# CRIME TYPE PREDICTION AND ANALYSIS USING MACHINE LEARNING

Mini Project Report

Submitted by

**PUNNYA.P.S.**

*Submitted in partial fulfillment of the requirements for the award of the degree of*

***Master of Computer Applications***
***Of***
***A P J Abdul Kalam Technological University***


**FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®**
**ANGAMALY-683577, ERNAKULAM(DIST)**
**MARCH 2022**

# DECLARATION

I **PUNNYA.P.S.**, hereby declare that the report of this project work, submitted to the Department of Computer Applications, Federal Institute of Science and Technology (**FISAT**), Angamaly in partial fulfillment of the award of the degree of Master of Computer Application is an authentic record of our original work.

The report has not been submitted for the award of any degree of this university or any other university.

**Date : 04-03-2022**

**Place: Angamaly**

# FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY (FISAT)®

**ANGAMALY, ERNAKULAM-683577**

**DEPARTMENT OF COMPUTER APPLICATIONS**


## CERTIFICATE

This is to certify that the project report titled **"crime type prediction and analysis using machine learning"** submitted by **PUNNYA.P.S.** towards partial fulfillment of the requirements for the award of the degree of Master of Computer Applications is a record of bonafide work carried out by them during the year 2022.

**Project Guide** **Head of the Department**

Submitted for the viva-voice held on ..... at .....

**Examiner1 :**

**Examiner2 :**

# ACKNOWLEDGEMENT

Gratitude is a feeling which is more eloquent than words, more silent than silence. To complete this project work we needed the direction, assistance and cooperation of various individuals, which is received in abundance with the grace of God.

We hereby express our deep sense of gratitude to **Dr. Manoge George**, Principal of FISAT and **Dr. C Sheela**, Vice principal of FISAT, for allowing us to utilize all the facilities of the college.

Our sincere thanks to **Dr. Deepa Mary Mathew**, Head of the department of Computer Applications FISAT and scrum master and our Internal guide for this project **Ms.Anju.L**for giving as valuable guidance, constructive suggestions and comment during our project work. We also express our boundless gratitude to all the lab faculty members for their guidance.

Finally we wish to express a whole heart-ed thanks to our parents, friends and well-wishers who extended their help in one way or other in preparation of our project. Besides all, We thank GOD for everything.

# ABSTRACT

Crime analysis and prediction is a systematic approach for analysing and identifying different patterns, relationships and trends in crime. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime in a particular area. The regions with high probability of occurrence of crime is also predicted by the system.

The system developed will help to speed up the process of solving crimes for the law enforcement agencies. In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analyzing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement.The government has to spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. The extraction of new information is predicted using the existing datasets.

The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime in a particular area. This study imposes one such crime pattern analysis by using crime data obtained from kaggle open source which in turn used for the prediction of most recently occurring crimes . it consists of crime information like location description, type of crime, date, time, etc... .Based on this information the officials can take charge and try to reduce the crime rate. the system will also convert crime information into a regression problem, so that it will help detectives in solving crimes faster.

In this project decision tree algorithm and Naive bayes algorithm are used. For prediction we are using the decision tree concept.The main advantages include its robust nature and also it works well with large data sets. This feature helps the algorithms to make better decisions about variables.For classification we are using an algorithm called naive bayes which is a supervised learning method as well as

a statistical method for classification.

# Contents

# Chapter 1

# INTRODUCTION

Crimes have a negative effect on any society both socially and ecnomically.Law enforcement bodies face numerous challenges while trying to prevent crimes.Governments spends lot of money and efforts to analysis and stop crimes from taking place.. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities.This system can predict which crime is happening frequently and region which have high probability for crime occurrences .And analyzing the dataset which consist of numerous crimes and predicting the type of crime in a particular area.System is mainly predict the type of crimes such as Robbery,motorvechicle theft,theft,narcotics,weapons violation.

The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Based on this information the officials can take charge and try to reduce the crime rate.The Machine learning algorithms are used here. The algorithms used here is Decision tree classifier and naive bayes algorithm. Decision tree algorithm is used here for prediction.And for classification naive bayes is used.

The occurrences of crime depended on several factors such as intelligence of criminals, security of a location,etc The work has followed the steps that used in data analysis, in which the important phases are Data collection ,data classification,

pattern identification, prediction and visualization. The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm.

# Chapter 2

# PROOF OF CONCEPT

## 2.1 Existing System

Crimes are increasing predominantly.finding out the crime patterns that a particular criminal is likely to perform will speed up the process of law system and reduce the crimes in society.For accessing of any information officials can refer the bulk of hard copies.more time needed for searching the required information.The large amount data maintenance is the big problem.searching and predicting a perticular data manually is not possible from the files.

## 2.2 Proposed System

A system which can analysis, detect, and predict various crime probability in given region. And we can find the areas with a high concentration of crime using machine learning techniques.The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime in a particular area.It estimates which type of crime contributes the most along with time period and location where it has happened.

## 2.3 Objectives

The main objective of the project is to predict the crime rate and analyze the crime rate to be happened in future. Based on this information the officials can take charge and try to reduce the crime rate.The concept of multi linear regression is used for predicting the graph between the types of crimes (independent variable) and the year (dependent variable) .the system will also convert crime information into a regression problem, so that it will help detectives in solving crimes faster.Crime analysis based on available information to extract crime patterns. Using various multi linear regression techniques, frequency of occurring crime can be predicted based on territorial distribution of existing data and crime recognition.

