Development of Machine Learning Models for Crime Prediction using Historical Data

Submitted by

Nakib Uddin Ahmad	**160104137**
Abu Ubaida Akash	**170104060**
Md Atique Ahmed Ziad	**170104087**
Faisal Tareque Shohan	**170104105**

Supervised by

Prof. Dr. Mohammad Shafiul Alam

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

January, 2022

CANDIDATE'S DECLARATION

We, hereby, declare that the thesis presented in this report is the outcome of the investigation performed by us under the supervision of Prof. Dr. Mohammad Shafiul Alam, Professor, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh. The work was spread over two final year courses, CSE4100: Project and Thesis I and CSE4250: Project and Thesis II, in accordance with the course curriculum of the Department for the Bachelor of Science in Computer Science and Engineering program.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Nakib Uddin Ahmad
160104137

Abu Ubaida Akash
170104060

Md Atique Ahmed Ziad
170104087

Faisal Tareque Shohan
170104105

CERTIFICATION

This thesis titled, **"Development of Machine Learning Models for Crime Prediction using Historical Data"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in January, 2022.

Group Members:

Nakib Uddin Ahmad	**160104137**
Abu Ubaida Akash	**170104060**
Md Atique Ahmed Ziad	**170104087**
Faisal Tareque Shohan	**170104105**

Prof. Dr. Mohammad Shafiul Alam
Supervisor
Professor & Head
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology

ACKNOWLEDGEMENT

All gratitude to the almighty ALLAH for his grace, which enabled us to complete our academic thesis in excellent health.At the very first, we would like to show our deepest gratitude to our supervisor Prof. Dr. Mohammad Shafiul Alam, Professor, Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, who offered continuing support and constant encouragement, which empowered us to complete our thesis. His guidance, comments, suggestions, constructive criticism, enthusiastic and steady oversight all through our work were invaluable. We also thank all the faculty members of the Computer Science and Engineering department, AUST for helping us in many ways whenever needed. Finally, we would also like to express our gratitude to our beloved parents and family for their moral support and warm encouragement.

Dhaka
January, 2022

Nakib Uddin Ahmad

Abu Ubaida Akash

Md Atique Ahmed Ziad

Faisal Tareque Shohan

ABSTRACT

A crime is an illegal act that is punishable under the law. Understanding crime is essential for preventing criminal activity. Due to poverty, overpopulation, political influence, a faulty education system, and other factors, Bangladesh has a high rate of crime. Despite the high crime rate, the scarcity of publicly available real-time crime records motivated us to create a dataset based on newspaper crime reports. Our dataset is built from the scratch consisting handpicked data from crime reports published in Bangladesh's most widely circulated English newspaper, The Daily Star [1]. One significant accomplishment of this thesis was the creation of a dataset containing over 6000 criminal records. However, after a great deal of effort, it was possible to collect approximately 6600 criminal records from across the country from 2019 to 2012. To the best of our knowledge, this is the first dataset in Bangladesh comprising spatio-temporal features related to crime. On those 6600 criminal records, the necessary preprocessing steps were taken. Following feature engineering, several versions of this dataset were prepared. Logistic Regression, Naive Bayes, KNN, and Decision Tree were tested as Supervised Classifiers. Ensamble learning was tested using Random Forest, Extra Tree, Ada Boost, and XGBoost. XGBoost yielded the highest accuracy of 41.50 %. Because the models are still in the early stages of development, no major tuning was performed, and no oversampling or undersampling was performed on this imbalanced dataset. However, after collecting more crime data, we intend to use better preprocessing and feature engineering. We believe that with more testing and inquiry, better results can be achieved.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Crimes are common social issues that affect a country's quality of life, economic development and credibility. Crimes are one of the key factors influencing the various important decisions of an individual's life, such as moving to a new location, roaming at the right time and right place, avoiding dangerous areas, etc. Crimes are influencing and defaming the reputation of a community. Crimes often influence the economy of a country by imposing a financial burden on the government due to the need for additional police forces, courts, etc. Because crime rate is increasing dramatically, society is at an alarming stage in order to reduce it at even faster rate. Crime prediction is a law enforcement technique that uses data and statistical analysis for the identification of crimes most likely to occur. This field has been subject to continued research in many parts of the world.

1.2 Motivation

Crime rates can be significantly reduced by real-time crime forecasting and mass surveillance, which are helpful in saving lives that is the most valuable thing. Proper analysis of previous crime data can help to predict crime and reduce the crime rate. Crimes can be predicted as the criminals are active and operate in their comfort zones [8]. Once successful they try to replicate the crime under similar circumstances. Criminals generally find similar location and time for attempting next crime. Although it may not be true for all the cases, but the possibility of repetitions is high, as per studies [9], and this makes the crimes predictable. As a densely populated country, lots of crimes are being committed in Bangladesh everyday. But there is no day to day crime dataset available to work on. As a result, it was decided to concentrate on this research gap in order to solve the problem.

1.3 Objective

Criminal activities take place all over the world and law enforcement agencies have to deal with them effectively and efficiently. If enforcement agencies have a prior assumption of the class of the crime, it would give them tactical advantages and help resolve cases faster. Also, an overall study of criminal activity in a geographic area helps to understand the underlying pattern of the crime the area suffers from. The purpose of this research is to find answers to the following questions:

- Is geographical location, weather, and time, basic details of the criminal activity have enough indicators to predict a type of crime?
- Given just a geographic location, time, and weather, how accurately can we classify the crime?
- Explore different techniques to improve the results.

1.4 Problem Description

The details of the crime were extracted from the newspaper articles, and the information was organized into the desired format. All crime news in newspapers was divided into six categories.

- Murder : Unlawful killing of another human without justification or valid excuse.
- Rape : Sexual assault usually involving sexual intercourse or other forms of sexual penetration carried out against a person without that person's consent.
- Assault : Act of inflicting physical harm or unwanted physical contact upon a person or, in some specific legal definitions, a threat or attempt to commit such an action.
- Robbery : Taking or attempting to take anything of value by force, threat of force, or by putting the victim in fear.
- Kidnapping : Unlawful transportation, asportation and confinement of a person against their will.
- Body Found : A dead corpse was discovered in an area where criminal information was lacking, with little information regarding the victim, the incident area, and the time.

Based on the location, time, and weather, this study attempts to predict the type of crime that is likely to occur.

1.5 Organization of the Thesis

This thesis is divided into chapters, with each chapter focusing on a different aspect of the research. The following are some of the topics covered by different chapters:

Chapter 2
A brief discussion about different papers related to our work is presented in this chapter.

Chapter 3
This chapter summarizes the lessons learned about crime, as well as Machine Learning Knowledge and its principles of operation.

Chapter 4
The steps taken for this research are briefly described in this Chapter.

Chapter 5
The various aspects of the dataset preparation process are covered in this chapter..

Chapter 6
Chapter 6 explains all of the steps involved in cleaning up the dataset.

Chapter 7
The process of feature addition, feature extraction, and encoding is described in this chapter.

Chapter 8
Results of proposed methods are presented. This chapter also shows a comparative analysis of different models.

Chapter 9
The conclusion of the overall research is presented in the final chapter. Besides, challenges and future plans of the work are mentioned.

Chapter 2

Literature Survey

Researchers have tried a variety of methods to accurately predict crime over the years. Many algorithms have been used in this domain, and many new techniques are still being tested. Many machine learning and deep learning approaches have recently been used by researchers as a result of the rise of AI, and they have already revolutionized crime prediction. Decision Trees, Random Forests, SVMs, Gradient Boosting, K-Nearest Neighbours, Naive Bayes, and AdaBoost are some of the most popular approaches. Previous research on Crime Prediction is discussed in this chapter.

2.1 Existing Works

This paper [2] uses a dataset from Chicago, which involves extraordinary classes of crimes occurring, based totally on several factors along with places and crimes over the 16 years. They applie dmachine learning procedures to find a criminal sample with the aid of its category with the given precise time and location.

This paper uses dataset which contains both the mixture of categorical and numeric values. Thus, the paper focuses on those algorithms which can work on the combination of both categorical and numeric values. Their selection of algorithm performs well for classification problem. Therefore, several algorithms they choose to serve the purpose such as

- Decision Tree
- Random Forest
- Bagging
- AdaBoost

- ExtraTree Classifier

Table 2.1: Result analysis of Predicting Crime Using Time and Location Data [2]

Algorithm	Accuracy
Random Forest	95.99
Decision Tree	99.88
AdaBoost	74.78
Bagging	99.92
Extra Tree	97.10

Highest accuracy is acquired with the implement of Bagging because ensemble method combines several tree classifiers and gives much better predictive results. The main reason of difference in actual and predicted values are noise, variance, and bias. Bagging gives an accuracy of 99.92 %, thus minimizing the maximum difference between actual and predicted values. To forecast the type of crime that could occur depending on time and place, this paper uses five different types of algorithms. The algorithm for trees shows that the estimated outcomes were far closer to the real outcomes. Thus, when applied with various tree classifiers, the dataset used produces the full accurate outcome with higher precision. The findings mentioned in this paper indicate that the system of bagging works well and that AdaBoost works least well for the prediction of crimes using time and place. When applied with tree-based algorithms, the outcomes in this paper have comparable effects. Therefore, as applied with other classifying algorithms in the future, this paper expects to obtain more variance in the results.

Dataset used in this paper [3] is acquired from Portland Police Bureau (PPB) and the public government source American FactFinder. Dataset contains features such as Category of crime, Call group, Final case type, Case description, Occurrence date, X and Y coordinate of the location of the crime and Census tract.

This paper explores several machine learning algorithms to build models to accomplish the classification task. The learning machines used include

- Support Vector Machine
- Random Forest
- Gradient Boosting Machines
- Multilayer Neural Networks

Table 2.2: Result analysis of Building a Learning Machine Classifier with Inadequate Data for Crime Prediction [3]

Algorithm	Accuracy
Random Forest	67.088
Support Vector Machine	67.095
Gradient Boosting Machines	76.42
Multilayer Neural Networks	50.2

This paper has started with preprocessing the dataset from Portland Police Bureau. Then, attempted to select some helpful features to represent the attributes of the samples in a proper manner. It divided the total forecast area into hotspots. Because of the large dataset and problem of imbalanced data, they employed under-sampling technique on its dataset to reduce the training set size to less than 20,000 samples and the accuracy of the model is affected by the accuracy of undersampling method.

Ensemble methods such as Random Forest or Gradient Boosting turned out to be the two best models when compared their performances with SVM. These two methods can handle with big training set and the training time is faster than the training time of SVM model.

In this paper [10], they designed and analyzed two area-specific models of crime prediction. They used hierarchical structures in the first approach and regularized multitask learning models in the second approach. Both methods solve the problem of sparsity by exchanging data throughout the learning process around smaller regions. They checked the hypothesis that their area-specific models could outperform global models using actual crime records from Chicago, Illinois, USA. Their experiments supported their assumption: Better prediction efficiency on 12 of 17 forms of crime is obtained by the area-specific models. This was the first recorded use and comparison of these approaches, so far they worked on it.

They extracted a variety of spatial characteristics (Major Streets, Police Stations, Hospitals, Bus stops, Parks, Water ways, Pedestrian ways) from the City of Chicago data portal. They also measured the spatial density of each density function. Finally, they segmented Chicago into its 61 current ZIP codes to construct region-specific models and grouped the crime data accordingly. They used two global models and three area-specific models:

- Global model
- Global Model with Area Indicator
- Pooled Model
- Hierarchical Model
- Multi-Task Model

The sparsity of crime in many areas complicates the application of area-specific predictive modeling. In this work, they developed and tested area-specific crime prediction models using hierarchical and multi-task learning. These approaches mitigate sparseness by sharing information across different areas, yet they retain the advantages of localized models in addressing non homogeneous crime patterns. They tested their models on crime data from the city of Chicago, Illinois, and they observed gains in many surveillance-feasible regions of the study area. In the future, they plan to investigate the use of area-grouping approaches to build spatial hierarchies. This would help in guiding the sharing of information across areas within the hierarchical and multi-task models. This would also highlight similar areas for uniform intervention, thus reducing resource expenditures.

The datasets used in this paper [4] were obtained from the open data catalog of the city of Vancouver. There are two datasets used for this project: crime and neighborhood. It provides information on the type of crime committed and the time and location of the offence.

As the objective of this paper is the classification of crime. There are many algorithms that can be used for the classification. This paper uses:

- K-Nearest Neighbour
- Boosted Decision Tree

Table 2.3: Result analysis of Crime Analysis Through Machine Learning [4]

Algorithm	Accuracy
K-Nearest Neighbour	39
Boosted Decision Tree	44

The accuracy, complexity, and training time of algorithms were slightly different for different approaches and algorithms. The prediction accuracy can be improved by tuning both the algorithm and the data for specific applications. Although this model has low accuracy as a prediction model, it provides a preliminary framework for further analysis.

The dataset used for the work [11] is reliable, real and authentic as data is acquired from the official site of the U.K. Police department. The data set contains a total of 11 attributes out of which 5 attributes were considered for the study, they are crime type, location, date, latitude, and longitude.In this phase, the history of crimes from the year 2015-17 was considered as the training dataset.

This paper used K-NN and Naïve Bayes classifier and presented the visualization techniques and classification algorithms that can be used for predicting the crimes and helps the law agencies. In future, there is a plan for applying other classification algorithms on the crime data and improving the accuracy in prediction.

This paper [5] used the Open data provided by San Francisco Police Department Crime Incident Reporting System. It provides information on crime incidents that occurred in San Francisco from 1/1/2003 to 5/13/2015. This dataset is a csv file which contains 8,78,049 rows. There are 39 types of crimes in the San Francisco Crime Dataset. They have observed that summer and winter have less criminal activities compared to other seasons. Most of the crimes occur on Friday, where least crimes occur on Sunday. Crime rates almost gradually increase from Monday to Thursday. They have used supervised classification method for their experiment.

- Decision Tree- The best accuracy is 31.17 % when log loss is 3.31, when the parameters are entropy and 300 split. The lowest accuracy is 28.26 % when log loss is 8.41. Gini and entropy improves log loss using higher split.
- K-Nearest Neighbor- With neighbor =50, they achieved best accuracy 28.50 % when log loss is 5.04. The lowest accuracy was 27.91 % when neighbors = 500, and log loss is 2.62.
- Random Forest- With 10 trees accuracy was 31.22 % and log loss 2.34, with 50 trees accuracy was 31.70% and log loss 2.28, with 100 trees accuracy was 31.71 % and log loss 2.28.

Oversampling and Undersampling

They used SMOTE oversampling technique. In SMOTE, k-nearest neighbor were used for generating synthetic minority classes by operation on feature space.The further used scikit learn package imblearn for undersampling the imbalance data. ENN and Random underampling are used.

Table 2.4: Result after Oversampling and Undersampling [5]

Sampling	Method	Models	Accuracy	Log loss
Over sampling	SMOTE	Random Forest	73.89	0.58
Under Sampling	Random Under Sampling	Random forest	99.16	.17

As the classes were poorly imbalanced, machine learning agent failed to perform well in the original dataset. From the original dataset, machine learning agents managed to provide a poor accuracy score of 31.71 % that is pretty low . So they divided the 39 classes into two classes. One is the frequent class and the other one is rare class. The frequent class consists of most frequent crimes, and the rare one consists of least frequent crimes. Machine learning agents performed well in remodeled dataset and resulted accuracy is 68.03 %. To overcome the imbalanced problem, they used oversampling and undersampling methods. Machine learning agents were highly beneficial after using these two methods. With a accuracy of

99.16 %, random forest performed the best decision making classifier than other machine learning agents.

This paper [6] uses Crime dataset available in kaggle.com. They converted categorical attributes (Location, Block, Crime Type, Community Area) into numeric using Label Encoder. The date attribute was splitted into new attributes like month and hour which were used as feature for the model. The attributes used for feature selection are Block, Location, District, Community area, X coordinate , Y coordinate, Latitude , Longitude, Hour and month. They used the following supervised machine learning models for crime prediction:

- K-nearest Neighbor
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Bernouli Naive bayes
- SVC
- Decision tree

Table 2.5: Result analysis of Crime Prediction and Analysis Using Machine Learning [6]

Algorithm	Accuracy
K-nearest Neighbor	78.73
Gaussian Naive Bayes	64.60
Multinomial Naive Bayes	45.62
Bernouli Naive Bayes	31.35
SVC	31.35
Decision tree	78.60

As we can see from the results obtained from the table the algorithm which can be used for the predictive modeling will be KNN algorithms with accuracy of 78.73 % highest among the rest of the algorithm.The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning they have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with accuracy of 78.73 %.

In [12] they prepared the dataset by collecting crime data from all states in India. Collected data contains factual reports on murder, rape and theft, which are considered as main elements. Basically, crime records from NCRB (National Crime Record Bureau) were taken which are considered as easily accessible. All violent crimes in crime data have been recorded in India over the last 15 years from 2001 to 2015.

Six algorithms was been chosen for their analysis named as J48, Naïve Bayes, SMO, and together stacking learning Bagging classifier and Random Forest. Based On efficiency, the Classify-based Stacking Ensemble is the best giving 99.5 % accuracy while with other classifiers only 95.55 % and 97.21 % accuracy is reported. Bagging and random forest takes longer for a large number of samples, while J48 requires less time to estimate.

2.2 Summary of Reviewed Literature

Table 2.6: Summary of Reviewed Papers

Title	Best Model	Best Performance
Predicting crime using time and location [2]	Bagging	99.92 %
Building a learning machine classifier with inadequate data for crime prediction [3]	Gradient Boosting	76.42 %
Crime analysis through machine learning [4]	Boosted Decision Tree	44 %
Crime prediction using spatio-temporal data [5]	Random Forest	99.16 %
Crime prediction and analysis using machine learning [6]	K-Nearest Neighbout	78.73 %
Crime prediction and monitoring framework based on spatial analysis [11]	K-Nearest Neighbour	Not specified
An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach [12]	Stacking	99.5 %

Chapter 3

Background Study on Crime and Machine Learning

For this research, studies on both technical and nontechnical requirements related to crime, machine learning techniques, and other things were conducted. This chapter briefly summarizes all of those lessons, as well as their working principle.

3.1 Crime

3.1.1 Overview

Since the dawn of society, crime has been a challenging issue.There is hardly any society which is not beset with the problem of crime. The word "**Crime**" comes from the Latin word "**krinos**", which means accusation. It covers acts that are against social order and deserve society's disapproval and condemnation. In a broad sense, crime refers to acts that violate the law, rules, or regulations, or that harm or destroy human society or resources, or that cause problems in daily life.

The Bangladesh Penal Code does not define the term "crime." However, in a broad sense, it can be defined as an act of commission or omission that is harmful to society as a whole.

According to Oxford Advanced Learners Dictionary: Crime is an activitie that involve breaking the law. [13]

According to Black's Law Dictionary: Crime is an act that the law makes punishable; the breach of a legal duty treated as the subject matter of a criminal proceeding. [14]

According to Dr. Nurul Islam: Crime is an act which the group regards as sufficiently menacing to its fundamental interests of justify formal reaction to restrain violator. [15]

3.1.2 Types of Crime

Crimes are often classified based on their severity, such as the distinction between felony and misdemeanor offenses.

According to Bangladesh Penal Code:

Various offenses have been statistically classified into seven broad categories under the Penal Code [16]. They are:

- Offences against Person
- Offences against Property
- Offences relating to Documents
- Offences affecting Mental Order
- Offences against Public Tranquility
- Offences against State
- Offences relating to Public Servants

From the standpoint of criminal law and penal justice administration, this classification appears to be more rational and detailed.

According to Bangladesh Police:

Crimes are mainly two types [17]. They are:

i **Heinous crime**

- Robbery
- Killing
- Rape
- Attempt to murder
- Burglary
- Acid throwing
- Stealing of livestock
- Serious injury
- Preservation of illegal arms
- Sexual crime

ii **Non heinous crime**

- Stealing
- Terrorism
- Flattery
- Breaking of drug related law
- Gambling
- Violation of traffic rules
- Disorganized behave
- Breaking of rule 144

3.1.3 Causes of Crime

The term "crime" can be defined in a variety of ways. Different societies may use different definitions of crime. In general, however, crime can be defined as the violation of laws put into effect by the ruling authority of the land. There are numerous causes of crime, and multiple studies are being conducted all over the world to better understand and reduce criminal activity. Governments and law enforcement agencies all over the world are constantly working to reduce crime rates so that the world becomes a safer place to live.The fight against crime is not a new one in human history, and it has been attempted since the establishment of society. Here's some of the factors that contribute to crime.

i **Victim of unfair rulings and the correction system**:
People who have been the victims of unfair or incorrect court rulings are frequently drawn into a life of crime. It is not unusual for a person to be a victim of chance and fall victim to crimes. Aside from that, people are frequently falsely accused of crimes, which results in a court conviction. Because of the conditions in jails and prisons, people frequently become worse criminals. Criminals are rarely rehabilitated in correctional facilities, and they are frequently thrown in overcrowded jails with people who are either victims or perpetrators of crimes far more serious than their own. Declassification of inmates in prisons is another major source of crime.

ii **Depression and other social and mental disorders:**
Depression is also a major contributor to criminal behavior. Apart from depression, people with severe mental illnesses are more likely to commit crimes. Such individuals should be treated before their afflictions and tendencies become out of hand. A person suffering from depression or another serious mental illness can easily harm themselves.

iii **Family conditions:**
There are a lot of things that go on in families that often cause people to get into a life of crime. Abuse from family members during a person's formative years, as well as other similar acts, can lead to a criminal career. People who are neglected by their families and do not receive the love and attention they desire are more likely to engage in criminal activity. In many ways, family violence and other issues are linked to crime.

iv **Regionalism:**
Regionalism is a major source of crime and social unrest. People with strong regionalist feelings frequently go to great lengths to perpetrate crimes against other communities. This fact is frequently overlooked by individuals and administrative bodies, who are also engrossed in regional classifications. A victim of such regionalism is frequently influenced and drawn into the criminal world.

v **Racism:**
Discrimination against people of color is a serious problem all over the world. Every human being is racist in some way toward some people in some part of the world.Racism has caused a great deal of unrest in many parts of the world, and such crimes are usually the result of the actions of one or two idiots. It is a sad fact of life that we end up discriminating against something that is the same flesh and blood beneath the surface, even if the external appearance and origins are different.

vi **Politics:**
Politics is often a source of criminal activity. Many political organizations around the world are known to have their own mafias, which they use to manipulate and subjugate people. Political power is frequently abused to exploit weaker groups and individuals, and the ensuing discontent often forces victims to resort to criminal activity. Politics is more closely linked to large-scale and heinous crime than anything else.

vii **Poverty:**
Poverty, or economic deprivation, is a major cause of crime all over the world. Poverty often drives people to extremes of desperation, and it is a major cause of crime all over the world. The fact that such dissatisfaction exists is in and of itself a very dangerous thing for society as a whole, given the fact that global inflation has risen dramatically in recent years. Although it appears that the rich are getting richer and the poor are getting poorer in today's world.

viii **Overpopulation:**
Increased population is the leading cause of crime and the source of many of the world's concerns. Despite the fact that population growth is linked to each of the causes listed here, it must still be considered a source of crime. The increase in population causes a

dynamo effect in society, resulting in the creation of more people who are frustrated or resentful of society as a whole.

ix **Parental Relations:**

Children who have been neglected or abused are more likely than others to commit crimes later in life. Similarly, sexual abuse as a child frequently leads to victims becoming adult sexual predators. Many of the inmates on death row have a history of severe abuse. Children's neglect and abuse are often passed down through generations. Abuse, crime, and sociopaths are all part of a cycle that keeps repeating itself. Children who have been neglected or abused commit significantly more crimes later in life than children who have not been neglected or abused.

x **Education:**

A survey of inmates in state prisons in the late 1990s revealed very low educational levels, which matched Merton's earlier sociological theories. Many couldn't read or write beyond the first grade level, if at all. Robbery, burglary, automobile theft, drug trafficking, and shoplifting were the most common crimes committed by these inmates. Their employment histories were mostly low-wage jobs with frequent periods of unemployment due to their poor educational backgrounds.

