

**CRIME RATE PREDICTION AND SPOT DETECTION SYSTEM USING
MACHINE LEARNING AND DATA-MINING**

BY

SAKIB MAHMUD

ID: 162-15-7962

AND

MUSFIKA NUHA

ID: 162-15-7930

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. ABDUS SATTAR

Assistance Professor

Daffodil International University

Dhaka, Bangladesh



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2020

APPROVAL

This Project titled “**Crime Rate Prediction and Spot Detection System Using Machine Learning and Data Mining**”, submitted by Sakib Mahmud, ID No: 162-15-7962 and Musfika Nuha, ID No: 162-15-7930 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 07/03/2019.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Md. Sadekur Rahman
Assistant Professor

Department of Computer Science and Engineering
Daffodil International University

Internal Examiner



Saiful Islam
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Md. Motaharul Islam
Professor


Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Abdus Sattar, Assistance Professor, Daffodil International University, Dhaka. Bangladesh,** We, also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:



Mr. Abdus Sattar

Assistance Professor & Proctor
Dept. of Computer Science and Engineering
Daffodil International University

Co-Supervised by:



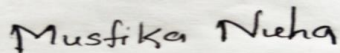
MD. Zahid Hasan

Assistance Professor
Dept. of Computer Science and Engineering
Daffodil International University

Submitted by:



Sakib Mahmud
ID: 162-15-7962
Department of CSE
Daffodil International University



Musfika Nuha
ID: 162-15-7930
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year research successfully.

We really grateful and wish our profound our indebtedness to **Mr. Abdus Sattar, Assistance Professor, Daffodil International University, Dhaka. Bangladesh.** Deep Knowledge & keen interest of our supervisor in the field of “*Machine learning and Data Mining*” to carry out this research project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to the Almighty Allah and our department Head of CSE Dr. Syed Akhter Hossain Sir, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Analysis of crime is a methodological approach to the identification and assessment of criminal patterns and trends. In a number of respects cost our community profoundly. We have to go many places every day for our daily purposes and many times in our everyday lives we face numerous safety problems such as hijack, kidnapping, harassment etc. In general, we see that when we need to go anywhere at first, we are searching for google maps, google maps show that one, two or more ways to get to the destination, but we always choose the shortcut route, but we don't understand the path situation correctly. Is it really secure or not that's why we face many unpleasant circumstances, in this job we use different clustering approaches of data mining to analyzing the crime rate of Bangladesh and we also used K-Nearest Neighbor(KNN) algorithm to train our dataset, For our job, we are using main and secondary data, By, analyzing the data, we find out for many places the prediction rate of different crime and use the algorithm to determine the prediction rate of the path. Finally, to find out our save route, we use the forecast rate. This job will assist individuals become aware of the crime area and discover their secure way to the destination.

TABLE OF CONTENTS

CONTENTS	PAGE NO
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-5
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the study	3
1.4 Research Question	3
1.5 Expected Output	4
1.6 Report Layout	4-5
CHAPTER 2: BACKGROUND	6-10
2.1 Introduction	5
2.2 Related Works	5-9
2.3 Research Summary	9-10

2.4 Scope of the Problem	10
2.5 Challenges	10
CHAPTER 3: RESEARCH METHODOLOGY	11-30
3.1 Introduction	11
3.2 Research Subject and Instrumentation	12
3.3 Data Collection Procedure	13-14
3.4 Data Set Description	14-15
3.5 Statistical Analysis	15-19
3.6 Algorithm	19-27
3.7 Field Survey Graph	28-3
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	31-32
4.1 Introduction	31
4.2 Experimental Results	31-32
CHAPTER 5: CONCLUSION, ACKNOWLEDGEMENT AND IMPLICATION FOR FUTURE RESEARCH	33-34
5.1 Conclusion	33
5.2 Acknowledgement	33
5.3 Implication for Future Study	34
CONTRIBUTION FROM THIS R&D PROJECT	34
REFERENCES	35-36

APPENDICES	37-38
Appendix A	37
Appendix B	38
PLAGARISM REPORT	39 -41

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1.1 Methodological Framework	12
Figure 3.4.1. Before removing outlier of Crime	16
Figure 3.4.2 After removing outlier of Crime	16
Figure 3.4.3 Workflow Diagram	17
Figure 3.6.1 Male vs Female Victim Rate	21
Figure 3.6.2 Monthly Crime Rate	22
Figure 3.6.3 Aged Based Crime Rate	23
Figure 3.6.4 Dhaka North vs South Crime Rate	24
Figure 3.7.1 Different Types of Crime Rate	28
Figure 3.7.2 Day vs Night Crime Occurs Rate	29
Figure 3.7.3 Experienced Based Victim People rate	30

LIST OF TABLES

TABLES	PAGE NO
Table 2.2.1 Summary of previous work	8-9
Table 3.3.3 Data-Set description	14-15
Table 3.5.1 Gender Description	18
Table 3.5.2 Month Description	19
Table 3.6.1 Age Range Description	22
Table 4.2.1 Accuracy Of result	32

CHAPTER 01

INTRODUCTION

1.1 Introduction

In this era of modern world, our popularity is increasing and citification carry enormous general, financial and environmental, while presenting challenges in urban management issues such as traffic resource planning, environment and safe water quality, public policy and public safety services. In addition, represent the most crime rates in larger cities, crime reducing is becoming one of the most important social issues in enormous metropolitan areas as it affects people security issues, youngster growth and person socio-economic status. Crime rate forecast is a scheme that uses different algorithms to determine the crime rate based on prior information. For our daily purposes we have to go many places every day and many times in our daily lives we face numerous security issues such as hijacking, kidnapping, harassment, etc. In general, we see that we are searching for google maps when we need to go anywhere at 1st, google maps show that one, two or more ways to get to the destination, but we always choose the shortcut route, but we don't comprehend the path condition properly. Is it really safe or not that's why we are faced with many unpleasant circumstances, this research introduces the design and execution of a strategy based on past crime data and analyzes the crime rate in past areas at distinct moments., for this work we uses primary data those are collect from the people based on their previous crime problem. In our train information collection, we used different algorithms to figure out the highest precision between the KNN algorithm that provides the greatest precision. In this paper, we use different models and table to show the different types of crime rate, mostly working data from last 3years of crime and showing the level of crime prediction in different issues described in section III. Section II in literature review section describes the previous work on crime rates. Section III describes the general paper information in the Methodology section, this chapter discusses the dataset, data processing, crime analysis, crime rate prediction and describes the multiple algorithms that we used in the crime database assessment and lastly demonstrates the greatest precision of the forecast rate and model

using python matplotlib and KNN algorithm. In chapter IV gives the conclusion and acknowledge about the total work, the and V, provided the references.

1.2 Motivation:

Prithula Islam is one of our colleagues, and she is a medical student at Dhaka Medical College Bangabandhu Sheik Mujib. She was a really good student and also a doctor-researcher based on your analysis she works with different kinds of organization of medical science. She has to go many medical institutions and many places according to her research works and those all places are not known to her, that's why she always used google map to find out the locations. She collects information also many times using different types of social platform. Most of the time she uses google maps to figure out where and google maps shows the various ways to get there. Between those ways she still prefers the shortcut or less traffic ways to reach the particular location but in this, moments she didn't know how convenient this route is for her. As a result, she typically spends a lot of time finding out about the place much of the time, but she did not find the exact location. Several times she encountered multiple forms of unpleasant occurrences such as hijacking, EVS-Teaching, theft etc. to find out the many new locations. This tale made us think about what happened to her and this proposed study came to mind from there. Our research goal is to find out the crime rate at various locations and to figure out the safe route for the people based on this crime rate. The system displays the crime rate relative to the previous crime rate and time-based in various paths. The machine generates a signal or displays the area in different colour depending on the crime rate such as green for hazardous healthy red etc. This also shows hospital, fire and emergency distance depending on the location of the scan. When using this system one can find the safety information about the new place quickly and find out the safe route as well.