# Chapter 3

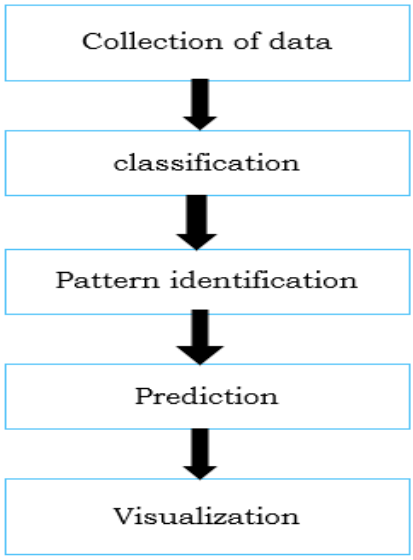
# IMPLEMENTATION

The dataset we took is a Chicago crime dataset from Kaggle.com. Shape of dataset (6000). We have isolated the label column from the data frame. Pre-processing steps are applied and then we split the dataset into train and test data set. Feature extraction is done after completing pre-processing on the given input data. The extracted features are then given as input to the Classification Algorithm.The output of this project gives type of crime based on the given inputs such as date,time,location,arrest,domestic,year etc.

## 3.1 TOOLS OR PROGRAMMING LANGUAGE

**FRONT END:**
HTML
**BACK END:**
Python

## 3.2 ARCHITECTURE




## 3.3 MODULES

**Data Preprocessing:**

This process includes methods to remove any null values or infinite values which may affect the accuracy of the system. The main steps includeFormatting, cleaning and sampling.Cleaning process is used for removal or fixing of some missing data there may be data that are incomplete. Sampling is the process where appropriate data are used which may reduce the running time for the algorithm.Using python, the preprocessing is done.

The null values are removed using df = df.dropna() where df is the data frame. The categorical attributes (Location, Block, Crime Type, Community Area) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like month and hour which can be used as feature for the model.

**Feature Extraction:**

Feature extraction helps to reduce the amount of redundant data from the data

set.in the end reduction of the data helps to build the model with less machine effort and also increase the speed of learning and generalization steps in the machine learning process. Feature extraction is used to identify key features in the data for coding by learning from the coding of the original data set to derive new ones.
Features selection is done which can be used to build the model. The attributes used for feature selection are Block, Location, District, Community area, X coordinate , Y coordinate, Latitude , Longitude, Hour and month.

**Training the model:**
A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is used to modify the model.
**Evaluation:**
Model evaluation techniques in machine learning are helping us to find a better model among all other models in machine learning. It is simply the selection of machine learning models or measuring the performance of machine learning models.

## 3.4 DATASET

The dataset used to test the efficiency of the model is produced by kaggle containing 9500 data.

**Attributes of dataset:**
primary type,Date,Month,Year,Arrest,Domestic,Location,Ward

| Date | Primary_Type | Location_Description | Arrest | Domestic | District | Ward | Community_A | Year |
|---|---|---|---|---|---|---|---|---|
| 29-Apr-05 | MOTOR VEHICLE THEFT | BANK | FALSE | FALSE | 25 | 29 | 19 | 2005 |
| 29-Apr-05 | MOTOR VEHICLE THEFT | BANK | FALSE | FALSE | 25 | 29 | 19 | 2005 |
| 26-Dec-05 | MOTOR VEHICLE THEFT | STREET | TRUE | FALSE | 7 | 3 | 68 | 2005 |
| 26-Dec-05 | MOTOR VEHICLE THEFT | STREET | TRUE | FALSE | 7 | 3 | 68 | 2005 |
| 24-Sep-05 | NARCOTICS | STREET | TRUE | FALSE | 6 | 17 | 71 | 2005 |
| 24-Sep-05 | NARCOTICS | STREET | TRUE | FALSE | 6 | 17 | 71 | 2005 |
| 08-Dec-05 | NARCOTICS | RESIDENCE | TRUE | TRUE | 8 | 15 | 66 | 2005 |
| 08-Dec-05 | NARCOTICS | RESIDENCE | TRUE | TRUE | 8 | 15 | 66 | 2005 |
| 01-Oct-05 | ROBBERY | APARTMENT | FALSE | TRUE | 12 | 27 | 27 | 2005 |
| 01-Oct-05 | ROBBERY | APARTMENT | FALSE | TRUE | 12 | 27 | 27 | 2005 |
| 01-Apr-05 | ROBBERY | RESIDENCE | TRUE | TRUE | 9 | 12 | 58 | 2005 |
| 01-Apr-05 | ROBBERY | RESIDENCE | TRUE | TRUE | 9 | 12 | 58 | 2005 |
| 11-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 25 | 29 | 19 | 2005 |
| 11-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 25 | 29 | 19 | 2005 |
| 10-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 2 | 28 | 2005 |
| 10-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 2 | 28 | 2005 |
| 11-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 15 | 28 | 25 | 2005 |
| 11-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 15 | 28 | 25 | 2005 |
| 12-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 27 | 2005 |
| 12-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 27 | 2005 |