Criminal activity is not deterred by employment at minimum wage or below the living wage. Even with government social services like public housing, food stamps, and medical care, a minimum wage household's income is still insufficient to meet basic needs. People must choose between long-term low income and the possibility of lucrative crime. Of course, continuing your education is an option, but classes can be costly and time consuming. While education can help people to get a better job, it can't always compensate for the effects of abuse, poverty, or other limiting factors.

xi **Peer influence:**

A person's peer group has a significant influence on his or her decision to commit a crime. For example, young boys and girls who do not meet expected academic achievement standards or participate in sports or social programs may become Crack cocaine pipes displayed by police. Drugs and alcohol impair judgment and lower inhibitions, making a person more willing to commit a crime. Children from low-income families who cannot afford adequate clothing or school supplies are at risk of falling into the same trap. Researchers believe that these youth may abandon schoolmates in favor of criminal gangs because gang membership earns respect and status in a different way. Antisocial and criminal behavior earns respect and street credibility in gangs.

xii **Drugs and alcohol:**

Some social factors have a particularly strong influence on a person's ability to make decisions. One such factor is drug and alcohol abuse. The desire to commit a crime in order to support a drug habit has a significant impact on the decision-making process. Drugs and alcohol both impair judgment and lower inhibitions (socially defined rules of behavior), giving a person more courage to commit a crime. Long prison sentences, for example, have little meaning when a person is high or drunk.

Criminologists estimate that the attacker's use of alcohol or drugs is responsible for 30 to 50 percent of violent crime, such as murder, sexual assault, and robbery. Furthermore, drugs or alcohol may make the victim a more vulnerable target for a criminal by making the victim less attentive to activities around them and possibly visiting a poorly lit or secluded area not normally frequented to purchase drugs.

3.1.4 Criminal Activities in Bangladesh

Previously, crime was very simple; however, crime is becoming more complex. On the other hand, some types of crime, such as sexual assault, murder, rape are evidenced by the statistics below. Source Dhaka Metropolitan Police Website. [7]

Table 3.1: Comparative Crime statistics [7]

Year	2011	2012	2013	2014	2015	2016	2017	2018
Dacoity	650	593	613	651	492	408	336	262
Robbery	1069	964	1021	1155	933	722	657	562
Murder	3966	4114	4393	4514	4037	3591	3549	3830
Speedy Trial	1863	1907	1896	1716	1549	1052	1045	922
Riot	109	94	172	79	96	53	23	26
Woman and Child Repression	21389	20947	19601	21291	21210	18446	17073	16253
Kidnapping	792	850	879	920	805	639	509	444
Burglary	3134	2927	2762	2809	2495	2213	2163	2137
Theft	8873	8598	7882	7660	6821	6110	5833	5561
Total	41845	40994	39219	40795	38435	33234	31188	29997

Some aspects of crime that occurred in Bangladesh are discussed below.

Wife burn to death by husbands: [18]
A housewife, set on fire allegedly by her husband in Majhipara area of the district town.The deceased is Sheuly Akter, 27, wife of Sujon Mia, 35, of the village and mother of two children. Shiblu Mia, brother of the deceased said, Sujon used to torture Sheully since their marriage several years ago over family trifling matters. As a sequel of their feud, both were engaged in an altercation and at one stage Sujon beat up Sheuly mercilessly and poured kerosene on her body and set her on fire.

Teenage girl gang-raped in Uttara: [19]
A teenage working girl was reportedly gang-raped by some men, one of whom was her colleague, in the capital's Uttara area. Her brother-in-law said she left work around 8:30pm on Thursday, and her co-worker was accompanying her on her way home when all of a sudden two men joined them and the three men forcibly took her to an under-construction building nearby. There they took turns at raping her until around 11pm, when they let her go home.

The brutal killing of Rajon: [20]
On July 8, 2015 13 year old Samiul Alam Rajon was beaten to death on suspicion of stealing a rickshaw van. Rajon was forced to leave school and sell vegetables due to poverty. The perpetrators also videoed the incident which was later circulated on the internet. It was learnt that the rickshaw van owner, Muhit Alam and his brother Kamrul Islam took Rajon to Kumargaon bus stand and beat him for three hours. Rajon died of the injuries he sustained. After the death of Rajon, they tried to hide the child's body. Rajon died of brain haemorrhage and around 64 injury marks were found in his body.

Killing of blogger Avijit Roy: [21]
Blogger Avijit Roy and his wife Rafida Ahmed Bonnya were attacked by two criminals while they were waiting for tea at a road side stall in front of Suhrawardy Uddan adjacent to TSC in the Dhaka University campus, after coming out of the Ekushey Book Fair Two armed criminals attacked them in the presence of police and fled the scene after stabbing them indiscriminately with sharp weapons. Despite the tight security presence the criminals were able to Attack the couple and escape.

Publisher Faisal Arefin Deepan killing: [22]
A publisher of secular books has been hacked to death in the Bangladeshi capital, police have said. In a separate incident in Dhaka, two other writers and a publisher were stabbed and shot at a publishing house, according to police. Fears of Islamist violence have grown in Bangladesh this year, following the assassination of at least four atheist bloggers. Police

have linked the attacks to domestic Islamist extremists, while the Islamic State has claimed responsibility for three others. Inside his office, the body of Faisal Abedin Deepan of the Jagriti Prokashoni publishing house was discovered.

Oishee killed her parents: [23]
Oishee Rahman had murdered her parents Special Branch Inspector Mahfuzur Rahman and his wife Swapna Begum. Their bodies bore multiple stab wounds, which forensic experts believed, were the work of amateurs. Police believed Oishee and her 'drug addict' friends had committed the murder.

Murder of Sagar Sarowar and Meherun Runi: [24]
The married Bangladeshi couple, Sagar and Runi, lived with their 5-year-old child on the fourth floor of a five-story building in the West Raja Bazar neighbourhood (mahallah) of Dhaka. Neighbors say that Sarowar and Runi had more than one person in the apartment as guests before they were killed. From information gathered from a security guard, police believe the couple was killed some time after Sarowar arrived home and before the dawn Fajr prayer, which occurs before sunrise. Each victim died of multiple stabbing wounds, and sources said Sarowar's limbs were tied and he had the most stab wounds. Their five-year-old son woke up at around 7 a.m. and discovered his parents dead in a pool of blood and called Runi's mother sometime around 7:30 a.m. by a cell phone.

Murder of Abrar Fahad: [25]
Abrar Fahad was a second year student of electrical and electronic engineering (EEE) department of the Bangladesh University of Engineering and Technology (BUET). On 6 October 2019, Sunday night, he was tortured and killed by BUET's Chhatra League leaders inside BUET's Sher-e-Bangla Hall. Police had recovered Abrar's body at 3 in the early hours of Monday at the ground floor of Sher-e-Bangla Hall. Abrar was pronounced dead around 3 am by BUET Medical Officer Dr Md Mashuk Elahi. An autopsy stated that Fahad was beaten to death by a blunt object. Footage captured by a closed-circuit camera installed on the residential hall's 2nd floor showed a few people dragging Abrar down the corridor by his hands and feet. There is speculation that he was killed over his views that condemned the recent India-Bangladesh deal (September–October 2019).

3.1.5 Impact of Crime

Crime has a significant impact on people's lives in every way. It has a negative impact on people's overall life.

- **Social impact of crime:**
  In every society, crime is a significant factor. Its costs and consequences have an impact on almost everyone. There are many different kinds of costs and effects. Furthermore, some expenses are temporary, whereas others are permanent. The ultimate price, of course, is human life. Medical bills, property losses, and lost wages are all possible expenses for victims. To avoid being a victim, a large sum of money is spent. A victim or a person afraid of crime moving to a new neighborhood. Some of the costs of crime are intangible (not easily or precisely identified). Pain and suffering, as well as a lower quality of life, are examples of such costs. There are also the traumatic effects on friends and family members. Crime can permanently alter and shape behavior, whether it's weighing the risks of visiting certain locations or the fear of making new friends.

- **Economic impact of crime:**
  Crime not only has a negative impact on economic productivity when victims miss work, but it also has a negative impact on communities due to lost tourism and retail sales. Even the ostensibly victimless crimes of prostitution, drug abuse, and gambling have significant social ramifications. Drug abuse reduces worker productivity, consumes public funds for drug treatment and medical care, and leads to criminal activity to fund a drug habit's expenses. Police departments, prisons and jails, courts, and treatment programs all use public funds, as do the salaries of prosecutors, judges, public defenders, social workers, security guards, and probation officers. The amount of time spent in court by victims, offenders, their families, and juries also reduces community productivity.

- **Psychological impact of crime:**
  Crime and violence have a psychological impact on people, whether they are directly exposed, such as when it involves ourselves, a family member, or a friend, or indirectly exposed, such as through our residence in the community/society or media coverage. It is natural for us to have strong feelings and effects after a family member or friend is killed or injured, or after being indirectly exposed to crime. Stress, anxiety, fear, and shock are all natural psychological reactions. Our sense of safety has also been shattered; as a result, we feel unsafe, insecure, vulnerable, helpless, and powerless, as well as anger and outrage. People may also have nightmares and flashbacks, relive the incident over and over, have bad dreams and difficulty sleeping, feel tense, startle easily, feel numb or hyper-vigilant, have memory blocks about the incident, lose

interest in activities, avoid places or things that remind us of the incident, and have angry outbursts. Withdrawal, disassociation, amnesia, and depression are all possible reactions. Our ability to eat, sleep, think, and concentrate is affected.

- **Political instability:**
  Criminal activities create an unstable situation in our country, hampered our economic development, and people are unable to exercise their basic human rights, such as throwing petrol boma on a vehicle, as a result of which many people are vulnerable. These criminal activities are not tolerated by society and are legally prohibited.

- **Familial disorder:**
  In Bangladesh's rural and urban societies, crime causes family strife. If one of them engages in criminal activity such as petty theft, rape, murder, or other similar crimes. As a result, the family's situation becomes chaotic.

3.1.6 Preventative Measures

- **Necessary economic measures:**
  In a developing country like Bangladesh, poverty and economic problems are one of the leading causes of crime. The implementation of some economic measures is critical for establishing an effective crime prevention strategy.

  i **Ensuring adequate job opportunities:**
  The root of all problems is unemployment, which requires special attention from both the government and private stakeholders. In most cases, unemployment leads to poverty, which leads to crime. By ensuring adequate job opportunities, a large portion of the unemployed can be re-employed, bringing stability to their lives and preventing a significant portion of crime before it occurs.

  ii **Enriching the standard of living:**
  Emphasis is required to improve the bare minimum of living conditions. Facilities should be managed by the government and other service providers to ensure a minimum standard of living.

  iii **Perfect economic framework:**
  It may appear that a particular economic system is favourable to criminals and that they are protected by that economic system. It is sometimes argued that the capitalist economic system is well equipped for crime prevention because it inspires the class system. Because the rich are becoming richer and the poor are becoming poorer in this system, a specific group is being exploited and, as a result, tending to commit crime in order to meet basic needs. As a result, the economic system must be determined in accordance with assets, cultures, human

beliefs, religion, and educational proportionality in order for every citizen to be treated equally and proportionally to what they deserve.

- **Necessary Social Measures:**
  Humans are residents of society, and their human psychology and behavior are nourished by it. Crime is a common occurrence in societies where there is inequality, injustice, and a lack of a proper social structure. As a result, some social measures must be implemented in order to develop an effective and long-term crime prevention strategy.

  i **Population control:**
  In some ways, a country's population is an asset. However, in the absence of other accessories to support the territorial limitation, it became a liability rather than an asset. Population control is required. Otherwise, meeting the other requirements will not yield the desired results. Maximum population is much more important in bringing about positive change in society as well as developing an effective crime prevention strategy.

  ii **Standard childhood:**
  Childhood is the most appropriate period for forming a human being's personality, and various research studies have shown that a person's crime-prone mentality is the result of their childhood experiences. As a result, it is our unavoidable responsibility to provide every child with a productive and positive childhood. The burden falls primarily on parents, whose assistance has a significant influence on their children. Developing a positive anti-crime mindset from the start will be extremely beneficial in terms of avoiding criminal activity in general. These lead to a movement of not participating in criminal activities while keeping moral considerations in mind.

  iii **Role of Educational institution:**
  Educational institutions are regarded as one of the most sacred places in the world, where basic human characteristics are taught to students in order to mold them into proper human beings. It is the responsibility of teachers to nourish their students with the utmost care and to spread the anti-crime mentality in order to bring peace to humanity. Teachers are students' idols, and their speeches are always followed by their peers. So, in educational institutions, teachers can implement some basic crime prevention measures to make students aware of the importance of crime prevention.

  iv **Role of Religious institution:**
  Religion and religious institutions are the most sacred places in the world, and the reflections spread by religious institutions are followed by their adherents. So, in order to make the world a more peaceful place, which is also a basic goal of

religion, religious institutions and religiously wise people can play an important role in this regard. It has the potential to make a significant difference in people's lives by educating them to be virtuous, principled, and anti-criminal.

- **Necessary legal measures:**
  The law is used to control criminals and criminal behavior. In a nutshell, legal measures are the most important means of preventing criminals. Legal administration also oversees our judicial and prison systems. In one sense, it is legal measures that can directly prevent crime and serve as a deterrent to criminals by imposing severe punishments. In another sense, crime can be greatly reduced by implementing some pre-cautionary measures mandated by law.

  i **Activism of law enforcing agencies:**
  The primary responsibilities for maintaining law and order are delegated to law enforcement agencies, the majority of whose activism focuses on crime prevention. All rules and regulations pertaining to crime prevention are vested in law enforcement agencies, and they play a critical role in crime prevention. If there is a problem in the law enforcement agencies, the entire strategy fails. So, if we want to live in a crime-free, nonviolent society, we must pay special attention to law enforcement agencies.

  ii **Effective and Speedy Court System:**
  By carrying out their duties in accordance with the rules and regulations, an effective and efficient judicial system can boost the crime prevention movement throughout the country. It can play a significant role in crime prevention by exercising proper judgment, not being biased toward any particular party, and not being corrupt.

  iii **Proper laws to be framed:**
  Proper laws should be legislated from the parliament for effective and sustainable crime prevention and community safety policy, leaving partisan views to criminals. Taking into account the need for crime prevention, legislators should be aware of the importance of enacting appropriate legislation.

  iv **Implementation of existing laws:**
  Only enacting laws is insufficient to prevent crime from occurring; rather, proper implementation is far more important than the legislative process. Law enforcement agencies should be given sufficient authority, with checks and balances, to carry out such laws. For example, in current Bangladesh, there is a law that prohibits smoking in public places, but it is not being enforced. We even see members of law enforcement agencies smoking in public places. To frame an actual crime prevention strategy and make the community safe, proper laws with

proper implementation are required.

3.2 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data-sets; such as examples, direct experience, or instruction; in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning algorithms are often categorized as the followings:

3.2.1 Supervised Learning

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training data-set, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly. Our work mainly focuses on supervised learning. It is a more popular approach.

3.2.2 Unsupervised Learning

Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system does not figure out the right output, but it explores the data and can draw inferences from data sets to describe hidden structures from unlabeled data.

3.2.3 Semi-supervised Learning

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning since they use both labeled and unlabeled data for training typically a small amount of labeled and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it or learn from it.

3.2.4 Reinforcement Learning

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. The key idea of reinforcement learning is based around a computer or robot called an “agent” that can observe the “environment”. The agent performs actions on the environment and then the environment will provide some sort of feedback in return. It’s up to the model to figure out how to perform the task to maximize the reward, starting from totally random trials and finishing with sophisticated tactics and superhuman skills. By leveraging the power of search and many trials, reinforcement learning is currently the most effective way to hint machine’s creativity. DeepMind’s AlphaGo is also a good example of Reinforcement Learning which made headline after beating world champion Ke Jie at the game of Go in 2017.

3.2.5 Classification and Regression

Regression and classification are categorized under the same umbrella of supervised machine learning. Both share the same concept of utilizing known datasets (referred to as training datasets) to make predictions.

In supervised learning, an algorithm is employed to learn the mapping function from the input variable *(x)* to the output variable *(y)*; that is $y = f(X)$. The objective of such a problem is to approximate the mapping function *(f)* as accurately as possible such that whenever there is new input data *(x)*, the output variable *(y)* for the dataset can be predicted.

Classification algorithms attempt to estimate the mapping function *(f)* from the input variables *(x)* to discrete or categorical output variables *(y)*. In this case, y is a category that the mapping function predicts. If provided with a single or several input variables, a classification model will attempt to predict the value of a single or several conclusions. Support vector machine, decision tree, KNN, random forest, logistic regression are some widely used classification algorithms.

On the other hand, regression algorithms attempt to estimate the mapping function *(f)* from the input variables *(x)* to continuous or numerical output variables *(y)*. Here, *(y)* is the output variable that can be an integer or floating-point. Regression is generally used in time series modeling, forecasting, or finding relationships between the variables. Common regression algorithms include linear regression, polynomial regression, ridge regression, lasso regression, etc.

3.3 Traditional Machine Learning Algorithims

Traditional machine learning algorithms learn from data and the choice of the feature is made by subject matter experts. This type of algorithm is being used to solve classification, regression, clustering, and dimensionality reduction problems. Some traditional machine learning algorithms include Decision tree, support vector machine, logistic regression, K-Nearest Neighbours, etc. In this sub-section, some widely used traditional machine learning algorithms are discussed.

3.3.1 Supervised

Logistic Regression

Logistic regression is a supervised learning algorithm, mainly used when there is a categorical outcome. Logistic regression is commonly used to estimate the probability of the target variable. If the estimated probability is greater than 50%, the model predicts that the instance belongs to that class, else it does not. This idea makes logistic regression a binary classifier. In other words, the target variable is binary in nature having data coded as either 1 (success/yes) or 0 (failure/no). Logistic regression is used in many classification problems such as spam detection, tumor detection, etc.

In logistic regression, sigmoid activation function is used to convert predicted probability into a categorical value. It is a mathematical function that takes any real number and maps it between 0 to 1 and represented by the letter “S”. The sigmoid function is also called a logistic function which can be defined by Figure 3.1 and the following equation:

$$\sigma(t) = \frac{1}{1 + e^{-t}} \tag{3.1}$$

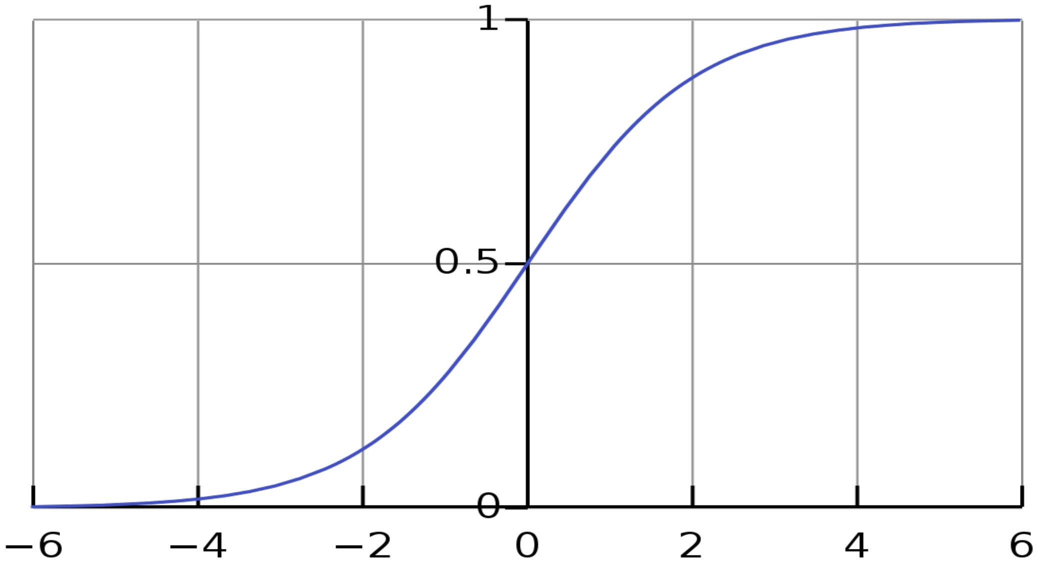


Figure 3.1: Sigmoid Function

Types of Logistic Regression There are three types of logistic regression. They are as follows:

i. Binary or Binomial: In binary logistic regression, the target variable will have only two possible outcomes either 1 and 0. So, the target variable will be dichotomous in nature. For instance, these variables can represent success or failure, yes or no, win or loss etc.

ii. Multinomial: In multinomial logistic regression, the target variable can have 3 or more possible types that are not ordered (i.e. types having no quantitative significance). For example, target variables may represent “class A” or “class B” or “class C”.

iii. Ordinal: In ordinal logistic regression, the target variable can have 3 or more possible types that are ordered (i.e. types having quantitative significance). For example, target variables may represent “poor” or “moderate”, “good”, “excellent”. Here, each category can be given scores like 0,1,2,3.

