JCBI
Journal of Computing & Biomedical Informatics Volume 08 Issue 02
ISSN: 2710 - 1606 2025

Research Article
https://doi.org/10.56979/802/2025

Crime Forecasting Using Data Analytics in Pakistan

Naeem Abbas[1], Rimla Anwar[2], Erssa Arif[2], Naila Nawaz[2], Muhammad Amjad[2], and Mudasir Zaheer[2*]

[1]Riphah College of Computing RIUF, Lahore, Pakistan.
[2]Riphah College of Computing RIUF, Faisalabad, Pakistan.
[*]Corresponding Author: Mudasir Zaheer. Email: mudasirzaheer@outlook.com

Received: January 03, 2025 **Accepted:** March 01, 2025

Abstract: According to the 2017 Census, Lahore has a population of 11,126,285, with a police-to-citizen ratio of 1:413—significantly higher than that of many other major metropolitan cities around the world. Given these limited resources, there is an urgent need to develop effective strategies to reduce and control crime in a city like Lahore. One viable solution lies in the ability to predict crimes using historical data. Machine learning and data mining techniques can play a crucial role in forecasting criminal activity and identifying common crime patterns. In this study, we applied regression analysis to a real 15-Lahore crime dataset provided by the Punjab Safe City Authority (PSCA) to predict various crime attributes, including location, time, and type. Additionally, association data mining and clustering techniques used to discover frequent patterns and categorize crimes into distinct clusters. The primary goal of this research is to enable the efficient use of existing resources by employing predictive analytics to support proactive crime prevention.

Keywords: PSCA; CAD; PRU; PPC.

1. Introduction

In early times when people were living without any written law and they just used to follow their religion to live their daily routine life, then the forbidden things in religion called sin, but in the present time there is law for every country for their citizens and the prohibited acts which could harm any individual, society or the state are called crimes which are punishable by the state.

In the 21st century crime is a big trouble for the citizens of the state. It is a duty of the state to protect its citizens from the different kinds of crime. Country like Pakistan having small budget to reduce crime is facing a big challenge now a days. Law enforcement agencies have to do more with less. It would be impossible without science and technology.

As we can find latest customer behaviors and shopping trends by analyzing purchasing data of the buyers, same crimes trends and patterns can be found by studying and analyzing criminal data of the past. With the help of Machine Learning and Big Data, we can predict the crimes' location and time by computing previous crime history. If we could know the time and location which is impossible to do it 100% but with some accuracy it can be done then we can make our forces to prevent crimes before they happen or can arrest the criminal by being at right location at right time. So predictive policing can make a huge difference to prevent or reduce the heinous crimes.

In this research thesis we wanted to implement some Machine Learning and Big Data algorithms on the crime data of Lahore for which we have collected a Lahore-15 crime data set from (PSCA) of the complete year 2018. (PSCA) uses Computer Aided Dispatch (CAD) Software to dispatch crime cases to relevant police station, Dolphin (a security force to control street crimes inspired by Istanbul police) and Police Response Unit (PRU)

force. Dolphin and PRU patrol in the city to fight against street crimes. In CAD there are almost XVI different sub categories under criminal offence but we shall take some of them most heinous crimes given below.

1.1. Types of Crimes

There are some heinous types of crimes and their short descriptions according To Pakistan Penal code (PPC Optimized Growth Conditions

1.1.1. Theft:

Steal something from owner without his/her permission. It's a criminal act.

1.1.2. Robbery\Snatching:

Act of taking valuables like gold and cash with the threat of pistol, gun or knife from citizen of the state.

1.1.3. Burglary:

Break into some one house, shop or property in the absence or without knowing the owner and stole valuables and households.

1.1.4. Vehicle Theft:

According to PPC it is an act of taking motor vehicle without knowing of the owner.

1.1.5. Vehicle Snatching:

Act of taking vehicle using force or without the will of the owner of the vehicle.

1.1.6. Bank Robbery:

It is the act of stealing money from bank using force and threat with the employees and customers of the bank.

1.1.7. Murder:

Murder is a most heinous crime. It's the act of intentional killing of citizen by other but accidental or suicidal deaths are not included in murder.

1.1.8. Sexual Offence:

Sex without consent of female, attempt to rape with force and child molestation are included in this category of crime.

1.1.9. Kidnapping:

According to (PPC) it's an act of abducting someone and keeps him as a captive. Minor kidnap, female kidnap and kidnapping for ransom are included in this category [1].

2. Literature Review

In paper [9] crime detection and prediction is done by using WEKA, an open source data mining tool. Dataset is provided by University of California-Irvine and the real time crime dataset of State of Mississippi which is collected from neighborhoodscout.com. Different techniques are applied to predict and analyses the crime rate like Linear Regression is applied on the given data set to predict some predictable attributes in data set using i.e. (x1,y1), (x2,y2), (x3,y3), (x4,y4), (x5,y5), (x6,y6)…….. Where xi belongs to X and yi belongs to Y the predictive attributes. Regression is describe as $Y = \alpha + \beta X$ where α and β are regression coefficient. Coefficients are determined using least square method by keeping y constant as given below [5].