1.3 Rationale of the Study:

We need to go many places every day for various purpose. Many times, lives we face numerous safety problems such as hijack, kidnapping, harassment etc. In general, if we need to go anywhere at first, we need to search the path in google map, google map show us one, two or more ways to get to the destination, but we always choose the shortcut route, but sometimes we don't understand the path situation correctly Is it really secure or not? that's why we face many unpleasant circumstances. In this project we use different clustering approaches of data mining to analyzing the crime rate of Bangladesh and we also used K-Nearest Neighbor (KNN) algorithm to train our dataset for our job, we are using main and secondary data. By analyzing the data, we find out the prediction rate of different crime of many places and use the algorithm to determine the prediction rate of the path. Finally, to find out our safe route, we use the forecast rate. This job will assist different individuals about awareness of the crime area and discover their secure way to the destination.

1.4 Research Questions:

Through-out this research, we have attempted to address the following questions concerning the impact of the crime in our daily life.

1. Does crime rate have an adverse impact on social life?
2. How does peoples affect by the crime?
3. Why people are being affected by the crime?
4. Who is more affected? Male or Female?
5. Is there any way of preventing those crime rate?

1.5 Expected Output:

The purpose of the study is to find out the adverse effects of crime rate in social life. However, the following specific outcomes have been expected to achieve the main goal of this paper:

1. Find out the correlation and intensity between different types of crimes in social life.
2. Investigating on how male and female are affected by the different types of crime in everyday life.
3. Finding effective solutions against the different types of crime.
4. Find out the safe route by analyzing the previous crimes rate.
5. Implementing an android and web-based application which will able to show the safe route in different times dependents on the crime data. This application will able to reduce the crime rate by finding the safe route and it's also increases the awareness among the peoples.

1.6 Report Layout:

The above report is split into six separate chapters to obviously present the study. There are several sections in each chapter that make the report simple to comprehend. Chapter 1: Introduction, is the first chapter of this report that helps a reader to know what we will truly explore throughout this research. This chapter clarifies the research issue and at the same time, emphasizes the significance of the current research. It was split into six sub-sections. Chapter 2: Background shows the root of the studied issue. In this chapter, existing literature's findings were discussed related to the current studies. To obtain the study goal, we included summaries of significant, appropriate research studies. We also discussed research gaps based on the previous study in this chapter. This chapter are divided into 5 sub-sections. Chapter 3: Research Methodology describes the methodological part of the study. This part demonstrated, in which way the study was conducted, how data was

collected, pre-processed, analysed and what statistical measures were used to acquire the purpose of the study were. This chapter also divided into 5 sub-sections. In Chapter 4: Experimental result and Discussion, the results of the investigation have been discussed. Experimental results and descriptive statistics were also represented in this chapter. This chapter is also divided into 5 sub-sections. Chapter 5: Summary, conclusion, Recommendation, Implication for future research, is the last chapter of the present report. This chapter demonstrates the ending part of the report. The ending part of the study, summery and future scope were discussed in this chapter. The above chapter was divided into 4 sub sections. References and appendices are included at the end of the study that will be useful in offering a more extensive knowledge of the study issue.

CHAPTER 02

BACKGROUND

2.1 Introduction

Analysis of crime is a methodological approach to the identification and assessment of criminal patterns and trends. In a number of respects cost our community profoundly. We have to go many places every day for our daily purposes and many times in our everyday lives we face numerous safety problems such as hijack, kidnapping, harassment etc. In general, we see that when we need to go anywhere at first, we are searching for google maps, google maps show that one, two or more ways to get to the destination, but we always choose the shortcut route, but we don't understand the path situation correctly. Is it really secure or not that's why we face many unpleasant circumstances, in this job we use different clustering approaches of data mining to analyzing the crime rate of Bangladesh and we also used K-Nearest Neighbor(KNN) algorithm to train our dataset, For our job, we are using main and secondary data, By, analyzing the data, we find out for many places the prediction rate of different crime and use the algorithm to determine the prediction rate of the path. Finally, to find out our save route, we use the forecast rate. This job will assist individuals become aware of the crime area and discover their secure way to the destination.

2.2 Related Works

For this research-based project we have studied the relationship between crime and different features in the criminology literature.

Reduce the crime and detect the techniques of crime and stop the crime before the author uses different techniques. Use Z-Crime Tools and Advanced ID3 algorithm with data mining technology to predict criminal activity. Identify the appropriate crime pattern and

statistical analysis of hidden linked detection algorithm used data. Forensic Tool Kit 4.0 is used to remove research and visualization of data.[1] The uses the K-Means Clustering algorithm for unsupervised learning to determine the crime rate. The model was then analyzed, preprocessed and implemented to taste the set of information and trained the algorithm. K-Means Clustering algorithm provided more than 75 percent. [2] The author used broken window theory, deep learning algorithm, random forest and naïve Bayes to reduce criminal activity and detect the crime zone. Prepare the data frame to train the model for recognition of images, pre-processing of information and detection of crime hotspot. The model tuned with deep learning provides 0.87 percent of the best accuracy. Machine learning offers methods of regression and classification used to predict rates of crimes. The author uses multi-linear regression to find the link between dependent and independent variables. K Nearest Neighbors is used for classification to single and multi-class variable. The Neural Network is used for the prediction's precision. The precision rate of the model is.60,.96 and.97 percent. [4] The Author presents a geographical analysis-based and self-regressive approach to automatically identify large danger urban crime areas and to represents crime patterns in each region reliably. Show the result of the algorithm of crime prediction system consisting of a collection of thick crime areas and a set of related crime forecaster. This operate primarily for the large region where the large amounts of individuals live and demonstrates that the suggested strategy achieves excellent precision over rolling time horizons in spatial and temporal crime forecasting. This paper's working process collects raw data the hotspot uses after splitting the data to create the new hotspot model and finally shows the predictive crime rate. [5] Shiju-Sathyadevan proposed Apriori algorithm for the identification of criminal trends and patterns. This algorithm is also used to identify association rules in the database that highlight general trends. This paper also suggested the naïve Bayes algorithm by training crime data to create the model. The result showed after testing that the Naive Bayes algorithm gave 90 percent precision. K. Zakir-Hussain et al. used the methods of information mining to analyze criminal conduct. This paper proposed tool for analyzing criminal investigation (CIA). Within the law enforcement community, this instrument was used to assist resolve violent offences. This

study is about the various type of crime scene. Both from an investigative and a behavioral perspective, the analysis was done. It provided insight into the unknown criminals as well as recommendation for investigation and interview and trial strategies. [9] Classification is one kind unique methods of information mining used to classify each object in an information set into one of the predefined classes or groups, The idea is to define the use of the Criteria for the segmentation of the entire database, once this is done, individual datasets can then fall naturally into one or more groups. By means of classification, existing datasets can be easily understood and it also helps to predict how new individual datasets will behave based on the classification. Datamining generates models of classification by observing classified data and discovering a predictive pattern between those data. Naive Bayes is a classification algorithm used to predict that it works on the principle of Bayesian.

