FMDB *Transactions on Sustainable Computer Letters*

Crime Prediction and Analysis Using Data Mining and Machine Learning: A Simple Approach that Helps Predictive Policing

G. Sivapriya[1,*], B. Vijay Ganesh[2], U.G. Pradeeshwar[3], Vishnu Dharshini[4], Muhammad Al-Amin[5]

[1,2,3,4]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram,
Chennai, Tamil Nadu, India.
[5]Department of International Relations, Sichuan University, Chengdu, China.
sg5741@srmist.edu.in[1], vb6055@srmist.edu.in[2], pg4893@srmist.edu.in[3], visnudhs@srmist.edu.in[4],
alamin2022@stu.scu.edu.cn[5]

Abstract: Crime is a dangerous and global social issue. Crimes influence a nation's economy, reputation, and quality of life. Innovative technology and novel approaches to crime analytics are needed to safeguard society from Crime. Predicting Crime in a chosen place and visualizing crime data can assist law enforcement in preventing Crime. This paper uses Vancouver crime statistics from crime study resources. As vital as the final prediction, data pre-processing comes first. This paper cleans and feeds data with feature selection, null removal, and label encoding. Pre-processed data is used to build a training model utilizing KNN, Decision trees, classification, linear regression, and random forest algorithms. Sklearn's matplotlib library predicts and analyses after model creation. The crime dataset is graphed. This work offers a reliable machine-learning criminal case prediction model. Systematic crime analysis and prediction identify crimes. Classification and resolution are difficult due to escalating criminal cases. Knowing local crime patterns helps crime-solving agencies. Machine learning and random forest can reveal local crime tendencies. This article predicts local crimes using crime statistics, speeding up criminal case classification and proceedings. This approach can anticipate high-crime locations and show crime-prone areas. Data mining finds useful information in unstructured data. Existing data anticipates new extraction. Using this method, we can analyze, detect, and estimate regional crime probability. It also discusses crime prediction and analysis using machine learning and data mining. Data mining and machine learning are becoming crucial in almost all fields, including crime prediction. Crime prediction and analysis are essential for detecting and reducing future crimes.

Keywords: Predictive Analysis; Machine Learning; Artificial Intelligence; Linear Regression; Data Mining; Classification Model; Neural Networks; Visualization Tools; Decision Trees; Crime data; Data Pre-processing.

Received on: 27/12/2022, **Revised on:** 25/02/2023, **Accepted on:** 29/03/2023, **Published on:** 05/04/2023

Cited by: G. Sivapriya. B. V. Ganesh, U.G. Pradeeshwar, V. Dharshini, M. Al-Amin, "Crime Prediction and Analysis Using Data Mining and Machine Learning: A Simple Approach that Helps Predictive Policing," *FMDB Transactions on Sustainable Computer Letters*., vol. 1, no. 2, pp. 64 –75, 2023.

1. Introduction

Crime rate is increasing daily, becoming a major concern, certainly hindering good governance. Kidnappings, murders, and other violent crimes have increased dramatically, according to data compiled by India's National Crime Records Bureau. As the crime rate has skyrocketed, so has the volume of data collected in connection with it, making it increasingly challenging to process. Intelligence agencies and local law enforcement cannot use this information to study crime trends or make predictions about future crimes [1]. As a result, there is a need for a powerful analytical tool that can rapidly and efficiently analyse crime data and provide some helpful crime patterns through the use of visualisation methods. In the instance of organised crime, the

[*]Corresponding author.

results of this analysis can be utilised to trace the criminal's network of accomplices as well [2]. Due to the rising crime rate, analysing the massive amounts of data collected and stored in data warehouses would be an arduous task if done manually. Concurrently, as criminals themselves become more technologically savvy, there is a pressing need to employ cutting-edge tools in order to stay one step ahead of them.

As the number of criminal cases in India rises, the backlog of cases in the justice system also grows. An in-depth internal investigation and analysis is required to solve a case based on specific evidence. Officials in India are hampered in their ability to analyse and resolve criminal cases by the sheer volume of data available to them at present [5]. Given the significance of this issue, this study is dedicated to developing a strategy for improving the judgement involved in criminal cases. [6]. Once everything is set up, the car will drive itself. To get to its target, an autonomous vehicle must repeatedly recognise, judge, and control its surroundings. Recognizing the trends of criminal activity in a given area is crucial for preventing further increases in crime, especially given the difficulty in categorising and solving the current spate of cases. Data mining and machine learning are presented as a means to organise the data and create reliable forecasts. Using techniques from statistics, artificial intelligence (AI), and database administration, data mining is the process of discovering patterns in big data sets. The fundamental goal of ML is to construct prediction modules [7]. Data collection, categorization, pattern identification, prediction, and visualisation are the core parts of this effort, which has followed the standard data analysis process [1]. The proposed framework makes use of a number of different visualisation methods to display crime trends and the many ways in which a machine learning algorithm might forecast crime. Our algorithms take in data such as the time of day, the month, the year, the coordinates of the location, and the type of crime that has been committed. For this purpose, HTML (Hypertext Markup Language) and CSS (Cascading Style Sheet) are crucial [3] technologies.

While HTML establishes the page's framework, CSS determines its visual and auditory presentation across a range of devices. HTML and CSS, along with graphics and coding, are the backbone upon which websites and web applications are built. Online documents written in HTML can include headings, body text, tables, lists, and even pictures. Hypertext links make it easy to quickly access data from the internet [4]. Design forms are available for use in searching for information, making purchases from remote services, etc. Documents can now incorporate media such as spreadsheets, videos, and audio files. Flask is the most popular framework of Python for web development and it is free, open source, and server-side [8]. The focus of this paper is on machine learning. Machine learning is frequently used for making forecasts. There are several prediction algorithms accessible in various libraries [9]. In this work, we will create a prediction model using historical data using several machine learning algorithms and classifiers, plot the results, and calculate the model's accuracy on the testing data. One component of the data [10] is the vast dataset used for building/training the model. However, the second step in applying machine learning in practise is integrating these models into existing software. We need to release it online so that others from all around the world can use it to make predictions based on the latest data [4]. The most in-demand technological talents of the moment are provided by this system [11]. It offers the user with lifesaving technologies and aids law enforcement in pinpointing the type of crime committed depending on its location [12].

2. Existing System

The application of data mining and machine learning to the study of criminology can be broken down into two broad categories: crime prevention and detection [13]. De Bruin presented a model for crime patterns that makes use of a new distance metric to compare and cluster people based on their profiles. E-governance innovations already in use by the Indian police force are highlighted by Parvatikar and Parasar [14]. In addition, he suggests a query-based interface for interactive crime analysis to aid law enforcement. Using crime data mining techniques like clustering, he developed an interface to access the NCRB's massive crime database and pull out actionable insights [15]. The usefulness of the proposed interface has been illustrated in Indian crime records. In his work [16], Rangineni and Marupaka analyse how cluster analysis might be used in the field of accounting, specifically for anomaly detection during audits. His research plans to look into how clustering might be used in audits to automate fraud filtering. When auditing claims for group life insurance, he employed cluster analysis to assist auditors prioritise their work [17].