| A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|
| 15-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 2 | 28 | 2005 | | |
| 15-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 2 | 28 | 2005 | | |
| 19-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 26 | 2005 | | |
| 19-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 26 | 2005 | | |
| 25-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 24 | 29 | 2005 | | |
| 25-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 24 | 29 | 2005 | | |
| 31-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 27 | 2005 | | |
| 31-Dec-05 | THEFT | OTHERS | FALSE | FALSE | 11 | 28 | 27 | 2005 | | |
| 18-Jun-05 | THEFT | RESIDENCE | FALSE | FALSE | 6 | 6 | 44 | 2005 | | |
| 18-Jun-05 | THEFT | RESIDENCE | FALSE | FALSE | 6 | 6 | 44 | 2005 | | |
| 15-Jun-05 | THEFT | HOTEL-MOTEL | FALSE | FALSE | 6 | 8 | 44 | 2005 | | |
| 15-Jun-05 | THEFT | HOTEL-MOTEL | FALSE | FALSE | 6 | 8 | 44 | 2005 | | |
| 13-Jun-05 | THEFT | STREET | FALSE | FALSE | 6 | 18 | 71 | 2005 | | |
| 13-Jun-05 | THEFT | STREET | FALSE | FALSE | 6 | 18 | 71 | 2005 | | |
| 11-Jun-05 | THEFT | APARTMENT | FALSE | FALSE | 6 | 17 | 71 | 2005 | | |
| 11-Jun-05 | THEFT | APARTMENT | FALSE | FALSE | 6 | 17 | 71 | 2005 | | |
| 09-Jun-05 | THEFT | STREET | FALSE | FALSE | 6 | 8 | 44 | 2005 | | |
| 09-Jun-05 | THEFT | STREET | FALSE | FALSE | 6 | 8 | 44 | 2005 | | |
| 08-Jun-05 | THEFT | STREET | FALSE | FALSE | 22 | 21 | 72 | 2005 | | |
| 08-Jun-05 | THEFT | STREET | FALSE | FALSE | 22 | 21 | 72 | 2005 | | |
| 11-Oct-05 | THEFT | APARTMENT | FALSE | TRUE | 10 | 25 | 31 | 2005 | | |
| 11-Oct-05 | THEFT | APARTMENT | FALSE | TRUE | 10 | 25 | 31 | 2005 | | |
| 12-Apr-05 | THEFT | SIDEWALK | FALSE | FALSE | 8 | 23 | 56 | 2005 | | |
| 12-Apr-05 | THEFT | SIDEWALK | FALSE | FALSE | 8 | 23 | 56 | 2005 | | |

## 3.5 FRAME WORK

**FLASK**

Flask is a micro web framework written in python. It is classified as a microframework because it does not require particular tools or libraries.It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, flask supports extensions that can add application features as if they were implemented in flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Flask is a web application framework written in python. It was developed by armin ronacher, who led a team of international python enthusiasts called poocco. Flask is based on the werkzeg WSGI toolkit and the jinja2 template engine.Both are pocco projects.

WSGI

The web server gateway interface (web server gateway interface, WSGI) has been used as a standard for python web application development. WSGI is the specification of a common interface between web servers and web applications.

Werkzeug

Werkzeug is a WSGI toolkit that implements requests, response objects, and utility functions. This enables a web frame to be built on it. The flask framework uses werkzeug as one of its bases.

jinja2

Jinja2 is a popular template engine for python.A web template system combines a template with a specific data source to render a dynamic web page.