Where mean x and mean y are the mean values of X and Y and other method which is used to predict crimes is additive regression using learning rate. By reducing learning rate over fitting is avoided [2]. Additive regression is done by using formula below.

$$\beta = \frac{\left[\sum (x_i - mean_x)(y_i - mean_y)\right]}{\left[\sum (x_i - mean_x)^2\right]} \quad \text{and } \alpha = mean_y - \beta * mean_x$$

$$\mathbf{E} = \left[Y \mid \vec{X} \models \vec{x}\right] = \alpha + \sum_{j=1}^{p} f_j(x_j)$$

Figure 1. Regression Formula

And other technique which is used to predict crime rate is decision tree. Mean absolute error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) are found

for above applied techniques and compared their results in tables given below.

Table 1. Results of crimes

Year	Homicide	Attempted murder	Child destruction	Causing death by careless driving
1990	10	19	0	7
1990	6	10	0	5
1990	6	8	0	9
1990	6	2	0	15
1990	10	5	0	1

In paper [10] crime analysis is done by using K- mean clustering on the dataset collected from England Police and Wales police forces from 1990 to 2011. Firstly some steps taken on dataset to get it ready for the processing missing values are filled in it to get it ready for processing. After filling the missing values dataset is normalized. After completing all preprocessing techniques on data k-mean clustering is applied on the dataset. In start number of cluster are defined in which data is going to be distributed. Initially centroids are determined. Every data point is assigned to its closest cluster [6]. After that centroids are computed again by taking the means of data points values and again every data point is distributed to its closest centroid and these iterations goes on until centroids keep changes.

In Paper [10] Association rule mining is applied on the dataset which is consist of socio- economic data of 1990 US census, Law enforcement data and FBI crime reports 1995.Data is preprocessed by removing the tuples of missing values and normalizing the dataset. Association rule mining works on Boolean values but in data there are some categorical values or number like numbers of murder for this fuzzy association rule mining is applied[3, 4]. Fuzzy association is in the form

(X is A) -> (Y is B) where X and Y are the attributes and A and B are their characteristics respectively. Support [7]

, Confidence and Lift is found by computing the dataset with formulas given below.

$$Fuzzy\ Confidence(< X, A >, < Y, B >) = \frac{Fuzzy\ Support\ of\ < XUY, AUB >}{Fuzzy\ Support\ of\ < X, A >}$$

$$number\ of\ records\ in\ D$$

$$Fuzzy\ Lift\ (X \rightarrow Y) = \frac{Fuzzy\ Confidence\ (X \rightarrow Y)}{Fuzzy\ Support(X \rightarrow Y)}$$

Figure 2. Fuzzy Computing dataset

Many rules are recognized 2 of them are pronounced below [8, 9]

[Employed (high)] & [Kids born to never married (High)] -> [Voilent crimes (high)], conf=0.67, lift=15.2, rel sup =1, sup=0.002

Four Regions			
Measure	**Support (%)**	**Confidence (%)**	**Lift**
Minimum	0.4777	60.0	0.59
Maximum	77.431	100.0	13.775
Average	11.295	70.837	1.71797
Three Regions			
Measure	**Support (%)**	**Confidence (%)**	**Lift**
Minimum	0.4753	60.0	0.576
Maximum	74.297	100.0	54.696
Average	11.407	66.47	2.0857
Two Regions			
Measure	**Support (%)**	**Confidence (%)**	**Lift**
Minimum	0.4755	60.0	0.586
Maximum	70.789	100.0	69.164
Average	11.111	67.41	2.0558

Figure 3. Max, min and average values of rules of subset of region

[Employed (High)] -> [Violent Crimes (Low)], conf=0.67, lift=1.2, rel sup=1.0, sup=0.297 by using these rules we can make very important decisions by studying these rules [10].

3. Methodology

Practical implantation is divided into two parts, first of which is consist of data collection and some preprocessing techniques which made our data set feasible so that some mathematical algorithm could implement on data set to get valuable information which could help to reduce the crime rate in Lahore city and the second part is consist of some algorithm implementation on dataset.

3.1. Pre-experimental

Pre experimental work is consist of Dataset collection and preprocessing techniques to make dataset practicable.

3.1.1. 15-Dataset Collection

To make our research valuable there was a need that dataset should be real and belong to the local community that's why we contacted to Punjab safe city Authority (PSCA) to collect the 15-Lahore Dataset which was never been an easy task because dataset has personal information of public and public have the right to privacy of the data as dataset contains the personal mobile number and addresses of the complainant.

So due to the privacy of personal data we didn't feel the necessity of personal mobile numbers of complainant which were not required for our research. We also signed dataset privacy agreement with PSCA and promised that dataset would be used for just research purpose and not used to do any illegal activity.

3.1.2. Dataset characteristics

15-Dataset contains 164603 rows and 8 columns which tell us about crime reporting time crime type police jurisdiction and coordinates of the occurred crimes. Dataset is consisting of all heinous crime like Robbery, theft, vehicle theft, vehicle snatching bank robbery and Assault reported in the whole year of 2018. Dataset contains the Governing units of police like Model Town division and sadder division and police jurisdictions like Bhaati Gate PS. It also contains the coordinates of the location where crime is occurred.