In addition, there are currently available options such as k-NN, RF, SVM, and Bayes models. Advanced data exploration with machine learning algorithms have been used in criminology studies, although prediction is still in its infancy. For reliable forecasting, more research is required [18]. Hidden Markov Model (HMM) parameters are chosen by mining the double-layered hidden states of past trajectories [19]. The double-layer hidden state sequences that are appropriate for the just-driven trajectory are located with the help of a Viterbi algorithm [20]. Last but not least, it suggests a brand-new approach for trajectory prediction using a hidden Markov model with double-layer hidden states. Next k-stage nearest neighbour location information is predicted. Research is being conducted in this area on a global scale, and numerous innovations have been made to address this critical issue. CrimeNet Explorer, which employs SNA methods, hierarchical clustering, and multidimensional scaling to establish stronger linkups among offenders, and PredPol, a company that provides crime prediction software to Police Departments in countries like the United States [21] are all examples of projects that use the idea of "concept space" to improve crime prediction. Predictive policing is now used by police agencies in a number of states, including California, Washington, South Carolina, Arizona, Tennessee, and Illinois [22]. Effectively identifying criminals requires analytical and predictive methods, which are employed in "predictive policing" [23].

While each of these systems and tools has its own unique methodology, set of capabilities, and scope, they all share the common goal of helping law enforcement better predict and prevent criminal activity. It's important to remember that the efficacy and ethics of these systems have been debated and analysed in different circles, therefore their application and use may differ from one jurisdiction to the next [24]. The employment of sophisticated tools and techniques is on the rise among criminals, and this is reflected in rising crime statistics [25]. The Crime Records Bureau reports a rise in burglary, arson, and other property crimes as well as an increase in murder, sex abuse, gang rap, and other violent crimes. There has been no practical application of this unstructured criminal data beyond the accumulation of files [26]. Except in exceptional circumstances, every crime follows a predictable formula. Careful examination of this crime trend reveals opportunities to close the gaps in our case solving. There is currently no crime prediction model, merely generated reports of already committed crimes. Machine learning prediction models are correct 70% of the time, according to studies [27]. So, we can use those forecasts to take preventative action. Several variables affected the frequency with which crimes were committed [28]. Till now, there have been systems that analyse Crime, gather crime data, and obtain system data when asked [29]. However, data mining techniques such as clustering and classification have not been combined with machine learning for use in crime detection [30].

3. Proposed System

In this research, we will anticipate criminal behaviour based on historical crime data using machine learning and data mining. The majority of the crime statistics come from the departments' own websites [31]. It is a jumble of data about crimes, including their locations, descriptions, dates, times, and coordinates. The raw data will be filtered and otherwise pre-processed before being used to train the model [32]. Following this, feature selection and scaling will be done to increase the accuracy produced. Cleaning the data, choosing the features to use, removing the zero values, and normalising and standardising the data are all part of the pre-processing phase. After data pre-processing, the absence of null values eliminates a potential source of error in the model [33]. To ensure that the model's accuracy is not compromised, feature selection is employed to narrow down the pool of potential characteristics and weed out those that aren't necessary [34].

Now that the data has been cleaned and organised, it may be used to train a model for crime prediction utilising techniques like Logistic Regression classification, Support vector machine, and other algorithms like Decision Tree and Random Forest [35]. As data pre-processing is complete, the models chosen, i.e., Logistic Regression, Decision Tree, and Random Forest, are trained by partitioning the data into train and test data. Classification models are utilised since the desired result is a binary value that cannot be changed [36]. The data forecasting is done in Python. Libraries essential to the operation of the system, such as numpy, can be imported with Python's help. Finally, the dataset will be visualised in the form of a graphical depiction of numerous examples, such as the time of day or month when criminal activity is most prevalent. We want to evaluate many different crime classification algorithms, including K-Nearest Neighbor (KNN), Decision Tree, and Random Forest. A system with greater query support will be utilised for training the proposed system [37]. The devised method is effective in detecting, predicting, and solving crimes at a considerably faster pace, which in turn reduces the crime rate. Depending on the accessibility of the dataset, this can be used in a variety of various contexts [38].

The suggested system is grounded in analysis performed using such sources of information. Predictions can be made about nearly all crimes [39] based on their location and the types of crimes that tend to occur there. Based on a review of related literature, we find that Linear Regression, Decision Tree, and Random Forest models all perform admirably when applied to crime prediction. The dataset used in this paper is Kaggle, a crime dataset of Vancouver [40]. The dataset includes various crimes committed there, classified by kind and other criteria. This study takes in data about the types of crimes that have been committed and outputs predicted maps of the neighbourhoods where such crimes have occurred. The goal of this research is to demonstrate the potential of machine learning to help law enforcement organisations combat crime [41]. This proposed model explains the required procedures and components to construct a complete crime prediction and analysis system [42]. At every stage of the project, it is essential to think about and act upon ethical concerns, data privacy, and justice. Successful system development also requires input from subject matter experts and other stakeholders [43].

4. Methodology

The methodology used in this paper is explained in the following area. Developing a crime prediction and analysis methodology involves defining a structured approach to collecting, analyzing, and utilizing data to predict and understand criminal activities. This system can be built using machine learning as well as deep Learning. Still, the preferred model is in machine learning because it reduces the complexity as deep Learning includes complex neural networks, and training time is more than machine learning. Machine learning is preferred as this project is solved with limited data, and the required output falls under classification and clustering models. Here, we employ methods to collect the data required to use that CSV data format for data pre-processing and feature extraction. Pre-processing prepares the data for training and enhances the performance of the model. Here, we plan to complete pre-processing using filtering techniques performed by certain functions after importing pandas. The categorical attributes for feature selection are Location, Block, Crime Type, and Community Area, which are converted into numeric using Label Encoder. Later, we plan to train and build the model to predict Crime before it happens by using machine learning algorithms such as linear regression, random forest, and clustering techniques (Fig.1).

