          df: pandas DataFrame
          size: vertical and horizontal size of the plot

        Displays:
          matrix of correlation between columns. Blue-cyan-yellow-red-darkred
          => less to more correlated
                                          0 ----------------------> 1
                                          Expect a yellow line running from top
                                          left to bottom right
        """

        corr = df.corr()
        fig,ax = plt.subplots(figsize(size, size))
        ax.matshow(corr)
        plt.xticks(range(len(corr.columns)), corr.c olumns)
        plt.yticks(range(len(corr.columns)), corr.c olumns)
        plt.setp(ax.get_xticklabels(), rotation=90, horizontalalignment='right')
```

Fig. 10. Giving the values for correlation

```
In [10]: plot_corr(cyber_data)
```

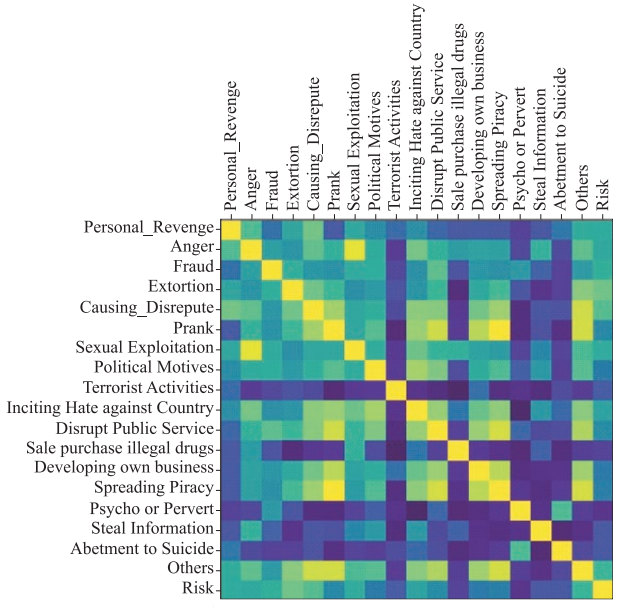

Fig. 11. **Correlation graph**

4. Training the Model

Splitting the cyber data into two sets one for training the model and the other for testing the model, about 70% of the data we have put in the training set and 30% of data in the testing set, after this, we have trained the algorithm with the training data and held the test data aside for evaluation. This training process produces a training model based on the logic in the algorithm and the values of the features in the training data. Care has taken not to use all the data to train since data drives training of the model. The library which handles machine learning, training and evaluation tasks in Python is Scikit learning, it provides a set of simple and efficient tools that can manage many of the tests in machine learning.

Scikit supports machine learning and it is built on Python libraries such as NumPy, SciPy and Matplotlib and supports these and panda's data frames. It is generally a toolset that makes training and evaluation tasks simple; these tasks involve splitting the data into training and test sets, preprocessing data before training, selecting the most important data features, creating train model, tuning the model for better performance.

```
In [11]:  x = cyber_data.drop(['Risk', 'State_UT'], axis=1) # Independent
          Variables
          y = cyber_data[['Risk']]  # Dependent Variable

In [12]:  X.shape

Out [12]: (36, 18)

In [24]:  # Splitting the data into train and test
          X_train, X_test, y_train, y_test = train_test_split(X, y, test_
          size=0.4, random_state=1)

In [14]:  # Building Linear Regression
          reg = LogisticRegression()
          reg.fit(X_train, _train)

          M: \Users\Puneet\anaconda3\lib\site-packages\sklearn\utils\
          validation.py:760: DataConversionWarning: A column-vector y was
          passe d when a 1d array was expected. Please change the shape of y
          to (n_samples, ), for example using ravel().
          y = column_or_1d(y, warn=True)
          M: \Users\ Puneet\anaconda3\lib\site-packages\sklearn\linear_
          model\_logistic.py:940: ConvergenceWarning: 1bfgs failed to
          converge (status=1):
          STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

Fig. 12. Applying regression algorithm

5. Checking the Accuracy

Explanation for code:

Since we know that through our research aimed to predict whether a particular State/UT is at a higher risk of cybercrime when using such variables as Personal revenge, Anger, Fraud, etc. In order to predict this relationship, we have used a statistical technique Logistic Regression. Before we move to modeling, we have to check if there is any correlation between the independent variables, in other words, we have to check if there is a relationship between the independent variables (example Personal revenge, Anger, Fraud etc.). In the correlation plot, we should ignore the diagonal block as the diagonal block in yellow represents the correlation with itself (i.e., Personal revenge and Personal revenge) in which we are not interested. Yellow color represents high correlation, light green color represents moderate correlation, dark green color represents low correlation, and complete dark color represents no correlation. So, from the plot we can say that there is a high correlation between Sexual exploitation and Anger, spreading piracy and prank etc. as if we see the block of these variables in the plot, they are yellow in color. There is a moderate correlation between Spreading piracy and Causing disrepute, Prank and Inciting hate against country etc. as if we see the block of these variables in the plot, they are yellow in color. Similarly, we can say that the variables with darker blocks have less of no correlation. We have divided the data into X and Y where X is the independent variable and y denotes the dependent variable. So are independent variables being Personal revenge, Anger, Fraud, etc. and our dependent variable is risk.

