Bangladesh Division Based Crime Rate Prediction Using Machine Learning and Deep Learning

Submitted By

Saimuna Sarker Tuli

ID: CSE 01906797

Ashraful Quader

ID: CSE 02006893

A thesis is submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering

Port City International University

7-14, Nikunja Housing Society, South Khulshi, Chattogram, Bangladesh

May 2023

Bangladesh Division Based Crime Rate Prediction Using Machine Learning and Deep Learning

Submitted By

Saimuna Sarker Tuli

ID: CSE 01906797

Ashraful Quader

ID: CSE 02006893

A thesis is submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering

Under the Supervision of

Umma Saima Rahman

Lecturer

Department of CSE

Port City International University

Department of Computer Science and Engineering

Port City International University

7-14, Nikunja Housing Society, South Khulshi, Chattorgram, Bangladesh

May 2023

DECLARATION OF ORGINALITY

We announce that to the best of the best of the author's knowledge, this thesis has not been submitted anywhere for the award of any degree. All references and acknowledgements to other researchers have been given as appropriate. We also ensure that we just used the materials that were mentioned. All formulations and notions taken directly or substantially from printed or unprinted literature, or the Internet, have been properly cited with footnotes or other exact references to the original source, as per standard scientific procedure. We're careful that delivering false information could achieve authentic ramifications.

(Signature of the candidate)

Saimuna Sarker Tuli
ID: CSE 01906797
Batch: 19-D-A
Department of CSE
Port City International University

(Signature of the candidate)

Ashraful Quader
ID: CSE 02006893
Batch: 20-E-A
Department of CSE
Port City International University

APPROVAL FOR SUBMISSION

This thesis titled **"Bangladesh Division Based Crime Rate Prediction Using Machine Learning & Deep Learning"** by Saimuna Sarker Tuli (CSE01906797) and Ashraful Quader (CSE 02006893) has been approved for submission to the Department of Computer Science and Engineering, Port City International University, in partial fulfillment of the requirement for the degree of Bachelor of Science (Engineering).

(Signature of the supervisor)

Umma Saima Rahman
Lecturer
Department of Computer Science and Engineering
Port City International University

DEDICATION

Our efforts are sincere tributes to our beloved parents and deserving teachers.

ACKNOWLEDGEMENT

We completed our thesis work with the help of Almighty Allah, who is the Most Gracious and Merciful. We really appreciate **Mrs. Umma Saima Rahman**, our thesis advisor, for her patient guidance and support throughout the thesis' conceptualization and execution. Our capacity to extend our viewpoints and discover significant direction is made possible by her extensive knowledge and willingness to spend time with us. Additionally, we would like to thank the entire faculty of the computer science and engineering department.

ABSTRACT

Predicting the different type of crime activity of each Division is paramount in order to prevent it. Law enforcement agencies can work effectively and respond faster if they have better knowledge about crime patterns in different geological points of a city. The aim of this paper is to use machine learning and deep learning techniques to prediction of crime rate based on division. The experiment is carried predicated on a dataset of around 2010–2019 crime records. Deep learning and machine learning regression models were employed to understand the prediction. CNN's model has the highest accuracy of all of them at 97%. Accuracy using ANN was 95%. The accuracy of various machine learning regression techniques, such as Random Forest, Decision Tree, Support Vector and Gradient Boosting, was approximately 93%, 91%, 89%, and 86%, respectively.

Keywords: Crime, Division, Prediction, Machine Learning, Deep Learning, Evaluation Matrices.

Table of Contents

List of Figures

List of Tables

CHAPTER 1
Introduction

1.1 Introduction

Crime refers to any conduct or activity that is deemed illegal or against the law. This can cover a wide range of actions, from less serious crimes like small theft and fatal violations to more serious ones like murder, robbery, kidnapping and theft. Depending on the gravity of the violation and the potential consequences, crimes are often divided into misdemeanors and felonies. Felons are more serious crimes that can result in lengthier prison sentences or even the death penalty in some circumstances, but misdemeanors are typically thought of as less serious acts and are punished by fines, community service, or short-term incarceration. The concept of crime varies from one country to another, and even within a country, from one state or province to another. In Bangladesh, crime has drastically escalated during the past several years. A few of the factors that affect the country's crime statistics are poverty, political upheaval, and the rise of organized criminal groups. Dacoity, Speedy Trial, Riot, Repression of Women and Children, Arms Act, Explosive, Narcotics, and Smuggling are common forms of crime in Bangladesh, particularly in metropolitan areas. Robberies and break-ins are also rather typical. In some regions of the country, there is an issue with violence towards individuals, such as murder and assault. Recently, the Bangladeshi government took steps to lower crime there. This calls for increasing police presence and toughening criminal penalties. The government has also taken action to combat corruption, which is seen to be a significant factor in the nation's crime problem.

In this paper worked on the crime rate of different divisions of Bangladesh. Its target was to identify which division had the most and the least crime. In this research we have worked on several types of crime in 2010 to 2019 years of Bangladesh .and the crime rates have been found out in different division of Bangladesh.

1.2 Problem Statement

To develop a reliable model that can accurately predict the crime rate in different regions of the country. In addition to endangering public safety, Bangladesh's high crime rate has a debilitating effect on the nation's economic and social structure. Forecasting crime rates accurately can help law enforcement agencies prevent crimes and maintain public safety by allowing them to take preventative action. It is challenging to predict crime rates due to the intricacy of crime and the large range of factors that affect it. In order to find a solution, it is required to evaluate a sizable amount of data from several sources, including crime statistics, demographics, economic indicators, and social factors. This will allow us to identify patterns and trends that may be utilized to accurately predict crime rates. Additionally, there are moral and legal issues that must be taken into consideration when gathering and evaluating data because crime is such a sensitive subject. Privacy concerns also need to be taken into consideration when dealing with individual data. Therefore, developing a model that can accurately predict crime rates while also addressing these concerns is a crucial challenge for researchers and policymakers in Bangladesh.

1.3 Motivation

To provide a more detailed understanding of crime patterns in various parts of the nation, division-based crime rate forecast in Bangladesh was created. Bangladesh is divided into eight administrative areas, each with its own distinct social, economic, and demographic characteristics. The overall crime rate in Bangladesh is alarming, although it can vary greatly between regions and even within a single city or town. Identifying patterns and trends in crime rates at the divisional level can help law enforcement agencies use resources more wisely and carry out targeted measures to deter crime.

The latest triumphs of machine learning and deep learning techniques serve as inspiration for developing an efficient model. For instance, if the model indicates that a certain division would have a high crime rate, policymakers might utilize this knowledge to direct resources and initiatives to that area in order to address the underlying causes of crime.

1.4 Objectives

The major objective of division-based crime rate prediction in Bangladesh is to create a model that can correctly forecast crime rates at the divisional level. This strategy will help law enforcement agencies and decision-makers implement focused initiatives to lower crime and maintain public safety. Finding patterns and trends in the divisional crime rates by processing and analyzing data from multiple sources. Locating locations with a high crime rate and recommending to law enforcement and decision-makers what particular measures might be taken to minimize crime there. Evaluating the effectiveness of the targeted initiatives implemented in high-crime regions and modifying the model and suggestions as necessary. The main goal is division-based crime rate prediction is to give law enforcement agencies and policymakers a more thorough and detailed understanding of crime patterns in Bangladesh and to assist them in taking targeted actions to reduce crime and ensure public safety at the divisional level.

1.5 Thesis Orientation

Chapter 1: Explain introduction about research works.

Chapter 2: Consist of literature review that describe about the prior work on crime rate prediction.

Chapter 3: Describe the implementation methodology which consist of the system structure, data collection, pre-processing, splitting, parameter tuning.

Chapter 4: Consist of the details about the hardware were used.

Chapter 5: Represent the outcomes, model testing and comparative analysis.

Chapter 6: Finalizes the thesis project with conclusion and discusses its potential future applications.

CHAPTER 2
Literature Review

Though crime rate prediction recently yet few accessible works already done by this topic. This session summarizes some of the previous work which uphold the strategies & enhance willingness to works. Consistently Approachable literature on this field represent the consequences of different process. So, here we hit off this literature into various useful crime rate prediction and analysis. Entirely, related work on crime rate prediction and challenges are represented.

2.1 Introduction

Considering all these aspects in this paper, we proposed deep learning model namely as CNN and ANN. As well as different types of machine learning model as like –Random Forest, Decision Tree .Support Vector, Gradient Boosting.

2.2 Related Work

In computer-assisted prediction process, [1] In order to predict crime rate, S. Mahmud, M. Nuha, and A. Sattar experiment basis on general people data. By using safe route they gather data and determine the suitable criminal pattern using statistical analysis of data. And they find out crime rate in different section like age based, gender based, area based and monthly. And also they wished to discover hot zone region by their precision and would like to use CNN model to analyze picture information and add the Google API for viewing hot zone. For their purpose they use Linear Regression, Naive Bayes, and K-Nearest Neighbor as classification methods. Accuracy for three methods are 73%, 69%, 76%. D. M. Raza and D. B. Victor [2] they work on 8 division of Bangladesh. According to their work they predict division which is based on crime numbers instead of opposite find crime rate based on division.

In [3] they use Bangladesh Police Statistics Bangladesh Police as the source for their dataset. This dataset contains crime data in a given division over the course of a year, i.e., in aggregate form. It lacks detailed spatio-temporal information. In addition, these authors do not consider demographics in their studies, which is an important feature to predict crime. [4] According their purpose they uses different approaches used for predicted features are caused crime in a locality or region. Basically they want find out the occurrence of crime. And their predictive model accuracy was 75.90% for Decision Tree, 83.39% for Random Forest, 77.64% for Naïve Bayes, 64.72% for Linear Regression. In [5], crime prediction and analysis methods are very important to detect the future crimes and reduce them. Their purpose is to highlight the worth and effectiveness of machine learning in predicting violent crimes occurring in a particular region .Their models accuracy was 88% for Decision Tree, 87 % for KNN, 88 % for Extratress, 16 % for ANN , 66% for SVM.

[6] In order to forecast crime trends in Bangladesh, Awal et al. Awal et al. [2016] investigate a linear regression model. The authors use some aggregate data from Bangladesh Police sources. After training the model, crime forecasting for robbery, murder, women and children Repression, kidnapping, and other crimes in the various regions of Bangladesh is attempted. Their experimental findings indicate that the majority of crimes are on the rise as the population increases. For the purpose of predicting crime areas of interest, [7] used a spatial-fleeting model that included KNN, Linear Discriminant Analysis (LDA), and factual techniques based on histograms. A major challenge regarding crime prediction is analyzing large crime datasets accurately and efficiently. Data mining is utilized to find hidden patterns in large crime datasets quickly and efficiently. The increased efficiency and reduced errors in crime data-mining techniques increase the accuracy of crime prediction. A general data-mining framework was developed in [8] based on the experience of the Coplink project, conducted at the University of Arizona. Most research in crime prediction is focused on identifying crime hotspots, which refers to the areas in which the crime rates are above the average level.