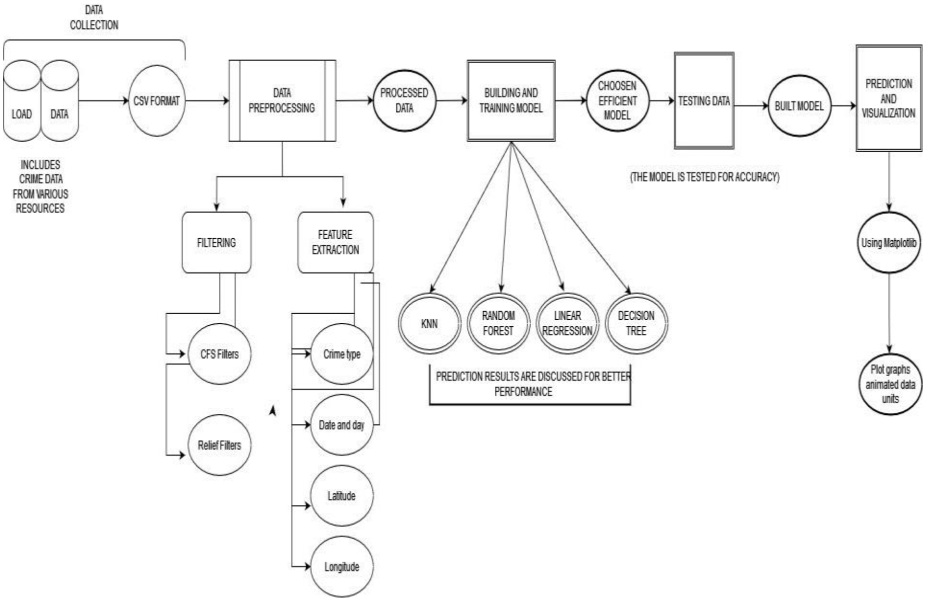


Figure 1: Architecture Diagram

K-nearest neighbours is a robust classification technique for use in pattern recognition. It uses a similarity metric to store and categorise all accessible examples (e.g., distance function). The system architectural design and data flow diagram are supplied, which suggests the design process for identifying the subsystems making up the system and the framework for subsystem control and communication. The purpose of the architectural design is to create a blueprint for the entire software application (Fig.2).

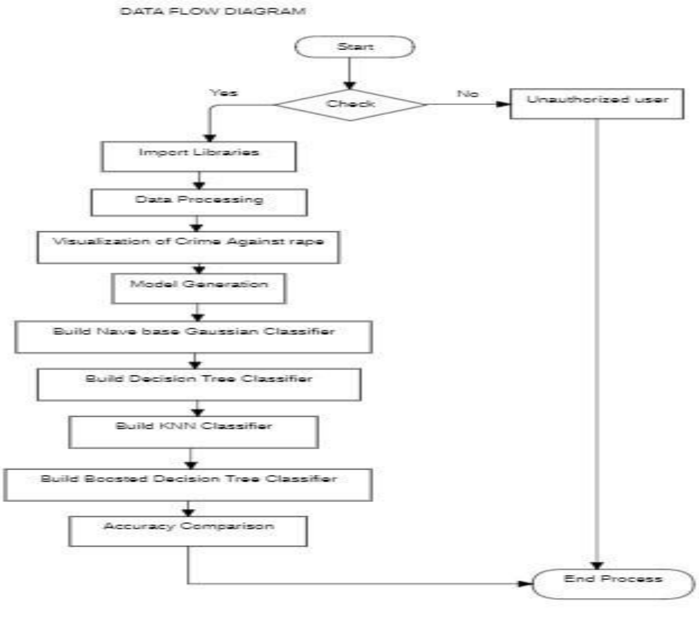


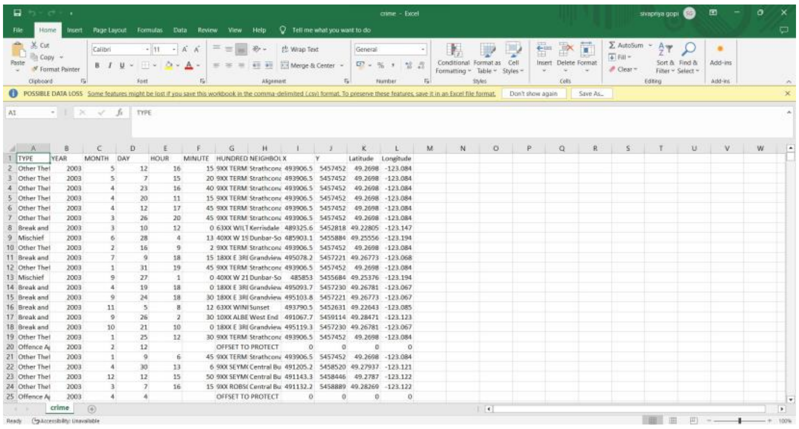
Figure 2: Dataflow Diagram

5. Module Description

The entire process is divided into four modules.

5.1. Module 1: Data collection module

The Kaggle crime dataset is converted to CSV and used. It includes 8,000 records of information about criminal activity in Vancouver. Data collected at the scene of a crime comprises of the officer's observations and judgments, which are recorded in a log. Inputs for this model include a list of victims and their addresses, a list of offenders and their addresses, the location of the crime, the nature of the offence, the time and date of the crime, the resources used, etc. The proposed research relies heavily on the collecting and pre-processing of data. In order to construct a robust and adaptable crime prediction system, we need access to a high-quality dataset of criminal activity. The data collection involves getting data files from various places, like online sites and law enforcement agencies. It's important to ensure the data you collect covers many details (Fig.3).



POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

A1 TYPE

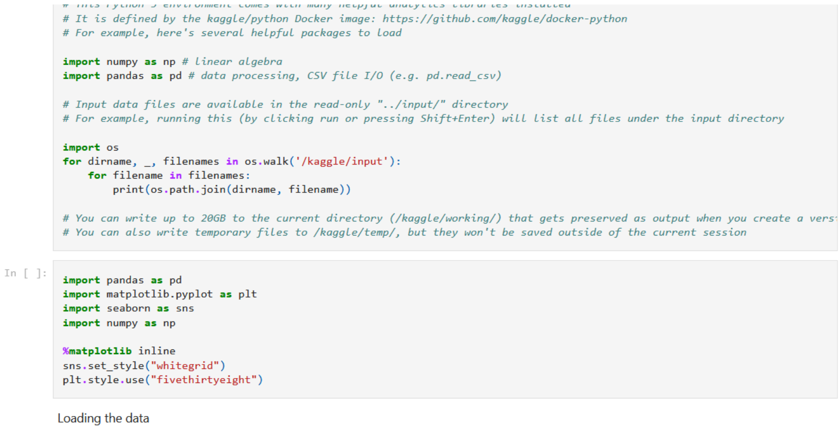
	A	B	C	D	E	F	G	H	I	J	K	L
1	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED	NEIGHBOI	X	Y	Latitude	Longitude
2	Other Thei	2003	5	12	16	15	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
3	Other Thei	2003	5	7	15	20	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
4	Other Thei	2003	4	23	16	40	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
5	Other Thei	2003	4	20	11	15	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
6	Other Thei	2003	4	12	17	45	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
7	Other Thei	2003	3	26	20	45	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
8	Break and	2003	3	10	12	0	63XX WILT	Kerrisdale	489325.6	5452818	49.22805	-123.147
9	Mischief	2003	6	28	4	13	40XX W 15	Dunbar-So	485903.1	5455884	49.25556	-123.194
10	Other Thei	2003	2	16	9	2	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
11	Break and	2003	7	9	18	15	18XX E 3RI	Grandview	495078.2	5457221	49.26773	-123.068
12	Other Thei	2003	1	31	19	45	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
13	Mischief	2003	9	27	1	0	40XX W 21	Dunbar-So	485853	5455684	49.25376	-123.194
14	Break and	2003	4	19	18	0	18XX E 3RI	Grandview	495093.7	5457230	49.26781	-123.067
15	Break and	2003	9	24	18	30	18XX E 3RI	Grandview	495103.8	5457221	49.26773	-123.067
16	Break and	2003	11	5	8	12	63XX WINI	Sunset	493790.5	5452631	49.22643	-123.085
17	Break and	2003	9	26	2	30	10XX ALBE	West End	491067.7	5459114	49.28471	-123.123
18	Break and	2003	10	21	10	0	18XX E 3RI	Grandview	495119.3	5457230	49.26781	-123.067
19	Other Thei	2003	1	25	12	30	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
20	Offence A	2003	2	12			OFFSET TO PROTECT		0	0	0	0
21	Other Thei	2003	1	9	6	45	9XX TERM	Strathcon	493906.5	5457452	49.2698	-123.084
22	Other Thei	2003	4	30	13	6	9XX SEYM	Central Bu	491205.2	5458520	49.27937	-123.121
23	Other Thei	2003	12	12	15	50	9XX SEYM	Central Bu	491143.3	5458446	49.2787	-123.122
24	Other Thei	2003	3	7	16	15	9XX ROBS	Central Bu	491132.2	5458889	49.28269	-123.122
25	Offence A	2003	4	4			OFFSET TO PROTECT		0	0	0	0