Table 2.2.1 Summary of the previous work:

Year	Context	Sample	Age	Country	Research Objective	Research findings	Ref.
2018	Crime Analysis	560,000	12-65	Canada	Machine learning based crime rate prediction.	By analyzing the 15 years data find out the crime prediction accuracy and provides a preliminary framework for further analysis.	[12]
2018	Analysis the kidnapping crime	226,000	10-40	Colombia	Soft computing technique to analysis the kidnapping crime.	Find out the relationship between kidnapping crime and military operation also investigate the kidnapping pattern and flows.	[09]
2018	Statistical analysis the crime rate	185,000	N/A	India	Identify the level of various crime based on statistical model.	Usages various statistical model and analysis the previous data find the future crime rate.	[07]

2017	Crime pattern detection	63780	N/A	India	To synthesize results on interactions between crimes type and crime pattern.	Detect the crime rate and generate the crime hotspot to help the police work.	[13]
2017	Criminal short List using Machine Learning	N/A	N/A	Colombo	Analysis the criminal activity and find out the criminal pattern based on criminal records.	Findings the criminal short list by using cross fold validation technique.	[02]
2017	Crime location and story detection.	N/A	N/A	Bangladesh	Crime news analysis based on news paper and location detection using data mining technique.	Find the news that is crime related or not and also finds the crime location to understand the crime pattern.	[15]
2017	Assist Crime Prevention	115670	N/A	Taiwan	This study aimed at evaluating the drug related crime rate and predict emerging crime hotspot for additional police attention.	By using machine learning algorithm predict the more accurate future crime rate and provided a predict hotspot to prevent the crime rate.	[24]
2018	LED policing and Crime Analysis	79870	N/A	Ukraine	Analyzing the criminal pattern by using data mining technique and developing the police activities to prevention and investigation crimes.	Find the model to create software-based crime rate tracker and automated search engine to find the crime rate based on those models.	[21]
2016	Criminal Network Tracker	N/A	N/A	Turkey	The purpose of this research was to clarify the correlations between various type of criminal and their networking activity.	Detect the crime location, time and similarity of crime committing methods between criminal and their others important factors.	[16]
2016	Crime Survey to Support Crime Profiling	N/A	16-55	UK	Identify the various type of statistical model and using data mining technique to find the crime rate based on previous year data.	In this research find the crime rate each year and proves a statistical model that's helps to developed CCJS (Center for Crime and justice Studies)	[22]

2.3 Research Summary

1. Most of the previous research find out the crime accuracy based on Machine Learning based algorithm.
2. Finding various types of crime pattern and core-relation between them using machine learning and data mining algorithm.
3. The main goal of the research was to find out how adverse crime effects impacts on the social life of the people.
4. Finding the statistical analysis of crime rate based on the crime pattern and various location.
5. Our findings demonstrate that male response by 78% on the other hand female are response by 83% in our survey analysis average in 3 years data.
6. Our study reveals that those people are conscious their daily out-work activity those are less affected by crime.
7. The research attempted to comprehend the reason for the adverse correlation various types of crime and finding the specific way how people can safe them to various types of crime and also create awareness among them.

2.4 Scope of the Problem

The scope of the current study is "Crime Rate Prediction and Spot Detection." We researched and attempted to investigate the effect of various types of crime in our social life on learners throughout the research. The hypothesis is that we can reduce the crime rate throughout by this research. As well as CRIME are a big concern for the people in their daily life and mental health.

2.5 Challenges

We faced several challenges to conduct the study. In several stages, we had to change our decisions to cope with those challenges.

- Identifying right research topic.
- Building a right research methodology.
- Identifying right target audience.
- Dealing with primary data (Data preprocessing)
- Identifying right algorithms to analyze those data.

CHAPTER 03

RESEARCH METHODOLOGY

3.1 Introduction

Methodological portion of this research enables the reader to assess the general validity and reliability of a research critically. The primary purpose of the present study is to explore the adverse effect of crime in our social life. To acquire the study's objective, both quantitative and qualitative approaches were followed. Figure 3.1.1 shows the methodological framework of the current study. A substantial literature review was performed at the beginning to identify the gap in the existing study. Simultaneously, significant study issues were found through literature review which assisted to collect relevant data. Authentic journal articles, book chapters, conference papers, MSc theses and PhD theses, newspapers, reports and other dissertations were gathered as a secondary source of data from numerous renowned online databases and reliable sources. Primary data was collected through field survey and online survey. In order to obtain the precise value, dataset was fully prepossessed until irrelevant, incomplete, inconsistent information were removed. Various statistical techniques have been pursued to obtain the objective of the research.

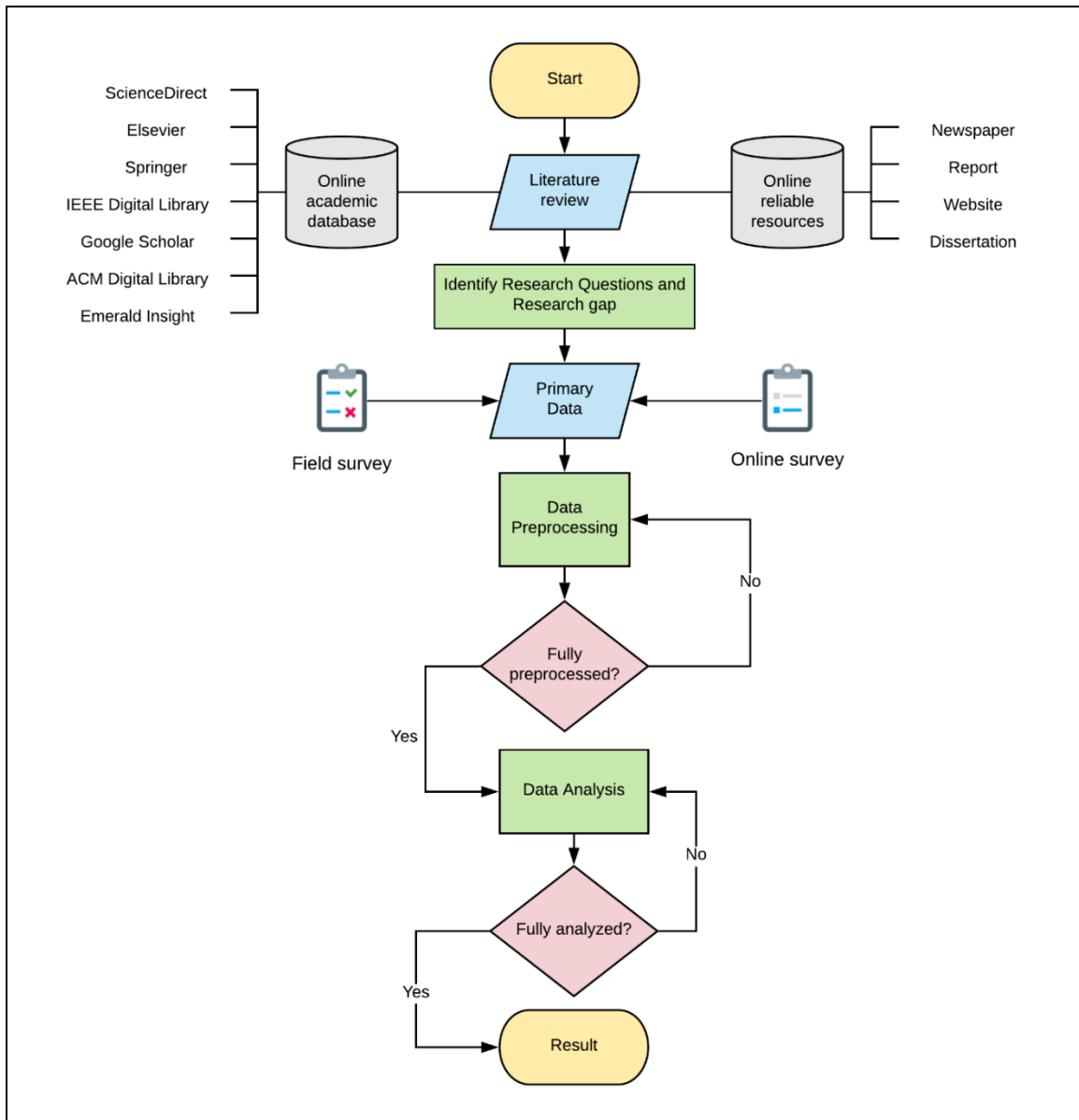


Figure 3.1.1 Methodological Framework.