Logistic regression is one of the simplest machine learning algorithms which is also easier to train and interpret. Also, it does not make any presumptions about distributions of classes in feature space. Logistic regression can be extended to multi-class classification and gives a probabilistic view of the class predictions. Logistic regression generally provides good accuracy for simpler data and data that can be linearly separable.

Decision Tree

A decision tree is a flowchart-like tree structure where the internal node represents a feature, the branch represents the decision rule, and each leaf node represents the outcome. The top node in a decision tree is known as the root. It figures out how to partition the tree based on the attribute value. It partitions the tree in a recursive manner called recursive dividing. This flowchart-like structure helps to make decisions, which effectively impersonates human level reasoning. That is the reason decision trees are straightforward and decipher.

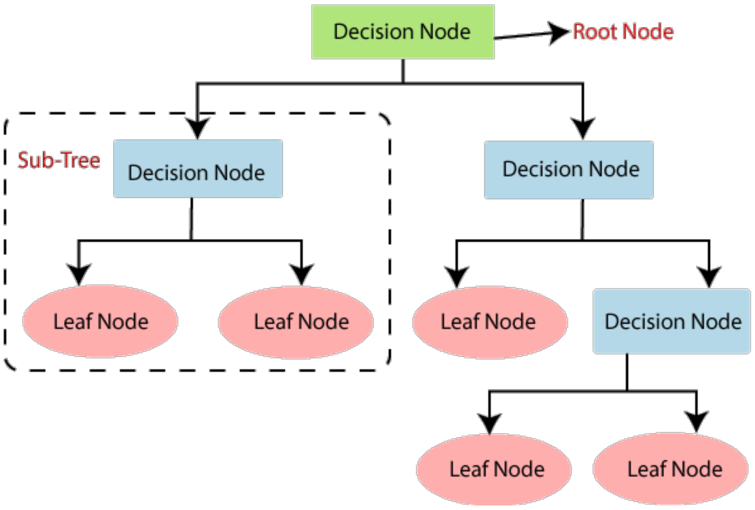


Figure 3.2: General structure of a decision tree

Decision Tree is a white box type machine learning algorithm. It shares internal decision logic, which isn't accessible in the black box type algorithm, for example, neural network. It is faster in terms of training time. The decision tree is a non-parametric method, which does not rely on probability distribution presumptions. Also, decision trees can deal with high dimensional information with great precision.
The fundamental idea behind the decision tree algorithm is as per the following:

i Select the best attribute using Attribute Selection Measures(ASM) to split the records.

ii Make that attribute a decision node and partition the dataset into smaller subsets.

iii Starts building the tree by repeating the following process recursively for each child node until any of the condition will match:

  - All the tuples belong to the same attribute value.
  - There are no more remaining attributes.
  - There are no more instances.

There are two main types of decision tree:

i **Classification Tree:** When the outcome of the decision variable is discrete or categorical, it is considered a classification tree.

ii **Regression Tree:** When the target variable of the decision tree is continuous, it is called a regression tree.

Attribute Selection Measures:
Attribute selection measure is a heuristic for choosing the splitting criterion that partitions data into the most ideal way. It is also called splitting rules since it helps to decide breakpoints for tuples on a given node. ASM gives a rank to each feature by explaining the dataset. Attribute with the best score will be chosen for splitting attribute. For attributes with continuous value, split points for branches need to define. Some well-known attribute selection measures are discussed below:

- **Information Gain:** Entropy measures the impurity of the input set. It alludes to the irregularity or randomness in the system. Information gain is a decrease in entropy. It computes the difference between entropy before the split and average entropy after splitting the dataset based on attribute values. ID3 (Iterative Dichotomiser) decision tree algorithm utilizes information gain. Mathematically, information gain can be written as:

$$Information\ Gain = Entropy(before) - \sum_{j=1}^{K} Entropy(i, ,after) \tag{3.2}$$

- **Gain Ratio:** C4.5, an improvement of ID3, uses an extension to information gain known as the gain ratio. Attribute with the highest gain ratio is chosen as the splitting attribute. Gain ratio handles the issue of bias by normalizing the information gain using Split Info. Gain ratio can be defined as:

$$GainRatio(A) = \frac{Gain(A)}{Splitinfo_A(D)} \tag{3.3}$$

- **Gini Index:** Gini method is used by the CART (classification and regression tree) algorithm to create split points. It is determined by taking away the amount of the squared probabilities of each class from one. It favors bigger partitions and is simple to execute though information gain favors smaller partitions with distinct values. The attribute with the minimum Gini index is chosen as the splitting attribute. Gini Index is defined as follows:

$$Gini = 1 - \sum_{i=1}^{C} (P_t)^2 \tag{3.4}$$

Decision tree is simple to understand and visualize. It requires fewer data pre-processing and no normalization is needed. It can also capture nonlinear patterns. Decision tree can also be used for feature selection. Because of the non-parametric nature of the algorithm, it does not require any assumption about the knowledge.

SVM

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data, the algorithm outputs an optimal hyperplane that categorizes new examples. In two-dimensional space, this hyperplane is a line dividing a plane in two parts where each class lay on either side. Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features.

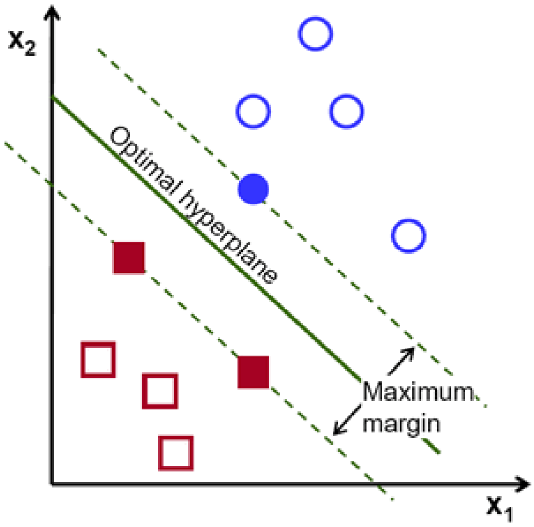


Figure 3.3: Classification by Finding Hyperplane

The classical SVM system requires that the dataset is linearly separable, i.e. there is a single hyperplane which can separate the two classes. For the data which can be separated linearly, we select two parallel hyperplanes that separate the two classes of data, so that distance between both the lines is maximum. The region b/w these two hyperplane is known as margin and maximum margin hyperplane is the one that lies in the middle of them.

$$\begin{cases} \overrightarrow{w} x_i - b \geq 1, & \Theta_i = 1 \\ \overrightarrow{w} x_i - b \leq 1, & \Theta_i = -1 \end{cases} \tag{3.5}$$

For non-linear datasets, a kernel function is used to map the data to a higher dimensional space in which it is linearly separable. In this higher dimensional feature space, the classical SVM system can then be used to construct a hyperplane. To deal with datasets with more than two classes, usually the dataset is reduced to a binary class dataset with which the SVM can work. There are two approaches for decomposing a multiclass classification problem to a binary classification problem: the one-vs-all and one-vs-one approach. There are many kernels that have been developed. Some standard kernels are:

- **Linear kernel:** The linear kernel function can be represented by the above expression. Where *k(x_i, x_j)* is a kernel function, x_i and x_j are vectors of feature space and d is the degree of function.

$$k(\overrightarrow{x_i}, \overrightarrow{x_j}) = (\overrightarrow{x_i}.\overrightarrow{x_j}) \tag{3.6}$$

- **Ploynomial kernel:** In the polynomial kernel, a constant term is also added. The constant term 'c' is also known as a free parameter. It influences the combination of features. x and y are vectors of feature space.

$$k(x, y) = (x^T y + c)^d \tag{3.7}$$

- **Radial Basis Function kernel:**: It is also known as RBF kernel. It is one of the most popular kernels. For distance metric squared euclidean distance is used here. It is used to draw completely non-linear hyperplanes.

$$k(x, x') = exp(-\frac{||x - x'||^2}{2\sigma^2}) \tag{3.8}$$

  where x are vectors of feature space. σ is a free parameter. The selection of parameters is a critical choice. Using a typical value of the parameter can lead to overfitting the data.

The main advantage of SVM is that it works effectively even if the number of features are greater than the number of samples. Also, non-linear data can also be classified using customized hyperplanes built by using kernel trick. Overall, it is a robust model to solveprediction problems since it maximizes margin.

Naive Bayes

This is a classification technique based on Bayes theorem with an assumption of independence between predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Bayes theorem is stated as:

$$P(y|X) = \frac{P(x|y)P(Y)}{P(X)} \tag{3.9}$$

here,

- P(y|X) is the probability of hypothesis y given the data X. This is called the posterior.
- P(X|y) is the probability of data X given that the hypothesis y was true. This is called the likelihood.
- P(y) is the probability of hypothesis y being true. This is called the prior probability of y.
- P(X) is the probability of the data

Naive Bayes is a classification algorithm for binary and multi-class classification problems.The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes or idiot Bayes because the calculation of the probabilities for each hypothesis are simplified to make their calculation tractable. There are some popular Naive Bayes classifiers. They are-

i **Multinomial Naive Bayes Classifier:**
In the multinomial document model, the document feature vectors capture the frequency of words, not just their presence or absence. Feature vectors represent the frequencies with which certain events have been generated by a multinomial distribution. This is the event model typically used for document classification. The likelihood of observing a histogram x is given by -

$$P(x|C_k) = \frac{(\prod_i x_i)!}{\prod_i x_i} \prod_i \mathrm{P}_{k^i}^{x^i} \tag{3.10}$$

Multinomial Naive Bayes classification algorithm tends to be a baseline solution for text classification tasks. Usually, Multinomial Naive Bayes is used when the multiple occurrences of the words matter a lot in the classification problem. Such an example is when we try to perform Topic Classification. The Binarized Multinomial Naive Bayes is used when the frequencies of the words do not play a key role in classification. Such an example is Sentiment Analysis, where it does not really matter how many times someone mentions the word "bad" but rather only the fact that he does. The algorithm for text classification is as follows.

ii **Bernoulli Naive Bayes Classifier:**
In the Bernoulli model a document is represented by a binary vector, which represents a point in the space of words. In the multivariate Bernoulli event model, features are independent booleans (binary variables) describing inputs. Like the multinomial model, this model is popular for document classification tasks, where binary term occurrence

(i.e. a word occurs in a document or not) features are used rather than term frequencies (i.e. frequency of a word in the document). If xi is a boolean expression the occurrence or absence of the i'th term from the vocabulary, then the likelihood of a document given a class Ck is given by,

$$P(x|C_k) = \prod_{i=1}^{n} \mathrm{P}_{k_i}^{x_i}(1-\mathrm{P}_{\mathrm{k_i}})^{(1-x_i)} \tag{3.11}$$

iii **Gaussian Naive Bayes Classifier:**

In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution. A Gaussian distribution is also called Normal distribution. When plotted, it gives a bell-shaped curve which is symmetric about the mean of the feature values as shown below: The likelihood of the features is assumed to be Gaussian, hence, conditional probability is given by:

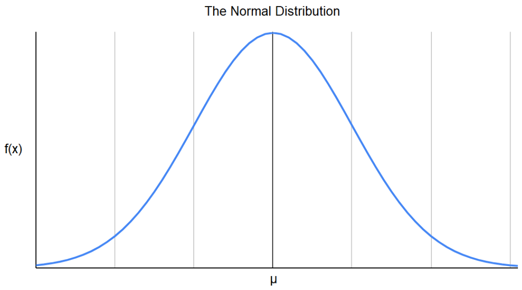


Figure 3.4: Gaussian Distribution Curve

$$P(x|Y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp(-\frac{(x_i-\mu_i)^2}{2\sigma_y^2}) \tag{3.12}$$

One of the major advantages that Naive Bayes has over other classification algorithms is its ability to handle an extremely large number of features. In our case, each word is treated as a feature and there are thousands of different words. Also, it performs well even with the presence of irrelevant features and is relatively unaffected by them. The other major advantage it has is its relative simplicity. Naive Bayes works well right out of the box and tuning it's parameters is rarely ever necessary, except usually in cases where the distribution of the data is known. It rarely ever overfits the data. Another important advantage is that its model training and prediction times are very fast for the amount of data it can handle.

K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a non-parametric algorithm used for classification and regression. Non-parametric means there is no assumption for underlying data distribution. KNN assumes that similar things remain in close proximity. In other words, similar things remain close to each other. Based on this idea of the feature similarity KNN predicts the value of a new data point which further means that the new data point will be assigned a value based on how closely it matches the points in the training set.

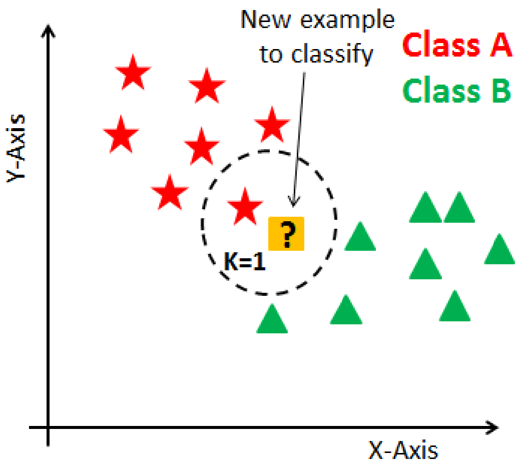


Figure 3.5: Example of K-Nearest Neighbor

KNN basically works in the following manner:

i Load training and test data

ii Choose the value of K i.e. the nearest data points. Here, K can be any integer.

iii For each sample in the test data do the following,

- Calculate the distance between the test sample and each sample of training data with any of the methods namely: Euclidean, Manhattan, or Hamming distance. Although, euclidean distance is the most common method used.
- Sort the calculated distance value in ascending order.
- Pick the top K rows from the sorted array.
- Finally, assign a class to the test sample based on the most frequent class of these rows. For regression return the average of the K labels.

Finally, assign a class to the test sample based on the most frequent class of these rows. For regression return the average of the K labels. Choosing the optimal K value is done by inspecting the data. In general, K value is chosen based on the model's ability to accurately predict the unseen data. In other words, the optimal K value is selected for the maximum accuracy of the test data. Cross-validation is another way to determine a good K value by using an independent dataset. Historically, optimal K value can be found within the range of 3-10.

KNN is preferred over many algorithms because it is much simpler and easy to understand. Also, it works well with the non-linear data as there is no presumption of data needed. There is no training required which makes it time-efficient. KNN can be used in both classification and regression.

3.4 Ensemble Learning

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. Supervised learning algorithms perform the task of searching through a hypothesis space to find a suitable hypothesis that will make good predictions with a particular problem. Even if the hypothesis space contains hypotheses that are very well-suited for a particular problem, it may be very difficult to find a good one. Ensemble methods combine the decisions from multiple models to improve the overall performance. Ensemble methods usually produces more accurate solutions than a single model would. Evaluating the prediction of an ensemble typically requires more computation than evaluating the prediction of a single model. In one sense, ensemble learning may be thought of as a way to compensate for poor learning algorithms by performing a lot of extra computation.

3.4.1 Ensemble Techniques

- **Basic Ensemble Techniques**

  i **Max Voting:** The max voting method is generally used for classification problems. Multiple models are used in this technique to make predictions for each data point. Each model's predictions are counted as a single 'vote.' The final prediction is made using the predictions obtained from the majority of the models.

  ii **Averaging:** In averaging, multiple predictions are made for each data point. We take the average of all the models' predictions and use it to make the final prediction in this method. In regression problems, averaging can be used to make predictions, and in classification problems, it can be used to calculate probabilities.

  iii **Weighted Average:** This is a variation on the averaging technique. Different weights are assigned to each model, indicating how important it is for prediction.

- **Advanced Ensemble techniques**

(a) **Stacking:** Stacking is an ensemble learning technique that uses predictions from multiple models (for example decision tree, knn or svm) to build a new model. This model is used for making predictions on the test set. Below is a step-wise explanation for a simple stacked ensemble:

(i) The train set is split into 10 parts.

(ii) A base model (suppose a decision tree) is fitted on 9 parts and predictions are made for the $10th$ part. This is done for each part of the train set.

(iii) The base model (in this case, decision tree) is then fitted on the whole train dataset.

(iv) Using this model, predictions are made on the test set.

(v) Steps II to IV are repeated for another base model (say knn) resulting in another set of predictions for the train set and test set.

(vi) The predictions from the train set are used as features to build a new model.

(vii) This model is used to make final predictions on the test prediction set.

(b) **Blending:** Blending is similar to stacking, but it only makes predictions using a holdout (validation) set from the train set. To put it another way, unlike stacking, the predictions are based solely on the holdout set. The holdout set and predictions are combined to create a model that is then tested on the test set. The following is a detailed description of the blending procedure:

(i) The train set is split into training and validation sets.

(ii) Model(s) are fitted on the training set.

(iii) The predictions are made on the validation set and the test set.

(iv) The validation set and its predictions are used as features to build a new model.

(v) This model is used to make final predictions on the test and meta-features.

(c) **Bagging:** Bagging is a technique for combining the results of multiple models (for example, all decision trees) to produce a more generalized result. One of the techniques used is bootstrapping. Bootstrapping is a sampling method that involves replacing subsets of observations from the original dataset. The size of the subsets is the same as the size of the original set. Bagging (or Bootstrap Aggregating) technique uses these subsets (bags) to get a fair idea of the distribution (complete set). The size of subsets created for bagging may be less than the original set. The steps are as follows:

(i) Multiple subsets are created from the original dataset, selecting observations with replacement.

(ii) A base model (weak model) is created on each of these subsets.

(iii) The models run in parallel and are independent of each other.

(iv) The final predictions are determined by combining the predictions from all the models.

(d) **Boosting:** Boosting is a sequential process in which each successive model attempts to correct the previous model's errors. The models that follow are reliant on the previous model. The following are the steps of boosting:
(i) A subset is created from the original dataset.
(ii) Initially, all data points are given equal weights.
(iii) A base model is created on this subset.
(iv) This model is used to make predictions on the whole dataset.
(v) Errors are calculated using the actual values and predicted values.
(vi) The observations which are incorrectly predicted, are given higher weights.
(vii) Another model is created and predictions are made on the dataset.
(viii) Similarly, multiple models are created, each correcting the errors of the previous model.
(ix) The final model (strong learner) is the weighted mean of all the models (weak learners).
As a result, the boosting algorithm combines several weak learners into a single strong learner. Individual models may not perform well across the entire dataset, but they do so for a portion of it. Thus, each model improves the ensemble's performance.

3.4.2 Algorithms based on Ensemble Techniques

Bagging and boosting are two of the most widely used machine learning techniques. The following is an overview of popular algorithms based on these two techniques:

Random Forest

Random forest is a supervised learning algorithm. It's a collection of decision trees that's typically trained using the "bagging" method. The bagging method's basic premise is that combining different learning models improves the overall result. To be more precise, random forest creates multiple decision trees and combines them to produce more accurate results. Voting determines which result is the best. A random forest is better than a single decision tree because it averages the results, which reduces over-fitting. It can be used for classification as well as regression, but it is most commonly used for classification.
Random forest algorithm works in the following manner:

- Select random samples from a given dataset

- Now, generate a decision tree for every sample and get the prediction result from every decision tree.
- For every predicted result perform voting
- Finally, select the most voted prediction result as the final result

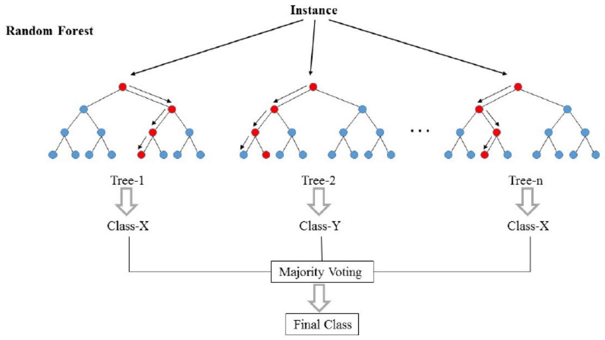


Figure 3.6: Example of Multiple Tree in Random Forest

Random forest adds additional randomness to the model while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

One of the best advantages of random forest is that it overcomes the problem of overfitting by combining the results of different decision trees. Again, random forest generally results in high accuracy and works efficiently on a large dataset. Important features can be extracted using random forest. Also, it has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.

AdaBoost

AdaBoost, or adaptive boosting, is one of the most basic boosting algorithms. Modeling is usually done with decision trees. Multiple sequential models are created, each one correcting the previous model's errors. AdaBoost assigns weights to the observations that are incorrectly predicted, and the following model works to correctly predict these values.
Below are the steps for performing the AdaBoost algorithm:

- Initially, all observations in the dataset are given equal weights.
- A model is built on a subset of data.
- Using this model, predictions are made on the whole dataset.

- Errors are calculated by comparing the predictions and actual values.
- While creating the next model, higher weights are given to the data points which were predicted incorrectly.
- Weights can be determined using the error value. For instance, higher the error more is the weight assigned to the observation.
- This process is repeated until the error function does not change, or the maximum limit of the number of estimators is reached.

Gradient Boosting

Boosting is a powerful machine learning technique used for classification and regression problems. Gradient boosting produces a model from a collection of classifiers or regressors. Gradient Boosting builds an ensemble of shallow trees in sequence with each tree learning and improving on the previous one. Although shallow trees by themselves are weak predictive models, they can be combined into a strong predictive model, when appropriately tuned, is often hard to beat with other algorithms.