Figure 3: Data collection

We'll also need to ensure you document the data's provenance so it's transparent and can be trusted. This step of data collection sets up a good base for the later stages of our project so you can analyze the data and train our model for prediction.

5.2. Module 2: Data Pre-processing Module and Feature Selection Module

Data pre-processing module: 8000 entries are present in the dataset. The null values are removed using df = df. Dropna (), where df is the data frame. Label Encoder is used to quantify the previously qualitative variables (Location, Block, Crime Type, Community Area). The date property is decomposed into smaller, more manageable pieces, such as month and hour, that can then be used as features in the model. In order to prepare the data for analysis and prediction, unnecessary details and gaps must be eliminated. When necessary values are absent, defaults are used (Fig.4).



```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a vers
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

```
In [ ]:
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

%matplotlib inline
sns.set_style("whitegrid")
plt.style.use("fivethirtyeight")
```

Loading the data

Figure 4: Data pre-processing module

In the pre-processing stage, files are converted into sequences of numbers, each number representing a prediction model. Cleaning and filtering are done to get rid of mistakes and inconsistencies. Data is normalized to make sure everything is formatted the same. Data augmentation techniques are used to make the dataset bigger and more diverse. This careful pre-processing process ensures that the data you feed into our model is high-quality and accurate and can help you learn how to generate prediction reports. The crime analysis and prediction project data will be carefully organized and stored with strong access controls. We'll also have a backup and recovery plan to protect against data loss. We'll keep track of all the data management steps and make sure everyone knows what's going on and who's responsible for it.

Feature selection module: Feature selection is done, which can be used to build the model. Feature selection is essential in building predictive models, data analysis, and machine learning. It involves choosing a subset of the most relevant and informative features (variables or attributes) from the original set of features. There are various techniques for feature selection, ranging from statistical methods (e.g., correlation analysis, mutual information) to model-based approaches (e.g., decision trees, L1 regularization) and domain knowledge-driven methods. The choice of method depends on the data, the problem, and the goals of the analysis or model building. The key is to balance retaining enough useful information and simplifying the model for practical use. The attributes used for feature selection are Block, Location, District, Community area, X coordinate, Y coordinate, Latitude, Longitude, Hour, and month.

5.3. Module 3: Building and Training Model

After feature selection, location and month attributes are used for training. Building and training a crime prediction module is a complex process that requires a multidisciplinary team with expertise in data science, machine learning, and domain knowledge related to criminology and law enforcement. Additionally, it's important to consider ethical and legal aspects throughout the development process. The dataset is divided into xtrain, train, and xtest, ytest pairs. The algorithm model is imported from sklearn. Building models is done using models. Fit (xtrain, ytrain). The model is built by keeping the base language as Python.

The programming language Python is a high-level, general-purpose interpreter. Its design philosophy emphasises code readability with its usage of heavy indentation. Its object-oriented design and language elements are made to facilitate the creation of clean, well-thought-out programmes for both simple and complex tasks. A predictive analysis report requires training and testing of this model. Classification, Regression, decision tree, random forest, and KNN are the mainstays of supervised Learning, which we employ here to build the model. Using a training set of data with observations (or instances) whose category membership is known, a machine learning classifier must determine which category (sub-population) a new observation belongs to (Figs.5 and 6).

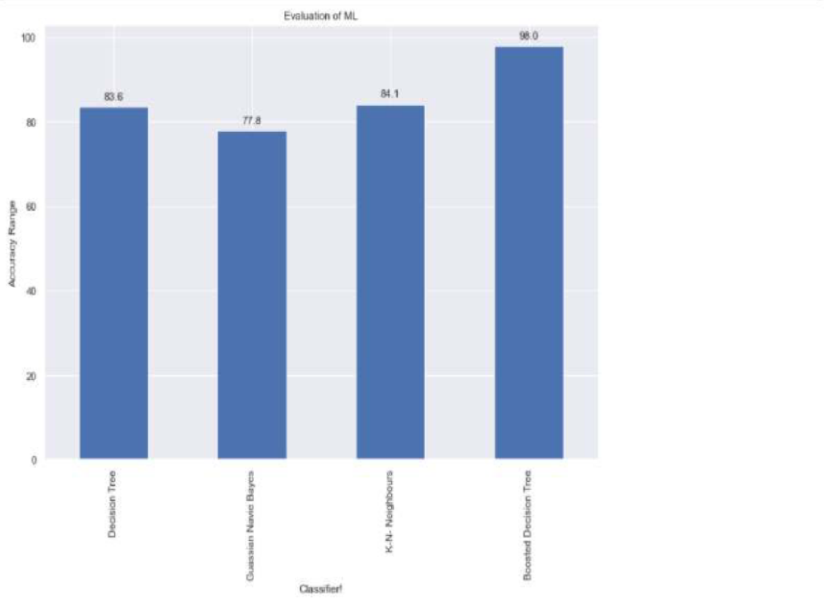
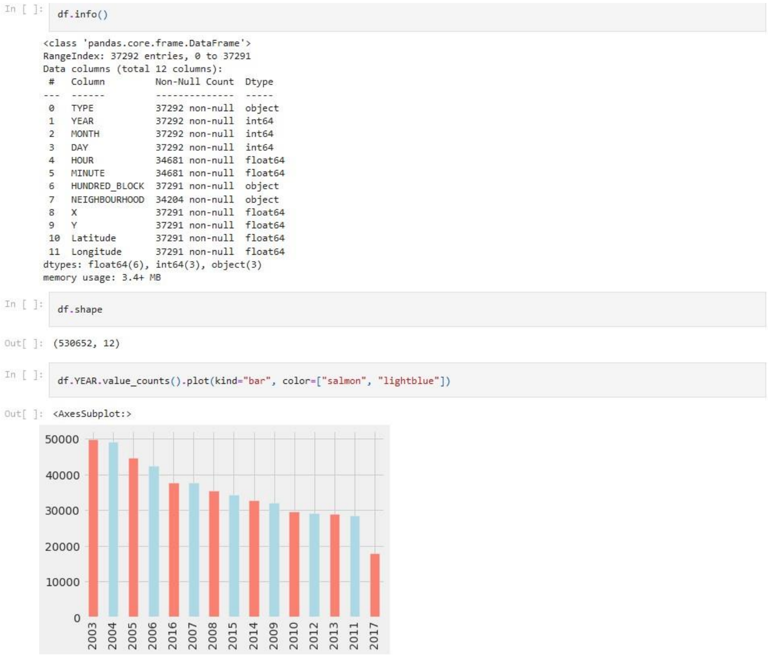


Figure 5: Different strategies used in Building model

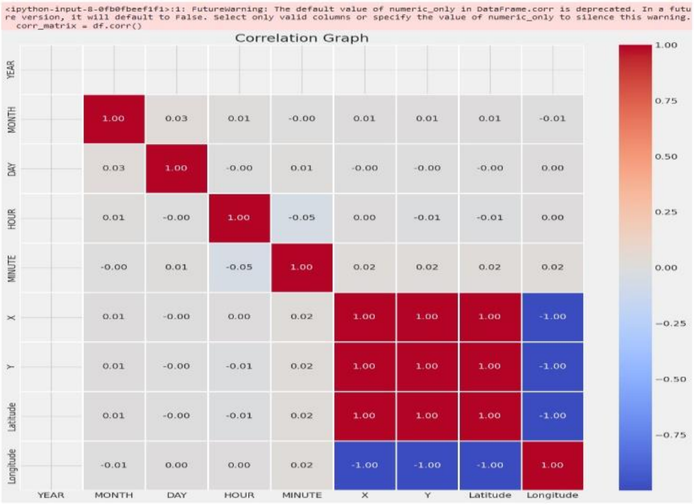