A trustworthy prediction model for predicting crime trends in metropolitan areas was developed in [9] using a methodology based on the Auto-Regressive Integrated Moving Average model (ARIMA). For forecasting crimes in the city of San Francisco, three algorithms—KNN, Parzen windows, and neural networks—were created, evaluated, and contrasted in [10]. A comparison of the violent crime trends from the communities and crime unnormalized dataset with the actual crime statistics for the state of Mississippi was done in [11]. On the dataset for communities and crime, linear regression, additive regression, and decision stump algorithms were used. Out of the three methods, linear regression had the best performance. [12] Parvez et al. Parvez et al. [2016] propose a spatio-temporal street crime prediction model that exploits street crime data of Dhaka City. Their dataset is obtained from Dhaka Metropolitan Police (DMP), which consists of the records of crimes from June 2013 to May 2014 but only in aggregate form.

CHAPTER 3
Methodology

In this chapter briefly explain the system design and discuss the overall the processes of operation of the system, which has been used to find out division wise crime rate of Bangladesh.

3.1 Problem Description

The main view of point is predict the crime rate which is based on division. This system will depend on the different types crime categories in each "**Division**" .As input we take crime categories to get to know which division is more in crime zone. So we need to find out the crime of so it is a regression tasks in this chapter we use machine learning and deep learning regression models.

3.2 Data Collection

We collected the dataset from the kaggle website.There were datasets from 2010 to 2019 and taken from this link https://www.police.gov.bd/en/crime_statistic/year/2019 . The crime dataset had crime information from 2019 to 2010 with 190 rows and 19 columns where there was a total of 3610 data. For our experiment, we merged and dropped some data.

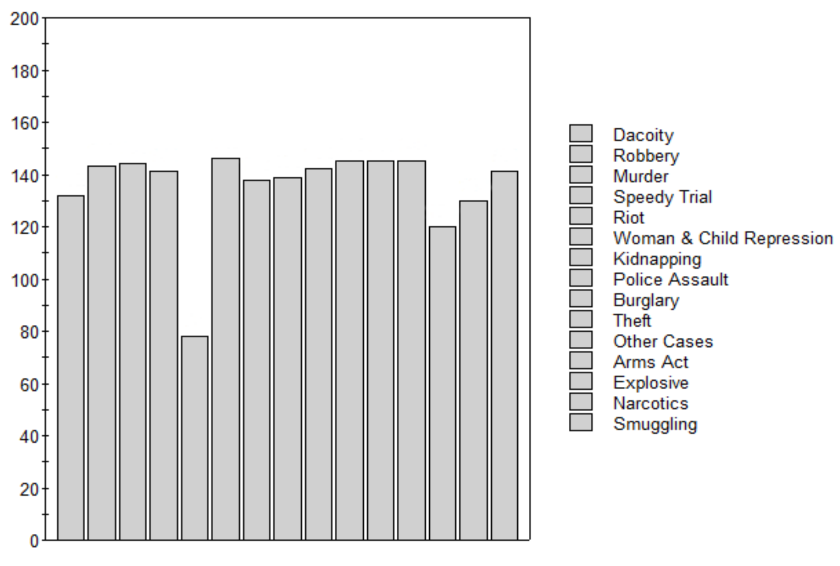

Figure 3.1: Sample data of crime categories from 2010 to 2019.

3.3 Attributes of Crime Dataset

Here, we display some representative information from the crime dataset, including Unit Name, Year, and Total Cases. We collaborate with these units such as DMP, CMP, KMP, RMP, BMP, SMP, Dhaka Range, Mymensingh Range , Chittagong Range , Sylhet Range , Khulna Range, Barisal Range, Rajshahi Range, Rangpur Range, GMP, RPMP.Our experiment was based on from 2010 to 2019 of crime statistics.Only the total cases for each unit for 2010 and 2019 are shown in Table 3.1.

Table 3.1: Sample features of crime data

Unit Name	Year	Total Cases
DMP	2010	23519
CMP	2010	4063
KMP	2010	1767
RMP	2010	1571
BMP	2010	1139
SMP	2010	1484
Dhaka Range	2010	3089
Mymensingh Range	2019	1004
Chittagong Range	2019	2396
Sylhet Range	2019	725
Khulna Range	2019	1585
Barisal Range	2019	895
Rajshahi Range	2019	2011
Rangpur Range	2019	1472
GMP	2019	244
RPMP	2019	121

This table provides the different types of crimes along with their respective total numbers for the sample years of 2010, 2015, and 2019.These are the input features used in our experiment. Dacoity, Robbery, Murder, Speedy Trial, Riot, and womenChildRepression, Kidnapping, Police Assault, Theft, Other Cases, Arms Act, Explosives, Drugs, and Smuggling.

Table 3.2: Sample features of crime data

Crime Name	Crime Number	Year
Dacoity	656	2010
Robbery	1059	2010
Murder	3988	2010
Speedy Trial	1666	2010
Riot	130	2010
WomenChildRepression	212110	2015
Kidnapping	805	2015
PoliceAssult	634	2015
Burglary	2495	2015
Theft	6821	2015
Other Cases	5428	2019
Arms Act	174	2019
Explosive	30	2019
Narcotics	9069	2019
Smuggling	361	2019

3.4 Workflow

We used machine learning and deep learning to make these predictions. We made several crime rate predictions over different types of crime. First we preprocessed the data and put all the data together on merge then I applied for the model. Even though the data was supposed to be preprocessed we checked if they had any doubts. After preprocessing then we split the data between train and test among these, 80% and 20% gave the best results. Then we tuning the parameters to see which performance was the highest value point.

3.5 System Architecture

The research's primary area of interest is system architecture. What is being done in this study is understandable to any researcher. I decided to clarify and make the system architecture easier to grasp. These crucial points are provided:

- ❑ Crime data collection
- ❑ Data pre-processing
- ❑ Data splitting
- ❑ Ml or Dl algorithm
- ❑ Result

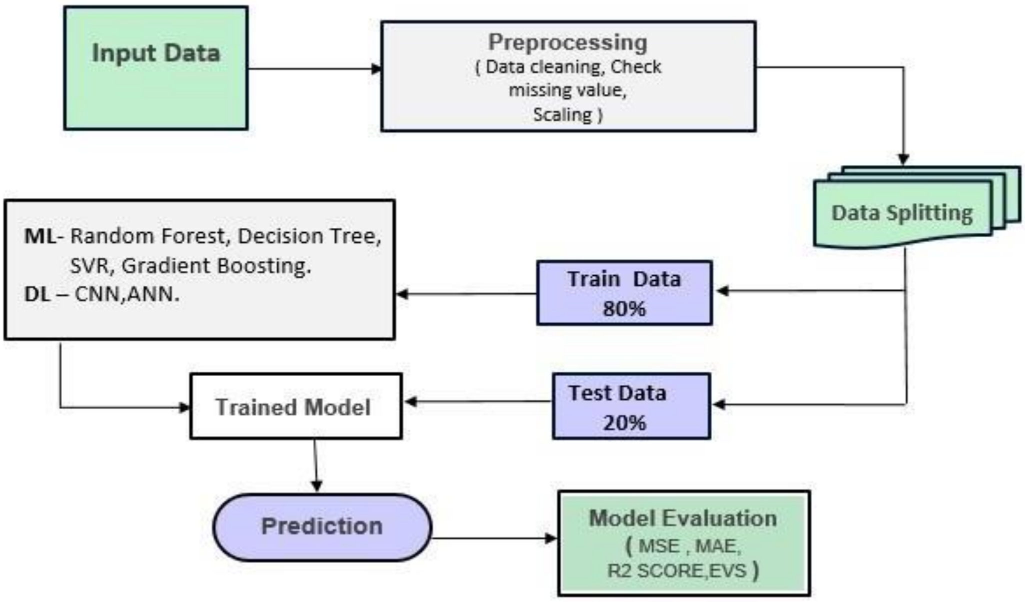

Figure 3.2: System Architecture

Each and every single important principle in the section must be followed by the student or researcher. Because every point results in a certain output. We have to do those things in that order.

3.6 Data Pre-processing

Data Cleaning: First, check data for missing values, duplicate data, or any outliers that might be skewing your results. Clean up y data by removing any problematic data points or filling in missing values.

Scaling Feature: It is an important step for numerical data as it helps to normalize the data so that all features have the same scale. We used MinMaxScaler method for scaling features.

MinMaxScaler: MinMaxScaler is a normalization technique that scales and translates features to a specified range. It is often used in machine learning algorithms to improve the stability and performance of the models. MinMaxScaler scales the data in a way that all features lie between a specified minimum and maximum values. The formula for MinMaxScaler is:

X_scaled = (X - X.min(axis=0)) / (X.max(axis=0) - X.min(axis=0))

where X is the input data, X.min(axis=0) and X.max(axis=0) are the minimum and maximum values of X along the specified axis, and X_scaled is the scaled data.

This normalization technique can be particularly useful when working with features that have different scales, or when working with algorithms that are sensitive to the scale of the input features. However, it's important to note that MinMaxScaler assumes that the distribution of the data is uniform, so it may not be appropriate for all types of data.

Feature Selection: Feature selection is the process of selecting the most important features for models. We can use techniques such as correlation analysis or feature importance to identify the most important features.

Data Visualization: Visualize the data to explore any patterns or trends in the data, identify any outliers, and gain insights into the distribution of the data.

Model Selection: Select an appropriate machine learning algorithm that can handle regression tasks. Some possible algorithms are Decision Trees Regression, Random Forest Regression, Support Vector Regression, and Neural Networks in regression.

Reshaping the Data: Convolutional Neural Networks and Artificial Neural Networks typically require data to be in a specific format, such as a 2D or 3D array. Therefore, you may need to reshape your data to fit the input shape of your model.

3.7 Splitting the Dataset

Divide the dataset into training and testing sets, and then train the chosen model using the trainig set. It's crucial to remember that the testing set should only be utilized for assessment after model We did several train test ratio like 70% for training 30% for testing, 75% for training 25% for testing, 80% for training 20% for testing. From them 80% train set and 20 % test set seems to be the best ratio as it giving the highest accuracy in deep learning model and for machine learning. We manually tuned the parameter in order to choose the optimum one. We experimented by altering the parameter and ultimately chose one that improved accuracy and decreased loss.

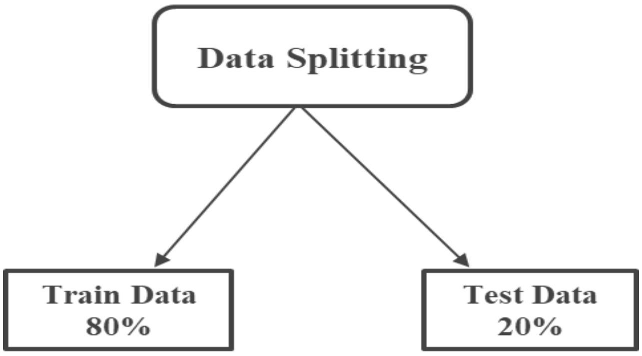

Figure 3.3: Splitting Dataset

3.8 Machine Learning

Computer systems may learn from input data and make predictions or judgments based on it thanks to the statistical models and algorithms developed in the field of machine learning. In essence, it is a technique for instructing computers to spot patterns in data and utilize those patterns to forecast the future or take action. Machine learning comes in a variety of forms, such as supervised learning, unsupervised learning, and reinforcement learning. A model is trained using supervised learning on a labeled dataset, where each input data point has a matching label or output. Unsupervised learning is training a model on an unlabeled dataset in which the model is left to its own devices to discover links and patterns in the data. Through reinforcement learning, a model is taught to make choices depending on input from its surroundings, such as game or simulation feedback.