Gradient boosting works by combining multiple models into an overall ensemble. Predictions are made by combining the predictions from the individual base models that make up the ensemble. Gradient boosting adds new models to the ensemble sequentially. It starts with a weak model (e.g., a decision tree with only a few splits) and sequentially boosts its performance by continuing to build new trees, where each new tree in the sequence tries to correct mistakes made by the previous tree. Gradient boosting iteratively improves any weak learning model. Weak learning models are those which have an error rate slightly better than random guessing. Generally, a decision tree is used as a base learner. The idea of gradient boosting is that each model in the sequence slightly improves upon the performance of the previous one. Trees are grown sequentially; with each tree is grown using information from previously grown trees to improve performance. Gradient boosting steps are as follows:

- Read the training data.
- Fit the decision tree to the data.
- Fit the next decision tree to the residuals of the previous.
- Finally, iterate through step 2 until some mechanism tells it to stop.

Gradient boosting is a powerful technique for developing predictive models. It can optimize different loss functions and provides several hyperparameters that make it flexible. Also, it does not require any preprocessing and works well with numerical and categorical values.

XGBoost

XGBoost is an ensemble learning method. Sometimes, it may not be sufficient to rely upon the results of just one machine learning model. Ensemble learning offers a systematic solution to combine the predictive power of multiple learners. The resultant is a single model which gives the aggregated output from several models.

The models that form the ensemble, also known as base learners, could be either from the same learning algorithm or different learning algorithms. Bagging and boosting are two widely used ensemble learners. Though these two techniques can be used with several statistical models, the most predominant usage has been with decision trees.
In contrast to bagging techniques like Random Forest, in which trees are grown to their maximum extent, boosting makes use of trees with fewer splits. Such small trees, which are not very deep, are highly interpretable. Parameters like the number of trees or iterations, the rate at which the gradient boosting learns, and the depth of the tree, could be optimally selected through validation techniques like k-fold cross validation. Having a large number of trees might lead to overfitting. So, it is necessary to carefully choose the stopping criteria for boosting.

3.5 Evaluation Metrics

Evaluation metrics are used to measure the quality of the statistical or machine learning model. Evaluating machine learning models is an essential part of building any project and often multiple evaluation metrics are used to measure the quality from different perspectives. Evaluation metrics are important to find the reliability of the model when dealing with the unknown data. There are many evaluation metrics available for both classification and regression problems. Here the most common evaluation metrics have been discussed.

3.5.1 Evaluation Metrics for Classification

Confusion Matrix: A confusion matrix is a table that is frequently used to depict the capacity of a classification model/classifier on a test data set information for which the actual values are known. Below is an example of a confusion matrix for a binary classifier:

The matrix tells that there are two potential predicted classes: "yes" and "no". If the presence of a cat in an image is being predicted, for example, "yes" would mean the image has a cat, and "no" would mean otherwise. It can be seen that the classifier made a total of 134 predictions and out of them, the classifier predicted "yes" 80 times, and "no" 54 times. But 77 of the images have the presence of a cat and the rest 57 do not.

True Positives (TP): These are the cases in which yes is predicted (the image has the presence of a cat), and the image does have the presence of a cat.

True Negatives (TN): No is predicted, and the image also does not have any cat in it.

False Negatives (FN): No is predicted, but the image does have the presence of a cat.

False Positives (FP): These are the cases in which yes is predicted (the image has the presence of a cat), but the image does not have any presence of a cat.

Accuracy: Accuracy is an essential classification metric. It is pretty straightforward. It is appropriate for both binary and multiclass problems.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \tag{3.13}$$

Precision: Precision finds out what proportion of positive identifications was actually correct.

$$Precision = TP/(TP + FP) \tag{3.14}$$

Recall: Recall finds out what proportion of actual positives was identified correctly.

$$Recall = TP/(TP + FN) \tag{3.15}$$

F1 Score: F1 score is the harmonic mean of precision and recall. It is a number between 0 and 1.

$$F1 = (2 \times precision \times recall)/(precision + recall) \tag{3.16}$$

3.5.2 Evaluation Metrics for Regression

Mean Absolute Error (MAE): Given any test data-set, the Mean Absolute Error (MAE) of a model alludes to the mean of the absolute values of each prediction error on all data of the test data-set. Prediction error is the difference between the actual value and the predicted value for that particular data. Statistically, Mean Absolute Error (MAE) refers to the results of measuring the difference between two continuous variables. When the performance is measured on continuous variable data then this metric is generally used. It averages the weighted individual differences equally and gives a linear result. The lower the value of this metric, the better is the model's performance.

$$MAE = (\sum_{i=1}^{n} |Predicted\ \ Value - Actual\ \ Value|)/n \tag{3.17}$$

where, n = Total number of data points

Mean Squared Error (MSE): Given any test data-set, the Mean Squared Error (MSE) of a model alludes to the sum, overall the data points, of the square of the difference between the predicted and actual target variables, divided by the number of data points. This metric is not very useful when the dataset contains a lot of noise. It is most handy when the dataset contains outliers, or unexpected values (too high or too low values). The lower the value of this metric, the better is the model's performance.

$$MSE = (\sum_{i=1}^{n} (Predicted\ \ Value - Actual\ \ Value)^2)/n \tag{3.18}$$

where, n= Total number of data points

Root Squared Error (RMSE): Root Squared Error (RMSE) is the standard deviation of the errors. This is the same as Mean Squared Error (MSE), but here the root of the difference between the actual value and the predicted value is considered. This metric is useful when there are large errors and they affect the model's accuracy drastically. Here also, the lower the value of this metric, the better is the model's performance.

$$RMSE = \sqrt{(\sum_{i=1}^{n} (Predicted\ \ Value - Actual\ \ Value)^2)/n} \tag{3.19}$$

where, n = Total number of data points

R Squared: R Squared is also known as the coefficient of determination. Given a dataset, this metric gives us an idea about how well a model fits the dataset. It shows how close the regression line is to the actual data values. The value of this metric is between 0 and 1 and the closer the value is to 1 the better is the model's performance. It is calculated by dividing the sum of squares of residuals from the regression model by the total sum of squares of errors from the average model and then subtracting the result from 1.

$$R\ \ Squared = 1 - (RSS/TSS) \tag{3.20}$$

$$RSS = \sum_{i=1}^{n} (Actual\ \ Value - Predicted\ \ Value)^2 \tag{3.21}$$

$$TSS = \sum_{i=1}^{n} (Actual\ \ Value - \overline{Y})^2 \tag{3.22}$$

where,

$\overline{Y}$ = sum of all the data points / n

n = Total number of data points

Chapter 4

Propossed Methodology

4.1 Introduction

Predicting crime can be done using a variety of methods and machine learning models. For developing any machine learning models, there are a few steps. Some of these necessary steps include defining the problem, gathering relevant data, choosing an appropriate evaluation protocol, preparing the data, correctly splitting the data, distinguishing between overfitting and underfitting, choosing appropriate machine learning algorithms, and tuning the models to achieve the best possible performance.
This chapter provides a quick overview of the proposed Crime Prediction model.

4.2 Data Collection

Data collection is the procedure of collecting, measuring, and analyzing accurate insights for research using standard validated techniques. In most cases, data collection is the primary and most important step for research. The approach of data collection is different for different fields of study, depending on the required information. Data for this study was collected from The Daily Star newspaper. First, crime news headlines and links were gathered after reading every headline in the newspaper's four sections (Front Page, Back Page, City, and Country). After reading each crime-related news article, the necessary crime information was gathered.

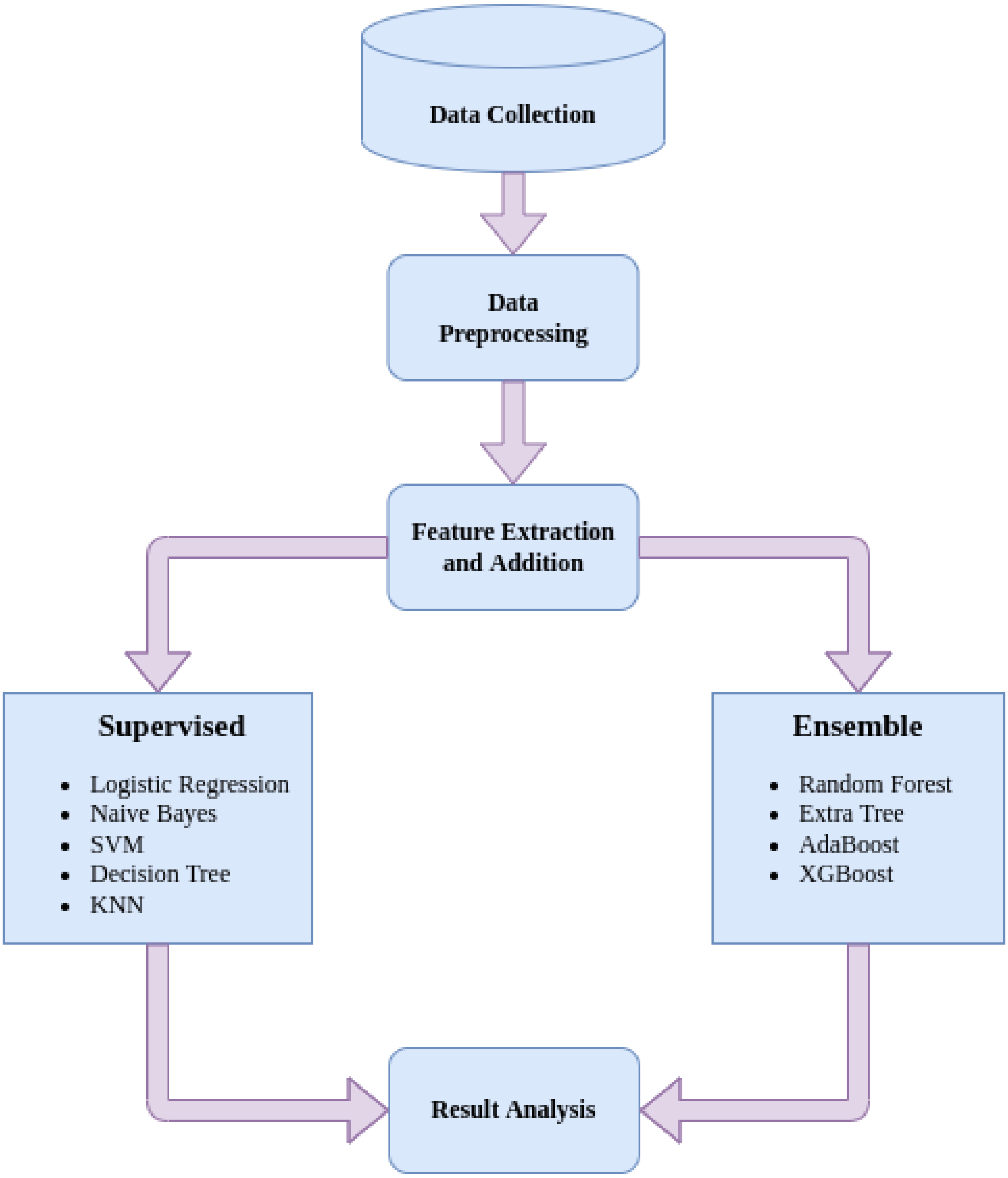


Figure 4.1: Proposed Crime Prediction Model

4.3 Data Preprocessing

Data preprocessing in Machine Learning is a crucial step that helps enhance the quality of data to promote the extraction of meaningful insights from the data. Data preprocessing in Machine Learning refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for building and training Machine Learning models. For this study, incident places found on the dataset were splitted into Upazila and District names by maintaining proper spelling from Wikipedia [26] and Bangladesh govt. website [27]. After converting incident date into date-time format incident year, month, week number, and weekday were then extracted from date-time format and added as feature. Label encoding was done after necessary features extraction.

4.4 Feature Extraction

Feature extraction is a process of remapping features and dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set. Incident Date, Part of the day of incident, Incident location were selected from dataset as features. Along with this Geo-code, Weather informations and season were added as feature. Geo-code informations were picked from Mapbox API and collected weather data through Positionstack API. Three types of seasons were merged with the dataset by analyzing the months.

4.5 Classification Strategy

Two approaches are proposed in this study. The first approach is supervised learning. Supervised learning is the first approach. Logistic Regression, Naive Bayes, SVM, KNN, and Decision Tree are instances of this approach.Ensemble learning is the Second approach. Popular algorithms such as Random Forest, Extratree, AdaBoost, and XGBoost will be included in this approach.

The following chapters demonstrate these critical steps of the proposed Crime Prediction Model.

Chapter 5

Dataset Preperation

5.1 Introduction

A dataset is any permanently stored collection of information for multiple survey instances that contains either case level data, aggregation of case level data, or statistical manipulations of either case level or aggregated survey data. It represents the contents of a single database table or statistical data matrix, in which each column of the table represents a specific variable and each row represents a specific dataset member.

The human brain is capable of performing a wide range of difficult tasks by combining logic, intellect, and a sense of humour. When it comes to machines or computers, however, they are not the same. They only understand binary languages and operate under specific conditions by following certain rules or algorithms. As a result, the dataset must be larger, more informative, and machine-readable. If no dataset exists or if the dataset is inadequate, it will not function properly and will be unable to perform the desired task, affecting the work's outcome.

Improving the dataset is one of the most effective ways to improve model performance.

5.2 Availability of Crime Dataset

Criminal activities is a major concern of many governments who are using different advanced technology to tackle such issues. Crime Analysis, a sub branch of criminology, studies the behavioral pattern of criminal activities and tries to identify the indicators of such events. To classify future crimes we need to understand previous crime and extract underlying pattern in those crime events. But unfortunately such day to day crime record is not openly available for study .

5.3 Possible source of data

In the United States, the United Kingdom, Canada, and other countries, there are numerous popular crime datasets to analyze [28] [29] [30] [31] [32]. Their Police Department has made the majority of these datasets accessible. However, we only have a statistical dataset for each crime's number of occurrences in a given year [33]. As a result, it was necessary to seek for other options.

5.3.1 Bangladesh Police

The Dhaka Metropolitan Police's Media and Public Relations Division was approached to see if such a crime record could be made public. To this day, however, no response has been received from them. After a while, it became clear that this was not a viable option.

5.3.2 Newspaper

As every newspaper carries criminal news, it might be an excellent source for gathering historical crime data. We started looking for the most appropriate newspaper with archived news where we may find earlier crime reports. The Daily Star [1] is Bangladesh's most widely circulated daily English-language newspaper. The website of the daily star newspaper also has an archived [34] news section.

5.4 Data Acquisition

One of the most difficult aspects of this study was gathering data. Since multiple papers can cover the same story, it was best to use a single newspaper for Crime News Source. Otherwise, there may be duplicate crime news.

5.4.1 Types of Crime News Found in Newspaper

The Daily Star [1] covers a wide range of crime articles. Murder, rape, assault, robbery, kidnapping, and corruption are the most common sorts of crime news, according to our findings. The frequency of corruption-related crime news is quite low. However, the number of stories in which an unknown person's body is discovered is extremely high. Even though Body Featured crime news is classified as Murder, as many of the features found in Murder news are not available in Body Found. Body Found was decided to be treated as a distinct type of crime.

5.4.2 News Link Collection

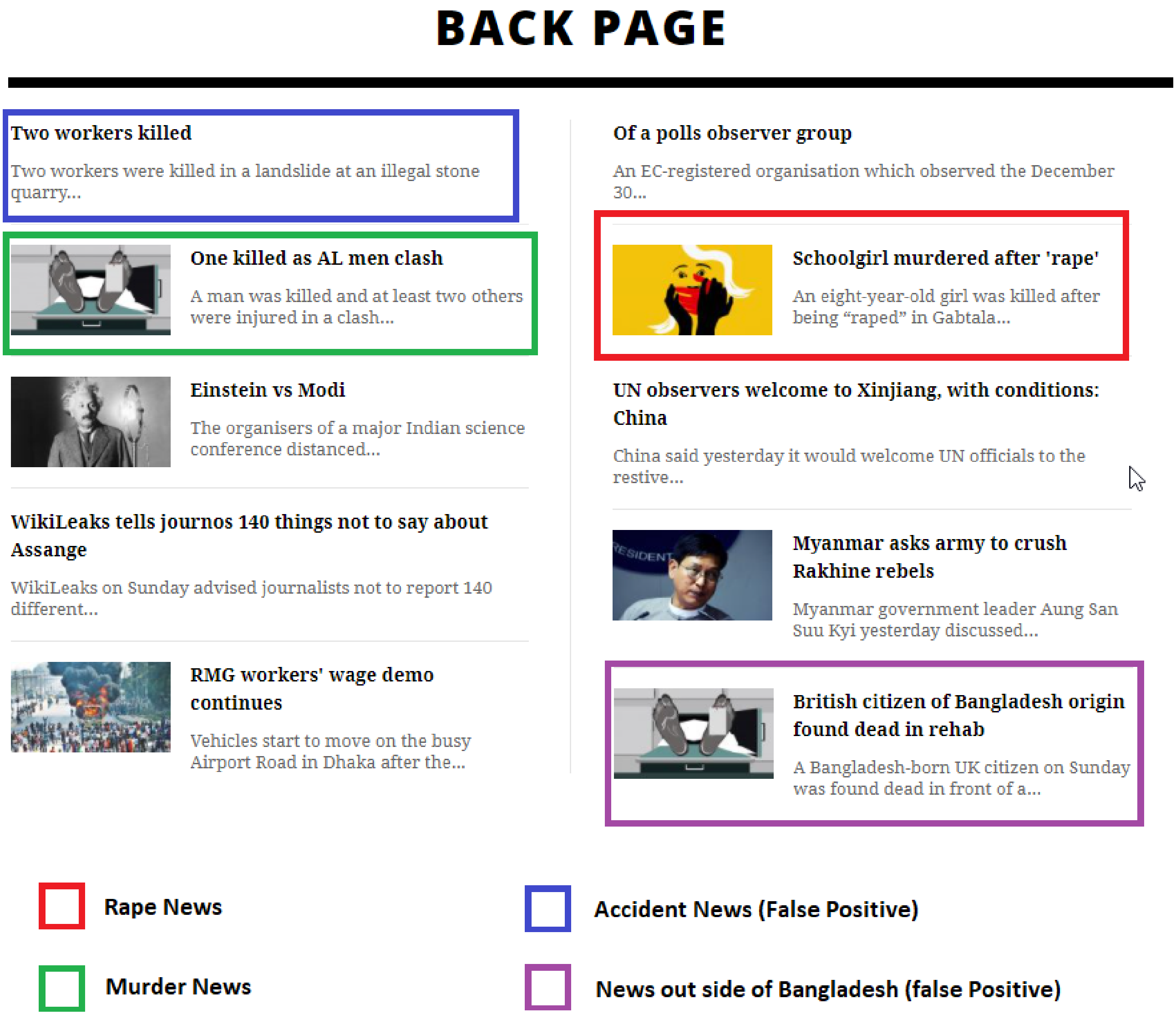
After deciding on The Daily Star as the source, the archival section of the newspaper was examined. All of the news that was published was divided into different sections. According to observation, the four sections that provide news about crime in our country are the Front Page, Back Page, City, and Country. However, once data collection concluded in August 2021, The architecture and design of the Daily Star Website have been changed. These four sections are no longer available. As news is now divided into different categories. However, the news links gathered during the data collection phase are still accessible.

Manual News Link Collection

There were few options for collecting crime news links from newspapers at first. The initial effort to collect data was a hands-on approach. The collection of crime news headlines and links began from January, 2019 news archive. After reading the headline, it is possible to determine whether the news is related to crime and which category it belongs to.

Every news item on the Front Page, Back Page, City, and Country was reviewed in order to identify crime-related news and the category to which it belongs. For example, if the news was about murder, the news link and headlines were categorized as murder. The process was continued on archive news from December 2018 to January 2019. Following a review of nearly 28000 news articles, a total of 2000 crime related news links were collected.

It is sometimes possible to misinterpret non-crime news as crime news based on the headline. For example, "Two Workers Killed" appears to be a crime based on the headline. However, after thoroughly reading the news, it is discovered that the news contains an accident report. The image below is a screenshot of the newspaper's Backpage section.



BACK PAGE

Two workers killed

Two workers were killed in a landslide at an illegal stone quarry...

One killed as AL men clash

A man was killed and at least two others were injured in a clash...

Einstein vs Modi

The organisers of a major Indian science conference distanced...

WikiLeaks tells journos 140 things not to say about Assange

WikiLeaks on Sunday advised journalists not to report 140 different...

RMG workers' wage demo continues

Vehicles start to move on the busy Airport Road in Dhaka after the...

Of a polls observer group

An EC-registered organisation which observed the December 30...

Schoolgirl murdered after 'rape'

An eight-year-old girl was killed after being "raped" in Gabtala...

UN observers welcome to Xinjiang, with conditions: China

China said yesterday it would welcome UN officials to the restive...

Myanmar asks army to crush Rakhine rebels

Myanmar government leader Aung San Suu Kyi yesterday discussed...

British citizen of Bangladesh origin found dead in rehab

A Bangladesh-born UK citizen on Sunday was found dead in front of a...

🟥 Rape News

🟦 Accident News (False Positive)

🟩 Murder News

🟪 News out side of Bangladesh (false Positive)

Figure 5.1: Crime news example from Backpage

However, collecting crime-related news after reading every news headline and news report from those four sections of the newspaper is time consuming, ineffective, and neurotic.

Automated News Link Collection

From 2018 to 2019, nearly 2000 crime news were gathered in various categories. Each category had enough headlines to analyze and extract keywords that could be used to identify which category any crime news belongs to. Using the BeautifulSoup4 [35] Python Library, nearly 60000 news links and headlines were scraped from The Daily Star [34] archival news from 2017-2012 Front Page, Back Page, City, and Country sections.

Beautiful Soup: Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. The Beautiful Soup library helps with isolating titles and links from webpages. It is designed for quick turnaround projects like screen-scraping. It can extract all the text from HTML tags, and alter the HTML in the document with which we are working. It commonly saves programmers hours or days of work. Some key features that make beautiful soup unique are:

- Beautiful Soup provides a few simple methods and Pythonic idioms for navigating, searching, and modifying a parse tree.
- Beautiful Soup automatically converts incoming documents to Unicode and outgoing documents to UTF-8.
- Beautiful Soup sits on top of popular Python parsers like lxml and html5lib, which allows us to try out different parsing strategies or trade speed for flexibility.