One year henious crimes data (1).xlsx

	A	B	C	D	E	F	G	H
1	Reporting Time	Location	Category	Sub-Category	Governing Unit	Jurisdiction	X Coordinate	Y Coordinate
2	2018-12-31 23:59:49	nad housing scheme, hanjarwal PS k pa	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Hanjarwal PS		
3	2018-12-31 23:59:37	basti rehmat near lidhar pind	ASSAULT	Firing/Aerial Firing	Police Cantt Div	Hair PS		
4	2018-12-31 23:59:36	ferozwala aniexy imamian colony	ASSAULT	Firing/Aerial Firing	Police City Div	Shahdara PS		
5	2018-12-31 23:59:34	Chung center	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Chung PS		
6	2018-12-31 23:59:29	johr town M block	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Johar Town PS		
7	2018-12-31 23:58:54	shan-e-sahaba wali khatak nala	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Sabzazar PS		
8	2018-12-31 23:58:54	Samnabad, Choudry Colony	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Samanabad PS		
9	2018-12-31 23:58:50	canal view	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Mustafa Town PS		
10	2018-12-31 23:58:43	saeed park main	ASSAULT	Firework	Police City Div	Shahdara PS		
11	2018-12-31 23:58:39	Nawan Kot k pass sy chowk py	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Nawan Kot PS		
12	2018-12-31 23:58:36	sadar digi mohallah	ASSAULT	Firing/Aerial Firing	Police Cantt Div	North Cantt PS		
13	2018-12-31 23:58:21	Muhammadi mohala wasanpora	ASSAULT	Firing/Aerial Firing	Police City Div	Shadbagh PS		
14	2018-12-31 23:58:12	kerriya wala gao	ASSAULT	Firing/Aerial Firing	olice Model Town Div	Nishter Colony PS		
15	2018-12-31 23:58:00	joy shah road pa	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Sanda PS		
16	2018-12-31 23:57:20	ichra kambol chowk	ASSAULT	Firing/Aerial Firing	olice Model Town Div	Ichhra PS		
17	2018-12-31 23:57:02	petrol pump shalimar link road daras r	ASSAULT	Firing/Aerial Firing	Police Civil Line Div	Shalimar PS		
18	2018-12-31 23:57:00	0, mujahidabad, ali hospital k sath wal	MISCELLANEOUS	Drunk Behavior	Police Civil Line Div	Mughalpura PS		
19	2018-12-31 23:56:55	shadewal rakhar	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Hanjarwal PS		
20	2018-12-31 23:56:42	anzil , makhdoomabad , chungi amarsa	ASSAULT	Firing/Aerial Firing	olice Model Town Div	Kot Lakhpat PS		
21	2018-12-31 23:56:19	wafaqi colony E1 block johar town	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Johar Town PS		
22	2018-12-31 23:56:14	bilal colony mohala drogha wala	ASSAULT	Firing/Aerial Firing	Police Cantt Div	Baghban Pura PS		
23	2018-12-31 23:56:09	bhatti pa data stop	OBBERY/SNATCHING	Snatching/Jhapatta	Police City Div	Bhatti Gate PS	4.306806168058	31.5792328334361
24	2018-12-31 23:56:03	b dogar chowk klharan stop punjab soc	ASSAULT	Firing/Aerial Firing	Police Cantt Div	Factory Area PS		

Figure 4. A screenshot of the Lahore crime dataset

The dataset provided above also has raw data which would have no use like coordinates columns have only 20% filled values other cells have empty value so only 20% of data could never represent the whole dataset. Dataset has also crime which are not so heinous like drunken behavior and some miscellaneous crimes. Data also have some missing values so there was a need to implement some data preprocessing techniques on data set to make it feasible. So we implemented some data preprocessing techniques which are described below in

details.

3.1.3. Tools used for Data preprocessing

To preprocess 15-Dataset we used different tools like Microsoft Excel to dell useless columns and rows of less heinous crimes. To convert nominal data to numeric so that mathematical algorithm could be implemented on dataset we use Rapid miner tool which is very useful for data science projects.