3.8.1 Supervised Learning

In supervised learning, a subset of machine learning, the algorithm is trained using a labeled dataset. The labeled dataset consists of input data and corresponding output data, or labels. During training, the approach tries to learn a mapping function that converts input data to the appropriate output or label. Once trained, the algorithm may produce predictions or conclusions based on fresh, unstudied data. Regression and classification are the two primary subtypes of supervised learning. In order to conduct our research, we used regression models.

3.8.2 Regression Models

Regression method is a supervised learning technique in machine learning that involves predicting a continuous numerical output or target variable based on a set of input or predictor variables. In this process, a model learns from a training dataset to identify the mathematical relationship between the predictor variables and the target variable. The output of a regression model is a continuous numerical value. The model is trained to identify the relationship between the input variables and the target variable, which is typically represented by a function or equation. Once the model is trained, it can be used to predict the target variable for new data points based on their input variables.

3.8.3 Machine Learning Regression Algorithms

Four machine learning regression methods applied in our problem.

- ❑ Random Forest Regression.
- ❑ Decision Tree Regression.
- ❑ Support Vector Regression.
- ❑ Gradient Boosting Regression.

Random Forest Regression

Random forest regression is a popular supervised learning algorithm used in machine learning for predicting continuous numerical output. It is a type of ensemble learning algorithm that combines multiple decision trees to make predictions. It is a development of decision tree regression, which makes use of numerous decision trees to increase the predictability and accuracy. Building several decision trees, each based on a distinct part of the training data and a separate subset of the predictor variables, is how random forest regression works. In order to form a decision tree during training, the algorithm randomly chooses a portion of the features and a subset of the training data. To create a "forest" of decision trees, this process is done several times. Random forest regression constructs a number of decision trees, each based on a different subset of the predictor variables and a different portion of the training data. The method selects a subset of the training data and a subset of the features at random to build a decision tree.

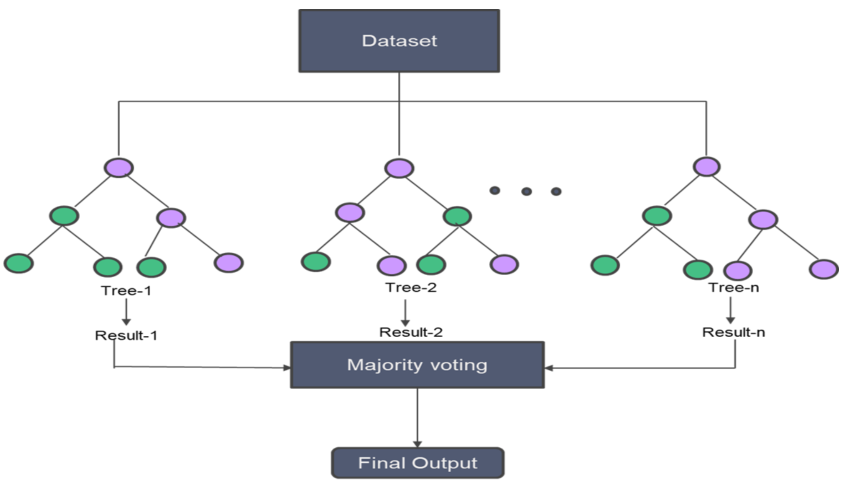

Figure 3.4: Random Forest Regression

Decision Tree Regression

A machine learning approach called Decision Tree Regression is utilized for regression tasks. It operates by fitting a straightforward model, such as a constant or linear function, to the target variable inside each smaller zone after recursively dividing the input space into smaller regions based on the values of input features. The finished model is a tree structure where each leaf denotes a prediction, each branch denotes a potential value for that feature, and each internal node denotes a feature. The decision tree regression algorithm divides the data into subsets depending on the feature that most effectively separates the data based on some standard, usually the mean squared error (MSE) or variance reduction.

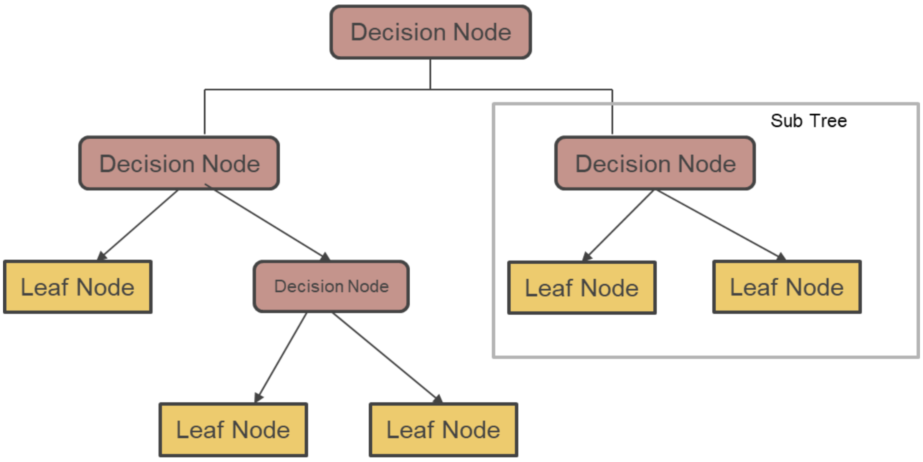

Figure 3.5: Decision Tree Regression

Support Vector Regression

Support Vector Regression (SVR) is a machine learning algorithm that is used for regression tasks. It is a type of Support Vector Machine (SVM) that uses a linear or nonlinear kernel function to map the input data into a higher-dimensional feature space, where the regression problem is easier to solve. The fundamental principle underlying SVR is to use a kernel function, like a radial basis function (RBF), to convert the input data into a high-dimensional space, and then train a linear regression model in this new space. A portion of the training data points known as support vectors, which are closest to the predicted values, constitute the hyperplane that the algorithm learns.

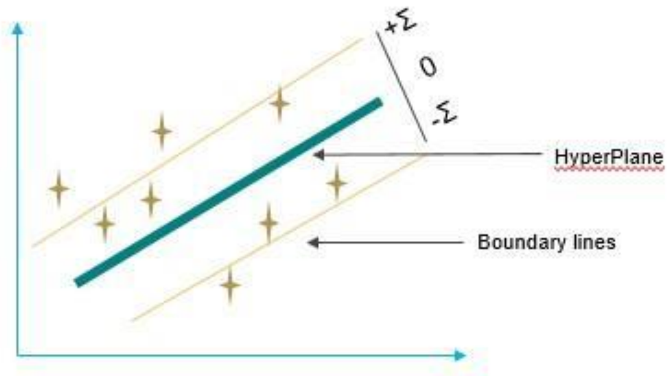

Figure 3.6: Support Vector Regression

Gradient Boosting Regression

Gradient Boosting Regression is a machine learning algorithm that combines multiple decision trees to create a strong ensemble model that can make accurate prediction. It is based on the idea of combining multiple weak learners, typically decision trees. The algorithm iteratively trains a sequence of decision trees to correct the errors of the previous trees and uses a loss function to measure the difference between the predicted values and the actual values. GBR can handle both regression and classification tasks and is known to produce highly accurate predictions.

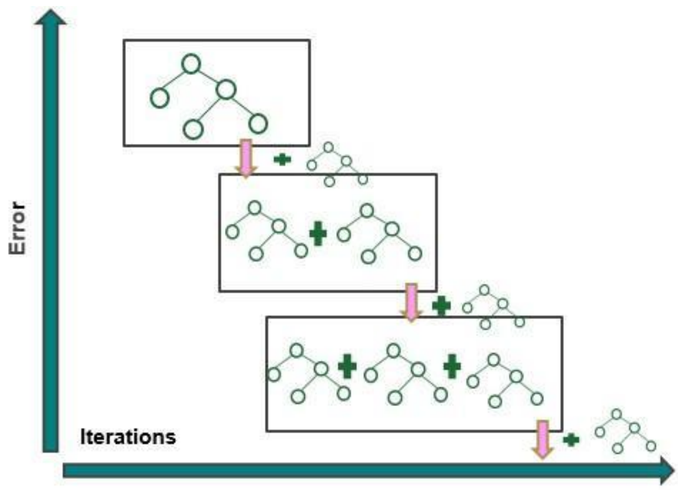

Figure 3.7: Gradient Boosting Regression

3.8.4 Parameter Tuning

Parameter tuning is an essential step in machine learning to improve the performance of the models. Here's a brief overview of parameter tuning for some popular regression algorithms:

Random Forest Regression

Random Forest is an ensemble learning method that creates multiple decision trees and aggregates their outputs to make a final prediction. Here are the key hyper parameters to tune in RF model:
n_estimators: the number of trees in the forest.
max_depth: the maximum depth of each decision tree.
min_samples_split: the minimum number of samples required to split an internal node.
min_samples_leaf: the minimum number of samples required to be at a leaf node.

Decision Tree Regression

Decision Tree is a popular decision-making algorithm that creates a tree-like model of decisions and their possible consequences. Here are the key hyper parameters to tune in DT model:
max_depth: the maximum depth of the tree.
min_samples_split: the minimum number of samples required to split an internal node.
min_samples_leaf: the minimum number of samples required to be at a leaf node.

Support Vector Regression (SVR)

Support Vector Regression is a regression algorithm that uses support vectors to find the optimal hyperplane that separates the data points into different classes. Here are the key hyper parameters to tune in SVR:
C: the regularization parameter that controls the trade-off between achieving a low training error and a low testing error.
kernel: the kernel function used to transform the data into a higher-dimensional space.
epsilon: the tolerance for errors in the prediction.

Gradient Boosting Regression (GBR)

Gradient Boosting is an ensemble learning method that uses decision trees as base learners and aggregates their outputs to make a final prediction. Here are the key hyper parameters to tune in GBR:
n_estimators: The number of trees in the forest.

learning_rate: The step size shrinkage used to prevent overfitting.
max_depth: The maximum depth of each decision tree.
min_samples_split: The minimum number of samples required to split an internal node.
min_samples_leaf: The minimum number of samples required to be at a leaf node.

3.8.5 Deep Learning

In order to evaluate and interpret complex data, deep learning includes training artificial neural networks with numerous layers. It draws inspiration from the way the human brain works, which is made up of linked neurons that process information.

Deep learning methods gradually extract higher-level features and patterns using a hierarchical structure of layers, each of which carries out a particular computation on the input data. Following the combination of these qualities, predictions or classifications concerning new data are made. The CNN and ANN models were utilized for the experiment.