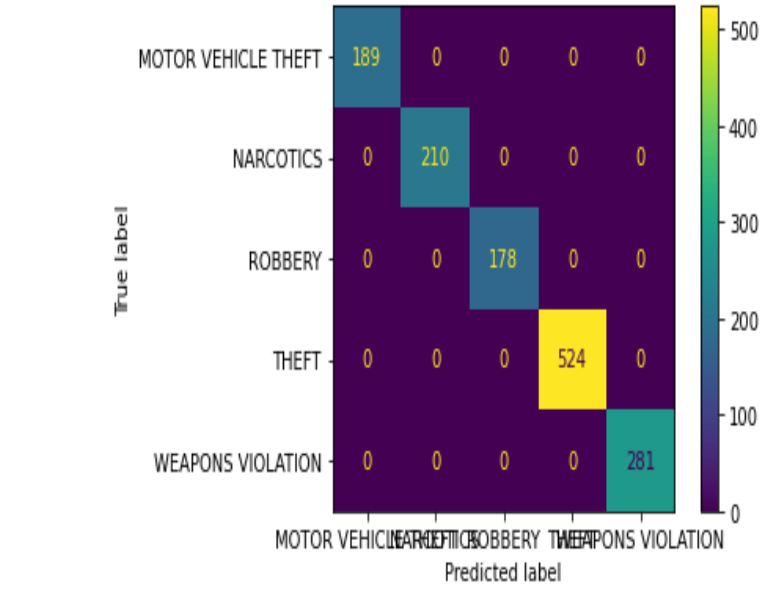
## 3.6 ALGORITHMS TO BE USED

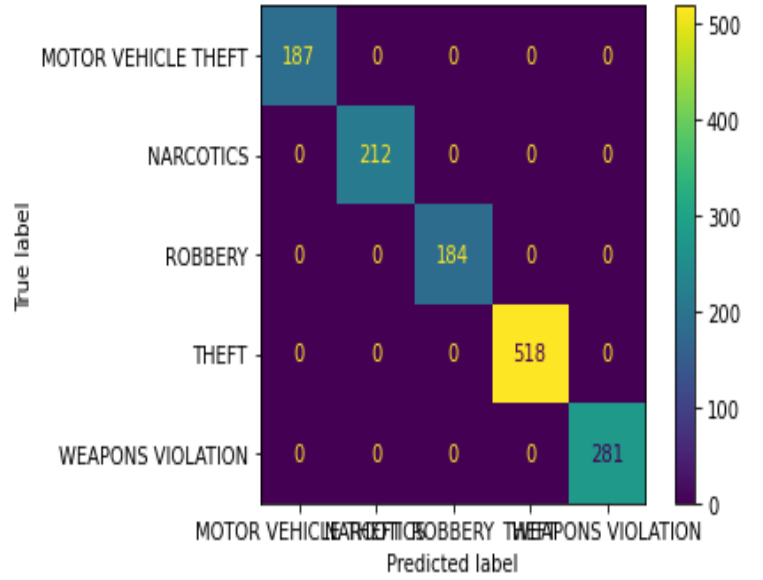
**Decision tree Algorithm**

For prediction we are using the decision tree concept. A decision tree is similar to a graph in which internal node represents test on an attribute, and each branch represents outcome of a test. The main advantage of using decision tree is that it is simple to understand and interpret. The other advantages include its robust nature and also it works well with large data sets. This feature helps the algorithms to make better decisions about variables.

**Naive Bayes**

For classification we are using an algorithm called naive bayes which is a supervised learning method as well as a statistical method for classification. Naive bayes classifier is a probabilistic classifier which when given an input gives a probability distribution of set of all classes rather than providing a single output. The advantage of using naive bayes classifier is that it is simple, and converges quicker than logistic regression.


MOTOR VEHICLE THEFT
NARCOTICS
ROBBERY
THEFT
WEAPONS VIOLATION
True label
189 0 0 0 0
0 210 0 0 0
0 0 178 0 0
0 0 0 524 0
0 0 0 0 281
500
400
300
200
100
0
Predicted label


MOTOR VEHICLE THEFT
NARCOTICS
ROBBERY
THEFT
WEAPONS VIOLATION
True label
187 0 0 0 0
0 212 0 0 0
0 0 184 0 0
0 0 0 518 0
0 0 0 0 281
500
400
300
200
100
0
Predicted label

# Chapter 4

# RESULT ANALYSIS

Accuracy is often the most used metric representing the percentage ofcorrectly predicted observations, either true or false. To calculate the accuracy of a model performance, the following equation can be used: In most cases, high accuracy value represents a good model, but considering the fact that we are training a classification model in our case, an article that was predicted as true while it was actually false (false positive) can have negative consequences; similarly, if an article was predicted as false while it contained factual data, this can create trust issues.

Accuarcy = TP+TN/TP+TN+FP+FN

**confusion matrix**

A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making.For a binary classification problem, we would have a 2 x 2 matrix as shown below with 4 values:

1. TP = True Positives
2. FP = False Positives
3. TN = True Negatives
4. FN = False Negatives

# Chapter 5

# CONCLUSION AND FUTURE SCOPE

## 5.1 Conclusion

Crime prediction is current trend in the society. It aims at reducing the crime occurrences by predicting the possible crimes that might occur in the future days.. The model predicts the type of crime and Data visualization helps in analysis of data set and prediction of crimes. Thus the crime that occur the most could be predicted and spotted using Naïve Bayesian Classification. The performance of the algorithm is also calculated by using some standard metrics. The crime rates in India are increasing day by day due to many factors such as increase in poverty, implementation, corruption, etc. The proposed model is very useful for both the investigating agencies and the police official in taking necessary steps to reduce crime.

## 5.2 Future Scope

Along with the present scope of our project ,which is prediction of the crime an individual criminal is likely to commit,we can also predict the estimated time for

the crime to take place as a future scope. Along with this, one can try to predict the location of the crime. In the future we can use apriori algorithm which will determine the next crime, a criminal is about to commit.