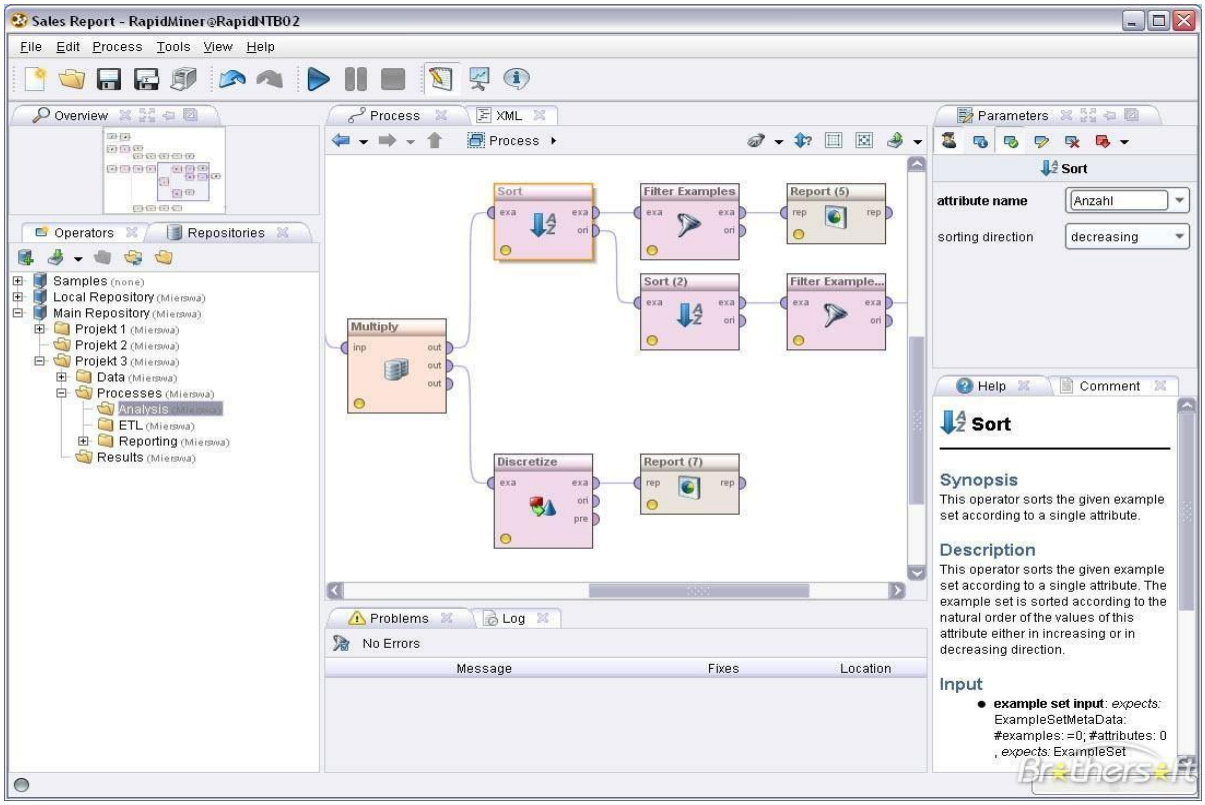


Figure 5. Rapid-miner

Rapid-miner is data science software which provides the environment to implement data mining, machine learning and deep learning techniques on the dataset.

3.1.4. *Preprocessing*

There are some preprocessing techniques implemented on dataset which are given below in details.

3.1.4.1. Feature selection

In the above dataset some features like X and Y coordinates which had only 20% filled values were of no use in data analysis though they were important could be more useful if they had not so much missing values. Location features\column which had unstructured and uncategorized data is also deleted using Microsoft excel tool. Now our dataset contains some necessary features like Reporting time, crime category, sub category, governing unit and jurisdiction. In Reporting time column there is time with date when caller called to complaint the crime which is occurred. Category column contains the heinous crime category in which crime

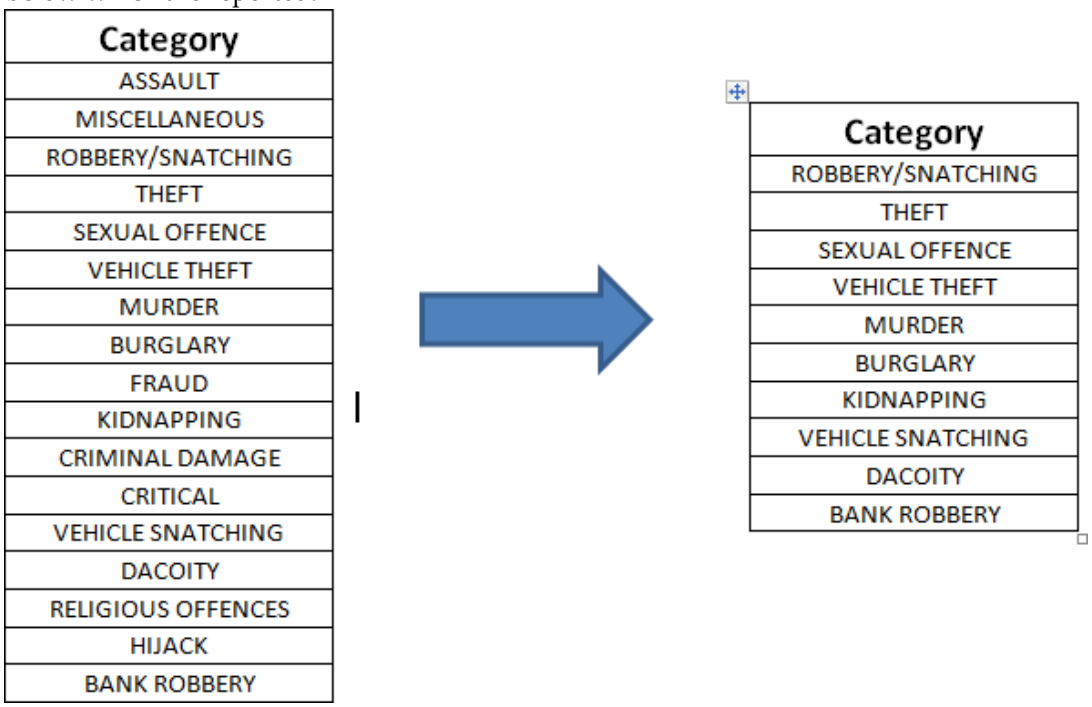
Reporting Time	Category	Sub-Category	Governing Unit	Jurisdiction
2018-12-31 23:59:49	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Hanjarwal PS
2018-12-31 23:59:37	ASSAULT	Firing/Aerial Firing	Police Cantt Div	Hair PS
2018-12-31 23:59:36	ASSAULT	Firing/Aerial Firing	Police City Div	Shahdara PS
2018-12-31 23:59:34	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Chung PS
2018-12-31 23:59:29	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Johar Town PS
2018-12-31 23:58:54	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Sabzazar PS
2018-12-31 23:58:54	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Samanabad PS
2018-12-31 23:58:50	ASSAULT	Firing/Aerial Firing	Police Sadar Div	Mustafa Town PS
2018-12-31 23:58:43	ASSAULT	Firework	Police City Div	Shahdara PS
2018-12-31 23:58:39	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Nawan Kot PS
2018-12-31 23:58:36	ASSAULT	Firing/Aerial Firing	Police Cantt Div	North Cantt PS
2018-12-31 23:58:21	ASSAULT	Firing/Aerial Firing	Police City Div	Shadbagh PS
2018-12-31 23:58:12	ASSAULT	Firing/Aerial Firing	Police Model Town Div	Nishter Colony PS
2018-12-31 23:58:00	ASSAULT	Firing/Aerial Firing	Police Iqbal Town Div	Sanda PS