3.2 Research Subject and Instrumentation

3.2.1 Research Subject:

We're working with a various type of crime in this research. This research examines the impacts of crime in our social life. The purpose of this research is to disclose the real reality about what crimes are actually affect and how crime affects us in various ways.

3.2.2 Research Instrumentation:

A survey was designed to collect data. We prepared questionnaire based on previous studies to acquire the purpose of the study. Five multiple-choice questions along with two general questions where respondents had to input number, were included in the survey questionnaires which was released at Daffodil International University, Dhaka, Bangladesh between March 2019 and January 2020. In order to fulfill the research objective and achieve the effective outcome, questionnaires were distributed to university students and they were asked “Did you ever face any unpleasant occurrence, “If you faced any unpleasant occurrence is it first time in your life?”, “What types of crime did you face?”, “Average spending time outside home?”, “In which area have you been faced that occurrence?” and At the same time, respondent’s ethical factors were carefully ensured that their privacy was kept strictly and confidentially secret.

3.3 Data Collection Procedure

To fulfil the purpose of the study both qualitative and quantitative methods were followed. Data from all appropriate sources were gathered to find responses to the study issue. The overall data collection procedure can be divided into two parts:

3.3.1 Primary data collection:

To acquire the aim of the study a survey was designed for collecting primary data. Both online and offline (field survey) were conducted. For online survey, Google form (Web based application) was used for data collection. Five multiple-choice questions along with two general questions where respondents had to input number, were included in the survey questionnaires which was released at Daffodil International University, Dhaka, Bangladesh between March 2019 and January 2020. In order to fulfill the research objective and achieve the effective outcome, questionnaires were distributed to university students.

3.3.2 Secondary data collection:

Secondary data basically collected from previous studies. We gathered previous studies and extract key information related to our study. Authentic journal articles, book chapters, conference papers, MSc theses and PhD theses, newspapers, reports and other dissertations were gathered as a secondary source of data from numerous renowned online databases and reliable sources.

3.3.3 Data Set Description:

The crime dataset is extracted from primary data collection based on field work. This dataset consists of about 2500 in 10 rows details. The key features such as Name, Years, Months, Crime Type, Crime Areas, Victim Genders, Victim Ages, Victim Areas, and Months are selected from the dataset as the system input features.

The characteristics Perpetrator Ages, Perpetrator Genders and Victims relation are select system's target variables.

Number	Details		
	Name	Type of Columns	Descriptions
1	Person_Id	Value Type	Person ID in Dataset
2	Name	String	Victim Person name
3	Year	Numeric	Crime Occurs Year
4	Ages	Numeric	The ages of the victim
5	Gender	String	Victims Neuter
6	Time	String	Time when the crime has occurred
7	Victim Area	String	Area where the crime has occurred.

Number	Details		
	Name	Type of Columns	Descriptions
8	Region	String	Region of the victim.
9	Home Town	String	Home town of the victim
10	Month	String	The year in which the crime has occurred

Table:3.3.3(Data-set Description)

3.4 Statistical Analysis

We divided the statistical analysis part into two following sub-parts:

3.4.1 Data Preprocessing

Total 3500 questionnaires were distributed and 2862 valid responses were collected including spontaneous male (78.8%) and female (83.2%) respondents. In order to obtain the precise value, dataset was fully preprocessed until irrelevant, incomplete, inconsistent information were removed. Our raw dataset had 2718 records. At first, we checked the missing values of records in our dataset. We dropped the records which were incomplete. After removing the incomplete record, we had 2604 records. As we collected various types of crime data, we wanted to make sure the range of the crime level should be in 0 to 4 range. So, all the records which had less than 0 and greater than 4 were dropped. After that we had 2586 records. After next, to discover outliers we used box plot and Interquartile Range (IQR) measures as a part of data preprocessing. First, we identified outlier for victim of the crime as shown in Figure 3.4.1.1.

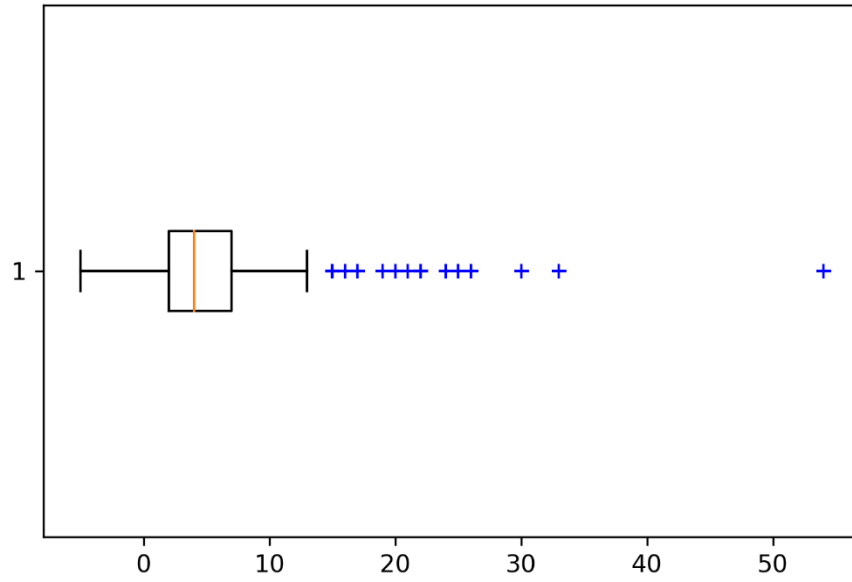


Figure 3.4.1.1 Before removing outlier of crime

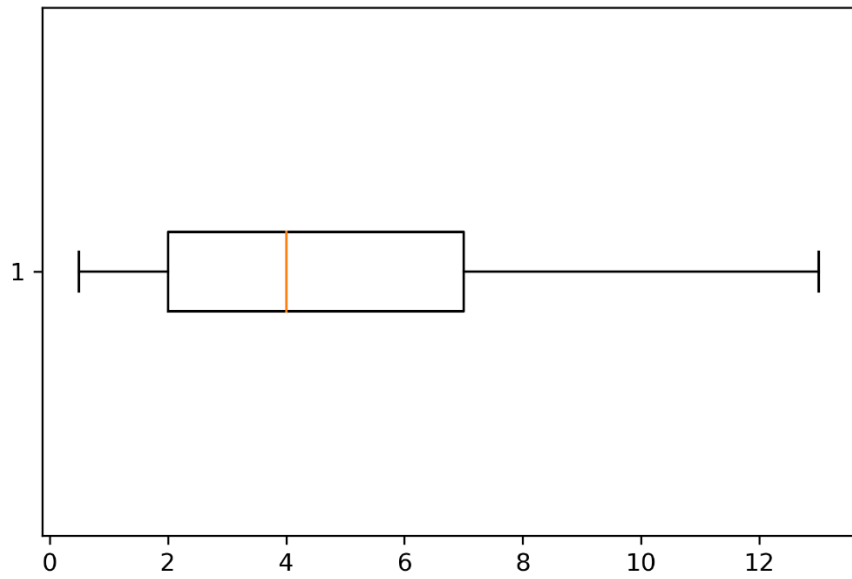


Figure 3.4.1.2 After removing outlier of crime

Since unknown is not a value to be considered but an indicator of unfinished papers, we decided to remove unknown values from the information set, Dates & times was recorded as a times window in MM / DD / YY & HH /MM" format from the documents as direct matching of these dates are complicated by the classification system to match date and time with each other. Classification of the date was considered to classify a particular date into three groups as weekends, weekdays and unaware. This classification is based on the features of the date time windows.



