1.2 PROJECT DESCRIPTION:

Day after day, the crime data rate has been increasing as a result of modern technologies and hi-tech methods as it helps criminals to perform illegal activities. Evidently, according to the Crime Record Bureau, crimes like burglary, arson, and so on have been increasing, while crimes like murder, sex abuse, gang rape, etc. have been increasing. Crime data is acquired from various blogs, news, and websites. A huge amount of data is utilized as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster. A vital component of existence is security. Our most basic needs cannot be satisfied unless all are safe. Therefore, having a sense of security is essential to achieving our objectives, whether they will be shared or personal. Criminal activity is a social issue that has a significant negative impact on our society. Both local authorities and residents are getting more concerned with the ability to spot crime and pinpoint the most recent crimes in a certain area. On the other hand, when residing in a bustling environment, people are constantly interested in enhancing safety and developing trustworthy connections with neighbors.

One of the biggest problems facing societies worldwide, especially those in urban areas, is the incidence of crime. While social crimes have been the subject of increasing research, social media has only been used in a limited number of studies involving crime and criminal behavior. As a result, the study attempts to propose a prediction model (algorithm) by using the machine-learning technique, which is intended to hold a high capability to forecast crimes by aspects of social media datasets using the Data Mining idea. Social media is the primary source of our data. The primary objective is to locate every hidden data source and forecast outcomes.

SCOPE OF THE PROJECT

Much of the current work is focused in two major directions:

- Predicting surges and hotspots of crime.
- Understanding patterns of criminal behavior could help in solving criminal investigations.

OBJECTIVE OF THE PROJECT

- The project's primary goal is to forecast crime rates and examine those that will actually occur in the future. The authorities can take responsibility and attempt to lower the crime rate based on this information.
- To forecast the graph between the types of Crimes (Independent Variable) and the Year, Multi Linear Regression is employed (Dependent Variable)
- To assist detectives in solving crimes more quickly, the system will examine how to turn criminal information into a regression problem.
- Crime analysis to identify trends in crime based on information already available. Based on the territorial distribution of the available data and crime recognition, the frequency of occurring crimes can be forecast using a variety of multi-linear regression algorithms.

CHAPTER 2
SYSTEM STUDY

2.1 EXISTING SYSTEM:

After finding and understanding various distinct methods used by the police forsurveillance purposes, the importance of each method has been demonstrated. Eachsurveillance method can perform well on its own and produce satisfactory results, although for only one specific characteristic, that is, Sting Ray can help us only when the suspect is using a phone, which should be switched on.

Thus, it is only useful when the information regarding the stakeout location is correct. Based on this information, the ever-evolving technology has yet again produced a smart way to conduct surveillance. The introduction of deep learning, ML,and computer vision techniques has provided us with a new perspective on ways to conduct surveillance.

2.2 PROPOSED SYSTEM:

An approach using SVM filtering methods to detect crime-related posts from the social media data set can be proposed. There are four main phases. In the first phase, social media text posts that relate to the crimes are extracted. In the second phase, the reprocessing techniques are applied to clean the data set.

Then, the TF- IDF values are calculated for each pre-processed post in the third phase. Finally, SVM- Based Filter is applied to remove non-related data. Then a random forest classifier is used for classification to categorize the data. The main steps include formatting, cleaning and sampling. The cleaning process is used for removal or fixing of some missing data . There may be data that is incomplete.

CHAPTER 3
SYSTEM SPECIFICATION

3.1 HARDWARE REQUIREMENT:

The following hardware facilities are utilized for this project work.

- Processor : Intel Core i5 @ 3.1 GHZ
- RAM : 4GB
- Hard disk : 500 GB

3.2 SOFTWARE REQUIREMENT:

- Operating system : Windows 8
- Front End : Python
- Back End : MySQL SERVER
- IDLE : Pycharm

3.3 ABOUT SOFTWARE:

Python is an interpreter, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding; make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages

Syntax

The" hello world" program is written in python

```
print ("Hello, Python!");
```

Python is a very simple language, and has a very straightforward syntax. It encourages programmers to program without boilerplate (prepared) code. The simplest directive in Python is the "print" directive - it simply prints out a line

Data types

Variables can store data of different types, and different types can do different things.Python has the following data type built-in by default, in these categories:

- Text Type : Str
- Numeric Types : int, float, complex
- Sequence Types: list, tuple, range
- Mapping Type : Dict
- Set Types : set, frozenset
- Boolean Type : Bool

- Binary Types : bytes, bytearray, memorview

Function

A function is a block of code which only runs when it is called. You can pass data, known as parameters, into a function. A function can return data as a result. A function is a set of statements that take inputs, do some specific computation and produces output. The idea is to put some commonly or repeatedly done task together and make a function, so that instead of writing the same code again and again for different inputs, we can call the function .Python provides built-in functions like print(), etc. but we can also create your own functions. These functions are called user-defined functions.