hold which is reported while jurisdiction and governing unit describes about the police station which would deal with that reported crime.

Figure 6. A screenshot of the Selected features of dataset

3.1.4.2. Instance selection

The dataset had some less heinous and unnecessary crime reported by caller dataset had total 164,403 rows which means 164,403 crime had been reported of specific category in the year 2018. Different categories are given below which are reported.



Category
ASSAULT
MISCELLANEOUS
ROBBERY/SNATCHING
THEFT
SEXUAL OFFENCE
VEHICLE THEFT
MURDER
BURGLARY
FRAUD
KIDNAPPING
CRIMINAL DAMAGE
CRITICAL
VEHICLE SNATCHING
DACOITY
RELIGIOUS OFFENCES
HIJACK
BANK ROBBERY

Category
ROBBERY/SNATCHING
THEFT
SEXUAL OFFENCE
VEHICLE THEFT
MURDER
BURGLARY
KIDNAPPING
VEHICLE SNATCHING
DACOITY
BANK ROBBERY

Figure 7. Instance Selection

Crime like Assault, Miscellaneous, Fraud, Criminal damage, Critical, Religious offences and Hijack are deleted as shown in above figure19.

3.1.4.3. Missing values

In this Dataset every column is completely filled except sub category which is needed to be filled or deleted. So we decided to delete the missing values entries which were few in the dataset. After deleting missing values and less heinous crime we have dataset of 70,333 rows and 5 columns. Dataset after some preprocessing is given below.

3.1.4.4. Date conversions

We had to implement some mathematical algorithm that's why data set needed to convert reporting time into months, date, days of weeks, and hour of the day so that we could do some meaningful analysis on dataset. Date conversion table is given.

Table 2. Date conversions

Date	Month	Day	Hour	Day
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday
12/31/2018	12	31	23	Monday

12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday
12/31/2018	12	31	22	Monday

3.1.4.5. Nominal to binary conversion

Given dataset had nominal values like category, sub category, governing unit and jurisdiction column. We have converted them into binary values like if the reported crime was of robbery category, model town division as governing unit and model town PS as jurisdiction then corresponding features would turn on as 1 and rest of features would turn off as 0.

Table 3. Binary conversion (a)

Category = THEFT	Category = VEHICLE THEFT	Category = MURDER	Category = BURGLARY
.0	.0	.0	.0
1.0	.0	.0	.0
1.0	.0	.0	.0
.0	.0	.0	.0
.0	.0	.0	.0
.0	1.0	.0	.0
.0	.0	1.0	.0
1.0	.0	.0	.0
.0	1.0	.0	.0
.0	1.0	.0	.0

Table 4. Binary conversions (b)

Sub-Category = Snatching/Jhapatta	Sub-Category = Other Theft	Sub-Category = Mobile Cash Theft	Sub-Category = Road Robbery
1.0	.0	.0	.0
.0	1.0	.0	.0
.0	.0	1.0	.0
.0	.0	.0	1.0
.0	.0	.0	1.0
.0	.0	.0	.0
.0	.0	.0	.0
.0	.0	1.0	.0

Table 5. Binary conversions (c)

Governing Unit = Police City Div.	Governing Unit = Police Model Town Div.	Governing Unit = Police Civil Line Div.	Governing Unit = Police Iqbal Town