Fig.3.4.1.3 Work flow Diagram

Explains the system's workflow. The workflow begins by extracting data from data collection, which is a dataset repository on different roles. The primary data will be pre-processed and transformed into a criminal data.

Four target variables are predicted.

1. Linear Regression is used to find the aged based crime accuracy.

2. Pearson r correlation analysis is used find out the various types of relation-based accuracy.

3. The sex is estimated using the classification of K- Nearest Neighbors.

4. The gender of the perpetrator is estimated using the classification of K- Nearest Neighbors.

5. The Final prediction rate for years based on age, sex, time, and year using K-NN algorithm.

3.5.1 Crime Data-Preprocessing

Gender	Male	Female
	0	1

Table:3.5.1(Gender)

Few property features of dataset like months, types of crime, victim gender, age, area and the relationship are in the above data set are qualitative from. This qualitative information should be classified as in order 0 or 1 to apply the mathematical models for prediction.

As follow the rules male contains 0 and female contains 1 in table 1. Also adding dummy column by using the crime data in different month of the year in table 3.

Months	D_A	D_B
Jan-April	1	0
May-August	0	1
Sept-Dec	0	0

Table:3.5.2(Month details)

This table 3 adds N-1 dummy columns in the given column for unique N values. The table above explains how data is pre-processed. Three months are selected from the Column state in the data table. In this case, in the column states, the number of unique values is N=3. It is necessary to add dummy columns to categorize the data N-1. Here are two types of dummy columns those are dummy A and dummy B.

1. 1 is assigned for column A and 0 for column B result between January to April.
2. If 0 is assigned for column A and 1 for column B result between May to August.
3. If both is assigned 0 for column A and column B result between September to December.
- 4.

3.6.1 Algorithm

Various statistical measures were followed to analyze the data. According to the purpose of the study, we had to find out the strength of relationship between “age, gender, various types of crime” and location. To find out the crime rate we used linear regression and Naïve Bayes and K-Nearest Neighbor algorithm and we use the Pearson r correlation to measure the degree of various types of relationship.

1. Linear regression: Simple linear regression is helpful in discovering a connection between two continuous variables. One variable is predictor or autonomous, and the other variable is variable reaction or dependent. It allows for a connection that is statistical but not deterministic.

Equation:

$$y = b_0 + b_1 * x$$

$$y = \text{SGPA}$$

$$x = \text{STISM}$$

$$b_0 = \text{y-intercept}$$

$$b_1 = \text{slope}$$

$$n = \text{number of records}$$

$$\text{Here, } b_0 = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$\text{And, } b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Linear regression is used in the crime prediction situation to represent the most likely perpetrator age given the crime scenes.

In this figure, the victim rate between male and female is shown, and the past information analyzes show that female victims are quickly increasing between males every day.

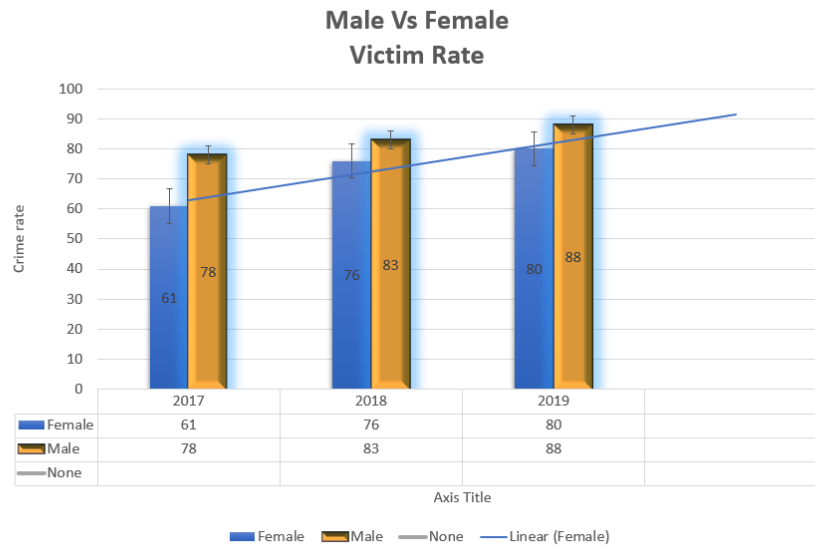


Fig 3.6.1: Male vs Female Rate

To find out the monthly crime rate we used KNN algorithm at the start we find out each month crime rate for every year. This figure shows the individual month crime rate for three years. We also find out the crime rate which time is more crime occurs in the year, finding that rate divides the years into three dimensions for type 1 from January to April, type 2 may occur to August and type 3 from September to December, and showing that type 1 and type 3 occurred the most crime in those months.

Monthly Crime Chart

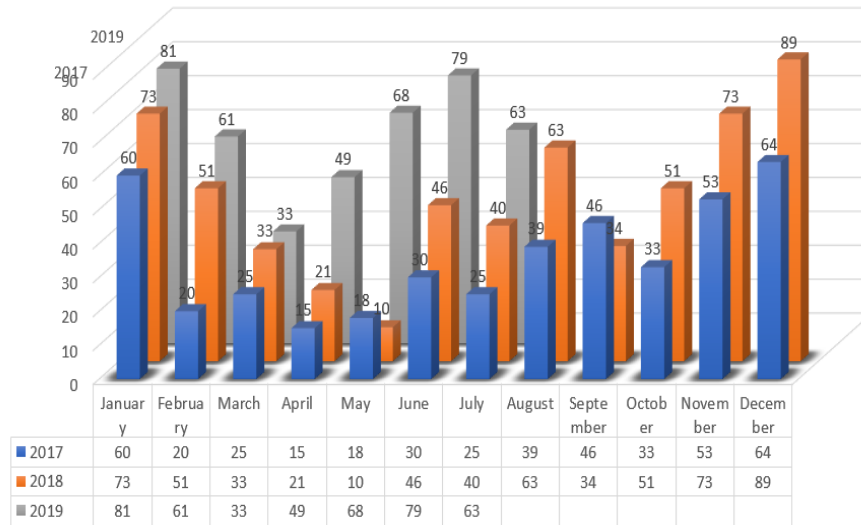


Fig 3.6.2: Monthly Crime Rate

By using the K-NN and Linear regression we find out the aged crime rate for 3 years. To find that rate we divide the age in three types those are Teenager, young and old.

By using the K-NN and Linear regression we find out the aged crime rate for 3 years. To find that rate we divide the age in three types those are Teenager, young and old.