To parse a document, it must be passed into the Beautiful Soup constructor. Beautiful Soup then parses the document using the best available parser. It will use an HTML parser unless you specifically tell it to use an XML parser. The Beautiful Soup object itself represents the document as a whole. For most purposes, it can treat as a Tag object. This means it supports most of the methods described in navigating the tree and searching for the tree.

5.4.3 Filtering Crime News From Scrapped News Link

Approximately 60000 news links were collected from the four sections of the newspaper using the BeautifulSoup Web scrapper. However, not all of them were about crime. It was tiresome and time consuming to select crime news after reading each headline and news from those news links. As a result, news headline analysis was performed on previously manually collected news links in order to identify any meaningful patterns among the headlines of various crime news categories. The results were then used to filter crime-related news from 60000 scraped news links. These are the most prevalent keywords present in each category after analyzing approximately 2000 crime news headlines.

Most frequent words on Assault news are: harass, attacked, assault, attack, torturing, tortur, stabbed, beat, brutally, unconscious, forced, boiling, shoot acid, thrown, molested, assaulted, burn, assaults, shot, stabbing, chained, tied, brutality, sexually, injur, harassed, abuse, brutalised, assailant, brutalise, attacks, forcibly, cuts, stalking, sexual, molestation, shave, throwing, cruelty, caned, wrath, abusing, burnt, hack, molest, mercilessly, resists, stab.

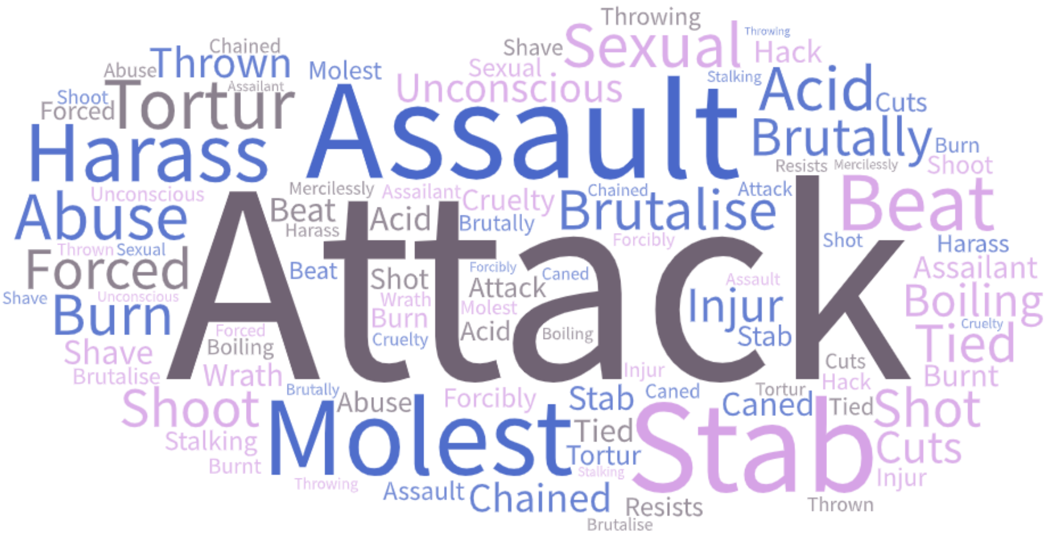


Figure 5.2: Frequent words in Assault news Headline

Most frequent words on Rape news are: raping, rape, raped, gang-raped, rapes, rapist, gang-rape.

Most frequent words on Kidnap news are: abduction, abduct, abducted, abducting, kidnap, rescued, missing, traceless.

Most frequent words on Robbery news are: mugger, mugged, robber, loot, snatch, robbed, robbery, looted.

Most frequent words on Body Found news are: body found, bodies found, body recovered, bodies recovered, found dead, found murdered, found hanging, found bodies, found body, murdered found, dead found, recovered body, recovered bodies, hanging body, hanging bodies, bullet-hit body, bullet-hit bodies, decomposed body, decomposed bodies.

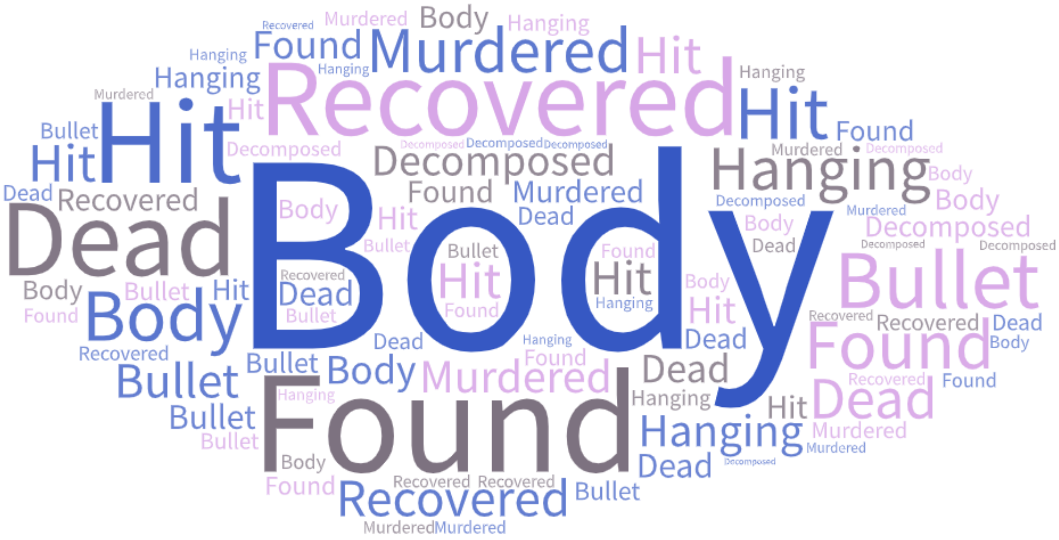


Figure 5.3: Frequent words in Body Found news Headline

5.4.4 Removing redundant news based on keyword not related to crime

The presence of a keyword in a news headline indicates that the news is not related to crime. Here are some of those words: trump, anniversary, obituary, top quote, war crimes, road crashes, road accident, road crash, India, US, UN, UK, obama, bumper, british, myanmar,sheikh hasina, bangabandhu, 'gunfight', landslide, passes away, passed away, human trafficking, IC, electrocu, messi, neymar, russia, collapse, seized, khaleda, protest, mexic, hearing, life term, korea, budget, BCIC, ICC, Japan, China, Bhutan, africa, taiwan, brazil, german, philippine, strike, ACC.

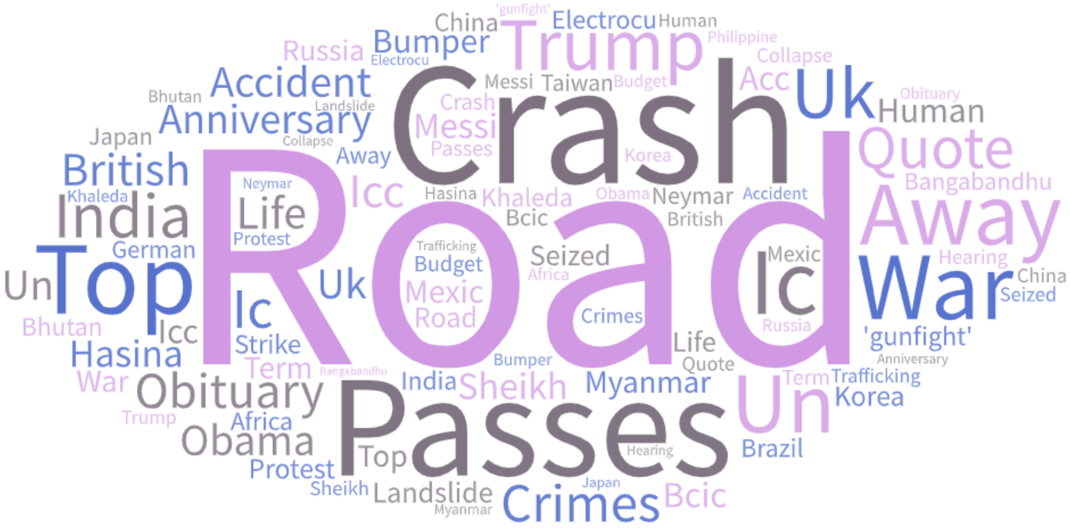
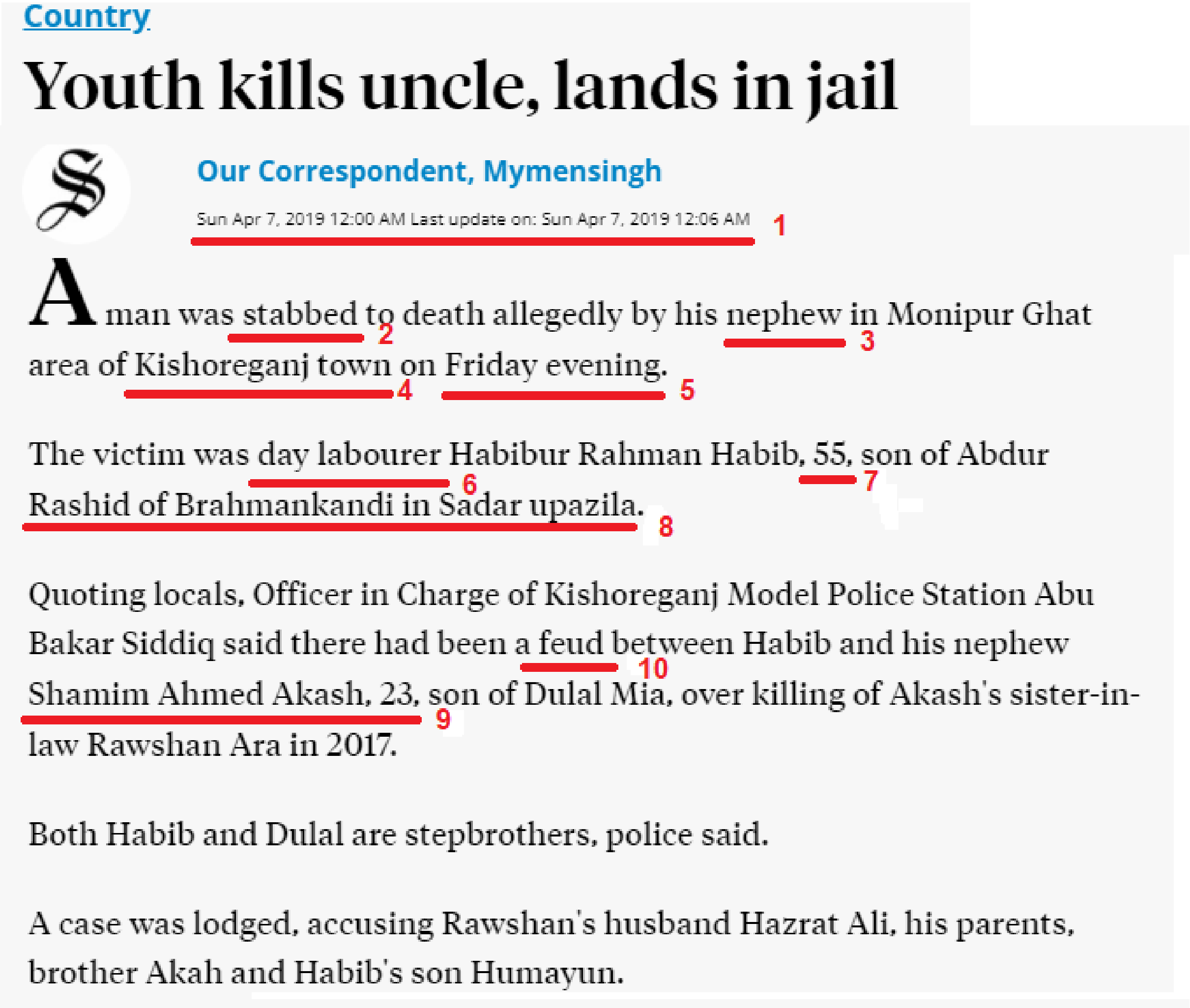


Figure 5.4: Frequent words in Redundant news Headline

Even after all of this filtering, there were news stories that appeared to be about a crime but were actually about an accident or natural cause. Further analysis were necessary to discover and remove such articles.

5.5 Selecting Feature: Information to be Extracted from News

Crime news from every category was examined to determine what information about crime are recorded in the news. The information typically found in crime news is graphically explained below, as well as how they were chosen as a feature for the respective crime category.



Country

Youth kills uncle, lands in jail

Our Correspondent, Mymensingh

Sun Apr 7, 2019 12:00 AM Last update on: Sun Apr 7, 2019 12:06 AM

A man was stabbed to death allegedly by his nephew in Monipur Ghat area of Kishoreganj town on Friday evening.

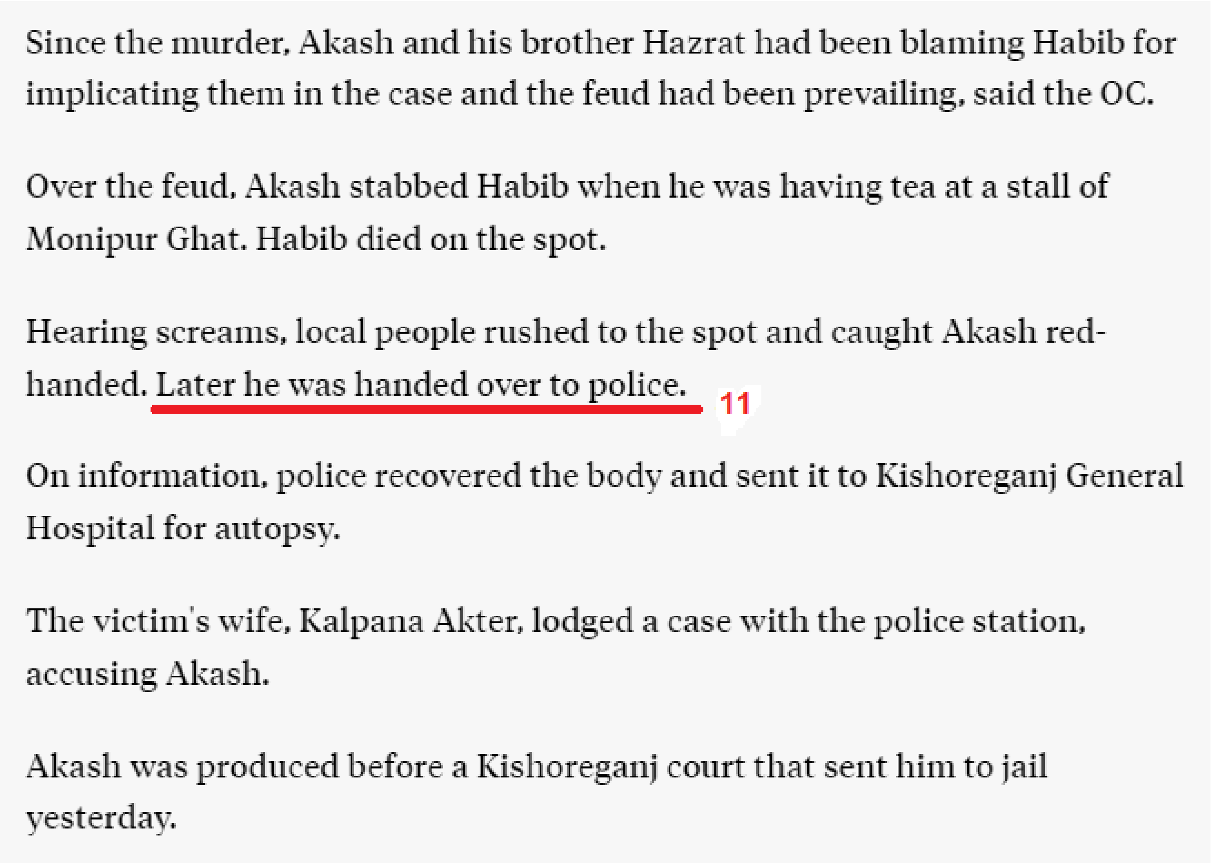
The victim was day labourer Habibur Rahman Habib, 55, son of Abdur Rashid of Brahmankandi in Sadar upazila.

Quoting locals, Officer in Charge of Kishoreganj Model Police Station Abu Bakar Siddiq said there had been a feud between Habib and his nephew Shamim Ahmed Akash, 23, son of Dulal Mia, over killing of Akash's sister-in-law Rawshan Ara in 2017.

Both Habib and Dulal are stepbrothers, police said.

A case was lodged, accusing Rawshan's husband Hazrat Ali, his parents, brother Akah and Habib's son Humayun.

Figure 5.5: Example of Murder News Part 1



Since the murder, Akash and his brother Hazrat had been blaming Habib for implicating them in the case and the feud had been prevailing, said the OC.

Over the feud, Akash stabbed Habib when he was having tea at a stall of Monipur Ghat. Habib died on the spot.

Hearing screams, local people rushed to the spot and caught Akash red-handed. Later he was handed over to police.

On information, police recovered the body and sent it to Kishoreganj General Hospital for autopsy.

The victim's wife, Kalpana Akter, lodged a case with the police station, accusing Akash.

Akash was produced before a Kishoreganj court that sent him to jail yesterday.

Figure 5.6: Example of Murder News Part 2

The information contained in a murder news can be seen in Figure5.5. They are as follows:

i News Date.

ii Crime approach.

iii Relation between Victim and Criminal.

iv Incident Place.

v Incident Time.

vi Victim Profession.

vii Victim Age.

viii Victim's Address.

ix Criminal Age.

x Motive behind the Crime.

xi If Criminal Arrested.

Along with these, information such as whether the victim or criminal was politically involved, as well as their religion, can be gathered.
News article from every type of crime were examined. Then, for each type of crime, the following features were chosen.

Information to be obtained from murder-related news:

News Date, Incident Date, If Arrested, Part of the day of Crime, Incident place, Murder approach, Murder Weapon, Motive, Victim Age, Victim Gender, Victim Profession, Victim Religion, Victim Address, Criminal Age, Criminal Gender, Criminal Profession, Criminal Religion, Relation between Victim and Criminal.

Information to be obtained from Rape related news:

News Date, Incident Date, If Arrested, Part of the day of Crime, Incident place, No of Victims, Victim Age, Victim Gender, Victim Profession, Victim Religion, Victim Address, Criminal Age, Criminal Gender, Criminal Profession, Criminal Religion, Relation between Victim and Criminal.

Information to be obtained from Assault related news:

News Date, Incident Date, If Arrested, Part of the day of Crime, Incident place, Motive, Victim Age, Victim Gender, Victim Profession, Victim Religion, Victim Address, Criminal Age, Criminal Gender, Criminal Profession, Criminal Religion, No of criminal, Relation between Victim and Criminal.

Information to be obtained from Robbery related news:

News Date, Incident Date, If Arrested, Part of the day of Crime, Murder place, Murder approach, Murder Weapon, Motive, Criminal Age, Criminal Gender, Criminal Profession, Criminal Religion, No of criminal.

Information to be obtained from Kidnap related news:

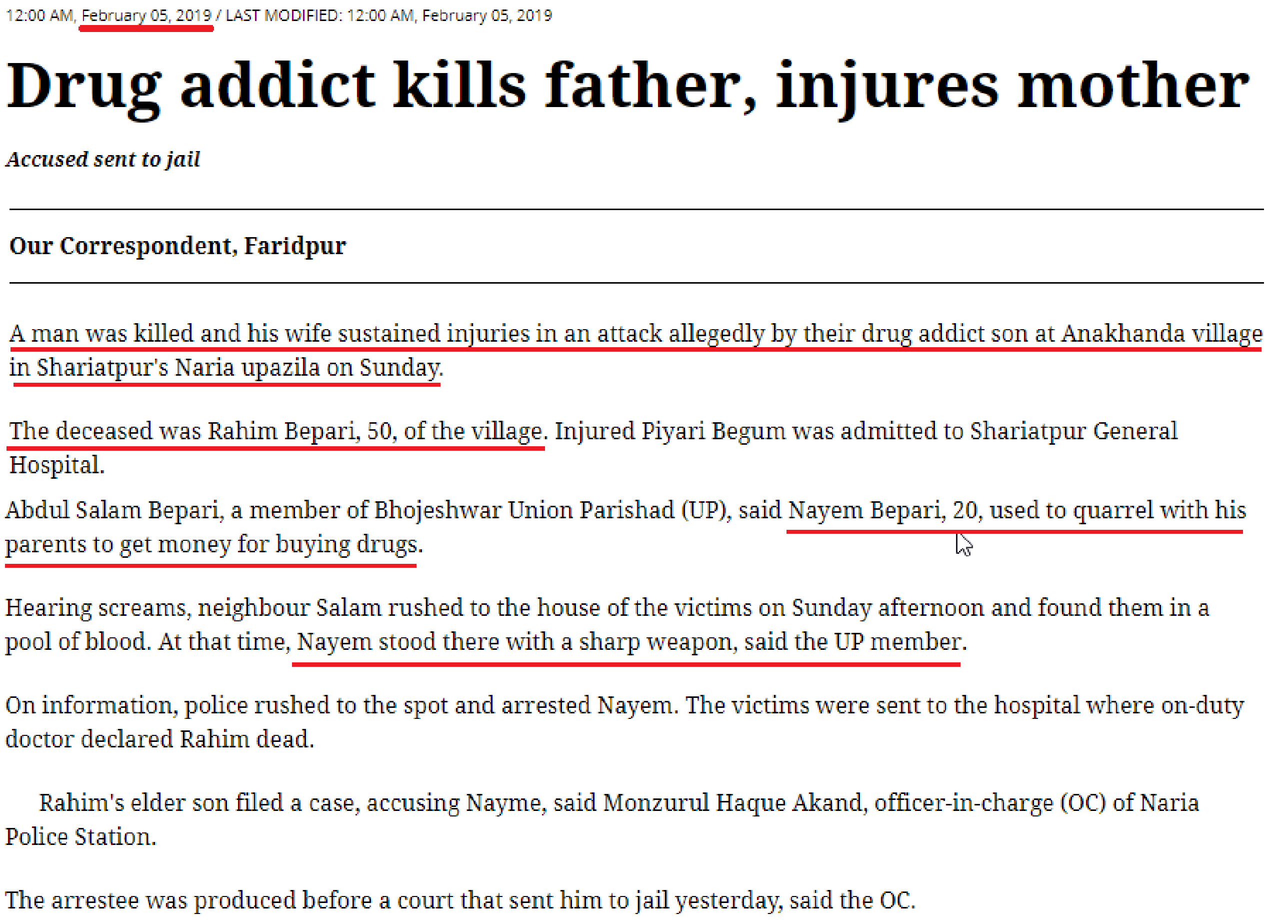
News Date, Abduction Date, Rescue Date, If Arrested, Part of the day of Crime, Incident place, Rescued Place, Victim Age, Victim Gender, Victim Profession, Victim Religion, Victim Address„ Victim Injured, Criminal Age, Criminal Gender, Criminal Profession, Criminal Religion,No of criminal, Relation between Victim and Criminal.