			Div.
1.0	.0	.0	.0
.0	1.0	.0	.0
.0	1.0	.0	.0
.0	1.0	.0	.0
.0	1.0	.0	.0
.0	.0	1.0	.0
.0	.0	.0	1.0
.0	.0	.0	.0
.0	1.0	.0	.0

Table 6. Binary conversion (d)

Jurisdiction = Bhatti Gate PS	Jurisdiction = Liaqat Abad PS	Jurisdiction = Nishter Colony PS	Jurisdiction = Mozang PS
1.0	.0	.0	.0
.0	1.0	.0	.0
.0	.0	1.0	.0
.0	.0	1.0	.0
.0	.0	1.0	.0
.0	.0	.0	1.0
.0	.0	.0	.0
.0	.0	.0	.0
.0	.0	.0	.0

3.1.4.6. Data visualization

Different aspects of dataset are visualized using Rapid miner tool. Crime rate in different months, different date of month, and different days of week and different hours of the day are visualized below. Crime is also visualized category and jurisdiction wise.

3.1.4.7. Category wise visualization n

Category wise crime is visualized in below. On Y- axis there is count of crime occurring according different categories and on X-axis there are different categories of crime. It can be visualized that crime of category theft and vehicle theft is occurring more than any other crime throughout the year 2018.

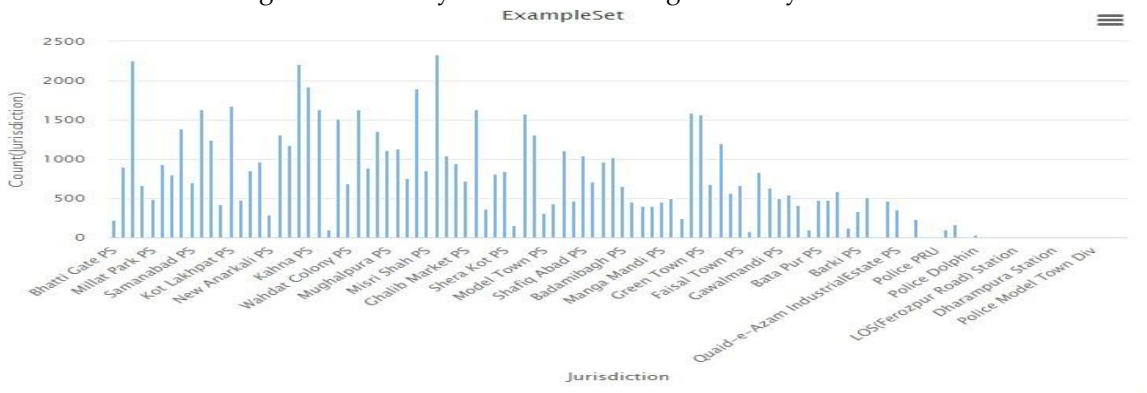


Figure 8. Category wise crime visualization

3.1.4.8. Jurisdiction Wise crime visualization

In given below figure crime is visualized jurisdiction wise. On Y-axis there is a count of crime of different jurisdiction and on X-axis there are different jurisdictions.

3.1.4.9. Month wise visualization

In figure given below crime rate is visualized month wise. On Y-axis there is a frequency of the crime while on X-axis there are different month of the year. It is noticed that in the month of December crime frequency is slightly higher that other months of the year.

3.1.4.10. Day wise Crime visualization

In the figure below crime is visualized day of the week wise. On Y-axis there is a crime frequency and on X-axis there are different days of the week. It can be noticed that on Monday which is a first day of week crime rate is high than other days of the week. Other than Monday crime trend is almost same.

3.1.4.11. Hour wise crime visualization

In figure below crime is visualized hour wise. On Y-axis there is frequency of the crime and on X-axis there are different hours of the day. It can be visualized in late hours of the day crime rate is high.

