

Crime Analysis for Bangladesh

Md. Imran Parvez (200101024), Md. Rasa Raj (200101054)

Course Code: CSE 4252, Course Title: Data Warehousing and Data Mining Sessional
Department of Computer Science and Engineering, Bangladesh Army University of Science and
Technology (BAUST)

Abstract

Bangladesh's crime analysis infrastructure faces challenges such as data fragmentation, limited resources, and outdated technology. This abstract highlights the need to centralize data, modernize technology, and introduce predictive analysis. While these steps offer a path to reducing crime rates, enhancing public safety, and optimizing resource allocation.

1. Introduction

Crime analysis is the methodical examination of crime data to find trends, patterns, and important information that re effectively and efficiently allocating resources as well as in preventing, solving, and reducing crime. Researchers, decision-makers, and the general public can all benefit from a greater knowledge of the state of crime and how it affects society through the use of crime analysis. The area of crime analysis in Bangladesh is still in its infancy and is beset by numerous obstacles. A few of the main challenges are the dearth of accurate and thorough crime data, the shortage of qualified and experienced crime analysts, and the absence of cutting-edge forensic equipment and methods. The purpose of this paper is to give a summary of the current. Crime analysis is the process of studying crime data and patterns to understand the nature, causes, and trends of criminal activity.

2. Problem Definition

Crime analysis has a big impact on public safety and efficient law enforcement, it's a major concern in Bangladesh. The incapacity of the nation to comprehend, predict, and counteract criminal activity is hampered by the inefficiency and insufficiency of the current crime analytic infrastructure. The absence of predictive analysis, antiquated technology, scarce resources, dispersed and inadequately centralized crime data, and inadequate training for law enforcement officers are some of the major issues. Because of these problems, crime control is done in a reactive manner, which raises crime rates and lowers public safety.

3. Dataset Description

Crime analysis datasets typically contain information related to various criminal incidents and activities. These datasets are essential for law enforcement agencies, researchers, and analysts to gain insights into crime patterns, trends, and hotspots (Incident details, Geographic information, Offender and victim characteristics, Arrest and clarence data, yearly information, Crime location and types).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Unit Name/Year	Sexual	Battery	Stolen	Sexual In Bat	Women & Children	Police	Assaults	Drugs	Other	Causes	Acc	Intox	Sexual	Intox	Intox	Intox	Intox
2	2009	2009	47	228	340	342	3	1216	126	125	125	1261	1218	128	82	12025	144	12020
3	2010	2010	55	359	54	11	7	405	37	38	121	114	1055	13	0	866	99	4061
4	2011	2011	1	9	24	21	9	213	12	8	82	11	555	39	1	700	11	1191
5	2012	2012	4	20	21	0	10	117	0	12	53	126	579	3	4	502	248	1717
6	2013	2013	8	12	19	11	9	112	6	8	34	88	557	17	0	155	117	1119
7	2014	2014	12	21	11	14	1	294	34	12	31	254	884	34	0	154	38	1444
8	2015	2015	161	129	1251	382	7	4332	171	71	643	1417	1864	895	90	4029	991	14174
9	2016	2016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	2017	2017	154	100	684	181	20	2165	111	87	424	998	12465	105	20	4700	612	14111
11	2018	2018	85	40	240	75	17	848	41	19	389	534	524	38	4	905	151	1448
12	2019	2019	39	125	111	120	15	1275	136	25	103	941	854	133	96	2070	1238	13331
13	2020	2020	58	37	387	111	9	983	46	15	234	324	692	33	1	104	33	9000
14	2021	2021	54	128	157	196	29	4085	128	20	442	1442	2854	134	20	3024	2028	14280
15	2022	2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	2023	2023	0	1	11	1	1	5	1	1	2	104	142	3	0	174	201	625
17	2024	2024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	2025	2025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	2026	2026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	2027	2027	44	294	329	534	9	1538	118	142	188	1550	674	104	17	1845	68	13540
21	2028	2028	18	48	81	43	15	408	30	19	121	155	1286	89	4	1542	28	4230
22	2029	2029	1	12	21	17	9	171	11	19	57	91	478	34	1	485	9	1471
23	2030	2030	1	14	38	3	3	348	16	11	55	102	665	8	1	255	293	1449
24	2031	2031	18	8	71	11	9	147	9	12	46	88	584	11	0	186	88	1308
25	2032	2032	0	26	17	11	1	158	16	11	119	857	8	0	179	32	1588	
26	2033	2033	180	174	1189	381	17	4838	144	117	681	1711	1880	388	21	7407	1076	18308
27	2034	2034	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	2035	2035	154	128	642	141	19	4387	128	128	498	1110	12846	102	20	4522	700	18211
29	2036	2036	80	40	241	58	18	867	40	27	171	481	1470	39	0	1448	188	1430
30	2037	2037	58	128	424	128	1	1786	128	10	105	901	1562	152	68	2338	1047	18821
31	2038	2038	11	40	184	41	8	1008	30	20	134	111	811	41	8	302	38	1443
32	2039	2039	41	148	188	176	20	4716	121	58	881	1888	25174	134	11	2707	1274	14131