Further checked the shape of X (just a sense check), then divided the variable into train and test, (we will use the X_train and y_train to train the logistic regression model and then test the model using X_test and y_test). Now we have used the function to build a logistic regression model using the data X_train and y_train. We are using this logistic regression when our dependent variable has dichotomous type, i.e., True/False, Absent/Present etc. Now having built a model, we have predicted the expected values of y using X_test. After predicting the expected values for y we will now check the accuracy of the model. The accuracy of the model depends on the number of cases we have predicted correctly, i.e., the number of times we have predicted that the State/UT is at risk. The state was actually at risk and the number of times we have predicted that the State/UT is not at risk and the state was not at risk. As seen in Fig. 11, 12, and 13, we can see that how the model behaves in predicting the accuracy of the threat in the states.

```
Out [14]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_
          intercept=True,
                    intercept_scaling=1, 11_ratio=None, max_iter=100,
                    multi_class='auto', n_jobs=None, penalty='12',
                    random_state=None, solver='1bfgs', tol=0.0001, verbose=0,
                    warm_start=False)

In  [15]: # Predicting  the cases
          y_pred = reg.predict(X_test)
          y_pred

Out [15]: array([0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0], dtype=int64)

In [18]:  Metrics.accuracy_score(y_test, y_pred)

Out [18]: 0.6666666666666666

In [19]:  cyber_data.head(5)

Out [19]:
```

Causing_ Disrepute	Prank	Sexual Exploitation	Political Motives	Terrorist Activities	Inciting Hate against Country	Disrupt Public Service	Sale purchase illegal drugs	Developing own business	Spreading	Psycho or Pervert	Steal Information	Abetment to Suicide	Others	Risk
7	0	92	12	1	1	1	0	2	14	2	0	1	236	0
0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
234	0	113	9	4	3	0	0	0	0	0	0	0	832	1
0	0	8	0	0	0	0	0	0	0	0	0	0	0	0
25	0	21	4	0	1	0	0	1	0	0	0	0	61	0

Fig. 13. Model predicting the accuracy

Conclusion

Through the study above, we can conclude that by trial and error of various algorithms, we could draw some crucial points with the help of the Logistic Regression Algorithm. This research would surely help the law enforcement agencies understand the root cause of the crime as if there was any political movement, natural crisis, or else massive dropouts in the particular state which led to a person committing the crime. As we cannot rely on this model completely in sentencing the accused, his/her parenting, upbringing, society, and teachings should also be gone through to understand the reason behind committing the crime, as we all know the law enforcement agencies or government can only bestow law upon us. Still, the root cause of this crime should be found out and eradicated. The bigger question is, how will technology shape the judicial function, and to what extent [Sourdin T., 2018. Judge v. Robot: Artificial Intelligence and Judicial Decision-Making. UNSWLJ, 41, pp: 1114], but it will surely benefit the judiciary system in

some or other way. The various sectors can benefit from this new technology provided that it is not used for somebody's harm for it to behave in unpredicted and potentially harmful ways [Cath C., 2018: 1–8]. Thus, the proper judicial monitoring of data fed can enjoy this model's beauty.

References

Alarie B., Niblett A. & Yoon A. (2018) How artificial intelligence will affect the practice of law. *University of Toronto Law Journal*, vol. 68, supplement 1, pp. 106–124.

Caldwell M. et al (2020) AI-enabled future crime. *Crime Science*, no 1, pp. 1–13.

Cath C. (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Phil.Trans. Royal. Society*, issue 2133, pp. 1–8.

Chen H. et al (2003) COPLINK Connect: information and knowledge management for law enforcement. *Decision support systems*, no 3, pp. 271–285.

Cui Y. (2020) Building AI-assisted rule of law for the future, seeking advantages and avoiding disadvantages to make AI better benefit mankind. In: *Artificial Intelligence and Judicial Modernization*. Singapore: Springer, pp. 187–191.

Dhillon K. (2011) The police and the criminal justice system in India. *The Police, State, and Society: Perspectives from India and France*. Pearson, pp. 27–59.

Dmitrik N. (2020) Digital State, Digital Citizen: Making Fair and Effective Rules for a Digital World. *Legal Issues in the Digital Age*, no 1, pp. 54–78.

Furtado V. et al (2010) Collective intelligence in law enforcement–The WikiCrimes system. *Information Sciences*, no 1, pp. 4–17.

Hauck R. et al (2002) Using Coplink to analyze criminal-justice data. *Computer*, no 3, pp. 30–37.

Hayward K., Maas M. (2020) Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, pp. 1–25.

Isaac W. (2017) Hope, hype, and fear: the promise and potential pitfalls of artificial intelligence in criminal justice. *Ohio St. J. Crim.* L., vol. 15, p. 543.

King T., Aggarwal N., Taddeo M., Floridi L. (2020) Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and engineering ethics*, no 1, pp. 89–120.

Marda V. (2018) Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences*, vol. 376, pp. 1–19.