# Chapter 6

# APPENDIX

In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
df = pd.read_excel("crime.xlsx") # ROWS , COL => DATAFRAME
dfhead=df.head()
```

In [3]:
```
df['day'] = df['Date'].dt.day
df['month'] = df['Date'].dt.month
df['year'] = df['Date'].dt.year
```

In [4]:
```
df.head()
```

Out[4]:

| | Date | Primary_Type | Location_Description | Arrest | Domestic | District | Ward | Community_Area | Year | day | month | year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2005-04-29 04:45:00 | MOTOR VEHICLE THEFT | BANK | False | False | 25 | 29 | 19 | 2005 | 29 | 4 | 2005 |
| 1 | 2005-04-29 04:45:00 | MOTOR VEHICLE THEFT | BANK | False | False | 25 | 29 | 19 | 2005 | 29 | 4 | 2005 |
| 2 | 2005-12-26 03:00:00 | MOTOR VEHICLE THEFT | STREET | True | False | 7 | 3 | 68 | 2005 | 26 | 12 | 2005 |
| 3 | 2005-12-26 03:00:00 | MOTOR VEHICLE THEFT | STREET | True | False | 7 | 3 | 68 | 2005 | 26 | 12 | 2005 |
| 4 | 2005-09-24 12:22:13 | NARCOTICS | STREET | True | False | 6 | 17 | 71 | 2005 | 24 | 9 | 2005 |

In [5]:
```
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
le.fit(df['Location_Description'])
df['desc']=le.transform(df['Location_Description'])
```

In [6]:
```
features = ["Arrest","Domestic","District","Ward","Community_Area","Year","day","month"]
target = "Primary_Type"
X = df.loc[:,features] # df.loc[1:5,icmp_features]
y = df.loc[:,target]
```

In [7]:
```
X.head()
```

Out[7]:

| | Arrest | Domestic | District | Ward | Community_Area | Year | day | month |
|---|---|---|---|---|---|---|---|---|
| 0 | False | False | 25 | 29 | 19 | 2005 | 29 | 4 |
| 1 | False | False | 25 | 29 | 19 | 2005 | 29 | 4 |
| 2 | True | False | 7 | 3 | 68 | 2005 | 26 | 12 |
| 3 | True | False | 7 | 3 | 68 | 2005 | 26 | 12 |
| 4 | True | False | 6 | 17 | 71 | 2005 | 24 | 9 |

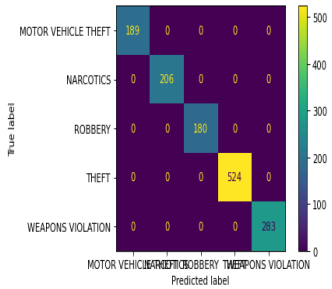
In [8]:
```
y.head()
```

Out[8]:
```
0    MOTOR VEHICLE THEFT
1    MOTOR VEHICLE THEFT
2    MOTOR VEHICLE THEFT
3    MOTOR VEHICLE THEFT
4              NARCOTICS
Name: Primary_Type, dtype: object
```

In [9]:
```
classes = np.unique(y)
classes
```

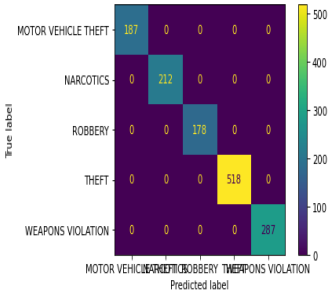
```
Out[9]: array(['MOTOR VEHICLE THEFT', 'NARCOTICS', 'ROBBERY', 'THEFT',
               'WEAPONS VIOLATION'], dtype=object)

In [10]: X.shape
         totsize=len(X)

In [11]: from sklearn.model_selection import train_test_split

In [12]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)

In [13]: X_train.shape
         trsize=len(X_train)

In [14]: tstsize=X_test.shape

In [15]: from sklearn.tree import DecisionTreeClassifier
         from sklearn.metrics import accuracy_score
         DT = DecisionTreeClassifier()
         y_pred = DT.fit(X_train, y_train)
         y_score=DT.predict(X_test)
         dacc=accuracy_score(y_test, y_score)

         dacc

Out[15]: 0.926917510853835

In [16]: from sklearn.metrics import plot_confusion_matrix
         import matplotlib.pyplot as plt
```

```
In [16]: from sklearn.metrics import plot_confusion_matrix
         import matplotlib.pyplot as plt
         plot_confusion_matrix(DT,X_test,y_score)
         plt.savefig('static\\dt.png')
```

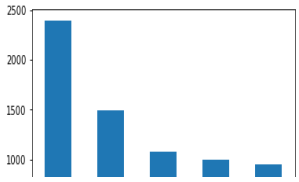



```
In [17]: from sklearn.naive_bayes import GaussianNB
         from sklearn.metrics import accuracy_score
         gnb = DecisionTreeClassifier()
         y_pred = gnb.fit(X_train, y_train)
         y_score=gnb.predict(X_test)
         nacc=accuracy_score(y_test, y_score)
         nacc