Age	Range
Teenager	13-19
Young	20-55
Old	56-100

Table:3.6.1 (Age Range)

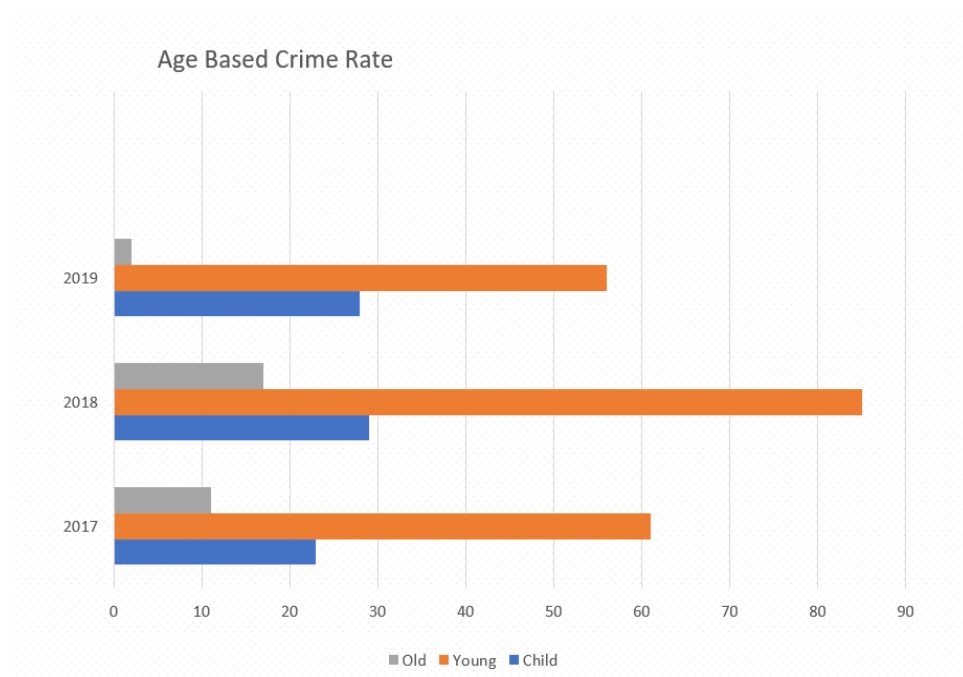


Fig 3.6.3: Age Based Crime Rate

In Dhaka city crimes rate we divide the city in two zone north and south and zone then plot the data by using KNN algorithm X axis for north and Y axis for south, finally show the result in fig:4 that is north cities people are more victim then south city in Dhaka.

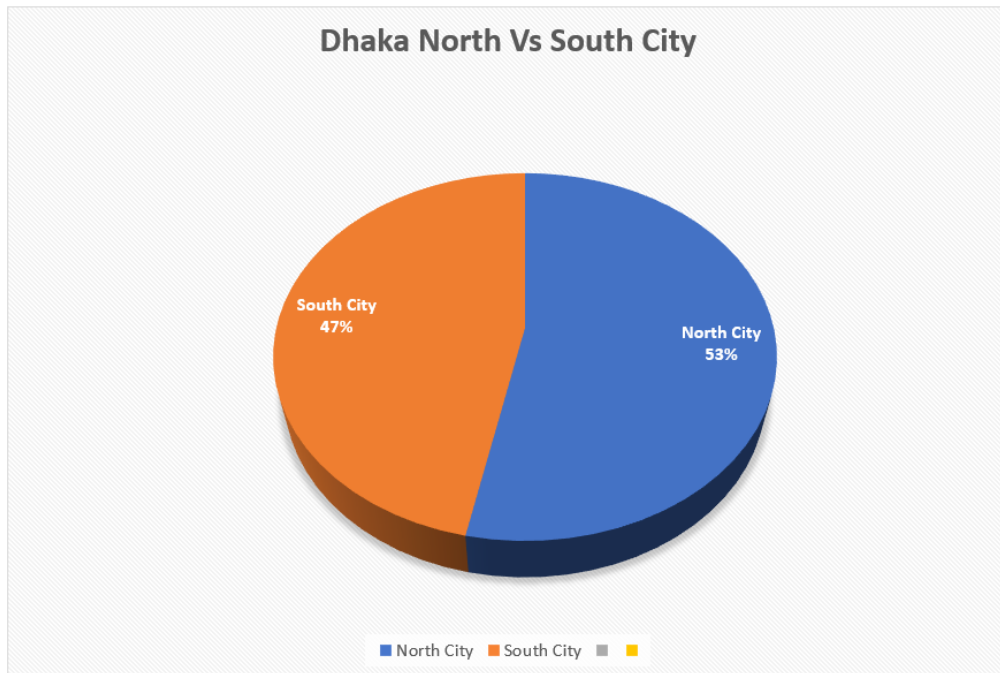


Fig 5: Dhaka City Crime Rate

2. K-Nearest Neighbors

K-Nearest Neighbors is used when the target variable must be classified in more than two classes. In this dataset, there are three classes of target variable perpetrator sex: male, female, and unknown. Similarly, three categories of young, old and kid are defined in age. To classify these target variables, K- Nearest Neighbors Classifier is used.

$$D(a, b) = \sum_{i=1}^n |b_i - a_i|^2 \dots \dots \dots (2).$$

Pseudo Code:

At first KNN Classifier (Data Entry)

Assign the Number of Cluster in K—

Choose a set of K instances to be cluster centers

Data points for each output

Calculate the Euclidean range

Assign next to the data point the cluster

Perpetually Calculate centroids and reassign the variables of the cluster

Repeat until you reach an appropriate cluster

Result, Give back the clusters and their values.

3. Naïve Bayes

Naive Bayes classifiers are a collection of Bayes' theorem-based classification algorithms. It is not a single algorithm, but an algorithm family in which all of them share a common principle, each combination of features being classified is separate from each other. Bayes' theorem considers the probability of a case occurring considering the probability of another case occurring already. The theorem of Bayes is indicated as the following equation:

$$p(A|B) = \frac{P(B|A)P(A)}{P(B)} \dots\dots\dots (3).$$

Naive Bayes algorithms are mostly used to determine emotions, delete spam, suggestions, etc. We are fast and easy to implement, but their biggest downside is the need for autonomous predictors. In most real-life situations, the predictors are dependent, hampering the output of the classifier. We do not use that algorithm to find the final result in our work for this problem, but in many previous cases its works are very good and given the best crime rate accuracy.

4. Pearson r correlation analysis:

Pearson r correlation is the most commonly used correlation statistics to assess the degree of the relationship between the linearly related variables. We use the Pearson r correlation to measure the degree of relationship between the gender, age, crimes type and location.

Equation:

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2] - [N \sum y^2 - (\sum y)^2]}}$$

Here,

N = Number of pairs

$\sum x^2$ = sum of squared x scores

$\sum y^2$ = sum of squared y scores

$\sum x$ = sum of x scores

$\sum y$ = sum of y scores

$\sum xy$ = sum of products of x and y scores

4.1 Spearman's rank-order correlation:

This is another method to calculate correlation. There are 2 methods are available to calculate this spearman's rank order correlation. As our data was tied ranks, we followed the following formula.

Equation:

$$\rho = \frac{\sum_i(x_i - x') (y_i - y')}{\sqrt{\sum_i(x_i - x')^2 \sum_i(y_i - y')^2}}$$

Here,

i = paired score

4.2 Kendall rank correlation:

It's a non-parametric measure of relationship between columns of ranked data. We also used this statistic to measure the relationship between Male and Female. The equation is given below.

Equation:

$$\text{Kendall's Tau} = (C-D/C+D)$$

Here,

C= number of concordant pairs

D= number of discordant pairs

4.3 Point –Biserial correlation:

It's another measure by which we are able to measure strength and direction of association that exist between one continuous variable and one dichotomous variable.

Equation:

$$r = \frac{(y'_1 - y'_2)\sqrt{pq}}{sd_y}$$

Here,

sd_y = Standard deviation of population

3.7.1 Field Survey Questions Based Graph:

In our data-set we find out the different types of crime, we see that 7 types of crime in our data set hijack, Harassment, robbery, pick-pocket, kidnapping, eve-teasing and murder. By analyzing the data set of 3 years 2017,2018 and 2019 we see that the crime rate of some crime types equal between 2017 and 2019 like hijack, murder, robbery and harassment. We also see that pick-pocket crime are decrease day by day, kidnapping and eve-teasing also decreasing.