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a well-known deep learning architecture that draws inspiration from the human brain's built-in visual perception system. t is a Deep Learning method that can take in an input image, give distinct aspects and objects in the image weights and biases that can be learned, and then distinguish between them. Without human assistance, CNN can recognize the primary feature. As CNN deals perfectly with computer vision and picture classification we have chosen this algorithm to over problem. In CNN there are mainly five layers - Input layer, Convolution layer, Pooling layer, fully connected layer, Output layer

Numerical data can be directly fed into the input layer of the CNN, which is followed by one or more convolutional layers that extract features from the data. The output of the convolutional layers is then flattened and fed into a fully connected layer that performs high-level feature extraction and produces the final output. CNNs can be used for both text and numerical data by converting them into a format that can be processed by the convolutional layers. For text data, word embedding's are used, while numerical data is directly fed into the input layer. In some cases, a hybrid model can be used to combine both types of data. The convolutional layers extract useful features from the data, which are then used to make accurate predictions or classifications. Convolutional Neural Network (CNN) can also be used for regression problems. Here's a brief overview of how CNNs can be used for regression:

Input Layer: The input layer takes in the data, which can be either numerical or image data, and prepares it for processing by the convolutional layers.

Convolution layer: To accomplish the convolution operation, a portion of the numerical data is joined to the Convolutional layer as well as determining the dot product between the receptive field and the filter. The convolutional layer consists of multiple filters, each of which extracts a specific feature from the input data. In 1D CNNs, these filters slide over the input data along the time axis, capturing local patterns. The output of each filter is called a feature map.
Additionally, the Convolutional layer has ReLU activation to set all negative values to zero. The ReLU (rectified linear unit) layer applies an activation function to the output of the convolutional layer. The activation function is a simple mathematical function that introduces non-linearity into the network, allowing it to learn more complex patterns. CNNs can be used for regression problems by modifying the output layer of the network to produce a continuous output instead of a categorical one. Here is a brief overview of how CNNs can be used for regression:

ReLU Layer: The ReLU (rectified linear unit) layer applies an activation function to the output of the convolutional layer, introducing non-linearity into the network and allowing it to learn more complex patterns.

Pooling Layer: The pooling layer reduces the spatial dimensionality of the feature maps by aggregating nearby values. Max-pooling is commonly used in CNNs for regression.

Flatten Layer: The flatten layer converts the 2D feature maps into a 1D vector that can be fed into a fully connected layer.

Drop Out: Dropout is a regularization technique commonly used in regression and Convolutional Neural Networks (CNNs). It helps prevent overfitting by randomly dropping out a fraction of the neurons during training. In regression tasks, dropout is typically applied to the fully connected layers of the network. It works by temporarily disabling a portion of the neurons, forcing the network to learn more robust and generalized representations. During training, each neuron has a probability of being "dropped out" or set to zero. This dropout rate is a hyperparameter that needs to be tuned.

Fully Connected Layer: The fully connected layer performs high-level feature extraction and produces the final output.

Linear activation: The linear activation function is often used in the output layer of a CNN model for regression tasks. In such cases, the goal is to predict a continuous numerical value

Output Layer: The output layer produces a continuous output value instead of a categorical one, which is suitable for regression problems. The output layer can be a single neuron or multiple neurons, depending on the complexity of the problem.

CNNs can be used for regression problems by modifying the output layer of the network to produce a continuous output instead of a categorical one. The convolutional layers extract local patterns and features from the input data, which are then used to make accurate predictions.

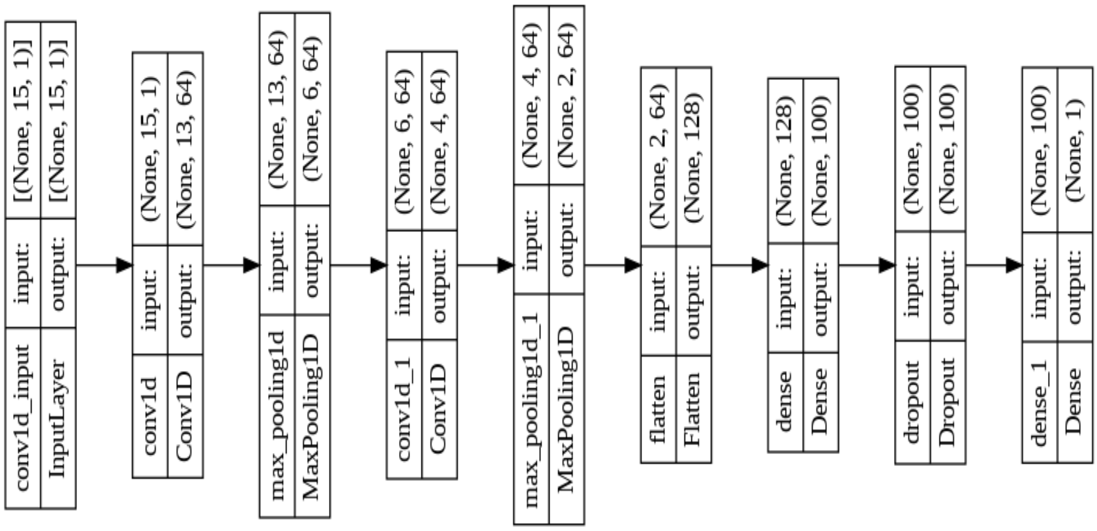

Figure 3.8: Block Diagram of CNN model.

Artificial Neural Networks (ANN)

ANN stands for Artificial Neural Network, which is a type of machine learning algorithm modeled after the structure and function of the human brain.

An ANN is composed of multiple layers of interconnected nodes, or "neurons," that receive input data, process it through a series of mathematical operations, and produce an output. Each neuron performs a simple calculation using the input it receives and a set of learned weights and biases. The output of one layer of neurons is used as input for the next layer, and the process repeats until the final output is produced.

Artificial Neural Networks (ANNs) can also be used for regression problems. Here's a brief overview of how ANNs can be used for regression:

Input Layer: The input layer of an ANN takes in the data, which can be either numerical or categorical, and prepares it for processing by the hidden layers.

Hidden Layers: The hidden layers consist of multiple neurons that perform non-linear transformations on the input data. The number of hidden layers and neurons in each layer can be adjusted based on the complexity of the problem.

Activation Function: Each neuron in the hidden layers applies an activation function to the input data, introducing non-linearity into the network. Common activation functions used in regression problems include the ReLU and linear functions.

Output Layer: The output layer of an ANN produces a continuous output value, which is suitable for regression problems. The output layer can have a single neuron or multiple neurons, depending on the complexity of the problem.

Loss Function: The loss function measures how well the network is performing on the regression task. Common loss functions used in regression problems include mean squared error (MSE) and mean absolute error (MAE).

Optimization: The optimization algorithm updates the weights and biases of the neurons in the network based on the loss function, with the goal of minimizing the error on the training data. Common optimization algorithms used in regression problems include stochastic gradient descent (SGD) and Adam.

Training and Validation: The network is trained on a subset of the data, called the training data, and validated on another subset of the data, called the validation data. This helps prevent overfitting and ensures that the network is generalizing well to new data.

ANNs can be used for regression problems by adjusting the output layer of the network to produce a continuous output value. The hidden layers perform non-linear transformations on the input data, and the network is trained to minimize the error on the training data using an optimization algorithm and a loss function. The network is then validated on a separate subset of the data to ensure that it is generalizing well to new data.

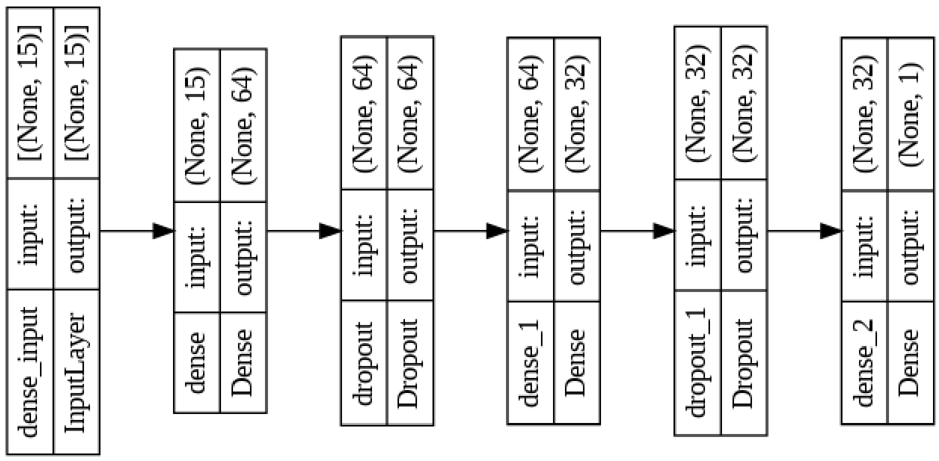

Figure 3.9: Block Diagram of ANN model.

CHAPTER 4
Hardware Implementation & Toolkit

4.1 Tools and Devices

The following tools are used to implement the intended framework: Python and Colab by Google.

4.1.1Colab

Google Research created colaboratory, sometimes referred to as "Colab." Everyone has the ability to create and execute incompatible Python code using Colab, which is best suited for AI, data analysis, and guidance. Additionally, Colab is an integrated Jupyter scratch pad structure that provides access to resources like GPUs without the requirement for a strategy. Google Colab is undoubtedly the simplest way to provide us access to robust GPU resources for your machine learning project. Data analysis using machine learning is a specialty of Colab. Through the browser, anybody may write and run arbitrary Python code.

4.1.2 Python

Python is one of the most user-friendly programming languages and general-purpose languages because it emphasizes natural language, has a simple syntax, isn't overly complex, and can be used to create software and websites, automate processes, and analyze data. As a result, it can be used to create a wide range of applications and isn't specifically designed to address any particular problems.

Compared to other programming languages, Python is one of the most well-liked and simple to learn and use, making it feasible to build and execute scripts rapidly. All of the following are included: classes, dynamic typing, very high level dynamic data types, exceptions, modules, and exception handling. The relevant Python libraries were utilized during this inquiry. Include –

Pandas: Pandas is largely used in Python for applications involving data analysis and machine learning. It is constructed on top of Numpy, a different package that supports multi-dimensional arrays. It is also built on top of Numpy, a different open-source program that supports multi-dimensional arrays. Another kind of information that Pandas offers are row labels. If we wish to arrange and refer to data naturally, this is helpful.

Numpy: NumPy is mostly used for working with numerical numbers because applying mathematical functions is made simple since it makes things simple. Matrix operations, the Fourier transform, and functions for working with linear algebra are also included.

Keras: A sophisticated deep learning API for creating neural networks is Google's Keras. It is used to make neural network implementation easy and is built in Python. Furthermore, it enables the calculation of several backend neural networks.

Tensor flow: One suitable toolset for numerical computing that speeds up and makes it easier to create neural networks and machine learning algorithms is TensorFlow. TensorFlow is used by many different technologies, including image identification, text-based applications, voice search, and many more.

Scikit-learn: A machine learning tool is Scikit-learn. Both supervised and unsupervised learning are supported by a Python program called Scikit-learn. It also provides a wide range of unique utilities, such as an information pre-processing tool, as well as tools for improvement, determination, and assessment.

Matplotlib: A Python module called Matplotlib may be used to produce 2D graphs and plots using Python programs. By letting you manage line styles, font attributes, axis formatting, and other features, its pyplot package makes plotting easier. It is a cross-platform toolkit for data visualization and graphical charting for Python and its numerical extension NumPy.

Seaborn: A Python data visualization package built on matplotlib is called Seaborn. It provides a sophisticated user interface for producing attractive and educational statistics visuals.