Out[17]: 0.9254703328509407
```

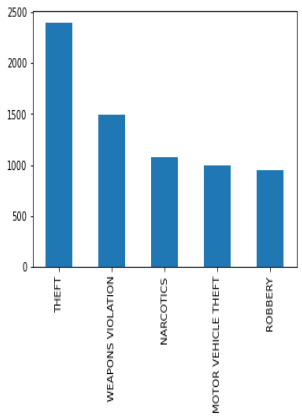
In [18]:
```
plot_confusion_matrix(gnb,X_test,y_score)
plt.savefig('static\\nb.png')
```




In [19]:
```
df['Primary_Type'].value_counts().plot(kind='bar');
```




In [19]:
```
df['Primary_Type'].value_counts().plot(kind='bar');
```




In [20]:
```
Arrest = 'False' #@param ["True", "False"]
Domestic='False' #@param ["True", "False"]
District= '25' #@param {type:"string"}
Ward= '29' #@param {type:"string"}
Area= '19' #@param {type:"string"}
```

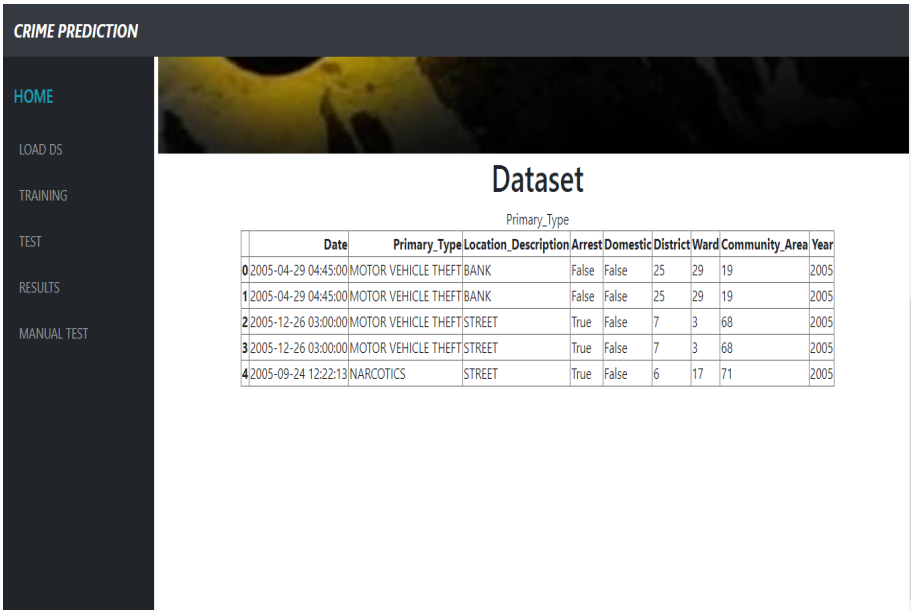
In [20]:
```
Arrest = 'False' #@param ["True", "False"]
Domestic='False' #@param ["True", "False"]
District= '25' #@param {type:"string"}
Ward= '29' #@param {type:"string"}
Area= '19' #@param {type:"string"}
Year= '2005' #@param {type:"string"}
Day= '10' #@param {type:"string"}
Month= '8' #@param {type:"string"}

tst=[bool(Arrest),bool(Domestic),District,Ward,Area,Year,Day,Month]
output=DT.predict([tst])
print("Result : "+output[0])
```

```
Result : WEAPONS VIOLATION
```

In [ ]:
```
from flask import Flask,redirect,render_template,request

app=Flask(__name__)

@app.route("/mantest",methods=['GET','POST'])
def mntest():
    if request.method=="POST":
            Arrest = request.form['Arrest']
            Domestic=request.form['dome'] #@param ["True", "False"]
            District= request.form['district'] #@param {type:"string"}
            Ward= request.form['ward'] #@param {type:"string"}
            Area= request.form['area'] #@param {type:"string"}
            Year= request.form['year'] #@param {type:"string"}
            Day= request.form['day'] #@param {type:"string"}
            Month= request.form['month'] #@param {type:"string"}

            tst=[bool(Arrest),bool(Domestic),District,Ward,Area,Year,Day,Month]
            output=DT.predict([tst])
            res="Result : "+output[0]
#           res="hello"

            return render_template("mntest.html",res=res,ar=Arrest,do=Domestic,di=District,w=Ward,are=Area,yr=Year,da=Day,mn=Mont
    return render_template("mntest.html",res="",ar="",do="",di="",w="",are="",yr="",da="",mn="")

@app.route("/loadds")
def loadds():
#     return "hello"
    return render_template("dataset.html",tables=[dfhead.to_html(classes='data')],titles=dfhead.columns.values)

@app.route("/training")
def training():
    return render_template("trainingview.html",ttsize=totsize,trnsize=trsize,tssize=tstsize)
@app.route("/test")
def test():
    return render_template("testview.html",tssize=tstsize,accdt=dacc,accnb=nacc)
@app.route("/result")
def result():
    return render_template("resultview.html")

@app.route("/")
def user():
    return render_template("dataset.html",tables=[dfhead.to_html(classes='data')],titles=dfhead.columns.values)
```

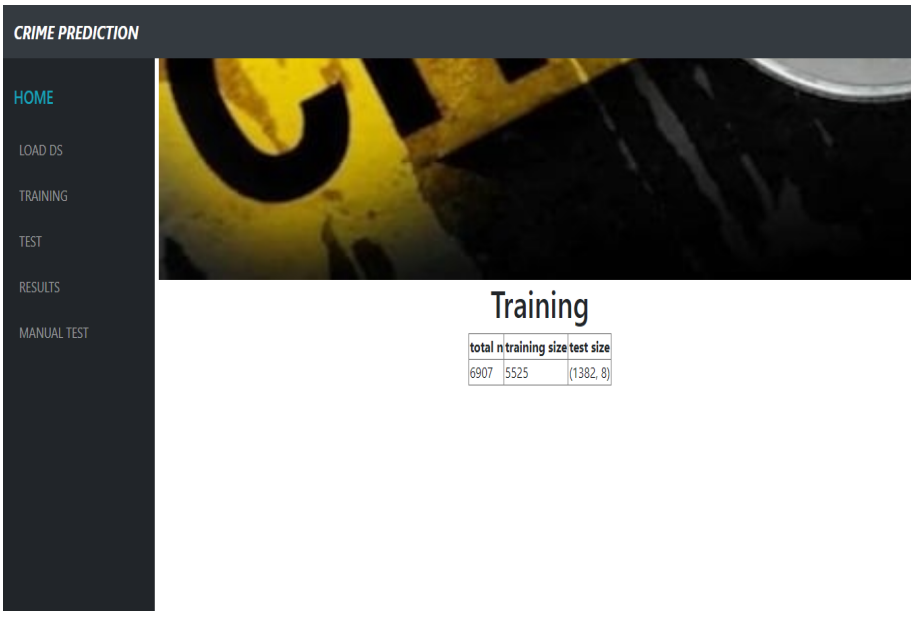
```
            res="Result : "+output[0]
#             res="hello"

            return render_template("mntest.html",res=res,ar=Arrest,do=Domestic,di=District,w=Ward,are=Area,yr=Year,da=Day,mn=Mont
    return render_template("mntest.html",res="",ar="",do="",di="",w="",are="",yr="",da="",mn="")