Different Types of Crime Rate Based on Data-Set



Fig: 3.7.1 Different types of crime rate

3.7.2 Field Survey Questions Based Graph:

Generally, we know that crimes rates are equally high in the night. We have survey questions in our appendix (In which time you spend most outside home?) by using the questions survey we create that graph. The graph result show that the crimes rates of day is 36% on the other hand night is 64%. Some, day time crimes rates are equally high then night time and also some night crimes occur rate are equally high then day time. We also see that some crimes are also happened in day time like harassment eve-teasing etc. This graph will show the crime occurs rate in day vs night in below:

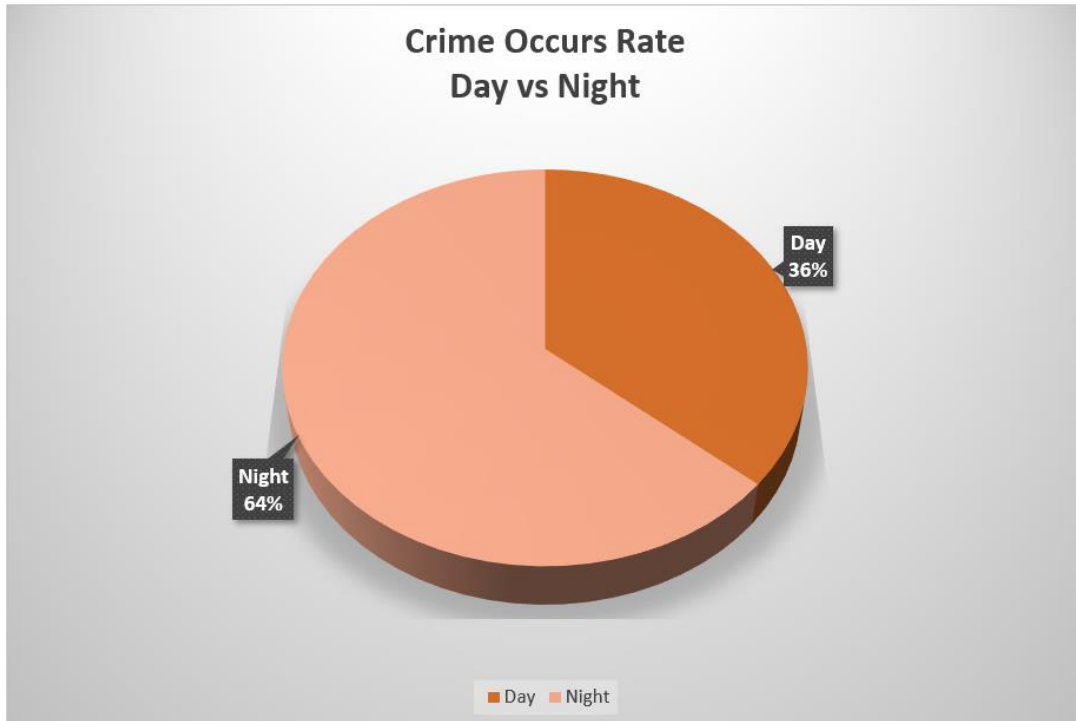


Fig: 3.7.3 Crime Occurs Rate Day vs Night

3.7.3 Field Survey Questions Based Graph:

One of our survey questions is that (Did you faced any unpleasant occurrence is it first time in your life?). By using this question, we want to find out that, those people are affected by the crime are they have knowledge about or not. From the data set we see that those have knowledge or already faced those types of problem those people are less affected. This graph will show the Crime Rate based on previous knowledge.

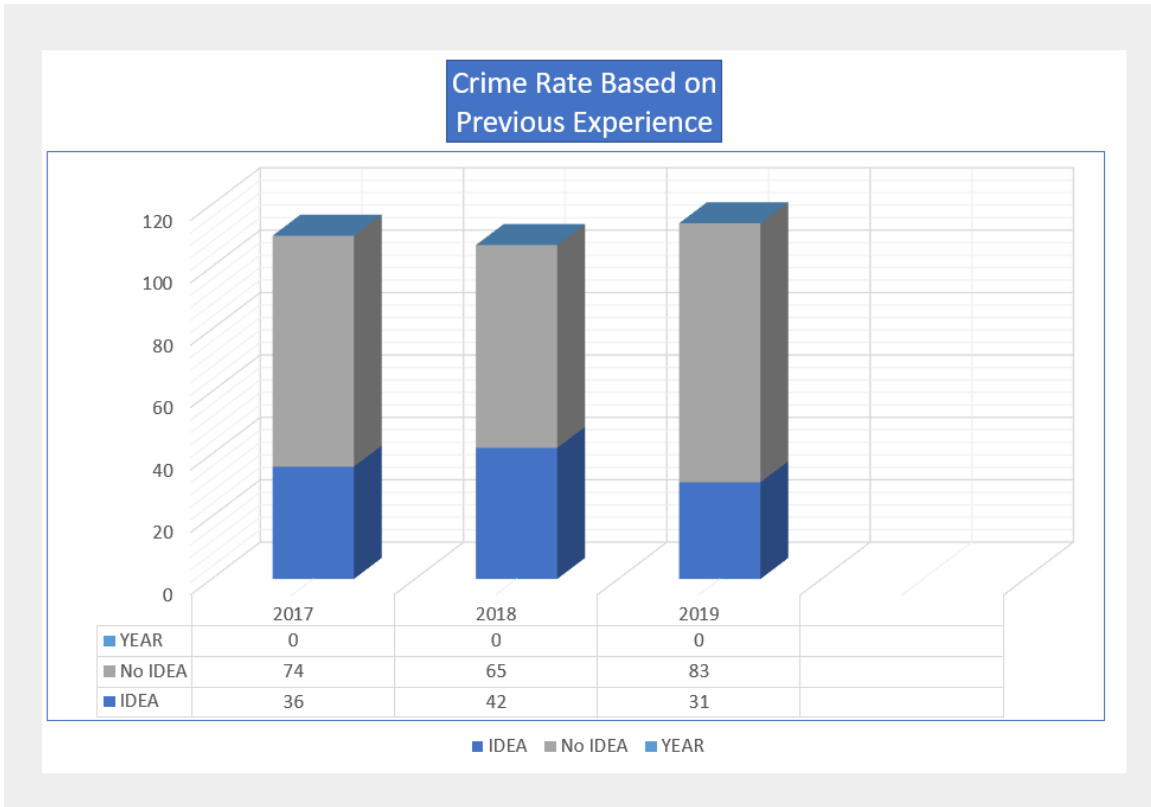


Fig: 3.7.3 Crime Rate Based on Previous Experience

CHAPTER 04

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

One of the most common things is that platforms for interacting with each other is social networking sites. Billions of everyday users have made this platform a main place to promote company. Most social media businesses hire attention engineers because they want to make sure their customers spend as much time on their websites and always attempt to create their product as addictive as possible. As a consequence, we can say that now a day's we are mostly depends on the Internet world. Daily we moved a lot of places and everyday a moved new place for our various working purposes. The main problem is that all places we don't know well as a result we need to help or uses various types of maps like google maps and many times we face various types of unpleasant occurrences. In this research purpose we are using main and secondary data, By, analyzing the data, we find out for many places the prediction rate of different crime and use the algorithm to determine the prediction rate of the path. Finally, to find out our save route, we use the forecast rate. This job will assist individuals become aware of the crime area and discover their secure way to the destination.

4.2 Experimental Results

In this part summarizes the paper and make aware about the future crime based on algorithms and crime data set we find out the crime rate in various section like age based, male vs female, area based and monthly crime rates. The data sources and methods used to guide forecasting include various type crime statistics, survey of the general people data, literature reviews and statistical models that extrapolate crime trends into the future. Algorithms models that describe the behavior of observe past values can be used to forecast future crime trends by projecting a time series analysis of crime trends into the future. Any predictive model endeavor to show a relationship between certain predictor and a dependent variable. To ensure the greater accuracy those models must identify and predict the scope and nature of a number of factors that will influence crime and victimization in the future. This research paper about future crime rate predictions are much more specific and precise. The accuracy table of different algorithm accuracy:

Accuracy table

Year	Algorithm	Accuracy
2017,18,19	Linear	73.61403
2017,18,19	Naïve Bayes	69.5087
2017,18,19	KNN	76.9298

Table 4.2 Result accuracy

By referring to these 3 algorithms, we demonstrate that K Nearest Neighbor provides the crime rate forecast system the greatest precision.