```
In [ ]:  df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37292 entries, 0 to 37291
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   TYPE           37292 non-null  object
 1   YEAR           37292 non-null  int64
 2   MONTH          37292 non-null  int64
 3   DAY            37292 non-null  int64
 4   HOUR           34681 non-null  float64
 5   MINUTE         34681 non-null  float64
 6   HUNDRED_BLOCK  37291 non-null  object
 7   NEIGHBOURHOOD  34204 non-null  object
 8   X              37291 non-null  float64
 9   Y              37291 non-null  float64
 10  Latitude       37291 non-null  float64
 11  Longitude      37291 non-null  float64
dtypes: float64(6), int64(3), object(3)
memory usage: 3.4+ MB

In [ ]:  df.shape

Out[ ]:  (530652, 12)

In [ ]:  df.YEAR.value_counts().plot(kind="bar", color=["salmon", "lightblue"])

Out[ ]:  <AxesSubplot:>
```

Figure 6: Building and training module

5.4. Module 4: Prediction Module and Visualization Module

Prediction module: After the model is built using the above process, a prediction is made using the model. predict(xtest). The accuracy is calculated using accuracy_score imported from metrics metrics.accuracy_score (ytest, predicted). The main objective of the prediction module is to develop predictive models using machine learning algorithms such as regression, time series analysis, clustering, or deep Learning. For accurate predictions, it considers factors like historical crime data, demographics, weather, and social events. A prediction module is a critical component in crime prediction and analysis systems, providing actionable insights and foresight to help prevent and respond to criminal activities effectively. It's important to note that the effectiveness of such modules can be influenced by the quality of the data, the choice of predictive models, and the ongoing refinement and validation of the predictions. Additionally, ethical considerations, fairness, and privacy must be considered when using predictive models in law enforcement and public safety contexts. A prediction module is a component of a larger system that is designed to make predictions or forecasts based on input data and a predictive model. In the context of crime prediction and analysis, a prediction module would generate predictions about future criminal activities or trends based on historical data and other relevant factors (Fig.7).