@app.route("/loadds")
def loadds():
#     return "hello"
    return render_template("dataset.html",tables=[dfhead.to_html(classes='data')],titles=dfhead.columns.values)

@app.route("/training")
def training():
    return render_template("trainingview.html",ttsize=totsize,trnsize=trsize,tssize=tstsize)
@app.route("/test")
def test():
    return render_template("testview.html",tssize=tstsize,accdt=dacc,accnb=nacc)
@app.route("/result")
def result():
    return render_template("resultview.html")

@app.route("/")
def user():
    return render_template("dataset.html",tables=[dfhead.to_html(classes='data')],titles=dfhead.columns.values)



if __name__=='__main__':
    from werkzeug.serving import run_simple
    run_simple('localhost',9000,app)
```


Figure 6.1: HOME PAGE




| | Date | Primary_Type | Location_Description | Arrest | Domestic | District | Ward | Community_Area | Year |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 2005-04-29 04:45:00 | MOTOR VEHICLE THEFT | BANK | False | False | 25 | 29 | 19 | 2005 |
| 1 | 2005-04-29 04:45:00 | MOTOR VEHICLE THEFT | BANK | False | False | 25 | 29 | 19 | 2005 |
| 2 | 2005-12-26 03:00:00 | MOTOR VEHICLE THEFT | STREET | True | False | 7 | 3 | 68 | 2005 |