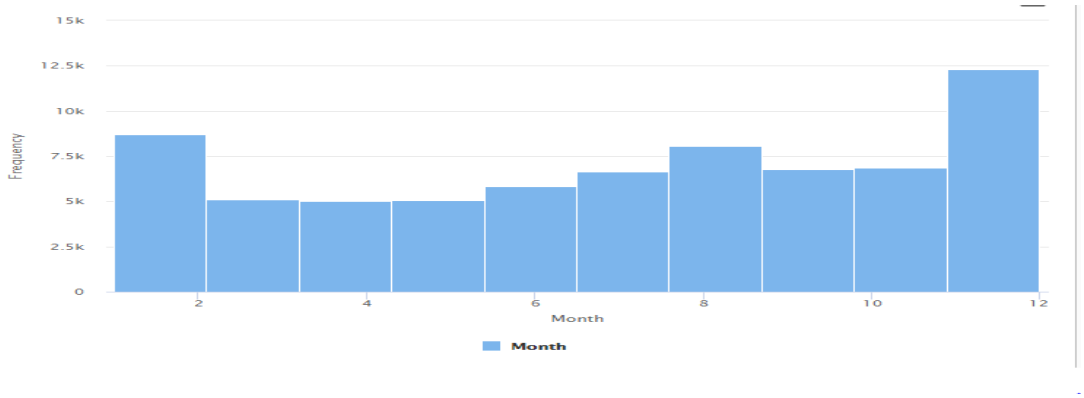


Figure 9. Month wise crime visualization

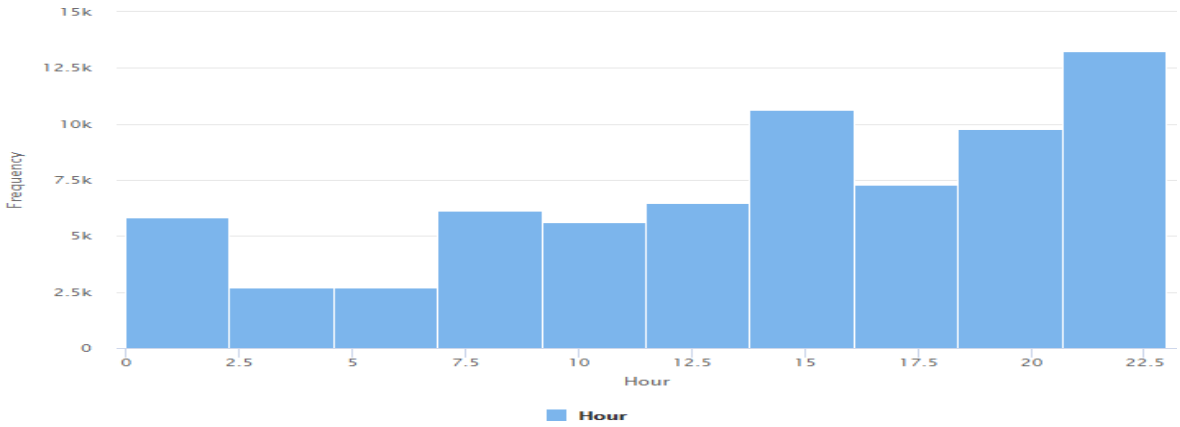


Figure 10. Hour wise crime visualization

3.2. Experimental

In experimental phase we have computed crime frequency and prediction model to predict the crime and have shown the results of our prediction.

Table 7. Crime Tale

Month	Date	Hour	Robbery\snatching	Model town Div.	Robbery in Model town div.
12	31	23	1	0	0
12	31	23	0	1	0
12	31	23	0	1	0

12	31	23	1	1	1
12	31	23	1	1	1
12	31	23	0	0	0
12	31	23	0	0	0
12	31	23	0	0	0
12	31	23	0	1	0
12	31	22	0	0	0
12	31	22	0	0	0
12	31	22	0	0	0

3.2.1. Crime Frequency

Using MATLAB tool we applied and operator between crime category and jurisdiction or governing unit to compute crime of specific category in specific jurisdiction i.e. robbery crime in model town division. As shown in given table:

After that we have calculated crime frequency of specific category in specific jurisdiction or governing unit. We calculated the frequency day wise which told us how much that crime occurred in that jurisdiction on a specific day of the year. Similarly we also have computed hourly frequency of specific crime in a specific jurisdiction or governing unit. By calculating frequency we can implement some prediction model to predict the crime. Actually frequency told about the crime trends. In given below tables it is shown the robbery crime frequency in model town division hourly and day wise.

Table 8. Robbery snatching in model town div. Date wise

Month	Date	Robber\snatching in Model town Div.
12	31	13
12	30	4
12	29	5
12	28	3
12	27	5
12	26	6
12	25	8
12	24	3
12	23	6
12	22	9
12	21	6
12	20	3
12	19	7

In above table we can see the number of robbery crimes which were reported in model town division in the different days of December.

Table 9. Robbery in model town hourly

Month	Date	Hour	Robbery in Model town div
12	31	23	2
12	31	22	3
12	31	21	3
12	31	20	2
12	31	19	0
12	31	18	1

12	31	17	0
12	31	16	0
12	31	15	1
12	31	14	0
12	31	13	0
12	31	12	0

In above table we can see the number of robbery crimes which were reported in model town division in the different hours of 31st, December.

3.2.2. Model Implementation

We have implemented moving average algorithm to predict the crime in python language.

In moving average algorithm we have computed the moving average of specific crime in specific jurisdiction of last 10 days which tell us about the recent trend of that crime in that specific area that's how we can predict the crime on specific day . Some experiments and their results are described below.

Experiment #1: In this experiment we have simply calculated the average crime of previous 10 days and predict the crime of current day. There have been given 2 results of this algorithm. We have predicted the Vehicle theft crime in Johar town Police station month 8 and date 11

```
In [27]: b = c_pred_cd(df, juris="Iqbal Town PS", crime="VEHICLE THEFT", month=8, date=11)

Ground Truth Crime: 2
Year Of Day: 223
Crime Records: [0, 0, 2, 0, 1, 3, 0, 2, 0, 5]
Our Prediction: 2
```

Figure 11. Result 1 experiment 1

In above figure we can see that ground truth crime of vehicle snatching in johar town Ps on that day is 2 and our model predict the crime that day is also 2 .