Information to be obtained from Body Found related news:

News Date, Incident Date, Part of the day of Crime, Incident place, Victim Age, Victim Gender, Victim Profession, Victim Religion, Victim Address, Body State.

5.6 Data Extraction from news

Data extraction from the news was also difficult. Reading each article and then entering the data into the appropriate feature is time consuming, exhausting, and tedious. However, there were no other solutions that could have met the goal. Scraping this information would be nearly impossible, and even if it could be done, the results would be disappointing. Even if scraping was used to extract data from news stories, it was necessary to double-check each one individually. It would take less time and effort to manually extract data from crime news by reading them than it would to check those results. Here are some examples of such process:



12:00 AM, February 05, 2019 / LAST MODIFIED: 12:00 AM, February 05, 2019

Drug addict kills father, injures mother

Accused sent to jail

Our Correspondent, Faridpur

A man was killed and his wife sustained injuries in an attack allegedly by their drug addict son at Anakhanda village in Shariatpur's Naria upazila on Sunday.

The deceased was Rahim Bepari, 50, of the village. Injured Piyari Begum was admitted to Shariatpur General Hospital.

Abdul Salam Bepari, a member of Bhojeshwar Union Parishad (UP), said Nayem Bepari, 20, used to quarrel with his parents to get money for buying drugs.

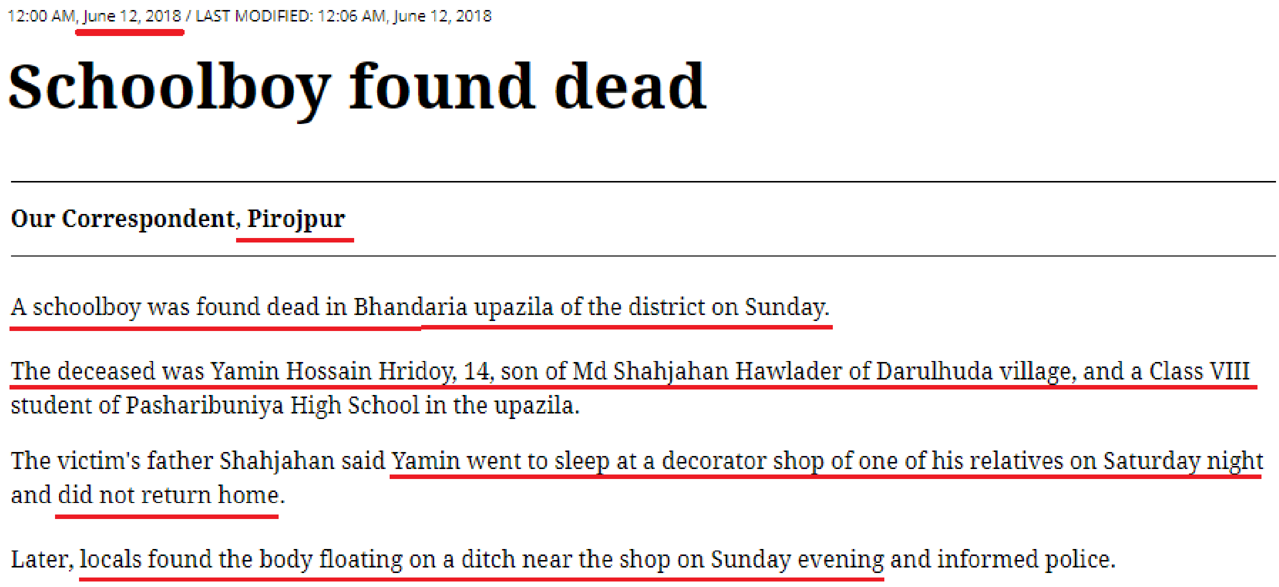
Hearing screams, neighbour Salam rushed to the house of the victims on Sunday afternoon and found them in a pool of blood. At that time, Nayem stood there with a sharp weapon, said the UP member.

On information, police rushed to the spot and arrested Nayem. The victims were sent to the hospital where on-duty doctor declared Rahim dead.

Rahim's elder son filed a case, accusing Nayme, said Monzurul Haque Akand, officer-in-charge (OC) of Naria Police Station.

The arrestee was produced before a court that sent him to jail yesterday, said the OC.

Figure 5.7: Example of Murder News



12:00 AM, June 12, 2018 / LAST MODIFIED: 12:06 AM, June 12, 2018

Schoolboy found dead

Our Correspondent, Pirojpur

A schoolboy was found dead in Bhandaria upazila of the district on Sunday.

The deceased was Yamin Hossain Hridoy, 14, son of Md Shahjahan Hawlader of Darulhuda village, and a Class VIII student of Pasharibuniya High School in the upazila.

The victim's father Shahjahan said Yamin went to sleep at a decorator shop of one of his relatives on Saturday night and did not return home.

Later, locals found the body floating on a ditch near the shop on Sunday evening and informed police.

Figure 5.8: Example of Body Found News

5.7 Obtained Dataset

It was possible to collect 6640 data points after 7 months of data extraction from news links. That was more than our original goal. The data points collected in each crime category are listed below.

Table 5.1: Data Points in each Crime Category

Crime	Data Points
Murder	1520
Rape	1210
Assault	1111
Robbery	602
Kidnap	668
Body Found	1529

Chapter 6

Data Preprocessing

Data preprocessing refers to the technique of preparing the raw data to make it suitable for building and training Machine Learning and Deep Learning models. In Machine Learning, data preprocessing is a technique for transforming raw data into a legible format.

Data from the real world is often incomplete, inconsistent, lacking in certain behaviors or trends, and prone to numerous errors. Preprocessing data is a tried and true method of resolving such problems. Preprocessing raw data prepares it for further processing.

6.1 Incident Place

In news articles, incident locations are presented in a variety of formats. When collecting data, these were added to the dataset in the order in which they were discovered. In some cases, only the district is available; in others, both Upazilla and the district are available; in still others, only the union is available. This type of variation is inappropriate for a dataset. As a result, incident location preprocessing was required.

6.1.1 Formatting Raw Location

In the news, the raw location names were in a different format. The names of the locations were then formatted. The rules that were followed were as follows:

- Each location is divided into three sections.
- The first part of the location contains the name of a specific location or union that was found in a news article.

- The second part of the item is the Upazilla name that was mentioned in the news.
- The third part is the District Name.
- The comma "," is used to separate the three parts.
- If a specific location or Union Name could not be found, that location only has Upazilla and District. If Upazilla was not found, that location only has Union and District. If some cases in only contains District as other information were not mentioned in the news.

6.1.2 Separating Formatted Location

After examining formatted locations, it was discovered that the majority of them lacked specific location or Union. The formatted location was split into two parts: Upazilla and District. If Upazilla isn't present, but Union is, keep Union's name. There were also few location where only District This was accomplished using the regular expression [36].

All of the locations were divided into Upazilla and District as a result of this.

Regular Expression

A regular expression is a sequence of characters that specifies a search pattern. String-searching algorithms typically use such patterns for string "find" or "find and replace" operations, as well as input validation. It is a formal language theory and theoretical computer science technique.
Regular expressions were first proposed in the 1950s by American mathematician Stephen Cole Kleene, who formalized the definition of a regular language. They were popularized by Unix text-processing utilities.
Regular expressions are used in search engines, word processors' search and replace dialogs, and text editors. Regex is built-in or available through libraries in many programming languages, as it is useful in a variety of situations. A regex processor converts the syntax of a regular expression into an internal representation that can be executed and matched against a string representing the text being searched in.

6.1.3 Replacing with Official Location Name

All the Union, Upazilla, and District names spelling were referenced from Wikipedia [26] and Bangladesh govt. websites [27]. When matching with the separated location it was found that the loation was eronoud. As the spelling of separated location and official names differs. Some of those cases are metioned below as example:

Table 6.1: Example of Spelling Difference in District Name

News article Location Spelling	Official Location Spelling
Netrakona	Netrokona
Jessore	Jashore
Bogra	Bogura
Barishal	Barisal
Cumilla	Comilla
Jhenaidha	Jhenaidah
Laxmipur	Lakshmipur
Chittagong	Chattogram
B'baria	Brahmanbaria
Moulovibazar	Moulvibazar

Table 6.2: Example of Spelling Difference in Upazilla Name

News article Location Spelling	Official Location Spelling
Zakiganj	Jakiganj
Zajira	Jajira
Ukhiya	Ukhia
Trishal	Trihsal
Taragonj	Taraganj
Sreemangal	Srimangal
Sreebordi	Sreebardi
Sonatala	Sonatola
Sonaimori	Sonaimuri
Singiar	Singair

Manually correcting these misspellings in location is inefficient and time-consuming. String matching was the best option for this job. The FuzzyWuzzy [37] String matching module was used for this.

FuzzyWuzzy

FuzzyWuzzy is a library of Python which is used for string matching. Fuzzy string matching is the process of finding strings that match a given pattern. The algorithm behind fuzzy string matching does not simply look at the equivalency of two strings but rather quantifies how close two strings are to one another. This is usually done using a distance metric known as 'edit distance'. This determines the closeness of two strings by identifying the minimum alterations needed to be done to convert one string into another. There are different types of edit distances that can be used like Levenshtein distance, Hamming distance, Jaro distance, etc. Fuzzywuzzy uses Levenshtein distance [38] to calculate similarity ratio between two sequences and returns the similarity percentage.

Levenshtein distance

The Levenshtein distance is a string metric to calculate the difference between two different strings. Soviet mathematician Vladimir Levenshtein formulated this method and it is named after him.

The Levenshtein distance between two strings *a,b* (of length $|a|$ and $|b|$ respectively) is given by *lev(a,b)* where,

$$lev(a,b) = \begin{cases} |a| & if\ |b| = 0, \\ |b| & if\ |a| = 0, \\ lev(tail(a), tail(b)) & if\ a[0] = b[0] \\ 1 + min \begin{cases} lev(tail(a), b) \\ lev(a, tail(b)) \\ lev(tail(a), tail(b)) \end{cases} & otherwise, \end{cases} \tag{6.1}$$

District and Upazilla Name Matching

With FuzzuWuzzy, string matching yields a percentage of similarity. Official names were matched to all Upazilla and District names found on news. The percentage of similarity between them was also recorded. The calculated results are as follows:

Table 6.3: Similarity Percentage Count of District Name

Similarity Percentage	Count
100%	6216
95%	8
90%	141
85%	22
80%	4
75%	4
Below 75%	226

Table 6.4: Similarity Percentage Count of Upazilla Name

Similarity Percentage	**Count**
100%	3992
95%	47
90%	1111
85%	111
80%	114
75%	114
Below 75%	1088

District Name

The majority of the names matched the official names, as shown in Table 6.3. The rest were then replaced with the official names that it most closely matched. Later, the district names that had been replaced were checked to see if there were any errors. However, after replacing all of the District names, it was discovered that they were all correct.

Upazilla Name

However, this was not the case with Upazilla's names. Only 3992 Upazilla matched the official names exactly. The rest were misspellings, not Upazilla names, specific place names, municipal area locations, and so on. Official names were used to replace misspelled Upazilla names. Others were dealt with in a variety of ways.

6.1.4 Correcting inaccurate Location

In some cases, the Upazilla and District did not correspond to the official listing. This means that Upazilla is not part of the Location District. Then they were manually checked to see if there was a porper Upazilla and District. Then they were replaced by the correct version.

6.1.5 Mapping Specific Places to corresponding Upazilla

Many of the news articles did not mention the Upazilla or Union names. Rather, they used other interpretations to specify location. Here are a few examples:

- Village name
- Road name

- Bus Station
- University
- Launch Terminal
- Town name
- School name
- Public place name
- Market name
- Police Station
- Hospital name
- Mega mall

All of these place names were later Googled to determine which Upazilla or municipal area they belonged to. They were then replaced with the name of the corresponding Upazilla or municipal area. This process were very time consuming and tiresome.

6.1.6 Mapping Places from Municipal Area

Crimes committed in municipal are do not have Upazilla Name. But as municipal area found in news articles are popular they can be used to obtain Latitude and Longitude responses from the API.

6.2 Incident Date

Incident date is a very important feature of this dataset. However, most incident dates found in news articles were in implicit form. Some of these forms are last Friday, yesterday, the day before, before an event, and so on. To determine the actual incident date, these types of data needed to be adjusted in relation to the news publication date. Other times, the news date was formatted as "29 March," "27 June 2018," "April 2nd," and so on. Before processing, these were manually formatted to "Day-Month-Year" format.
After these incident date were converted to datetime object using Date-time [39] module.

Datetime

Python has a module named datetime to work with dates and times. Python Datetime module supplies classes to work with date and time. These classes provide a number of functions to deal with dates, times and time intervals. Date and datetime are an object in Python, so it can be manipulated as specific object not as strings. Day number, Weekday, Week Number, etc information can be extracted from datetime.

6.3 Incident Time

The incident time is the time when the crime was committed. It is also an essential feature of this dataset. However, the time of the incident was not specified in any news article. Rather, the time of the crime was recorded as parts of the day. As a result, when processing incident time, the following table was used to categorize them into different parts of the day.

Table 6.5: Parts of the day time-frame

Timeframes	**Parts of the day**
6 am - 11:59 am	Morning
12 pm - 3:59 pm	Noon
4 pm - 5:59 pm	Afternoon
6 pm - 7:59 pm	Evening
8 pm - 5:59 am	Night

Chapter 7

Feature Engineering

Feature engineering is the process of using domain knowledge to extract features from raw data. The motivation is to use these extra features to improve the quality of results from a machine learning process, compared with supplying only the raw data to the machine learning process.

The most important characteristic of these large data sets is that they have a large number of variables. These variables require a lot of computing resources to process. So Feature engineering helps to get the best feature from those big data sets by select and combine variables into features, thus, effectively reducing the amount of data.

7.1 Feature Addition

Until now, the incident location has been divided into 2 parts: Upazilla (specific place name) and District. Another priority was to include weather data in this dataset. We need the latitude and longitude of a location to get weather information about it. Due to the presence of Upazilla and District in the dataset, these features were used to obtain the Latitude and Longitude of those specific locations. Then historical weather data was collected using these latitude and logitude, as well as incident dates. The procedure is further explained in the subsections that follow.

7.1.1 Latitude and Longitude

Geographic Coordinate System

The geographic coordinate system is a spherical or ellipsoidal coordinate system used to measure and communicate positions on the Earth as latitude and longitude. It is the most basic, oldest, and widely used of the thousands of spatial reference systems in use. [40]

Latitude: The angle between the equatorial plane and the straight line passing through that point and through (or close to) the center of the Earth is the "latitude" of a point on the Earth's surface. Lines connecting points of the same latitude trace circles on the Earth's surface known as parallels because they are parallel to the Equator and each other.

Longitude: The angle east or west of a reference meridian to another meridian that passes through a point on Earth's surface is called "longitude". All meridians are halves of large ellipses (also known as great circles) that meet at the North and South Poles.

The combination of these two components specifies the position of any location on the surface of Earth.

Geocoding

The process of taking a text-based description of a location, such as an address or the name of a place, and returning geographic coordinates, most commonly a latitude/longitude pair, to identify a location on the Earth's surface is known as address geocoding. A computer representation of address points, the street / road network, as well as postal and administrative boundaries, is used in geocoding. [41]

Mapbox

Mapbox's [42] web services APIs allow programmatic access to the company's tools and services. The Mapbox APIs are broken down into four categories: maps, navigation, search, and accounts. The geographic coordinates of a given location were obtained using the Mapbox Geocoding API from the Search service. There are two endpoints in the Geocoding API: mapbox.places and mapbox.places-permanent. Because the goal was to learn geographic co-ordinates of places, Mapbox.places was used. The forward geocoding query type allows you to look up a specific location by name and get its geographic coordinates back. Mapbox was choosen because of its easy to use and cost free service.

Latitude and Longitude from Mapbox

To do the forward geocoding following endpoint was used:
$https : //api.mapbox.com/geocoding/v5/mapbox.places/\{place - name\}.json$
Here place-name is a string value.
The response to a Geocoding API request is an object that contains the following properties:

- type: <string> : a GeoJSON type object.
- query: <array> : An array of space and punctuation-separated strings from the original query.
- features: <array> : Returned features are ordered by relevance.
- attribution: <string> : Attributes the results of the Mapbox Geocoding API to Mapbox.

Latitude and longitude for addresses were added to the dataset.

Not Found Response

For some addresses, Mapbox did not provide a response in terms of latitude and longitude. "Chapainawabganj" is one such example. Even though the official name is "Chapainawabganj," Geocoding services refer to it as "Nawabganj". "Jessore" is another example. The district airport is still known as "Jessore Airport", even though the official name has been changed to "Jashore" [43]. This kind of inconsistency led to some oddities. These cases were later dealt with on a case-by-case basis.

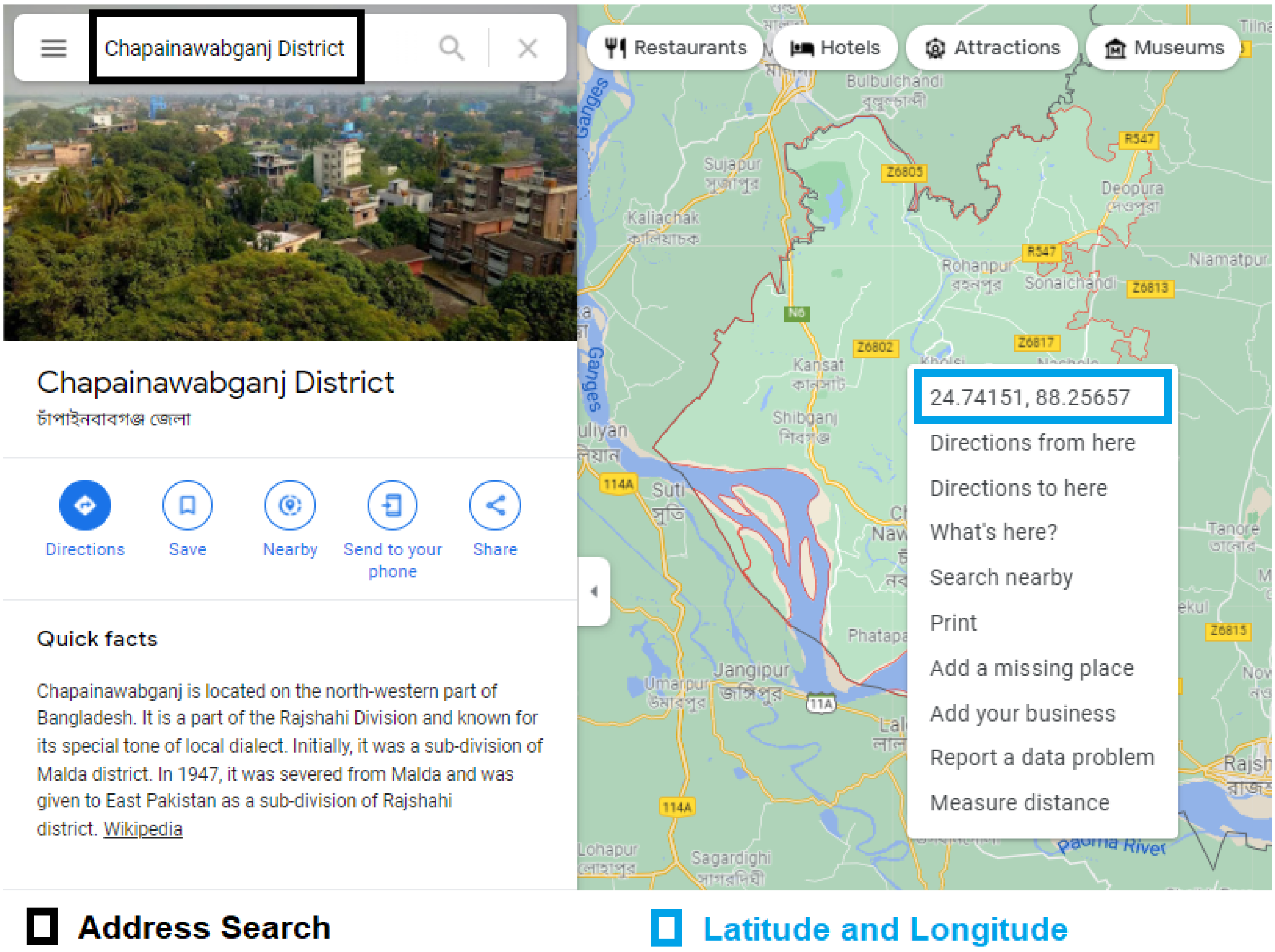
Incorrect Latitude and Longitude

After analyzing Mapbox responses, it was discovered that some of the Latitude and Longitude were incorrect. Meaning that the Latitude and Longitude found for a valid address in Bangladesh are actually outside of Bangladesh. The reason for such a response is still unknown. Then, using Bounding Box, all of the Latitude and Longitude were checked to see if the geocodes were correct.

Bounding Box: A bounding box [44] in geographic coordinates is an area defined by minimum and maximum longitudes and latitudes.

Manual Gecoding Address by Address

Every incident area where no response was received or where an incorrect response was received was searched on Google Map. The latitude and longitude were obtained from the results of such Google Maps search.