Unit Name/Year

2009201020112012201320142015201620172018201920202021202220232024202520262027202820292030203120322033203420352036203720382039

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
33	2010	2010	55	359	54	11	7	405	37	38	121	114	1055	13	0	866	99	4061
34	2011	2011	1	9	24	21	9	213	12	8	82	11	555	39	1	700	11	1191
35	2012	2012	4	20	21	0	10	117	0	12	53	126	579	3	4	502	248	1717
36	2013	2013	8	12	19	11	9	112	6	8	34	88	557	17	0	155	117	1119
37	2014	2014	12	21	11	14	1	294	34	12	31	254	884	34	0	154	38	1444
38	2015	2015	161	129	1251	382	7	4332	171	71	643	1417	1864	895	90	4029	991	14174
39	2016	2016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	2017	2017	154	100	684	181	20	2165	111	87	424	998	12465	105	20	4700	612	14111
41	2018	2018	85	40	240	75	17	848	41	19	389	534	524	38	4	905	151	1448
42	2019	2019	39	125	111	120	15	1275	136	25	103	941	854	133	96	2070	1238	13331
43	2020	2020	58	37	387	111	9	983	46	15	234	324	692	33	1	104	33	9000
44	2021	2021	54	128	157	196	29	4085	128	20	442	1442	2854	134	20	3024	2028	14280
45	2022	2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	2023	2023	0	1	11	1	1	5	1	1	2	104	142	3	0	174	201	625
47	2024	2024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	2025	2025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	2026	2026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	2027	2027	44	294	329	534	9	1538	118	142	188	1550	674	104	17	1845	68	13540
51	2028	2028	18	48	81	43	15	408	30	19	121	155	1286	89	4	1542	28	4230
52	2029	2029	1	12	21	17	9	171	11	19	57	91	478	34	1	485	9	1471
53	2030	2030	1	14	38	3	3	348	16	11	55	102	665	8	1	255	293	1449
54	2031	2031	18	8	71	11	9	147	9	12	46	88	584	11	0	186	88	1308
55	2032	2032	0	26	17	11	1	158	16	11	119	857	8	0	179	32	1588	
56	2033	2033	180	174	1189	381	17	4838	144	117	681	1711	1880	388	21	7407	1076	18308
57	2034	2034	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	2035	2035	154	128	642	141	19	4387	128	128	498	1110	12846	102	20	4522	700	18211
59	2036	2036	80	40	241	58	18	867	40	27	171	481	1470	39	0	1448	188	1430
60	2037	2037	58	128	424	128	1	1786	128	10	105	901	1562	152	68	2338	1047	18821
61	2038	2038	11	40	184	41	8	1008	30	20	134	111	811	41	8	302	38	1443
62	2039	2039	41	148	188	176	20	4716	121	58	881	1888	25174	134	11	2707	1274	14131

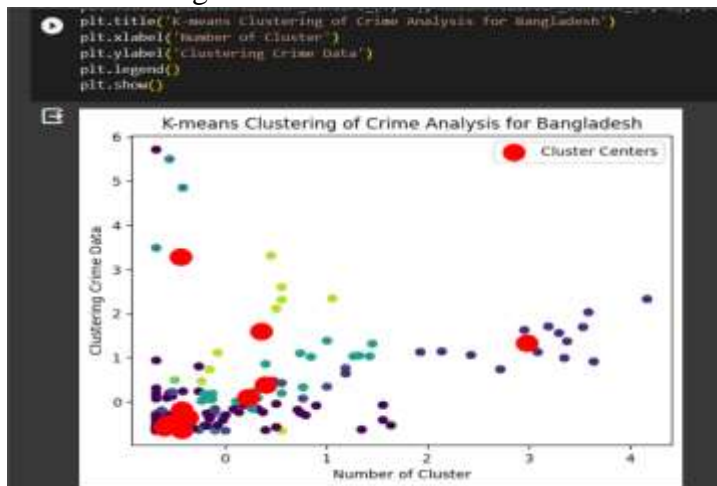
Unit Name/Year

2009201020112012201320142015201620172018201920202021202220232024202520262027202820292030203120322033203420352036203720382039