```
<ipython-input-8-0fb0fbeef1f1>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a futu
re version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  corr_matrix = df.corr()
```

	YEAR	MONTH	DAY	HOUR	MINUTE	X	Y	Latitude	Longitude
YEAR									
MONTH		1.00	0.03	0.01	-0.00	0.01	0.01	0.01	-0.01
DAY		0.03	1.00	-0.00	0.01	-0.00	-0.00	-0.00	0.00
HOUR		0.01	-0.00	1.00	-0.05	0.00	-0.01	-0.01	0.00
MINUTE		-0.00	0.01	-0.05	1.00	0.02	0.02	0.02	0.02
X		0.01	-0.00	0.00	0.02	1.00	1.00	1.00	-1.00
Y		0.01	-0.00	-0.01	0.02	1.00	1.00	1.00	-1.00
Latitude		0.01	-0.00	-0.01	0.02	1.00	1.00	1.00	-1.00
Longitude		-0.01	0.00	0.00	0.02	-1.00	-1.00	-1.00	1.00

Figure 7: Prediction module

Visualization module: Using matplotlib library from sklearn. Analysis of the crime dataset is done by plotting various graphs. The above network and its analysis can be visualized only when we input data that visualize bar graphs, maps, and data clusters to represent the input data and predict crime-specific reports. The module uses various visualization techniques to present the data and insights effectively. Common visualization techniques include Maps: Geographic information system (GIS) visualizations to show crime hotspots and trends on maps. Charts and Graphs: Bar charts, line charts, scatter plots, and histograms to illustrate patterns and correlations in crime data:

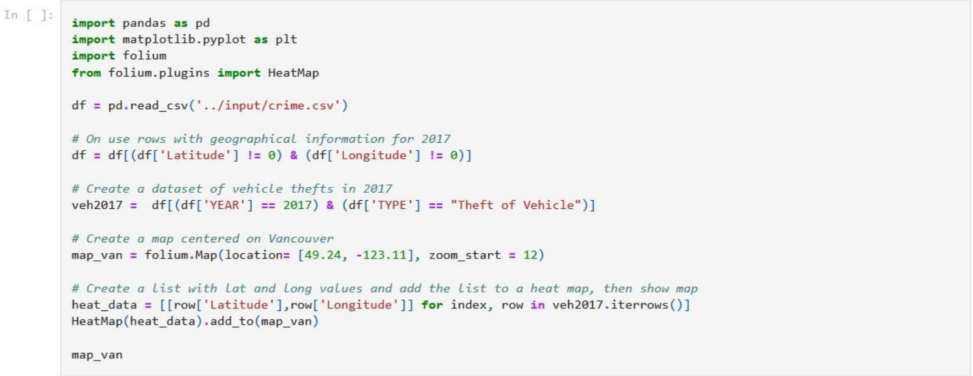
- Heatmaps: Displaying the intensity of crimes in specific geographic areas.
- Time Series Plots: Showing how crime rates change over time.
- Dashboards: Interactive dashboards with multiple visualizations for holistic analysis.

The visualization module is critical in helping law enforcement agencies, policymakers, and the community make data-informed decisions to effectively prevent and respond to Crime. It transforms complex data into actionable insights, facilitating the interpretation of crime patterns, resource allocation, and the development of targeted crime prevention strategies. Additionally, visualizations enhance transparency and engagement with the public, fostering trust and collaboration in crime prevention efforts

6. Efficiency of this Model

The efficiency of a crime prediction and analysis system is measured by its ability to generate easily understandable visuals of data quickly and with minimal resources. The proposed system, which uses various machine learning and data mining methods together, has several advantages over existing systems in terms of efficiency. First, the proposed system can generate graphs when thousands of input data are given. This contrasts existing systems, which just show the data asked. Generating these visuals for analysis makes the proposed system much more efficient. The second advantage of the proposed system is that it can generate visuals based on different parameters. The low number of parameters allows the proposed system to be more efficient than current systems, which often need a lot of parameters to produce an analyzed visual presentation of the dataset. Third, the system can employ this technique for various data sets irrespective of where and how much data is present. The prediction results are different for different algorithms, and the accuracy of the Boosted Decision Tree Classifier used in this is good, with an accuracy of 95.122%.

Waikato Environment for Knowledge Analysis, a free and open-source data mining programme, was used to conduct a study comparing reported vehicle thefts from the Communities and Crime Unnormalized Dataset with real crime statistics (WEKA). Communities and real-world crime records were used to develop three algorithms—linear regression, additive regression, and decision stump—using the same limited set of variables. Samples for the experiment were chosen at random. Among the three chosen methods, linear regression performed the best since it was able to account for some degree of randomness in the test samples. The goal of the study was to validate the usefulness of Machine Learning algorithms for forecasting the likelihood of vehicle thefts given a set of inputs. It shows the peak hours for thefts and the average number of thefts per hour. Knowledge Flow, a new graphical interface included within WEKA, can be used as a replacement for Internet Explorer. As part of the process orientation, it offers a more streamlined perspective on data mining by visually depicting the flow of information via individual learning components (represented by Java beans). The authors then detail a second graphical interface, the experimenter, whose stated goal is to evaluate and contrast the efficacy of various learning algorithms across a variety of datasets.



```
In [ ]:
import pandas as pd
import matplotlib.pyplot as plt
import folium
from folium.plugins import HeatMap

df = pd.read_csv('../input/crime.csv')

# On use rows with geographical information for 2017
df = df[(df['Latitude'] != 0) & (df['Longitude'] != 0)]

# Create a dataset of vehicle thefts in 2017
veh2017 =  df[(df['YEAR'] == 2017) & (df['TYPE'] == "Theft of Vehicle")]

# Create a map centered on Vancouver
map_van = folium.Map(location= [49.24, -123.11], zoom_start = 12)

# Create a list with lat and long values and add the list to a heat map, then show map
heat_data = [[row['Latitude'],row['Longitude']] for index, row in veh2017.iterrows()]
HeatMap(heat_data).add_to(map_van)