In other example we have changed the jurisdiction to Iqbal town PS crime vehicle theft month 9 and date 25 result is given below.

```
In [28]: b = c_pred_cd(df, juris="Iqbal Town PS", crime="VEHICLE THEFT", month=9, date=25)

Ground Truth Crime: 2
Year Of Day: 268
Crime Records: [2, 0, 1, 6, 6, 2, 4, 1, 2, 6]
Our Prediction: 3
```

Figure 12. Result 2 experiment # 1

So we can see that ground truth crimes on that day are 2 and our model predicted it as 3 which is almost near to the actual results so we can predict the crime almost to the actual crime by using this moving average algorithm on our 15-Lahore crime dataset.

Experiment #2: In this experiment we implemented moving average algorithm but in different manner. We have calculated the crime average of same day of the week of previous 3 weeks which tell us about the crime trend as week day wise so by using this technique we also can predict the crime. The results of this method are given below. In the example given below we wanted to predict the Robbery/Snatching crime in in Johar town PS on month 6 date15 and the results are given below.

```
In [29]: a = c_pred(df, juris="Johar Town PS", crime="ROBBERY/SNATCHING", month=7, date=28)

Ground Truth Crime: 1
Year Of Day: 209
Crime Records: [1, 1, 0]
Our Prediction: 1
```

Figure 13. Result#1 Eperiment#2

```
In [31]: a = c_pred(df, juris="Johar Town PS", crime="ROBBERY/SNATCHING", month=6, date=15)

Ground Truth Crime: 1
Year Of Day: 166
Crime Records: [0, 2, 2]
Our Prediction: 2
```

Figure 14. Result #2 Experiment#2

In above figure we can see that ground truth crime on that day is 1 and our model predicted crime is also 1. So our model predicts the crime with accuracy. In other example given below we wanted to predict the Robbery/Snatching crime in Johar town PS. So in above figure it is shown that ground truth crime is 1 but our model predicted it 2 which is almost near to the actual results.

My thoughts: In above experiments results are close to the actual values or equal to equal values because I have implemented the moving average. In dataset visualization we noticed that crime rate is different in different month of the year if we implanted the whole average of the data it may produce less accurate results because the trends are different in different month but moving average shows the current trends of the dataset that's why it predicts the results close to the actual values.

4. Conclusion

So from above preprocessing work and experimental work in which we have calculated the frequency of the crime day wise and hour wise we also have visualized the crime data and predicted the crime with some accuracy . So we draw a conclusion that by implementing mathematical and machine learning model on crime data set we can be able to predict the crime. So if we are able to predict the crime we eventually able to stop it before it occurring by utilizing our limited resources and smart work. We also have drawn the conclusion that by implementing algorithm we can find the trend of the crime like road robbery crimes are higher in specific jurisdiction so we can manage our best resources like Dolphin and PRU which are used to control road crime.

Through smart and precise work as above we can reduce the crime 10% to 15% using the available resources but using them efficiently. By analyzing criminal data we can predict the crime category and jurisdiction so we can be able to use our resources. We conclude that heinous crime rate is high at late hours of night so by patrolling in specified area.

If we somehow able to manage coordinates of reported crime we can also be able to produce heat map which could be very helpful to make police patrol efficient.

5. Future Work

In future dataset with more features can make our work more efficient. Different parameters on which crime rate depends if included in our dataset can make our model more efficient. Some features are given below.

1. Literacy rate of the area
2. Total number of police personal in the area.
3. Total population of the area
4. Location\coordinates of reported crime
5. Employment rate of the area
6. Social trends

In future by adding exact coordinates of reported crime we can generate heat map of crime and by using deep learning predicted heat map also can be generated which make our work more reliable.

References

1. J., García, S., Herrera, F. Luengo, On the choice of the best imputation methods for missing values considering three groups of classification methods. 2012.
2. Ramla Shah, Sidra Younas, Maram Sadiq," Crime Patterns in Pakistan: Analyzing Trends and Crime rates using statistical distributions". Feb 2025.
3. H., Motoda, H. Liu, "Feature Selection for Knowledge Discovery and Data Mining," 1998.
4. Faria Feroz, Tahir Hasan, Sajid Mahmood, Hira Asim, "Risk and Pattern analysis of Pakistani Crime Data using Unsupervised Learning techniques. 6 April 2022. V12 isue 7.
5. H., Motoda, H Liu, "Instance Selection and Construction for Data Mining. Kluwer Academic," 2001.
6. J., Sáez, J.A., López, V., Herrera, F Luengo, "A survey of discretization techniques," 2013.
7. Rizwan Hameed, Saman Aziz, Faisal Rehman, "Crime prediction using Advance deep learning techniques: A systematic Review". (2024), Vol 8 No4.
8. I., Derrac, J., García, S., Herrera, F Triguero, "A taxonomy and experimental study on prototype".
9. Lawrence McClendon and Natarajan Meghanathan, "USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA," March 2015.
10. Renuka Nagpal,Rajni Sehgal Jyoti Agarwal, "Crime Analysis using K-Means Clustering," 2013