Chapainawabganj District

Chapainawabganj District

চাঁপাইনবাবগঞ্জ জেলা

Quick facts

Chapainawabganj is located on the north-western part of Bangladesh. It is a part of the Rajshahi Division and known for its special tone of local dialect. Initially, it was a sub-division of Malda district. In 1947, it was severed from Malda and was given to East Pakistan as a sub-division of Rajshahi district. Wikipedia

24.74151, 88.25657

Figure 7.1: Collecting Latitude and Longitude from Google Map

7.1.2 Weather information

Weather is the state of the atmosphere, describing for example the degree to which it is hot or cold, wet or dry, calm or stormy, clear or cloudy. Climate refers to the averaging of atmospheric conditions over longer periods of time, whereas weather refers to day-to-day temperature, precipitation, and other atmospheric conditions.

Weatherstack

The Weatherstack [45] API delivers accurate weather data for any application and use case, from real-time and historical weather information all the way to 14-day weather forecasts, supporting all major programming languages. It covers global weather data across the board from a multi-year history all the way to live information and accurate weather forecasts. The Weatherstack provides the following services:

- Real-Time Weather API
- Historical Weather API
- Weather Forecasts API
- Location Autocomplete
- Bulk API Endpoint

Historical Weather Information from Weatherstack

Historical Weather API was chosen because the incident in the dataset dates back to 2012. Users can look up historical weather data from 2008 using the Historical Weather API. However, the free Weatherstack plan does not include access to the Historical Weather API. For access to the Historical Weather API service, a Weatherstack standard plan [46] was purchased.

To get the historical weather report, following endpoint was used:
$https : //api.weatherstack.com/historical?access - key = \{ACCESS - KEY\}\&query = \{co - ordinates\}\&historical - date = \{incident - date\}\&hourly = 1\&interval = 24$
Here,

- co-ordinates: <string> : The latitude and longitude of the incident place is provided as coma-sperated value, such as - "<latitude>, <longitude>"
- Incident-date: <string> : The date of the incident is given.
- hourly: <number :: 1 or 0> : Set this parameter to 1 to ask the API to
- return weather data split hourly.
- interval: <number> : Set this parameter to 24 to get the day average.

Response from Weatherstack

In addition to the requested historical weather data, a successful historical weather API call will also return the current weather in the location used for the request, as well as information about the API request and location. The following is the necessary information returned by a successful API call:

- Maximum Temperature
- Minimum Temperature
- Average Temperature
- Weather Code
- Weather Description
- Precipitation
- Humidity
- Visibility
- Cloud Coverage
- Heat Index

These weather data from Weatherstack were added to the dataset.

7.2 Feature Extraction

Feature extraction increases the accuracy of learned models by extracting features from the input data. This phase of the general framework reduces the dimensionality of data by removing redundant data. Of course, it increases training and inference speed. The methods of feature extraction obtain new generated features by doing the combinations and transformations of the original feature set.

A feature extraction mechanism computes numeric or symbolic information from the data, which are referred to as the features. Finally, the classifier is fed with the extracted features, and the decision is made by classifying these features.

7.2.1 Data Adjustment

For ease of use and finding unique values, the entire dataset was converted to lowercase. The precision of latitude and longitude ranged from 6 to 8 digits in floating point precision. All latitude and longitude values were rounded to the nearest six digit floating point precision.

7.2.2 Feature Value Type: Unique Values, Floating Numbers and Sequentiality

A categorical variable tends to have less unique values than a numerical one. It also presents no floating numbers compared to what can be observed in a numerical one.
A feature will considered to be categorical if:

- It has very few unique values (but not only)
- It contains a character or a string
- It contains integers (but not only)

A feature will considered to be numerical if:

- It has a significant amount of unique values
- It contains floating numbers
- It contains integers (but not only)

When a categorical variable is comprised of integers, the unique values are naturally sequential. A numerical variable will contain non sequential modalities.

7.2.3 Date

Using Python's Datetime module [39], the incident date was converted to a Datetime Object. The features listed below were extracted from this Datetime object.

- Year
- Month
- Day
- Weekday

These features were later included in the dataset.

7.2.4 Season

Season feature were introduced to the dataset. The feature was created with the help of Month feature. Season is another categorical feature. There are three types of categories in it. Below is a list of them.

- **Hot :** March-May
- **Rainy:** June-October
- **Winter:** November-February

7.2.5 Weekend

New feature Weekend were added to dataset. This feature is derived from weekday. It specifies whether the incident occurred on a weekend or not. If the incident occurred on a Friday or Saturday during the week, it was a weekend; otherwise, it was not. This feature is divided into two categories: True and False.

7.2.6 Weather

Weather information is contained in six features: precipitation, humidity, visibility, cloud cover, heat index, and average temperature. Strings were used to store the values of these features in the dataset. The values of all features were converted to Int and Float.

Correlation of the Weather Features

Correlation between the six weather features were calculated. The following figure describes the correlation between those features.

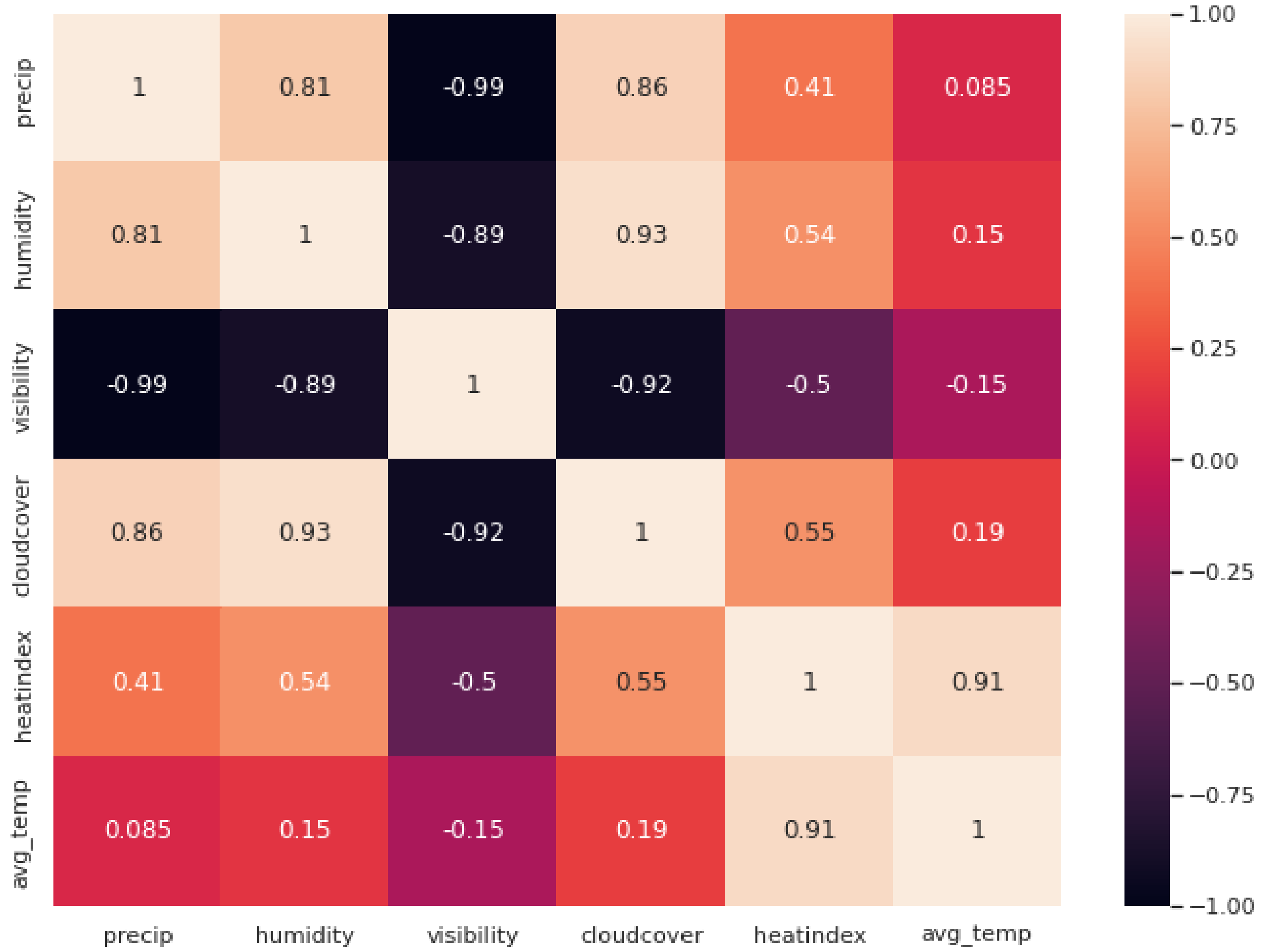


Figure 7.2: Correlation of Six Weather Features

The following are the conclusions drawn from Figure 7.2.

- Positive Correlation Between Average Temperature and Heat Index of 91.
- Positive Correlation Between Cloud Cover and Humidity of 93.
- Positive Correlation Between Cloud Cover and Precipitation of 86.
- Negative Correlation Between Cloud Cover and Visibility of -92.
- Negative Correlation Between Humidity and Visibility of -89.
- Negative Correlation Between Visibility and Precipitation of -99.

Weather features were grouped into three categories.

- Average Temperature and Heat Index.
- Cloud Cover and Humidity.
- Visibility and Precipitation.

Weather Feature Importance

The importance of each feature must be calculated before deciding which one should be chosen from these three categories. These feature values must be normalized before being used to calculate feature importance. Min-Max normalization was used for this purpose. The results of the feature importance calculation are shown below.

Table 7.1: Feature Importance of Weather Features

Feature	Importance
Cloud Cover	0.258
Humidity	0.206
Precipitation	0.202
Heat Index	0.166
Average Temperature	0.134
Visibility	0.031

Heat Index, Cloud Cover, and Precipitation were preserved in the dataset.

Weather Description

Categories on this feature were "sunny", "partly cloudy", "cloudy", "moderate or heavy rain shower", "thundery outbreaks possible", "patchy light rain with thunder", "torrential rain shower", "moderate rain at times", "overcast", "light rain shower", "patchy rain possible", "moderate rain", "mist", "moderate or heavy rain with thunder", "patchy light rain", "light drizzle", "heavy rain", "patchy light drizzle", "fog", "heavy rain at times", "light rain". These categories were reduced to Normal, Rainy, and Winter with the help of domain expertise.

7.2.7 Converting Features from Numerical to Categorical

Precipitation

Precipitation values were numerical. Rainfall intensity can classified according to the rate of precipitation, which depends on the considered time [47]. Precipitation values were converted to categorical values with the help of following classification.

Table 7.2: Rain intensity classification according to Precipitation

Intensity	Precipitation Rate
No Rain	rate = 0.0
Light Rain	0.0 <rate <2.5
Moderate Rain	2.5 <rate <10
Heavy Rain	10 <rate <50
Violent Rain	50 <rate

Cloud Cover

Cloud cover refers to the fraction of the sky obscured by clouds when observed from a particular location. This feature was converted to four categorical values. They are described as [48]:

Table 7.3: Different Types of cloud cover

Type	Cloud Cover
Clear	cover <10
Scattered	10 <cover <50
Broken	50 <cover <90
Overcast	90 <cover

Heat Index

The heat index (HI) is an index that combines air temperature and relative humidity, in shaded areas, to posit a human-perceived equivalent temperature, as how hot it would feel if the humidity were some other value in the shade. Effects of the heat index can be categorized as [49]:

Table 7.4: Heat Index Shade Values

Shade	Temperature (Celsius)
Normal	below 26
Cautious	26 - 32
Extreme Cautious	33 - 41
Danger	42 - 54
Extreme Danger	over 54

According to these shade values, the heat index was converted to categorical values.

7.2.8 Month

There are 12 categories in the month feature. When encoding is used, it becomes 12 distinct features. These categories must be reduced in order to achieve lower dinmentionality. The visualization of the Month feature category with respect to crime count is shown below.

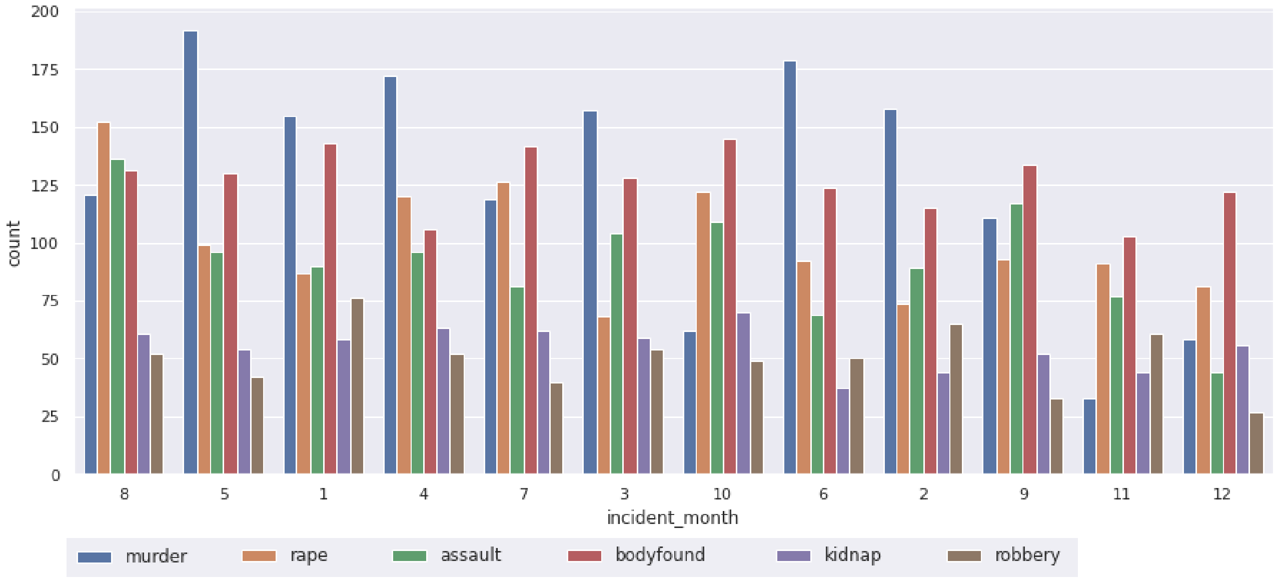


Figure 7.3: Crime Count per Month

Months 1 and 2 follow a similar pattern, as shown in figure 7.3. However, no other patterns were discovered between the months. So, based on the observation below, the Month feature was converted to the 11 category.

Table 7.5: Crime Frequency Order on Months

Class	Months	Frequency Order
1	8	Rape >Assault >Body Found >Murder >Kidnap >Robbery
2	5	Murder >Body Found >Rape >Assault >Kidnap >Robbery
3	1,2	Murder >Body Found >Assault >Rape >Robbery >Kidnap
4	4	Murder >Rape >Body Found >Assault >Kidnap >Robbery
5	7	Body Found >Rape >Murder >Assault >Kidnap >Robbery
6	3	Murder >Body Found >Assault >Rape >Kidnap >Robbery
7	10	Body Found >Rape >Assault >Kidnap >Murder >Robbery
8	6	Murder >Body Found >Rape >Assault >Robbery >Kidnap
9	9	Body Found >Assault >Murder >Rape >Kidnap >Robbery
10	11	Body Found >Rape >Assault >Robbery >Kidnap >Murder
11	12	Body Found >Rape >Murder >Kidnap >Assault >Robbery

7.2.9 Weekday

Reducing category helps with lower dimentionality after encoding. The visualization of the Weekday feature category with respect to crime count is shown below.

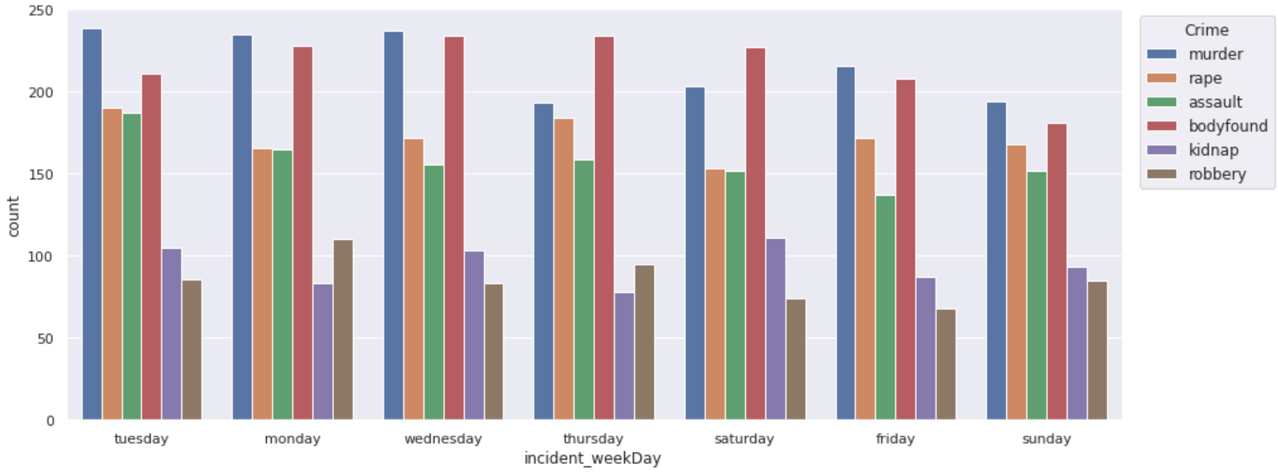


Figure 7.4: Crime Count per Weekday

Tuesday, Wednesday, Friday, Sunday have similar pattern. Rest weekdays do not have similar pattern with other weekdays. Based on this observation Weekday category were reduced to 4.

Table 7.6: Crime Frequency order on Weekdays

Class	Weekdays	Frequency Order
1	Tuesday, Wednesday Friday, Sunday	Murder >Body Found >Rape >Assault >Kidnap >Robbery
2	Monday	Murder >Body Found >Rape >Assault >Robbery >Kidnap
3	Thursday	Body Found >Murder >Rape >Assault >Robbery >Kidnap
4	Saturday	Body Found >Murder >Rape >Assault >Kidnap >Robbery

7.2.10 Angular Distance for Weekday and Month

One-time encoding can be used to generate a boolean feature for each day of the week. This solution provides information on the day of the week, but it does not provide any relationships between the days. In this case, the sequence of the days is irrelevant.

If the data is to be used to train machine learning models, this is not the correct way to encode days of the wee. Saturday is actually closer to Monday than Wednesday. The sense of data is altered when days of the week are encoded as numbers.

The information about the circular nature of weeks, months, and the actual distance between the days should not be lost. As a result, the day of week feature can be encoded as "points" on a circle, with 0° representing Monday, 51.5° representing Tuesday, and so on. There is one problem. Because it is a circle, the difference between Sunday and Monday for a machine learning model is 308.5° rather than 51.5°. That is incorrect.

The cosine and sine values of the degree must be calculated in order to solve the problem. For different inputs, both functions produce duplicate outputs, but when used together, unique pairs of values are guaranteed.

Following the exploration of the explained idea, a new circular feature from weekday and another circular feature from month were added to the dataset.

7.2.11 Distance Incident area, District, Division

As a new feature, the distance between the incident area and the corresponding District city was added to the dataset. Also the distances between the District city and the corresponding Divisional city were introduced to the dataset. Latitude and Longitude values were used to calculate the distance. The Haversine formula was used to calculate distance from latitude and longitude.

Haversine formula

The haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes. Important in navigation, it is a special case of a more general formula in spherical trigonometry, the law of haversines, that relates the sides and angles of spherical triangles.

7.3 Encoding

Encoding is a technique of converting categorical variables into numerical values so that they could be easily fitted to a machine learning model. These are different encoding techniques.

7.3.1 One Hot Encoding

In this method, we map each category to a vector that contains 1 and 0 denoting the presence of the feature or not. The number of vectors depends on the categories which we want to keep. For high cardinality features, this method produces a lot of columns that slows down the learning significantly.

7.3.2 Label Encoding

Label Encoding is a popular encoding technique for handling categorical variables. Label Encoding refers to converting the labels into a numeric form so as to convert them into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning.

7.3.3 Mean Encoding

Mean encoding is similar to label encoding, except here labels are correlated directly with the target. For example, in mean target encoding for each category in the feature label is decided with the mean value of the target variable on a training data. The advantages of the mean target encoding are that it does not affect the volume of the data and helps in faster learning.

Chapter 8

Result and Performance Analysis

In the final step, crime predictions were made using test data. For a more in-depth analysis, predicted values are compared to original values using an evaluation function. The results of the proposed models are presented and evaluated in this chapter using various evaluation metrics. This chapter is divided into three sections for convenience. The first section presents the performance of supervised classifiers, while the second section analyzes the performance of ensemble learning. In addition, in the third section, a comparative analysis of various models has been discussed.