map_van
```

Figure 8: Sample code

The raw criminal data is cleaned and organised so that information about car thefts can be extracted from a larger dataset. The data is then utilised to make inferences about the hidden relationships between time of day and theft rates. The crime statistics utilised in this analysis are public information that may be found online in either CSV or Google format. Based on this data, a model is developed that may estimate the likelihood of thefts occurring in the area. Figure 8 depicts the code used to make vehicle theft predictions, while Figure 9 displays the final product.

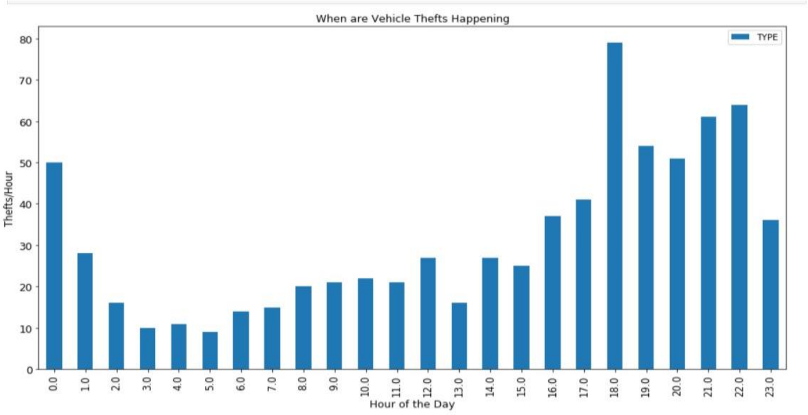


Figure 9: Implementation output

7. Discussions

The current volume of crime data in India makes it impossible for authorities to effectively analyse the situation and make appropriate decisions. Crime decision-making is the focus of this study since it has been identified as a significant issue. The crime rate in a country has a direct correlation with the standard of living there, as well as the country's GDP and international standing. Intending to secure society from crimes, there is a need for modern technologies and novel ways for improving crime analytics to protect their communities. Here, the proposed system may examine, detect, and foretell a wide range of regional crime risks. Predicting and analysing criminal behaviour is possible with the right methodology. Crime hotspots can be identified and visualised with the help of this technique.

Due to the predictive nature of this paper, data mining techniques are employed. Using ML, we can quickly and easily construct a prediction model for a suggested location by mining existing information. The KNN, DECISION TREE, and RANDOM FOREST algorithms are emphasised throughout this study. Crimes such as homicide, rape, and theft were all predicted by the algorithms. In the presented findings, the algorithm with the highest correlation coefficient and the lowest error values for each characteristic is highlighted. Input a crime summary, and the system will provide a prediction as to what kind of crime it is. The system's central hub presents the predicted crime types from the provided crime summary. The probabilities are displayed as a bar chart in the prediction results. The most likely categories of criminal activity are shown. The expected results allow field agents to swiftly determine the sort of crime. The outcome of this forecast is also shown in real time. The platform can forecast crime types and present the prediction results in real-time; consequently, field staff, such as police officers, may easily review predictive information about crimes obtained through the platform.

8. Conclusion

In this research, we develop models to forecast monthly crime rates by category in a city where such data exists. Many reasons, including rising poverty and corruption, are contributing to India's steadily rising crime rates. The proposed methodology is extremely helpful for law enforcement in their efforts to reduce crime. Crime analysts can use the project's interactive visualisations to better understand these networks. In this study, criminal hotspots are predicted using machine learning-trained bots. Enhanced prediction can be achieved by combining state-of-the-art machine learning methods with those of data mining. For better prediction, it is possible to strengthen data privacy, data reliability, and data accuracy. Machine learning agents can use signs found in the most basic information about criminal behaviour in an area to determine what type of crime occurred on a given day and where it occurred. The training agent suffers from imbalanced dataset categories and has been ready to address the problem by oversampling and under-sampling the dataset. This study proposes a crime data prediction using Jupyter Notebook with Python as its main language and Python's built-in tools, such as Pandas and Numpy, to expedite the process. The inputs are the sorts of crimes, and the outputs are the patterns in which these crimes are committed. Scikit gives comprehensive instructions for utilising Python's many library options. As a result, this concept aids the judicial system as a whole by strengthening the analytical department's ability to make more accurate predictions based on crime datasets.

8.1. Future Enhancements

Positive findings lead us to hope that improving crime data analysis and prediction will help law enforcement and intelligence agencies do their jobs better. Crime patterns can be used to produce more visual and intuitive methods of criminal and intelligence investigation. Data mining techniques such as clustering and classification are now within our reach thanks to our work in this area. Moreover, we may examine a wide range of statistics, including business survey data, poverty data, aid effectiveness data, and so on. Although the focus of this paper was to demonstrate the efficacy and accuracy of machine learning algorithms in predicting violent crimes, data mining has many other applications in the field of law enforcement, including mapping crime "hot spots," developing detailed profiles of individual offenders, and gaining insight into broader patterns of criminal behaviour. For law enforcement authorities, sifting through massive amounts of data using data mining programmes may be a time-consuming and difficult process. Yet the pinpoint accuracy with which one may infer and develop novel insights into methods of reducing speed It's not worth risking people's safety and security to prevent crime.

There is a lot of activity in the subject of criminology, with some machine learning and graph mining approaches being applied to analyse co-offender networks, uncover crime patterns, and predict future Crime in an effort to alleviate Crime-related issues. In the highly charged field of criminology, only a reliable model and effective algorithms can aid in the search for answers and provide insight into who might be committing crimes in the future. The driving force behind this effort is the need to bridge the gap between modern police tools and technology. Today, with the expanding mass of data in law enforcement and intelligence agencies, the largest problem they are facing is to accurately and efficiently assess that massive data. The time and money spent on recruiting, training, and supervising additional staff can be saved with the help of computer-aided analysis. Crime can be prevented in the first place by using a predictive method in crime analysis. These days, prediction methods and network mining are utilised all over the globe. This method, which makes use of data and graph mining tools, can also be applied to the study of crime patterns across India. We can utilise a method to examine the network of co-offenders in India and forecast the likely future network of offenders in light of the country's rising crime rate and number of criminals.

Acknowledgment: We thank our family whose prayers have got us here. Their belief in our vision helped us achieve our goal. We are also extremely grateful to our friends who stood by us patiently during the research.

Data Availability Statement: The data, graphs, and code supporting this study’s findings are available from the corresponding author upon reasonable request.

Funding Statement: No funding was received to conduct the research.

Conflicts of Interest Statement: Authors collectively produce this work where they all agree with the work’s points, issues, and findings.

Ethics and Consent Statement: Authors of the work unanimously consent to make this publication available to all interested people for reading, teaching, and Learning.

References

1. S. Sathyadevan, M. S. Devan and S. S. Gangadharan, "Crime analysis and prediction using data mining," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, India, 2014, pp. 406-412, doi: 10.1109/CNSC.2014.6906719.
2. L. McClendon and N. Meghanathan, “Using machine learning algorithms to analyze crime data,” Mach. Learn. Appl. Int. J., vol. 2, no. 1, pp. 1-12, 2015.
3. S. Kaur and W. Singh, “Systematic review of crime data mining,” International Journal of Advanced Research in computer science, vol. 8, no. 5, pp. 1336-1342, 2017.
4. T. Sonawanev, S. Shaikh, R. Shinde, and A. Sayyad, “Crime Pattern Analysis, Visualization and Prediction Using Data Mining,” Indian Journal of Computer Science and Engineering, vol.1, no.4, pp.681-686, 2015.