OS (Operating System): Python's OS module offers a variety of methods that let programmers communicate with the operating system they are using. In this tutorial, we'll study the principles of file management as well as how to create and remove directories and folders, as well as how to rename them.

4.3 Hardware Implementation

We utilized the following hardware instruments for implementation:

- Windows 10 Pro supports an operating system in 64 bit and 8 GB of RAM.
- The 2.40GHz and 2.50GHz processors in Intel Core i3 models.
- 12 GB of colab's virtual memory.

After reading Deep Learning on Windows, one will be able to develop deep learning models and web apps on the Windows 10 pro 64-bit operating system. Artificial intelligence enthusiasts and developers who want to work on Windows.

RAM is widely used to store temporary copies of data or files, allowing for rapid and simple access. 8 GB of RAM is provided by a general-purpose CPU. It is straightforward to fit a tiny dataset into the system RAM while training a model. Insufficient system RAM will prevent a model from being trained on a To train Windows 10 Pro supports an operating system in 64 bit and 8 GB of RAM.

CHAPTER 5
Results and Performance Analysis

5.1 Introduction

This chapter contains the discussion of the result of the experiment. The model's accuracy, evaluation methods, actual & predicted graph, the Accuracy of the training and validation, loss charting will all be included. We must first understand evaluation metrics in order to comprehend the key concepts of my research. The following are some crucial evaluation metric points.

5.2 Evaluation Metrics

Four types of evaluation metric used to measure the performance of regression models.

- ❑ Mean Squared Error (MSE),
- ❑ Mean Absolute Error (MAE),
- ❑ R-Squared (R^2) Score,
- ❑ Explained Variance Score(EVS)

Mean Squared Error (MSE)

.Mean Squared Error (MSE) is a common evaluation metric used in regression models to measure the average squared difference between the predicted values and the actual values. It provides a measure of the overall accuracy and goodness of fit of the regression model.

The formula for Mean Squared Error is:

$$\mathbf{MSE = mean((y\_true - y\_pred)^2)}$$

Where:

- y_true is the true values of the target variable.
- y_pred is the predicted values of the target variable.
- ^{2} is the squared function.
- mean() is the mean function.

To calculate the MSE, we first take the squared difference between the actual and predicted values for each observation. Then, we take the mean of these squared differences to obtain the MSE score.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is another commonly used evaluation metric in regression models to measure the average absolute difference between the predicted values and the actual values. It also provides a measure of the overall accuracy and goodness of fit of the regression model. The formula for Mean Absolute Error is:

MAE = mean (abs (y_true - y_pred))

Where:

- y_true is the true values of the target variable.
- y_pred is the predicted values of the target variable.
- abs() is the absolute value function.
- mean() is the mean function.

To calculate the MAE, we first take the absolute difference between the actual and predicted values for each observation. Then, we take the mean of these absolute differences to obtain the MAE score.

R-Squared (R^2) Score

R-squared (R^2) score is a commonly used evaluation metric in regression models to measure the goodness of fit of the model. It is a statistical measure that indicates the proportion of the variance in the dependent variable (i.e., the target variable) that is explained by the independent variables (i.e., the features) included in the model. The R^2 score ranges from 0 to 1, where 0 indicates that the model does not explain any of the variance in the dependent variable, and 1 indicates that the model explains all of the variance in the dependent variable. A higher R^2 score indicates a better fit of the model to the data, as it signifies that a larger proportion of the variance in the dependent variable is explained by the independent variables included in the model.

The formula for R^2 score is:

$$R^2 = 1 - (sum((y\_true - y\_pred)^2) / sum((y\_true - mean(y\_true))^2))$$

Where:

- y_true is the true values of the target variable.
- y_pred is the predicted values of the target variable.
- mean (y_true) is the mean value of the target variable.

To calculate the R^2 score, we first calculate the total sum of squares (TSS), which is the sum of the squared differences between the actual values and the mean value of the target variable. Then, we calculate the residual sum of squares (RSS), which is the sum of the squared differences between the actual and predicted values. Finally, we use these values to calculate the R^2 score, which represents the proportion of the variance in the target variable that is explained by the independent variables in the model.

Explained Variance Score (EVS)

The explained variance score (EVS) is a metric used to evaluate the performance of regression models. It measures the proportion of variance in the target variable (i.e., the dependent variable) that is explained by the independent variables (i.e., the features) included in the model.

The EVS score ranges from 0 to 1, where 1 indicates a perfect prediction by the model and 0 indicates that the model does not explain any of the variance in the target variable. A higher EVS score indicates a better fit of the model to the data, as it signifies that a larger proportion of the variance in the target variable is explained by the independent variables included in the model.

The formula for EVS is:

EVS = 1 - (var(y_true - y_pred) / var(y_true))

Where:

- y_true is the true values of the target variable.
- y_pred is the predicted values of the target variable.
- var() is the variance function.

To calculate the EVS score, we take the ratio of the explained variance to the total variance. A higher EVS score indicates a better fit of the model to the data

5.3 Results of ML

❑ Data Splitting of ML

The performance of machine learning models that split data into train and test is seen in the following table. After dividing the data into segments of 70% and 30%, 75% and 25%, 80% and 20%, we can observe 80% and 20% delivers the maximum accuracy of these ML models.

Table 5.1: Performance-based on the split ratio (ML)

Model Name	Train	Test	Accuracy
Random Forest Regression	70%	30%	0.82
	75%	25%	0.86
	80%	20%	0.93
Decision Tree Regression	70%	30%	0.86
	75%	25%	0.87
	80%	20%	0.91
Support Vector Regression	70%	30%	0.78
	75%	25%	0.87
	80%	20%	0.89
Gradient Boosting Regression	70%	30%	0.75
	75%	25%	0.80
	80%	20%	0.85

❑ Result of ML Models

In Table 5.2 and Table 5.3, we have used four types of metrics to measure performance which is Mean Squared Error, Mean Absolute Error, R-Squared Score, explained Variance Score.

Table 5.2: Performance-based on the Metrics (ML)

Model Name	Evaluation Metric	Metric Score
Random Forest Regression	Mean Squared Error	8.6423
	Mean Absolute Error	1809.4
	R-Squared Score	0.9309
	Explained Variance Score	0.9327
Decision Tree Regression	Mean Squared Error	1.1043
	Mean Absolute Error	2804.2
	R-Squared Score	0.9117
	Explained Variance Score	0.9155

Table 5.3: Performance-based on the Metrics (ML)

Model Name	Evaluation Metric	Metric Score
Support Vector Regression	Mean Squared Error	1.6952
	Mean Absolute Error	2306.0
	R-Squared Score	0.8996
	Explained Variance Score	0.9214
Gradient Boosting Regression	Mean Squared Error	1.7958
	Mean Absolute Error	3890.1
	R-Squared Score	0.8564
	Explained Variance Score	0.8633

❑ Result Comparison of Machine Learning Models

These scatter plot will visually show how well the Random Forest Regression and Decision Tree Regression models predict the crime values. If the data points are close to the diagonal line, it indicates accurate predictions. However, if the points are scattered far from the diagonal line, it suggests the models may have difficulty accurately predicting the crime values.

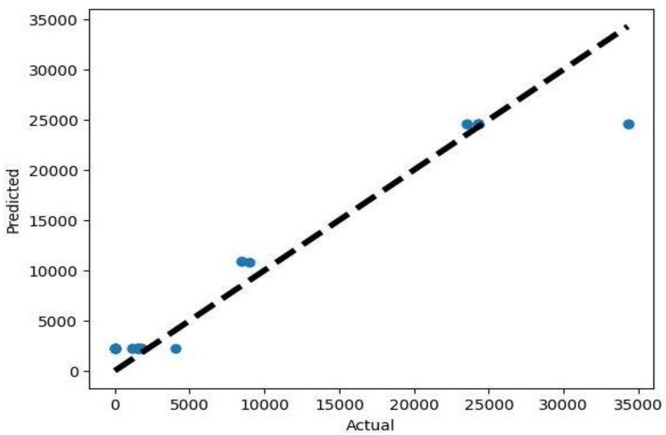

Figure 5.1: Random Forest Regression

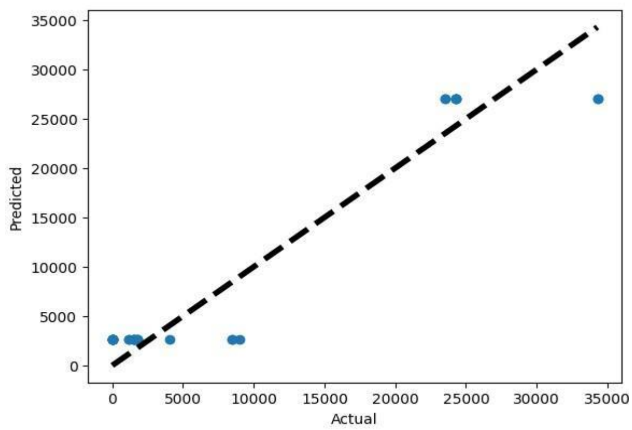

Figure 5.2: Decision Tree Regression

As well as, These scatter plot will visually show how well the Support Vector Regression and Gradient Boosting Regression models predict the crime values. If the data points are close to the diagonal line, it indicates accurate predictions. However, if the points are scattered far from the diagonal line, it suggests the models may have difficulty accurately predicting the crime values.