8.1 Supervised Classifier

8.1.1 Logistic regression

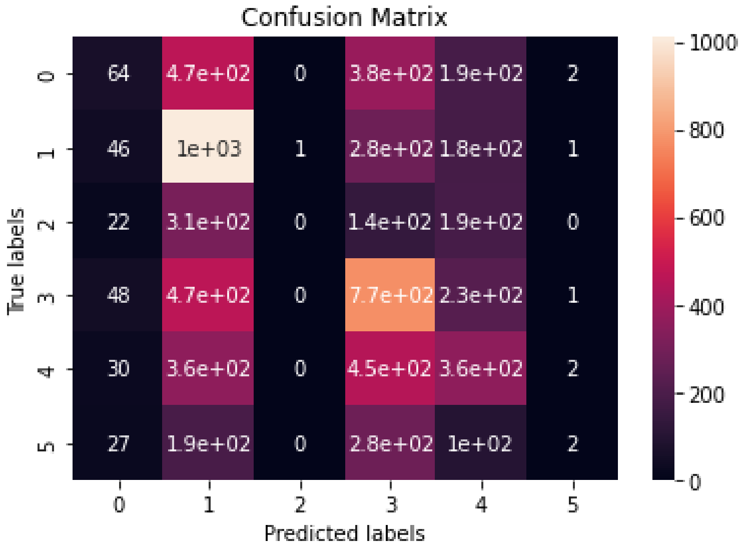


Figure 8.1: Confusion Matrix for Logistic Regression

Table 8.1: Performance of Logistic Regression

Crime	Precision	Recall	F1	Accoracy
Assault	0.27	0.06	0.10	37.97
Body Found	0.36	0.66	.047	
Kidnap	0.00	0.00	0.00	
Murder	0.33	0.51	0.40	
Rape	0.29	0.30	0.30	
Robbery	0.25	0.00	0.01	

8.1.2 Naive Bayes

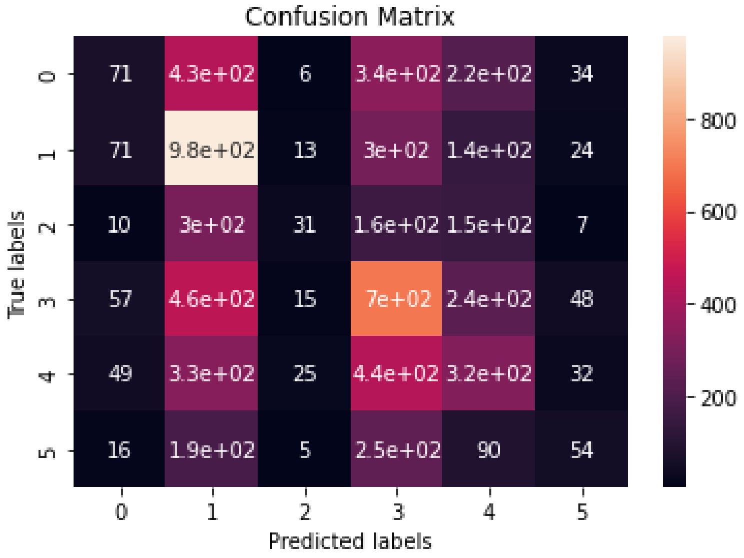


Figure 8.2: Confusion Matrix for Naive Bayes

Table 8.2: Performance of Naive Bayes

Crime	Precision	Recall	F1	Accuracy
Assault	0.26	0.06	0.10	36.00
Body Found	0.36	0.64	0.46	
Kidnap	0.33	0.05	0.08	
Murder	0.32	0.46	0.38	
Rape	0.28	0.27	0.27	
Robbery	0.27	0.09	0.14	

8.1.3 SVM

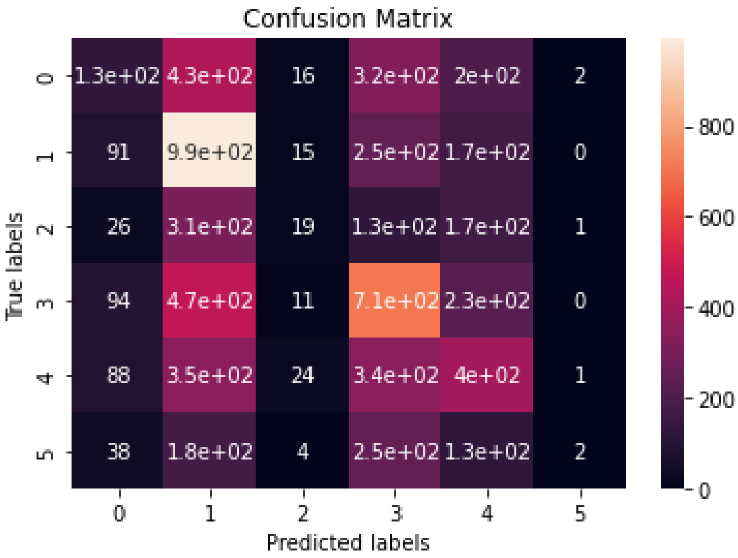


Figure 8.3: Confusion Matrix for SVM

Table 8.3: Performance of SVM

Crime	Precision	Recall	F1	Accuracy
Assault	0.28	0.12	0.17	38.12
Body Found	0.36	0.65	0.47	
Kidnap	0.21	0.03	0.05	
Murder	0.35	0.47	0.40	
Rape	0.31	0.33	0.32	
Robbery	0.33	0.00	0.01	

8.1.4 Decision Tree

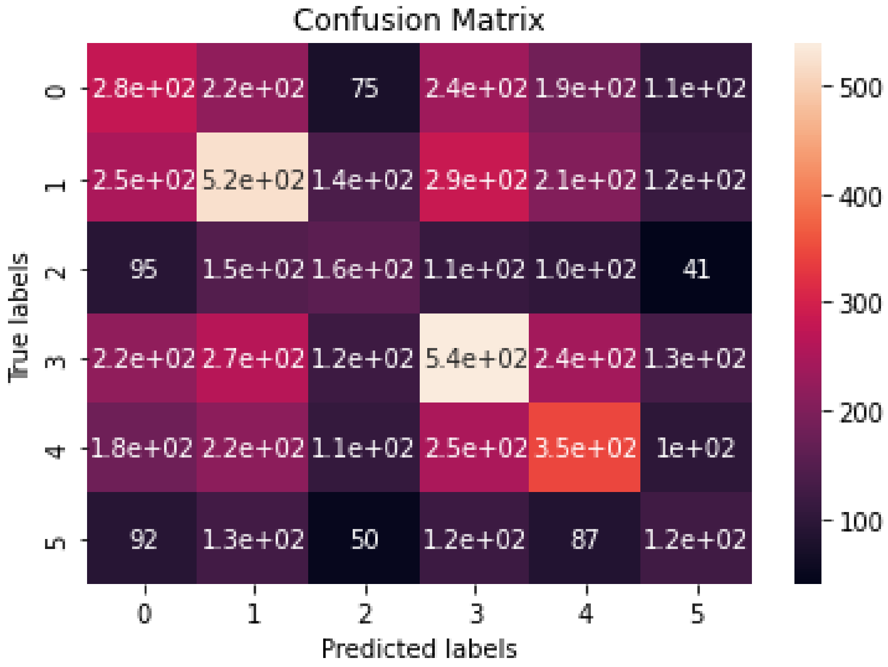


Figure 8.4: Confusion Matrix for Decision Tree

Table 8.4: Performance of Decision Tree

Crime	Precision	Recall	F1	Accuracy
Assault	0.25	0.25	0.25	34.74
Body Found	0.36	0.35	0.36	
Kidnap	0.24	0.24	0.24	
Murder	0.35	0.35	0.35	
Rape	0.28	0.28	0.28	
Robbery	0.20	0.22	0.21	

8.1.5 KNN

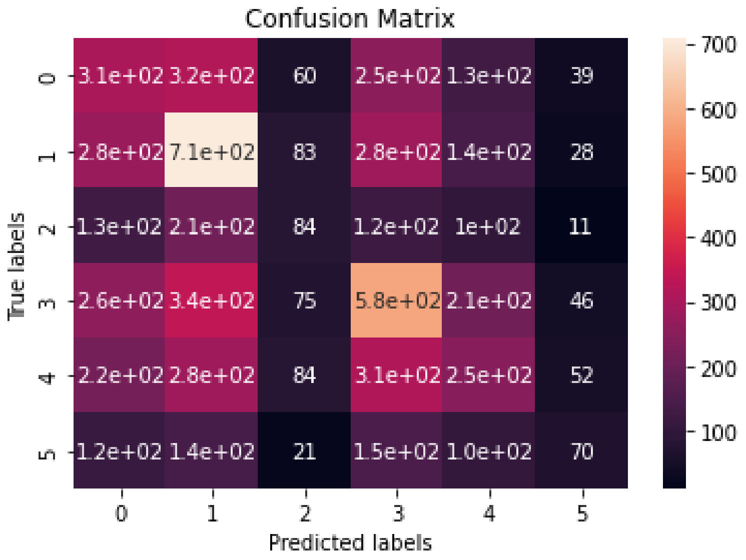


Figure 8.5: Confusion Matrix for KNN

Table 8.5: Performance of KNN

Crime	Precision	Recall	F1	Accuracy
Assault	0.23	0.28	0.25	33.05
Body Found	0.35	0.46	0.40	
Kidnap	0.21	0.13	0.16	
Murder	0.34	0.38	0.36	
Rape	0.26	0.21	0.23	
Robbery	0.28	0.12	0.17	

8.2 Ensemble Learning

8.2.1 Random Forest

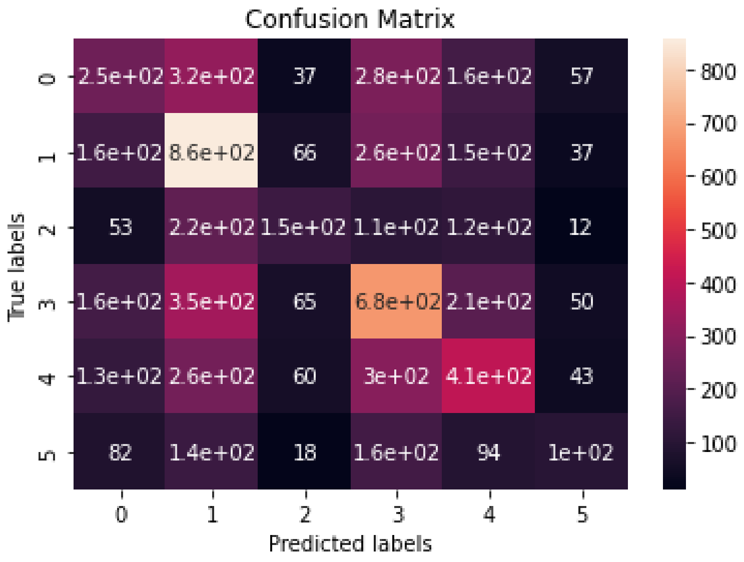


Figure 8.6: Confusion Matrix for Random Forest

Table 8.6: Performance of Random Forest

Crime	Precision	Recall	F1	Accuracy
Assault	0.30	0.23	0.26	40.33
Body Found	0.40	0.56	0.46	
Kidnap	0.36	0.21	0.27	
Murder	0.38	0.45	0.41	
Rape	0.34	0.32	0.33	
Robbery	0.33	0.17	0.23	

8.2.2 Extra Tree

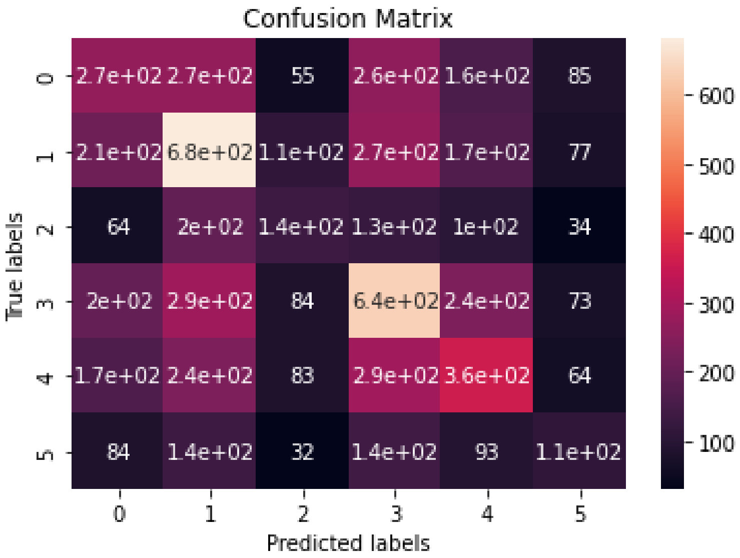


Figure 8.7: Confusion Matrix for Extra Tree

Table 8.7: Performance of Extra Tree

Crime	Precision	Recall	F1	Accuracy
Assault	0.28	0.26	0.27	36.60
Body Found	0.37	0.43	0.40	
Kidnap	0.28	0.22	0.24	
Murder	0.36	0.42	0.39	
Rape	0.31	0.28	0.29	
Robbery	0.25	0.19	0.21	

8.2.3 Adaboost

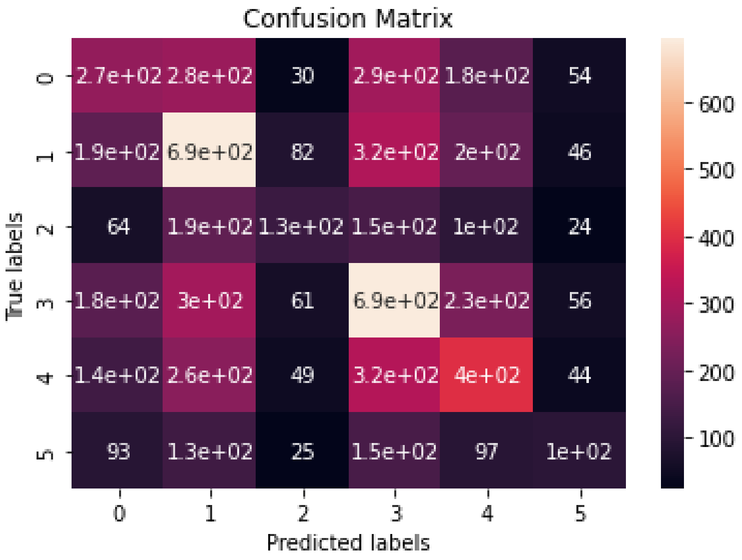


Figure 8.8: Confusion Matrix for AdaBoost

Table 8.8: Performance of AdaBoost

Crime	Precision	Recall	F1	Accuracy
Assault	0.30	0.24	0.27	39.57
Body Found	0.38	0.52	0.44	
Kidnap	0.42	0.17	0.24	
Murder	0.37	0.49	0.42	
Rape	0.37	0.34	0.35	
Robbery	0.37	0.16	0.23	

8.2.4 XGBoost

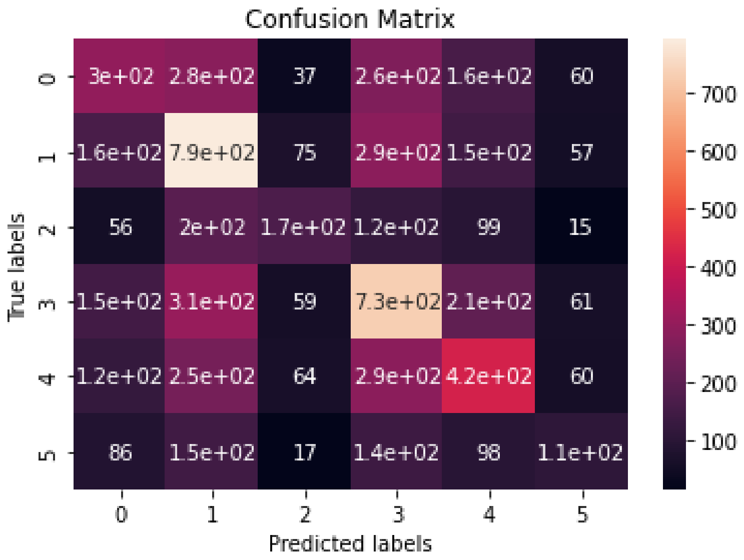


Figure 8.9: Confusion Matrix for XGBoost

Table 8.9: Performance of XGBoost

Crime	Precision	Recall	F1	Accuracy
Assault	0.34	0.27	0.30	41.50
Body Found	0.40	0.52	0.45	
Kidnap	0.40	0.26	0.32	
Murder	0.40	0.48	0.44	
Rape	0.37	0.35	0.36	
Robbery	0.30	0.18	0.23	

8.3 Comparative analysis

Table 8.10: Comparative performance analysis of different algorithms

Algorithm	Accuracy
Logistic Regression	37.97
Naive Bayes	36.00
SVM	38.12
Decision Tree	34.74
KNN	33.05
Random Forest	40.33
Extra Tree	36.60
AdaBoost	39.57
XGBoost	41.50

Chapter 9

Conclution and Future Works

9.1 Conclusion

It is now much easier to find relationships and patterns among various data sets thanks to machine learning technology. The main goal of this study was to predict the type of crime that is likely to occur given spatio temporal features. Nine different types of algorithms were examined in order to predict the type of crime in this study. With a precision of 41.50 %, Random Forest was the most accurate. The models are still in the developmental stages. When combined with other classifying algorithms and more data in the future, it is expected to produce better results.

9.2 Limitations

Our work was not without its difficulties; we encountered some roadblocks along the way like those of many other scientific studies, our work have some limitations. They are briefly described as follows:

- There is a dataset limitation. There are only 6600 criminal records in this dataset. Only those crimes that were reported in the newspaper were gathered.
- There are only a few features related to crime. It was difficult to collect socioeconomic data.
- This dataset is imbalanced. From 2019 to 2012, there were approximately 600 kidnap and robbery criminal records available. However, around 1500 murder criminal records were added during the same time period.

9.3 Future Works

- Increase the number of criminal records in this dataset. Include human trafficking, narcotics, smuggling, and other criminal records.
- Applied models will be fine-tuned and customized more.
- Explore algorithms like CatBoost, LogitBoost, LGBM, Bagging, and others.
- Handleling missing values in other ways, such as by clustering them or making predictions based on dataset.
- Make the dataset more balanced by using oversampling and undersampling techniques.

References

[1] "The daily star." https://www.thedailystar.net/.

[2] J. Q. Yuki, M. M. Q. Sakib, Z. Zamal, K. M. Habibullah, and A. K. Das, "Predicting crime using time and location data," in *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, pp. 124–128, 2019.

[3] T. T. Nguyen, A. Hatua, and A. H. Sung, "Building a learning machine classifier with inadequate data for crime prediction," *Journal of Advances in Information Technology Vol*, vol. 8, no. 2, 2017.

[4] S. Kim, P. Joshi, P. S. Kalsi, and P. Taheri, "Crime analysis through machine learning," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 415–420, IEEE, 2018.

[5] S. Hossain, A. Abtahee, I. Kashem, M. M. Hoque, and I. H. Sarker, "Crime prediction using spatio-temporal data," in *International Conference on Computing Science, Communication and Security*, pp. 277–289, Springer, 2020.

[6] A. Bharati and D. S. RA, "Crime prediction and analysis using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 09, 2018.

[7] "Bangladesh police crime statictics." https://www.police.gov.bd/en/crime_statistic/year/2018.

[8] "Geographic profiling." https://en.wikipedia.org/wiki/Geographic_profiling.

[9] "Recidivism." https://en.wikipedia.org/wiki/Recidivism.

[10] M. Al Boni and M. S. Gerber, "Area-specific crime prediction models," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 671–676, IEEE, 2016.

[11] H. K. R. ToppiReddy, B. Saini, and G. Mahajan, "Crime prediction & monitoring framework based on spatial analysis," *Procedia computer science*, vol. 132, pp. 696–705, 2018.

[12] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim, and G. Sinha, "An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach," *IEEE Access*, vol. 9, pp. 67488–67500, 2021.

[13] A. S. Hornby and J. Turnbull, *Oxford advanced learner's dictionary of current English*. Oxford: Oxford University Press, 2010.

[14] B. A. Garner and H. C. Black, *Black's Law Dictionary*. St. Paul, MN: West, 2009.

[15] Dr. Md. N. Islam, *Crime and society*. Tasmia publication, Dhaka, 2013.

[16] "Bangladesh penal code." http://bdlaws.minlaw.gov.bd/act-11.html.

[17] "Bangladesh police act." http://bdlaws.minlaw.gov.bd/act-12.html.

[18] "Wife burned to death by husband." https://www.thedailystar.net/wife-burned-to-death-by-husband-29229.

[19] "Teenage girl gang-raped in uttara." https://www.dhakatribune.com/uncategorized/2015/07/31/teenage-girl-gang-raped-in-uttara.

[20] "13-yr-old boy beaten to death." https://icrfoundation.org/home/13-yr-old-boy-beaten-to-death-video-shows-brutality-stirs-outra

[21] "Blogger avijit roy killing." https://www.thedailystar.net/tags/blogger-avijit-roy.

[22] "Bangladeshi secular publisher hacked to death." https://www.bbc.com/news/world-asia-34688245.

[23] "Oishee killed her parents." https://www.thedailystar.net/tags/oishee-rahman.

[24] "Journalist couple killed." https://www.thedailystar.net/news-detail-222110.

[25] "Murder of abrar fahad." https://www.thedailystar.net/frontpage/news/buet-student-beaten-death-critical-fb-post-costs-

[26] "Upazilas in bangladesh." https://en.wikipedia.org/wiki/Upazila.

[27] "Bangladesh govt portal." http://ddm.portal.gov.bd.

[28] "San francisco crime dataset." https://www.kaggle.com/c/sf-crime/data.

[29] "Denver crime dataset." https://www.kaggle.com/paultimothymooney/denver-crime-data.

[30] "London crime dataset." https://www.kaggle.com/LondonDataStore/london-crime.

[31] "Vancouver crime dataset." https://www.kaggle.com/wosaku/crime-in-vancouver.

[32] "Los angeles crime dataset." https://www.kaggle.com/cityofLA/crime-in-los-angeles.

[33] "Bangladesh police crime statictics." https://www.police.gov.bd/en/crime_statistic/year/2019.

[34] "The daily star news archive." https://www.thedailystar.net/newspaper.

[35] "Beautifulsoup4." https://pypi.org/project/beautifulsoup4/.

[36] "Regular expression." https://en.wikipedia.org/wiki/Regular_expression.

[37] "Fuzzywuzzy." https://pypi.org/project/fuzzywuzzy/.

[38] "Levenshtein distance." https://en.wikipedia.org/wiki/Levenshtein_distance.

[39] "Datetime." https://docs.python.org/3/library/datetime.html.

[40] "Geographic coordinate system." https://en.wikipedia.org/wiki/Geographic_coordinate_system.

[41] "Geocoding." https://en.wikipedia.org/wiki/Address_geocoding.

[42] "Mapbox." https://www.mapbox.com/.

[43] "English spellings changes of five district." https://www.dhakatribune.com/bangladesh/2018/04/02/english-spellings-chittagong-comilla-barisal-jessore-bogra-chang

[44] “Bounding box.” https://wiki.openstreetmap.org/wiki/Bounding_Box.

[45] “Weatherstack.” https://weatherstack.com/.

[46] “Weatherstack plan.” https://weatherstack.com/product.

[47] “Rain.” https://en.wikipedia.org/wiki/Rain.

[48] “Cloud coverage.” http://sciencenetlinks.com/lessons/measuring-cloud-coverage/.

[49] “Heat index.” https://en.wikipedia.org/wiki/Heat_index.