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
63	2010	2010	55	359	54	11	7	405	37	38	121	114	1055	13	0	866	99	4061
64	2011	2011	1	9	24	21	9	213	12	8	82	11	555	39	1	700	11	1191
65	2012	2012	4	20	21	0	10	117	0	12	53	126	579	3	4	502	248	1717
66	2013	2013	8	12	19	11	9	112	6	8	34	88	557	17	0	155	117	1119
67	2014	2014	12	21	11	14	1	294	34	12	31	254	884	34	0	154	38	1444
68	2015	2015	161	129	1251	382	7	4332	171	71	643	1417	1864	895	90	4029	991	14174
69	2016	2016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
70	2017	2017	154	100	684	181	20	2165	111	87	424	998	12465	105	20	4700	612	14111
71	2018	2018	85	40	240	75	17	848	41	19	389	534	524	38	4	905	151	1448
72	2019	2019	39	125	111	120	15	1275	136	25	103	941	854	133	96	2070	1238	13331
73	2020	2020	58	37	387	111	9	983	46	15	234	324	692	33	1	104	33	9000
74	2021	2021	54	128	157	196	29	4085	128	20	442	1442	2854	134	20	3024	2028	14280
75	2022	2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76	2023	2023	0	1	11	1	1	5	1	1	2	104	142	3	0	174	201	625
77	2024	2024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
78	2025	2025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
79	2026	2026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	2027	2027	44	294	329	534	9	1538	118	142	188	1550	674	104	17	1845	68	13540
81	2028	2028	18	48	81	43	15	408	30	19	121	155	1286	89	4	1542	28	4230
82	2029	2029	1	12	21	17	9	171	11									

4. Code Implementation

A. Data Clustering.



K-means clustering data of crime analysis for Bangladesh (label and without label)