McCarthy J. et al (2006) A proposal for the Dartmouth summer research project on artificial intelligence. *AI magazine*, no 4, pp. 12–14.

McGuire M., Holt T. (eds.) (2017) The Routledge Handbook of Technology, Crime and Justice. L.: Taylor & Francis, pp. 1–722.

Nath S. (2006) Crime pattern detection using data mining. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent, pp. 41–44.

Perkel J. (2018) Why Jupyter is data scientists' computational notebook of choice. *Nature*, vol. 563, pp. 145–147.

Raschka S. et al (2020) Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, no 4, p. 193.

Sourdin T. (2018) Judge v. Robot: Artificial Intelligence and Judicial Decision-Making. UNSW *Law Journal*, vol. 41, pp. 11–14.

Završnik A. (2020) Criminal justice, artificial intelligence systems, and human rights. *ERA Forum Springer*, no 4, pp. 567–583.

Expression Through Socialising Media in India: Why Fixing the Existing Legal Dilemmas Is Critical?

Meera Mathew

Assistant Professor, Symbiosis Law School, Deemed University, PhD. Address: NOIDA Sec-62, Block-A, 47 & 48, NOIDA (PIN-201301), Uttar Pradesh, India. E-mail: meera@symlaw.edu.in

Abstract

The emergence of the social media and its virtual communication space has enabled people at large to interact and communicate from the conventional mode of one-to-one to many-to-many. It exploded onto the technology in the last decades for commercial and entertainment purpose and rapidly it had become very much prevalent globally. Initiated as a friend-finder it went on to the extend encompassing every features of media where the users had a dominant role. When mass media and digital media was through certain modes, social media not only changed the mode but the creators and audience. From passive news listeners, it became active creators and sharers of contents in the form of information. With the enablement of technology, anybody with an internet access and own opinion can be part of social media. Under the guise of user-generated content, be it in sharing of news or opinion or images or videos and now even the live video promoting political, social, cultural aspects, social media do not hold any accountability because only users are producing contents. Also, being an intermediary, it is free from any liability for the user generated data under Indian Information Technology Act, 2008 and the existing global consensus under safe harbour doctrine. The law in this area is still relatively unsettled. The misuse of social media got reported with various incidents of such as impersonation, anonymity, profile account hacking, privacy threats, sexual or aggressive solicitation, cyber-bullying, and many such related serious issues. However, in all these matters, social media was provided with a benefit for its passive involvement of choosing the users or the contents posted. The liability was always on the content producers. It is certain degree of due diligence social media platform needs to observe that too very minimal! This paper endeavours to question the existing privilege available to social media at par with conventional media and also highlights the social-legal dilemma it put forth with unprecedented use of data. It further dwells upon the legal impediments in challenges that social media pose for the lack of legislation- especially for data protection and user profile anonymity detection. It thus attempts to find out whether social media is to be equated like media or should it be viewed as mere platform for people to express. If it is just a platform to express, whether the current Indian legal framework is sufficient enough, to deal with the ramifications arising out of social media especially when most of them are social media companies incorporated and registered under foreign jurisdictions.

Keywords
expression, social media, legal issues, Indian legal system, Information technology.

For citation: Meera M. (2020) Expression through "Socialising" Media in India: Why Fixing the Existing Legal Dilemmas is Critical? // Legal Issues in the Digital Age, no 3, pp. 97–124.

DOI: 10.17323/2713-2749.2020.3.97.124

Introduction

The internet service websites, blog pages, mobile technologies, social media and networking sites web have entirely altered previously prevailed communication model. The internet, digitalization and social media are transforming news from its traditional practice from its original notions of press and media. The degree at which exchange of communication existed had been multi-folded with sudden increase in information collected and circulated. Today every news-media has its social media webpage including *Twitter* handles or *Facebook* pages thus stories are searched on internet service providers to know if any user has uploaded anything that became 'viral'. Moreover, it has become a necessity for mainstream print media to have their websites, live videos, journalists' blogs, invited newsrooms debates where invitation is extended to community participation [Knutson A., 2009: 437–474].

The bloggers consider themselves as journalists and break *scoops* and stories. With notable shift to mobile news access news has now become omnipresent-available on every platform at any time. Regardless of their professions, resources or training today, *netizens* are disseminating news to the public themselves. Personalized and participatory stories having maximum views or shares are now converted as news.

Further the technological changes and ongoing perception of news", its practices of reporting are greatly influencing at its quantity, quality and nature of reporting, whether online or in print. While print media still have a noteworthy readership, the digital media and new media sites have clearly had a fading impact on the print medium. Social media has divulged in innovative ways to interconnect and collaborate the population through technology. Smart-phones and tablets have redefined customer computing and provide instantaneous access to information from any locality. For instance, observe the development and muti-fold uses of a smart phone [McPeak A., 2015: 235–292]. On it, one can listen to music, phone people, text, watch videos, send and receive emails, surf the internet, play games, watch videos, store pictures and plan the travel with calendar and many other things. Instead of carrying disc-man, walk-man, laptop, diary, camera, telephone today all in one is possible. This is the convergence where all contents and in-