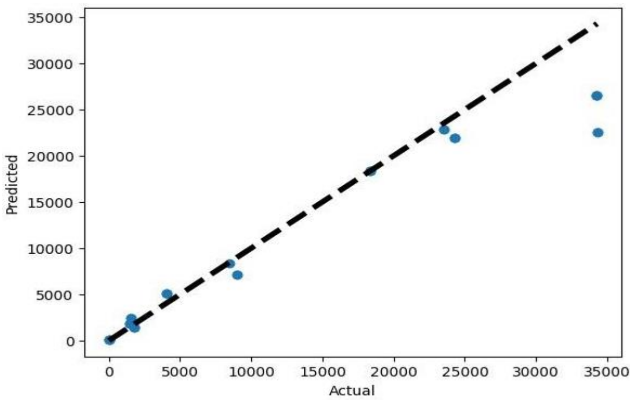

Figure 5.3: Support Vector Regression

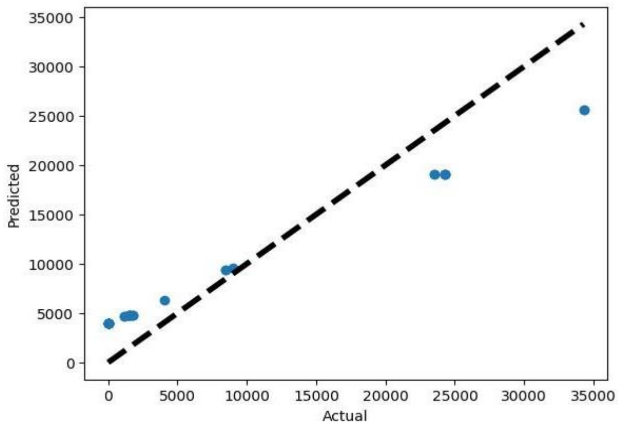

Figure 5.3: Support Vector Regression

5.4Predicted Crime Rate for ML Models

Random Forest Regression

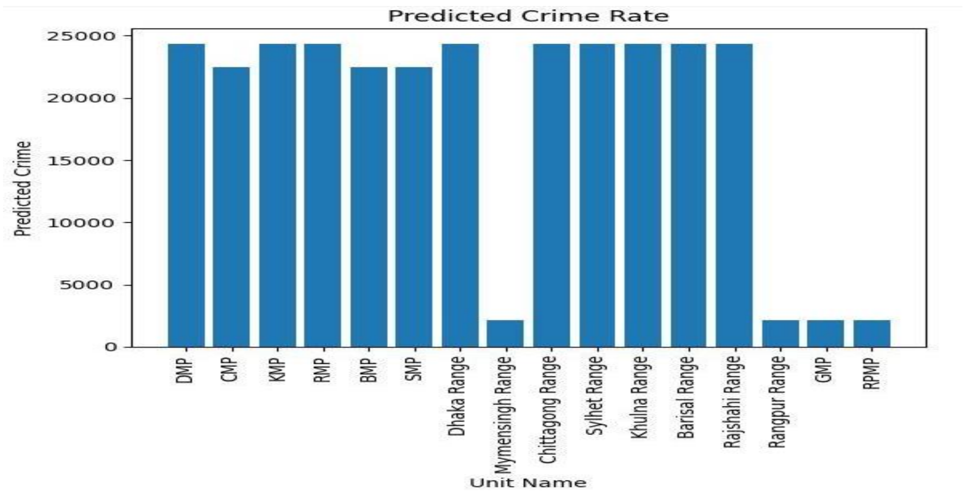

Figure 5.5: Predicted Chart of Random Forest Regression

We can see that the DMP, KMP, RMP, Dhaka Range, Chittagong Range, Sylhet Range, Khulna Range, Barisal Range, and Rajshahi Range are the areas predicted by our RF model to have the highest crime rates. On these units, the highest predicted crime rate is 24347.89.The predicted crime rate for the medium-rated areas of CMP, BMP, and SMP is 22495.77.Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rates; the estimated crime rate for these areas is 2143.94.

Decision Tree Regression

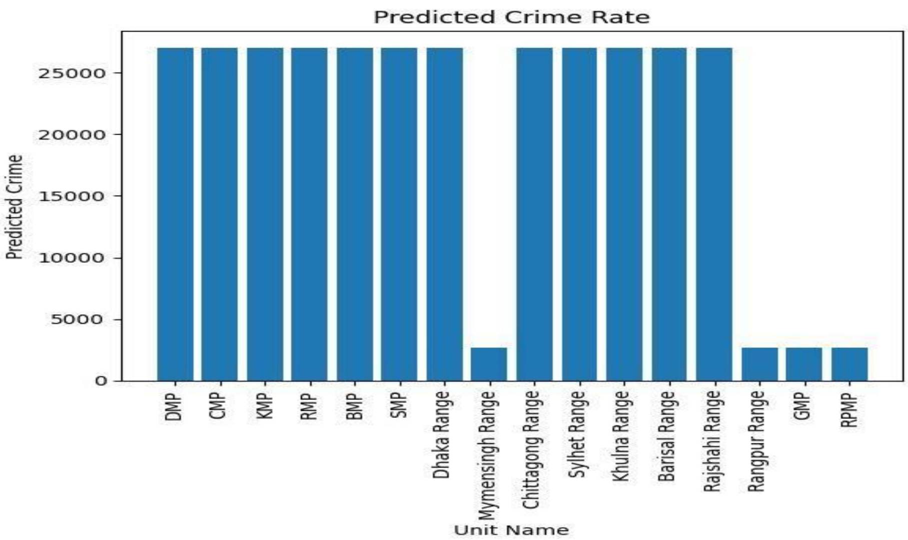

Figure 5.6: Predicted Chart of Decision Tree Regression

We can see that the DMP, CMP, KMP, RMP, BMP, SMP, Dhaka Range, Chittagong Range, Sylhet Range, Khulna Range, Barisal Range, and Rajshahi Range are the areas predicted by our DT model to have the highest crime rates. On these units, the highest predicted crime rate is 27016.38. Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rates; the estimated crime rate for these areas is 2670.81

Support Vector Regression

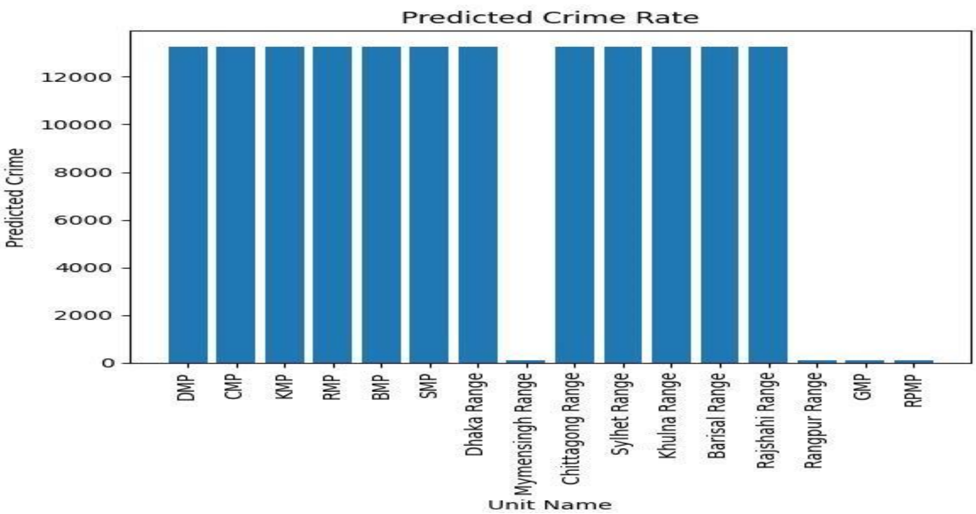

Figure 5.7: Predicted Chart of Support Vector Regression

We can see that the DMP, CMP, KMP, RMP, BMP, SMP, Dhaka Range, Chittagong Range, Sylhet Range, Khulna Range, Barisal Range, and Rajshahi Range are the areas predicted by our SVR model to have the highest crime rates. On these units, the highest predicted crime rate is 13261.54 Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rates; the estimated crime rate for these areas is 113.17

Gradient Boosting Regression

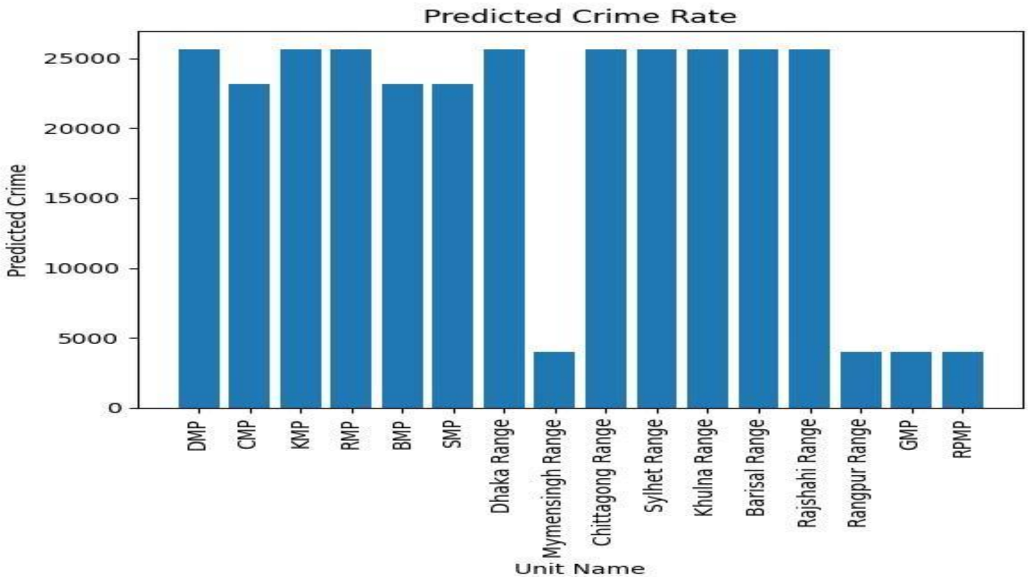

Figure 5.8: Predicted Chart of Gradient Boosting Regression

We can see that the DMP, KMP, RMP, Dhaka Range, Chittagong Range, Sylhet Range, Khulna Range, Barisal Range, and Rajshahi Range are the areas predicted by our RF model to have the highest crime rates. On these units, the highest predicted crime rate is 25633.50 .The predicted crime rate for the medium-rated areas of CMP, BMP, and SMP is 23163.54. Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rates; the estimated crime rate for these areas is 3999.40

5.5 Results of DL

❑ Data Splitting of DL

The performance of deep learning models that split data into train and test is seen in the following table. After dividing the data into segments of 70% and 30%, 75% and 25%, 80% and 20%, we can observe 80% and 20% delivers the maximum accuracy of these DL models.

Table 5.4: Performance-based on the split ratio (DL)

Model Name	Train	Test	Accuracy
Convolutional Neural Network (CNN)	70%	30%	0.95
	75%	25%	0.96
	80%	20%	0.97
Artificial Neural Network (ANN)	70%	30%	0.93
	75%	25%	0.92
	80%	20%	0.95

❑ Result of DL Models

Four different metrics, including Mean Squared Error, Mean Absolute Error, R-Squared Score, and Explained Variance Score, have been utilized to assess performance in Table 5.5.

Table 5.5: Performance-based on the Metrics (DL)

Model Name	Evaluation Metric	Metric Score
Convolutional Neural Network (CNN)	Mean Squared Error	3.15180
	Mean Absolute Error	1255.4
	R-Squared Score	0.9748
	Explained Variance Score	0.9752
Artificial Neural Network (ANN)	Mean Squared Error	6.0923
	Mean Absolute Error	2031.7
	R-Squared Score	0.9513
	Explained Variance Score	0.9624

❑ Parameter Tuning of DL

Dropout is a regularization technique commonly used in neural networks, including Convolutional Neural Networks (CNNs). It involves randomly disabling a fraction of the neurons during training. Dropout helps prevent overfitting and improves generalization by forcing the network to learn more robust representations. Higher dropout rates, such as 0.2, 0.3, and 0.4, indicate that a larger proportion of neurons are dropped out during training.

Table 5.6: Performance based on dropout rate

Dropout Rate	Accuracy
0.2	0.97
0.3	0.81
0.4	0.75

Each activation function has its own characteristics and may perform differently depending on the specific regression problem and dataset. It is common to experiment with different activation functions to find the one that yields the best results for our particular task.

Table 5.7: Selecting Activation function

Activation Function	Accuracy
ReLu	0.95
Selu	0.94
Linear	0.97

Both Adam and RMSprop aim to improve the convergence speed and performance of the model during training. However, the choice between them depends on the specific problem, dataset, and the behavior of the gradients in our regression task. It is recommended to experiment with different optimization algorithms to find the one that yields the best results for our particular scenario.