B. Machine Learning Algorithm

```

model = LinearRegression()
training_score = 0.693082099677791
predictions = [3684.749878134, 506.72942451, 2224.11775755, 672.5417471, 381.1594734,
233.53826645, 46.38548112, 110.18354882, 4538.72961981, 346.357719051,
1881.87061945, 1511.60062186, 850.81511882, 368.50984257, 1181.67771251,
5786.68786134, 1171.89958134, 252.19821186, 218.72288706, 3227.45257812,
387.11898631, 1189.62740654, 5145.54418884, 2148.80188711, 186.4689962,
926.46498861, 128.13888941, 155.84882083, 1314.98892928, 165.46988771,
1358.46271189, 342.44322837, 2816.2586976, 1517.82129618, 1419.88976197,
882.85882191, 3877.90725251, 546.75834388, 1476.12875939, 104.23275688,
346.15964441, 3854.17442556, 6343.8565672 ]
model = DecisionTreeRegressor()
training_score = 1.8
predictions = [1.814e+01 4.154e+02 2.730e+01 2.488e+02 1.466e+02 5.488e+00 1.070e+02
1.228e+02 5.761e+05 9.508e+01 1.120e+02 1.770e+03 2.488e+02 1.650e+02
1.508e+02 1.400e+02 4.150e+02 1.974e+03 2.350e+02 4.350e+02 6.570e+02
6.150e+02 2.825e+02 2.821e+02 2.825e+05 1.116e+03 1.100e+02 1.800e+00
1.238e+02 1.974e+03 4.350e+02 1.810e+02 1.470e+02 1.453e+03 4.350e+02
1.238e+02 5.788e+03 8.080e+01 1.011e+03 1.530e+02 1.810e+02 1.300e+01
5.842e+03]
model = RandomForestRegressor()
training_score = 0.9466308257812412
predictions = [1882.51, 666.87, 2551.43, 586.52, 225.4, 167.2, 456.44, 152.05, 4812.89,
148.39, 3160.53, 2224.41, 469.87, 536.72, 895.68, 1652.55, 1341.3, 836.89,
241.41, 1775.98, 1139.36, 1082.13, 1478.61, 1845.96, 1658.49, 3892.63, 192.59,
282.63, 1585.61, 823.2, 1251.78, 112.32, 973.41, 1229.4, 1595.04, 1043.36,
4415.81, 931.93, 1670.85, 759.66, 756.72, 857.84, 8936.28]

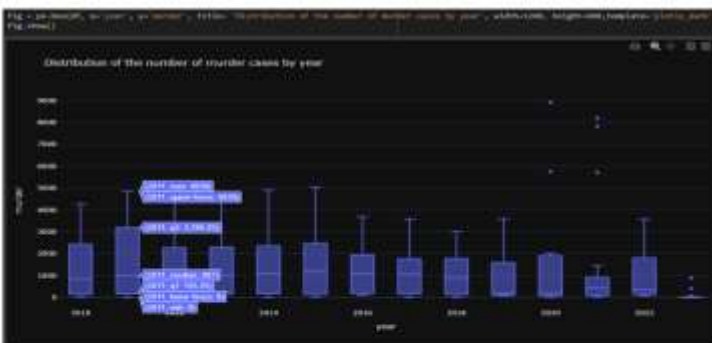
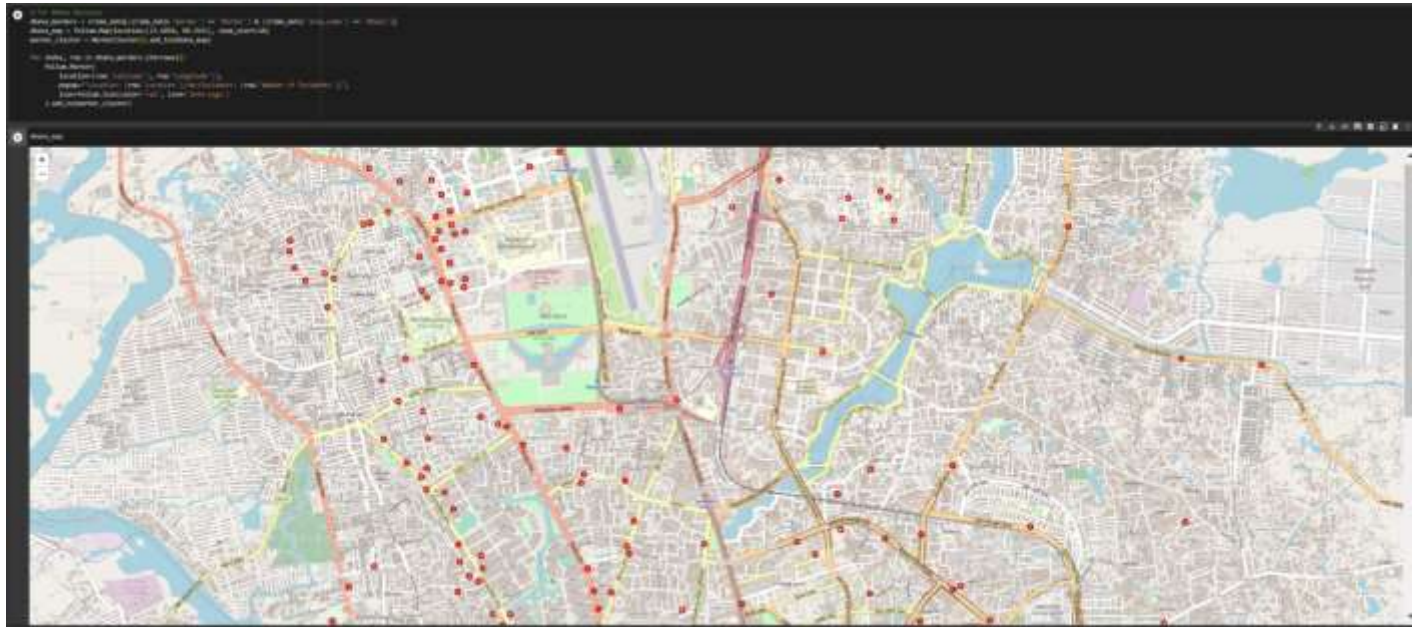
```

We are using 3 machine learning algorithms for this project and show the result.

1. Linear Regression
2. Decision Tree
3. Random Forest

5. Result and Analysis

Most of the places of murder in Dhaka area are identified here by maps



Here is the value of a box plot for murder, here present at (maximum value, minimum value, median, etc)

Values of all recovery cases are shown through pie charts



Clustering has been important in the work we have done so far, because we have done more work inside, such as preparing something hard work. Then we tried to analyze each crime separately and then make predictions using machine learning algorithms after taking out the value and graphing it, the almost assurance of winning is like 96%