5. G. Saltos and M. Cocea, “An exploration of crime prediction using data mining on open data,” Int. J. Inf. Technol. Decis. Mak., vol. 16, no. 05, pp. 1155-1181, 2017.
6. H. Benjamin Fredrick David and A. Suruliandi, Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft Computing, vol.7, no.3, pp. 1459-1466, 2017.

7. O. Llaha, "Crime Analysis and Prediction using Machine Learning," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 496-501, doi: 10.23919/MIPRO48935.2020.9245120.
8. A. Khushabu, P. Tisksha, D. S. Kakade, C. G. Tumasare, and B. E. Wadhai, “Crime Detection Techniques Using Data Mining and K-Means,” International Journal of Engineering Research & Technology, vol.7, no.02, pp. 223-226, 2018.
9. A. B. Naeem, B. Senapati, M. S. Islam Sudman, K. Bashir, and A. E. M. Ahmed, “Intelligent road management system for autonomous, non-autonomous, and VIP vehicles,” World Electric Veh. J, vol. 14, no. 9, pp.32-42, 2023.
10. M. Sabugaa, B. Senapati, Y. Kupriyanov, Y. Danilova, S. Irgasheva, and E. Potekhina, “Evaluation of the prognostic significance and accuracy of screening tests for alcohol dependence based on the results of building a multilayer perceptron,” in Artificial Intelligence Application in Networks and Systems, Cham: Springer International Publishing, 2023, pp. 240–245.
11. D. Parasar, I. Sahi, S. Jain, and A. Thampuran, “Music recommendation system based on emotion detection,” in Artificial Intelligence and Sustainable Computing, Singapore: Springer Nature Singapore, 2022, pp. 29–43.
12. V. H. Patil, N. Dey, P. N. Mahalle, M. Shafi Pathan, and V. V. Kimbahune, Eds., Proceeding of first doctoral symposium on natural computing research: DSNCR 2020. Singapore: Springer Singapore, 2021.
13. S. Nandan Mohanty, S. K. Saxena, S. Satpathy, and J. M. Chatterjee, Eds., Applications of Artificial Intelligence in COVID-19. Singapore: Springer Singapore, 2021.
14. S. Parvatikar and D. Parasar, “Categorization of plant leaf using CNN,” in Intelligent Computing and Networking, Singapore: Springer Singapore, 2021, pp. 79–89.
15. R. Bora, D. Parasar, and S. Charhate, “A detection of tomato plant diseases using deep learning MNDLNN classifier,” Signal Image Video Process., vol. 17, no. 7, pp. 3255–3263, 2023.
16. S. Rangineni and D. Marupaka, “Data Mining Techniques Appropriate for the Evaluation of Procedure Information,” International Journal of Management, IT & Engineering, vol. 13, no. 9, pp. 12–25, 2023.
17. K. Bhanushali, K. Sivagnanam, B. K. Singh, L. T. Mittapally, and P. Reddi, “Analysis of Breast Cancer Prediction Using Multiple Machine Learning Methodologies”,” Int J Intell Syst Appl Eng, vol. 11, no. 3, pp. 1077–1084, 2023.
18. S. Parate, H. P. Josyula, and L. T. Reddi, “Digital Identity Verification: Transforming Kyc Processes In Banking Through Advanced Technology And Enhanced Security Measures,” International Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 9, pp. 128–137, 2023.
19. K. Peddireddy and D. Banga, “Enhancing Customer Experience through Kafka Data Steams for Driven Machine Learning for Complaint Management,” International Journal of Computer Trends and Technology, vol. 71, no. 3, pp. 7–13, 2023.
20. R. Kandepu, “Leveraging FileNet Technology for Enhanced Efficiency and Security in Banking and Insurance Applications and its future with Artificial Intelligence (AI) and Machine Learning,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 12, no. 8, pp. 20–26, 2023.
21. A. A. Alarood, M. Faheem, M. A. Al-Khasawneh, A. I. A. Alzahrani, and A. A. Alshdadi, “Secure medical image transmission using deep neural network in e-health applications,” Healthc. Technol. Lett., vol. 10, no. 4, pp. 87–98, 2023.
22. S. Markkandeyan et al., “Deep learning based semantic segmentation approach for automatic detection of brain tumor,” International Journal of Computers Communications & Control, vol. 18, no. 4, 2023.
23. M. A. Al-Khasawneh, A. Alzahrani, and A. Alarood, “Alzheimer’s Disease Diagnosis Using MRI Images,” in Data Analysis for Neurodegenerative Disorders, Singapore; Singapore: Springer Nature, 2023, pp. 195–212.
24. M. D. M. Akhtar, A. S. A. Shatat, S. A. H. Ahamad, S. Dilshad, and F. Samdani, Eds., “Optimized cascaded CNN for intelligent rainfall prediction model: a research towards Statistic based machine learning,” Theoretical Issues in Ergonomics Science, vol. 24, no. 5, pp. 564–2022.
25. M. Md et al., “Stock market prediction based on statistical data using machine learning algorithms”,” Journal of King Saud University - Science, vol. 34, no. 2, 2022.
26. M. D. M. Akhtar, R. S. Ali, A. S. A. Shatat, and S. A. Hameed, Eds., IoMT-based smart healthcare monitoring system using adaptive wavelet entropy deep feature fusion and improved RNN”, Multimedia Tools and Applications. Springer Nature.
27. D. K. Sharma, B. Singh, R. Regin, R. Steffi, and M. K. Chakravarthi, “Efficient Classification for Neural Machines Interpretations based on Mathematical models,” in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021.
28. F. Arslan, B. Singh, D. K. Sharma, R. Regin, R. Steffi, and S. Suman Rajest, “Optimization technique approach to resolve food sustainability problems,” in 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2021.
29. G. A. Ogunmola, B. Singh, D. K. Sharma, R. Regin, S. S. Rajest, and N. Singh, “Involvement of distance measure in assessing and resolving efficiency environmental obstacles,” in 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2021.

30. A. Garg, A. Ghosh, and P. Chakrabarti, “Gain and bandwidth modification of microstrip patch antenna using DGS,” in Proc. International Conference on Innovations in Control, Communication and Information Systems (ICICCI-2017), India, 2017.
31. A. K. Sharma, A. Panwar, P. Chakrabarti, and S. Vishwakarma, “Categorization of ICMR using feature extraction strategy and MIR with ensemble learning,” Procedia Comput. Sci., vol. 57, pp. 686–694, 2015.
32. A. Prasad, D. Gupta, and P. Chakrabarti, “Monitoring Users in Cloud Computing : Evaluating the Centralized Approach,” in Proc. 2nd International Conference on Advanced Computing, Networking and Security (ADCONS) India, 2013.
33. A. Sarwar Zamani et al., “Cloud Network Design and Requirements for the Virtualization System for IoT Networks”,” IJCSNS International Journal of Computer Science and Network Security, vol. 22, no. 11, 2022.
34. A. Singh and P. Chakrabarti, “Ant based resource discovery and mobility aware trust management for Mobile Grid systems,” in 2013 3rd IEEE International Advance Computing Conference (IACC), 2013.
35. D. K. Sharma, B. Singh, M. Raja, R. Regin, and S. S. Rajest, “An Efficient Python Approach for Simulation of Poisson Distribution,” in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021.
36. D. K. Sharma, N. A. Jalil, R. Regin, S. S. Rajest, R. K. Tummala, and Thangadurai, “Predicting network congestion with machine learning,” in 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 2021
37. K. Sharma, B. Singh, E. Herman, R. Regine, S. S. Rajest, and V. P. Mishra, “Maximum information measure policies in reinforcement learning with deep energy-based model,” in 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2021.
38. M. A. Al-Khasawneh, A. Alzahrani, and A. Alarood, “An Artificial Intelligence Based Effective Diagnosis of Parkinson Disease Using EEG Signal,” in Data Analysis for Neurodegenerative Disorders, Singapore; Singapore: Springer Nature, 2023, pp. 239–251.
39. M. A. S. Al-Khasawneh, M. Faheem, E. A. Aldhahri, A. Alzahrani, and A. A. Alarood, “A MapReduce based approach for secure batch satellite image encryption,” IEEE Access, vol. 11, pp. 62865–62878, 2023.
40. M. D. M. Akhtar, D. Ahamad, A. S. A. Shatat, and E. M. Abdalrahman, Eds., “Enhanced heuristic algorithm-based energy-aware resource optimization for cooperative IoT”,” International Journal of Computers and Applications, vol. 44, no. 10, 2022.
41. M. D. M. Akhtar, D. Ahamad, and A. S. A. Shatat, Eds., “A novel hybrid meta-heuristic concept for green communication in IoT networks: An intelligent clustering model”,” in International journal communication systems, vol. 35, wiley, 2021.
42. P. Kumar, A. S. Hati, S. Padmanaban, Z. Leonowicz, and P. Chakrabarti, “Amalgamation of transfer learning and deep convolutional neural network for multiple fault detection in SCIM,” in 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 2020.
43. S. Viswakarma, P. Chakrabarti, D. Bhatnagar, and A. K. Sharma, “Phrase Term Static Index Pruning Based on the Term Cohesiveness,” in Proc. International Conference on Computational Intelligence and Communication Networks (CICN) India, 2014.