Table 5.8: Performance based on optimizer

Optimizers	Accuracy
Adam	**0.97**
RMSprop	0.95

When training a regression model, the learning rate is a hyperparameter that determines the step size taken during each iteration of the optimization algorithm. A higher learning rate like 0.01 may provide faster initial convergence but carries a higher risk of overshooting. A lower learning rate such as 0.001 leads to slower convergence but provides more stability and accuracy in reaching the optimal solution.

Table 5.9: Performance-based on learning rate

Learning rate	Accuracy
0.01	**0.97**
0.001	**0.95**

The linear loss function uses a linear connection to calculate the difference between expected and actual values. Using the Poisson distribution, the Poisson loss function measures the discrepancy between expected and actual counts.

Table 5.10: Performance-based on loss function

Loss Function	Accuracy
mean _ squared _ error	0.97
poisson	0.95

It is common for models to improve their performance as the number of training epochs increases. This is because the model has more opportunities to learn from the data and adjust its parameters. Increasing the number of epochs generally improves the model's performance, and you observed higher accuracy with 50 epochs. It is important to monitor the learning curve and check for signs of overfitting when training for a larger number of epochs.

Table 5.11: Performance-based on epoch

Epochs	Accuracy
10	0.93
20	0.95
30	0.94
40	0.95
50	0.97

Batch size refers to the number of samples processed in each iteration during training. The choice of batch size can have an impact on the model's performance and training dynamics. Smaller batch sizes, such as 12 in our case, can lead to more frequent updates of the model's parameters and potentially improve convergence.Smaller batch sizes may also introduce more noise in the gradient estimation due to the limited sample size per batch.

Table 5.12: Performance-based on batch size

Batch Size	Accuracy
12	0.97
16	0.96
32	0.95
42	0.93
64	0.94

❑ Result Comparison of Deep Learning Models

In this section the outcomes of every model will be represented as the screenshots.

Model accuracy and loss graph of CNN

This graph shows the model's accuracy and loss on the y-axis and epoch on the x-axis .I used 50 epochs in my regression task. The orange curve indicates test, whereas training indicated by the blue curve.

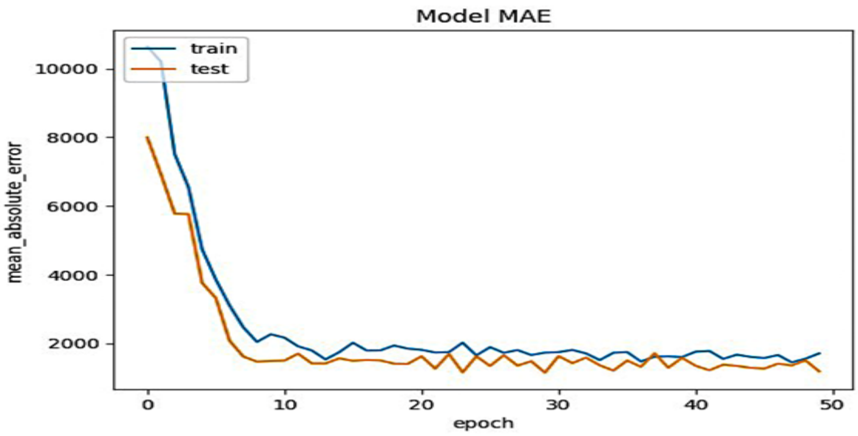

Figure 5.9: CNN model’s accuracy graph

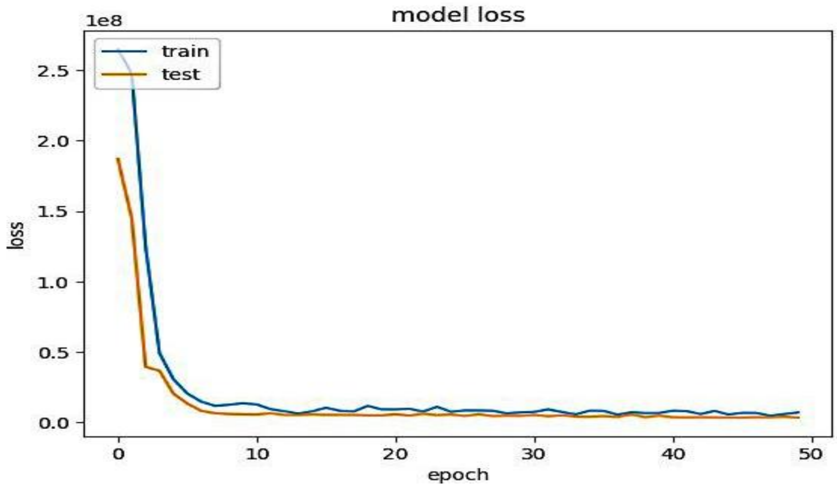

Figure 5.10: CNN model’s loss graph

Model accuracy and loss graph ANN

This graph shows the model's accuracy and loss on the y-axis and epoch on the x-axis .I used 50 epochs in my regression task. The orange curve indicates test, whereas training indicated by the blue curve.

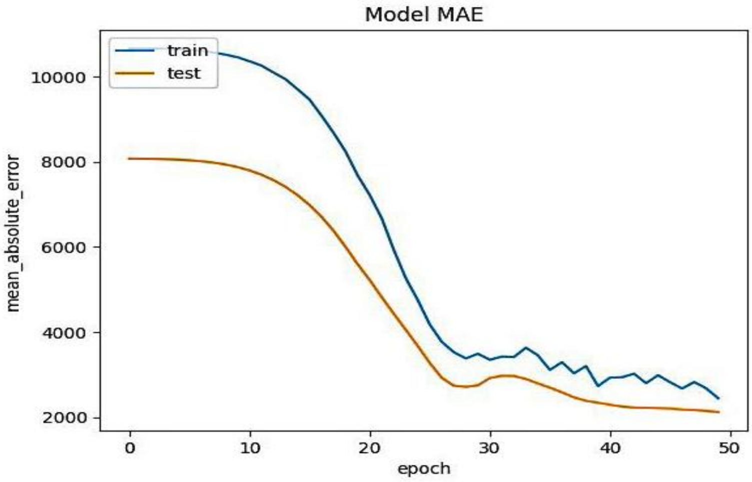

Figure 5.11: ANN model’s accuracy graph

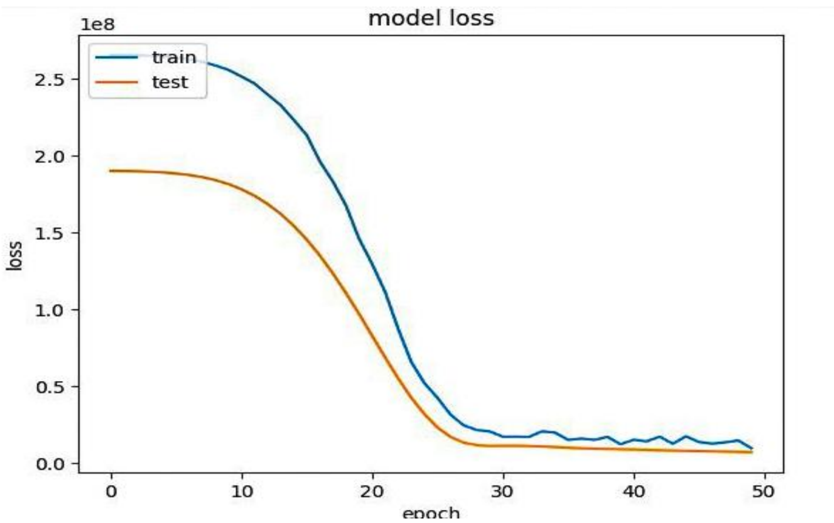

Figure 5.12: ANN model’s loss graph

❑ Result Comparison of Deep Learning Models

These scatter plot will visually show how well the CNN and ANN Regression models predict the crime values. If the data points are close to the diagonal line, it indicates accurate predictions. However, if the points are scattered far from the diagonal line, it suggests the models may have difficulty accurately predicting the crime values.