| 3 | 2005-12-26 03:00:00 | MOTOR VEHICLE THEFT | STREET | True | False | 7 | 3 | 68 | 2005 |
| 4 | 2005-09-24 12:22:13 | NARCOTICS | STREET | True | False | 6 | 17 | 71 | 2005 |

Figure 6.2: DATASET VIEW




| total n | training size | test size |
|---|---|---|
| 6907 | 5525 | (1382, 8) |

Figure 6.3: TRAINING SET

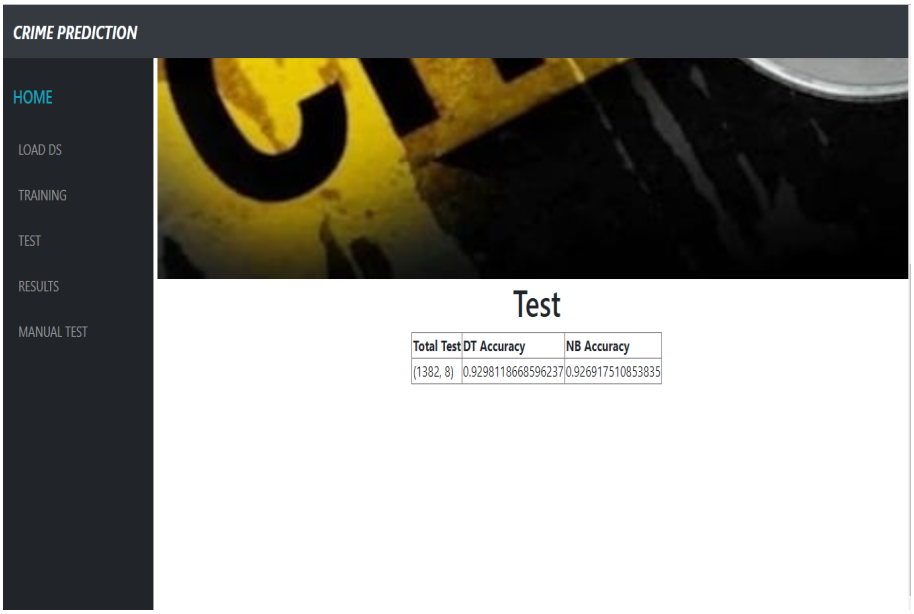



Figure 6.4: TEST SET

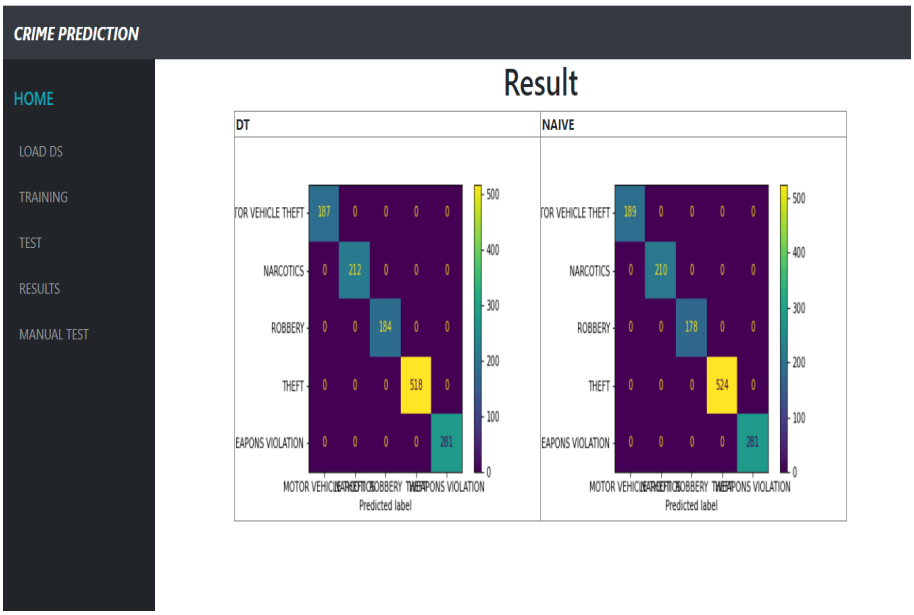



Figure 6.5: RESULT

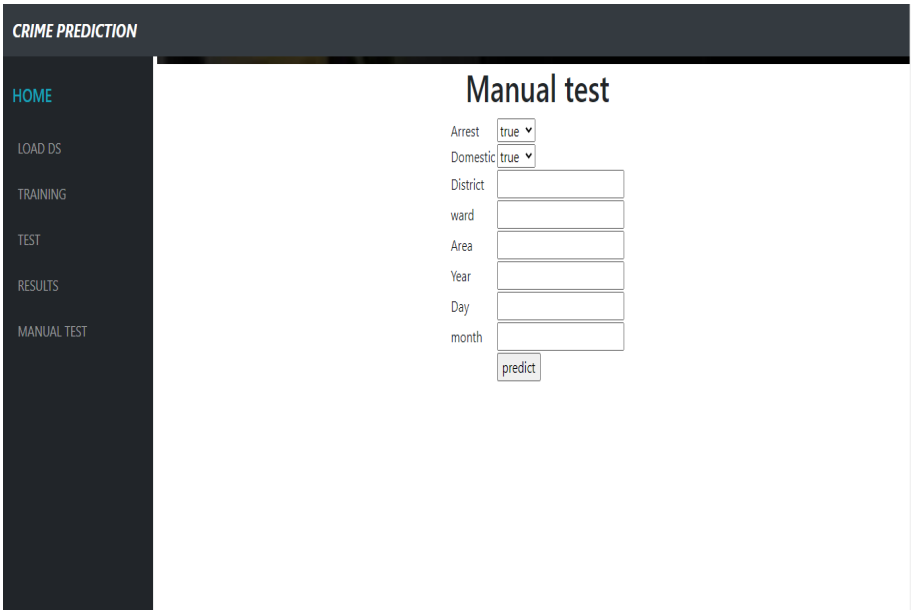



Figure 6.6: MANUAL TEST

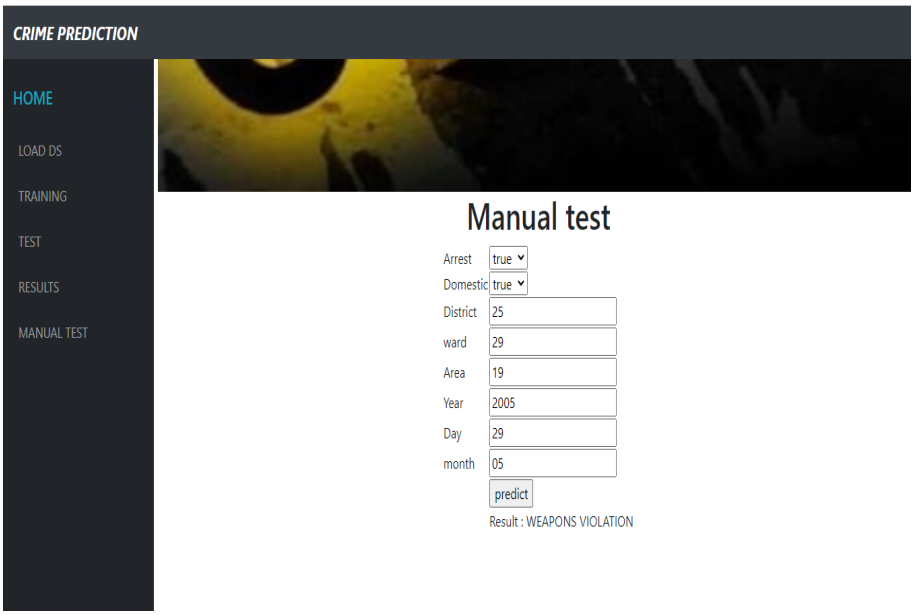



Figure 6.7: MANUAL TEST

# Chapter 7

# REFERENCES


- www.youtube.com
- www.wikipedia.com

1 malathi. A and dr. S. Santhosh baboo. Article:an enhanced algorithm to predict a future crime using data mining. International journal of computer applications, 21(1):1-6, may 2011. Published by foundation of computer science

[2] @Articlesathyadevan2014crimeaa, title=crime analysis and prediction using data mining, author=shiju sathyadevan and mary scinthia james devan and S. Surya gangadharan, journal=2014 first international conference on networks & soft computing (ICNSC2014), year=2014, pages=406-412