CHAPTER 05

CONCLUSION, ACKNOWLEDGEMENT AND IMPLICATION FOR FUTURE RESEARCH

5.1 CONCLUSION

The sparsity of crime in many areas complicates the application of the prediction rate area-specific modeling. We used the Machine Learning algorithm in that work to create and test age, sex, year, moment, month prediction of crime. In that job we use three types machine learning algorithms Linear regression, Naïve Bayes and K Nearest Neighbor among which we discover distinct precision in different instances some linear operates good and provides better precision but the general situation K Nearest Neighbor provides the appreciated accuracy other than that's why we use K Nearest Neighbor for our Crime Prediction scheme.

By using these predict systems we will discover the stronger precision in the future and also by using this precision we will identify and discover the hot zone region in crime rate. In order to finish this job, we would like to use the CNN algorithm to analyze picture information and add the Google API for viewing the hot zone.

5.2 ACKNOWLEDGE

We are very grateful to my Daffodil International University for offering us the chance to be part of the independent research study that contributes to this work being developed.

Many thanks to Mr. Abdus Sattar sir for innumerable debates and feedback that helped me effectively finish the job.

5.3 Implication for Future Study

In this research we use various types of crime data to find out the crime accuracy based on time, year, sex, age to find the accuracy we use uses various types of machine learning and data mining algorithm.

By using these predict systems we will discover the stronger precision in the future and also by using this precision we will identify and discover the hot zone region in crime rate. We show the normal map and identify the hot zone in this map by using this map people can easily identify the safe route to complete this task we use Google API for viewing those zone area and secure path. We also add the picture searching option to find out the specific crime rate of that area we use CNN algorithm to finish the task. This project also able to show the hospital, fire service and police station location based on the user search destination or specific area that he/she wants to go.

CONTRIBUTION FROM THIS R&D PROJECT

1. International publication-1: Scopus indexed

Sakib Mahmud, Musfika Nuha and Md. Abdus Sattar, " Crime Rate Prediction System Using Machine Learning and Data Mining ", Accepted in Conference on Computing and Communication in Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majhitar, East Sikkim from 20th – 21st March, 2020.

2. International publication-2: Scopus indexed

Sakib Mahmud, Musfika Nuha and Md. Abdus Sattar, " Fraud Rate Detection System Using Machine Learning and Data Mining algorithm ", Accepted in Conference on Computing and Communication in Computer Science and Engineering, Sikkim Manipal Institute of Technology, Majhitar, East Sikkim from 20th – 21st March, 2020.

REFERENCES

- [1] Lin, Y., Chen, T. and Yu, L. (2017). Using Machine Learning to Assist Crime Prevention. 2017 sixth IIAI International Congress on Advanced Applied science (IIAI-AAI).
- [2] Munasinghe, Malith & Perera, Harsha & Udeshini, Shanika & Weerasinghe, Ruvan. (2015). Machine Learning Based Criminal Short Listing Using Modus Operandi Features. 10.1109/ICTER.2015.7377669.
- [3] Chauhan, Chhaya & Sehgal, Smriti. (2017). A review: Crime analysis exploitation data processing techniques and algorithms. 21-25. 10.1109/CCAA.2017.8229823.
- [4] Munasinghe, Malith & Perera, Harsha & Udeshini, Shanika & Weerasinghe, Ruvan. (2015). Machine Learning Based Criminal Short Listing Using Modus Operandi Features. 10.1109/ICTER.2015.7377669.
- [5] Chauhan, Chhaya & Sehgal, Smriti. (2017). A review: Crime analysis using data mining techniques and algorithms. 21-25. 10.1109/CCAA.2017.8229823.
- [6] Catlett, C., Cesario, E., Talia, D. and Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments.
- [7] Anon, (2019). [online] Available at: https://www.researchgate.net/publication/322000460_A_review_Crime_analysis_using_data_mining_techniques_and_algorithms [Accessed 30 Aug. 2019].
- [8] Quick, M., Li, G. and Brunton-Smith, I. (2018). Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, 58, pp.22-32.
- [9] Ghosh, D., Ae Chun, S., Shafiq, B. and Adam, N. (2019). *Big Data-based Smart City Platform*.
- [10] McClendon, L. and Meghanathan, N. (2015). Using Machine Learning Algorithms to Analyze Crime Data. *Machine Learning and Applications: An International Journal*, 2(1), pp.1-12.

- [11] He, L., Páez, A. and Liu, D. (2016). Persistence of Crime Hot Spots: An Ordered Probit Analysis. *Geographical Analysis*, 49(1), pp.3-22.
- [12] Sangani, A., Sampat, C. and Pinjarkar, V. (2019). Crime Prediction and Analysis. *SSRN Electronic Journal*.
- [13] E-education.psu.edu. (2019). [online] Available at: https://www.e-education.psu.edu/geog885/sites/www.e-education.psu.edu/geog885/files/geog885q/file/Lesson_04/Predicting_Crime_groff.pdf [Accessed 30 Aug. 2019].
- [14] W. Osaku, "EDA of crime in Vancouver (2003-2017)," *Kaggle*, 2018. [Online] Available: <https://www.kaggle.com/wosaku/eda-of-crime-invancouver-2003-2017>.
- [15] J. Kerr, "Vancouver police go high tech to predict and prevent crime before it happens," *Vancouver Courier*, July 23, 2017. [Online] Available: <https://www.vancourier.com/news/vancouver-police-go-high-tech-to-predict-and-prevent-crime-before-it-happens-1.21295288>. [Accessed: 09- Aug- 2018]
- [16] R. Marchant, S. Haan, G. Clancey, and S. Cripps, "Applying machine learning to criminology: semiparametric spatialdemographic Bayesian regression," *Security Inform.*, vol. 7, no. 1, Dec. 2018.
- [17] N. Shama, A machine learning approach to predict crime using time and location data, B.Sc. Thesis, BRAC University, 2017.
- [18] Lin, Y., Chen, T. and Yu, L. (2017). Using Machine Learning to Assist Crime Prevention. *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*.
- [19] Catlett, C., Cesario, E., Talia, D. and Vinci, A. (2019). Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments.

APPENDICES

Appendix A: Field survey

Crime Rate Prediction and Spot Detection

Daffodil International University

Daffodil Tower

4/2, Sobhanbag, Mirpur Road, Dhanmondi, Dhaka-1207

*Required

Survey Description

We (Sakib Mahmud and Musfika Nuha) are from daffodil international university. We are currently conducting research on Crime Rate Prediction and Spot Detection System. We need your help. Please complete the following survey. You Can fill or tick mark to give your opinion.

1. Gender *

Tick all that apply.

- Male
- Female
- Others

2. Email

3. Age *

4. Present Address

5. Home Town

6. Did you ever face any unpleasant occurrence? *

Tick all that apply.

- Yes
 No

7. If you faced any unpleasant occurrence is it first time in your life? *

Tick all that apply.

- Yes
 No

8. What types of crime did you face? *

9. In which area have you been faced that occurrence? *

10. In which month you been faced that occurrence? *

11. In which time you spend most outside the home? *

Tick all that apply.

- Day
 Night

This content is neither created nor endorsed by Google.

Appendix B: Online survey link

<https://forms.gle/N3QQXshm7ji4fEnT8>

PLAGARISM REPORT