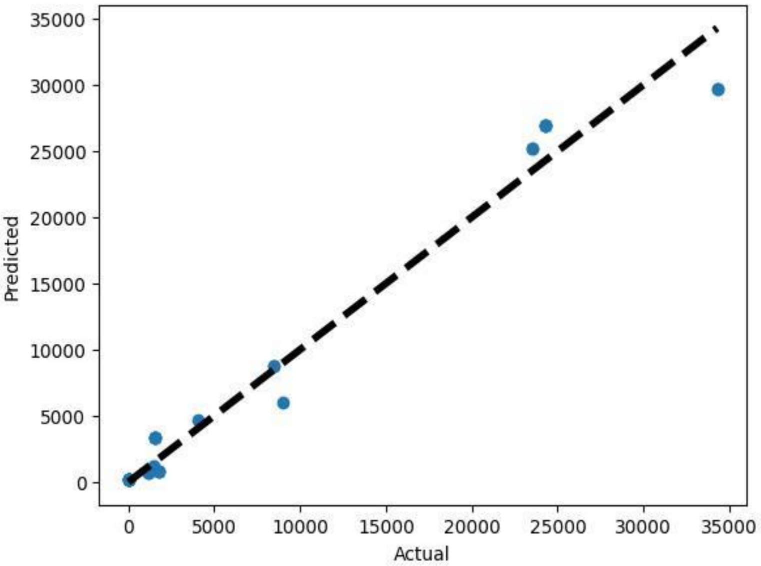

Figure 5.13: CNN model

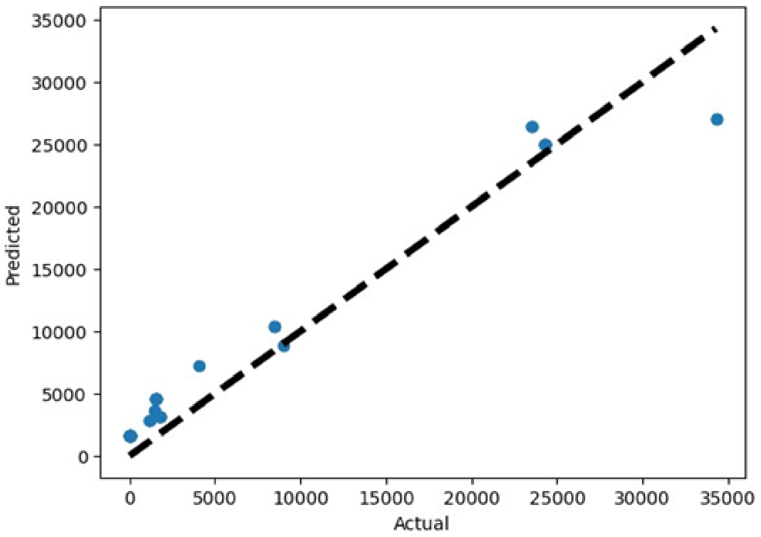

Figure 5.14: ANN model

5.7 Predicted Crime Rate of DL

Convolutional Neural Network

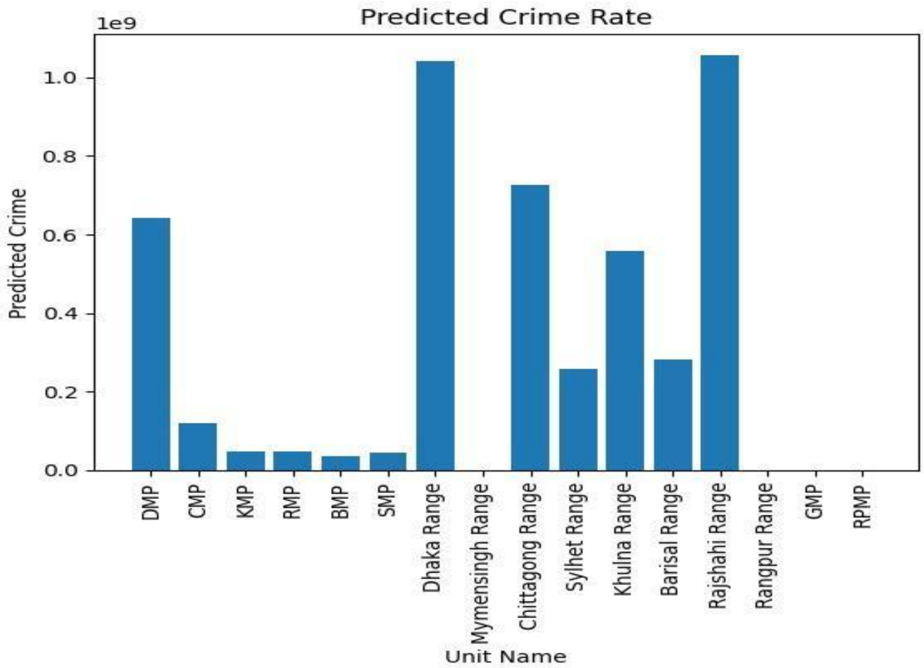

Figure 5.15: Predicted Chart of CNN model

We can see that the Dhaka Range, Chittagong Range, and Rajshahi Range are the areas predicted by our CNN model to have the highest crime rates. And medium high area are DMP, Khulna Range, Barisal Range ,Sylhet Range, The predicted crime rate for the medium low-rated areas are CMP , KMP, RMP, BMP, and SMP. Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rate area.

Artificial Neural Network

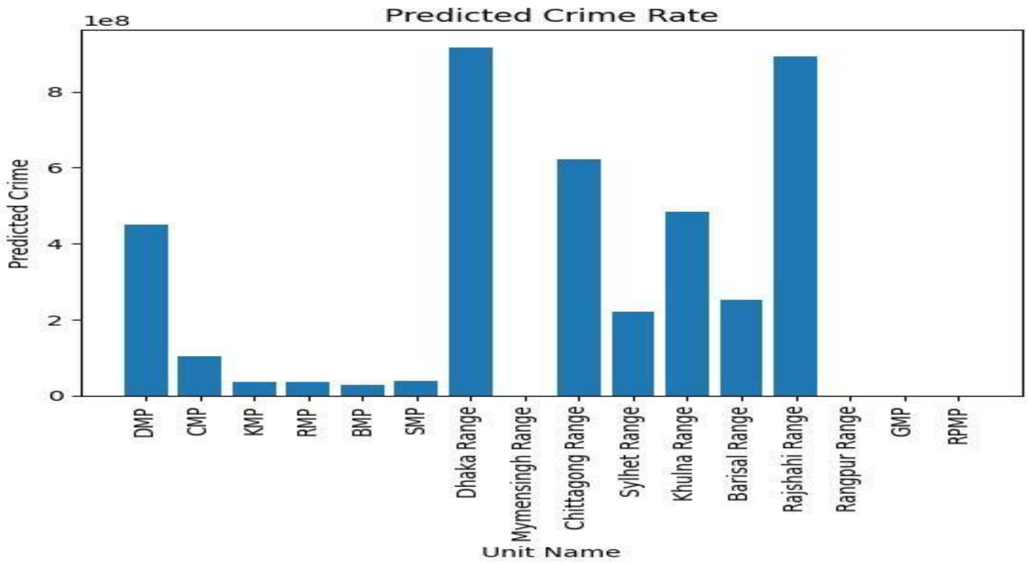

Figure 5.16: Predicted Chart of ANN model

We can see that the Dhaka Range, and Rajshahi Range are the areas predicted by our ANN model to have the highest crime rates. And medium high area are Chittagong Range, DMP, Khulna Range, Barisal Range ,Sylhet Range, The predicted crime rate for the medium low-rated areas are CMP , KMP, RMP, BMP, and SMP. Mymensingh Range, Rangpur Range, GMP, and RPMP have the lowest crime rate area.

5.8 Comparative Analysis

Both table 5.12 and 5.13 clearly illustrates the effectiveness of the deep learning method. All other machine learning algorithms pale in comparison to deep learning algorithms in terms of accuracy. Furthermore, we find that when it comes to model prediction, deep learning outperforms machine learning.

Table 5.13: Machine learning results

Model Name	**Accuracy**
Random Forest Regression	**93.09%**
Decision Tree Regression	**91.17%**
Support Vector Regression	**89.96%**
Gradient Boosting Regression	**85.64%**

Table 5.14: Deep learning results

Model Name	**Accuracy**
CNN	**97.48%**
ANN	**95.13%**

CHAPTER 6
Conclusion & Future Work

In this section of our document, we aggregate our present and future goals in an effort to make the entire process swift and simple to grasp for everyone.

6.1 Conclusion

This research's primary objective is to simplify a system for computer aided prediction system so that everyone may quickly integrate it. In here, we predict crime rate in 16 type of unit of Bangladesh. For crime rate we used Machine Learning model such as Random Forest Regression, Decision Tree Regression, SVR & GBR & Deep Learning model such as CNN & ANN.
Our dataset was located on the Kaggle website, which contains data from 2010 to 2019.Therefore, using crime data for 16 units of Bangladesh, we forecast crime rates and identify high-crime areas.

6.2 Future Work

We might also want to continue the work on ambiguity, but this would need for a better reliable set of data to build models. In the meanwhile, various researchers have worked on it but since its related to our country so, let's conduct additional research. The researchers may expand many pattern of crime activity, such as adding more attribute to a publicly accessible dataset to help find out accurately identify the crime rate .Using Neural Network & ML model to understand which models shows the better performance. Additionally, we are aware that crime is inhuman and unlawful activity which is never accepted by the law.

We made an effort to identify high crime regions in accordance with prior projections of crime rate. As a result, we may create any computer aid tool by utilizing AI in order to first determine the crime rate. Due to a recent year dataset gap, it is impossible to anticipate the most current criminal activities. Future research will involve additional deep learning methods.

References

[1] S. Mahmud, M. Nuha, and A. Sattar, “Crime rate prediction using machine learning and Data Mining,” *Advances in Intelligent Systems and Computing*, pp. 59–69, 2020.

[2] D. M. Raza and D. B. Victor, “Data Mining and region prediction based on crime using Random Forest,” *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021.

[3] A. A. Biswas and S. Basak, “Forecasting the trends and patterns of crime in Bangladesh using Machine Learning Model,” *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, 2019.

[4] Yerpude, Prajakta, Predictive Modelling of Crime Data Set Using Data Mining (July 21, 2020). International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.7, No.4, July 2017.

[5] Pratibha, A. Gahalot, Uprant, S. Dhiman, and L. Chouhan, “Crime prediction and analysis,” *2nd International Conference on Data, Engineering and Applications (IDEA)*, 2020.

[6] Md Awal, Jakaria Rabbi, Sk Hossain, and M.M.A. Hashem. Using linear regression to forecast future trends in crime of Bangladesh. 05 2016. Awal et al. - 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV) – 2016

[7] N. Mahmud, K. Ibn Zinnah, Y. Ar Rahman, and N. Ahmed, "CRIMECAST: a crime prediction and strategy direction service," IEEE 19th Intl. Conf. on Compute. and Inform. Technol., Dhaka, Bangladesh, Dec. 2016

[8] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," IEEE Computer, vol. 37, no. 4, pp. 50-56, Apr. 2004.

[9] M. A. Tayebi, U. Glässer, and P. L. Brantingham, "Learning where to inspect: location learning for crime prediction," IEEE Intl. Conf. on Intel. and Security Inform. (ISI), Baltimore, MD, USA, May 2015

[10] M. V. Barnadas, Machine learning applied to crime prediction, Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, September. 2016.

[11] S. Shojaee, A. Mustapha, F. Sidi, M. A. Jabar, “A study on classification learning algorithms to predict crime status”, In: International Journal of Digital Content Technology and its Applications (JDCTA) 7.9 (May 2013), pp. 361–369, issn: 1975-9339.

[12] Md. Rizwan Parvez, Turash Mosharraf, and Mohammed Eunus Ali. A novel approach to identify spatio-temporal crime pattern in Dhaka city. In Proceedings of the Eighth International Conference on Information and Communication Technologies and Development, ICTD ’16, New York, NY, USA, 2016. Association for Computing Machinery.

[13] L. McClendon, N. Meghanathan, “Using Machine Learning Algorithms to Analyze Crime Data”, Machine Learning and Applications: An International Journal (MLAIJ) vol. 2, no. 1, March 2015.

[14] C. H. Yu, M. W. Ward, M. Morabito, W. Ding, “Crime Forecasting Using Data Mining

Techniques", In Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW' 11), pp. 779-786, 2011.

[15] E. Cesario, C. Catlett, and D. Talia, "Forecasting crimes using autoregressive models," IEEE 14th Intl. Conf. on Dependable, Auton. and Secure Comput., Auckland, New Zealand, Aug. 2016.

[16] Md. Rizwan Parvez, Turash Mosharraf, and Mohammed Eunus Ali. A novel approach to identify spatio-temporal crime pattern in Dhaka city. In Proceedings of the Eighth International Conference on Information and Communication Technologies and Development, ICTD '16, New York, NY, USA, 2016.

[17] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. S. Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," Indian Journal of Science and Technology, vol. 6, no. 3, pp. 4219-4225, 2013

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

[19] W. Gorr, A. Olligschlaeger, and Y. Thompson, "Short-term forecasting of crime," *Int. J. Forecast.*, vol. 19, no. 4, pp. 579–594, 2003.

[20] R. Liao, X. Wang, L. Li, and Z. Qin, "A novel serial crime prediction model based on Bayesian learning theory," *2010 Int. Conf. Mach. Learn. Cybern .* , no. July , pp. 1757–1762, 2010.

[21] J. Phillips and K. C. Land, "The link between unemployment and crime rate fluctuations: An analysis at the county, state, and national levels," *Soc. Sci. Res.*, vol. 41, no. 3, pp. 681–694, 2012.

[22] S. A. Salleh, N. S. Mansor, Z. Yusoff, and R. A. Nasir, "The Crime Ecology: Ambient Temperature vs. Spatial Setting of Crime (Burglary)," *Procedia - Soc. Behav. Sci.*, vol. 42, no. July 2010, pp. 212–222, 2012.

[23] A. Stec and D. Klabjan, "Forecasting crime with deep learning," arXiv preprint arXiv:1806.01486, 2018.

[24] D. K. Tayal et al , "Crime detection and criminal identification in india using data mining techniques," AI & SOCIETY, vol. 30, no. 1, pp. 117– 127, 2015

[25] N. Yadav, A. Kumar, R. Bhatnagar, and V. K. Verma, "City crime mapping using machine learning techniques," The International Conference on Advanced Machine Learning Technologies and Applications, Springer, pp. 656-668, 2019