Object

Python is an object oriented programming language. Unlike procedure oriented programming, where the main emphasis is on functions, object oriented programming stress on objects. Object is simply a collection of data (variables) and methods (functions) that act on those data. And, class is a blueprint for the object.

Features

- Easy to Learn and Use. Python is easy to learn and use.
- Expressive Language. Python language is more expressive means that it is more understandable and readable.
- Interpreted Language.
- Cross-platform Language.
- Free and Open Source.
- Object-Oriented Language.
- Extensible.
- Large Standard Library.

Advantages

Presence of third-party modules. Extensive support libraries (NumPy for numerical calculations, Pandas for data analytics etc).Open source and community development. Easy to learn. User-friendly data structures. High-level language.

MYSQL

MySQL is a relational database management system. It is open-source and free. Our MySQL tutorial includes all topics of MySQL database such as insert record, update record, delete record, select record, create table, drop table etc.

Features

MySQL is offered under two different editions: the open source MySQL Community Server and the proprietary Enterprise Server. MySQL Enterprise Server is differentiated by a series of proprietary extensions which install asserver plug-in, but otherwise shares the version numbering system and is built from the same code base.

- Major features as available in MySQL

• A broad subset of ANSI SQL 99, as well as extensions

• Cross-platform support

• Stored procedures, using a procedural language that closely adheres

to SQL/PSM[76]

• Triggers

• Cursors

• Updatable views

• Online Data Definition Language (DDL) when using the in no DB

Storage Engine.

• Information schema

Data types

• Integers.

• Booleans.

• Characters.

• Floating-point numbers.

Function

Stored function is a set of SQL statements that perform some operation and return a single value. Just like MySQL in-built function, it can be called from within a MySQL statement. By default, the stored function is associated with the default database.

CHAPTER 4
MODULE

Module Description:

Data Collection:

- Social media posts are collected through the Search API.
- The search of the social media posts must be based on a set of keywords that can be used to classify the crime situations.
- Thus, in the first filter, use the main crime-related keywords according to crime categories.

Data Cleaning:

- Data cleaning is a critically important step in any machine learning project.
- In this module, data cleaning is done to prepare the data for analysis by removing or modifying the data that may be incorrect, incomplete, duplicated or improperly formatted. In tabular data, there are many different statistical analysis and data visualization techniques, so it is used to explore the data in order to identify data cleaning operations.

Data Pre-processing:

- As the next step, it is very important to apply the pre-processing techniques to the extracted data set. Because, there may be typos, unwanted content like URLs, and stop words in the social media post.
- Thus, data which is obtained from social media is highly unstructured and noisy.
- Pre-processing techniques will generate clean tweet data that will be used for the next process.

- First, remove the stop words such as is, then, which, have, etc. The words do not convey any positive or negative meaning. So, we can easily remove the stop word without affecting the meaning of the message.
- Then, remove the URLs, hashtags, symbols, usernames, expressions, quotes, etc. Next, combine words are split by applying tokenizing techniques.
- Finally, apply a stemming algorithm to reduce a word to its word stem that affixes to suffixes and prefixes or the roots of words.

Feature Extraction:

- This is done to reduce the number of attributes in the dataset, hence providing advantages like speeding up the training and accuracy improvements.
- In machine learning, pattern recognition, and image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases, leading to better human interpretations. Feature extraction is related to dimensionality reduction
- When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meter, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector).
- Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Data Preparation:

- After completing the pre-processing, social media posts are transformed into vectors to generate the feature vectors. The vectors are used in the learning phase for machine learning algorithms.
- This research, used the term frequency-inverse document frequency (TF–IDF) values to create the vectors.
- TF–IDF value reflects the importance of a term in a document to the collection of documents.

Second Stage Filtering:

- Despite a large amount of data collected via keyword filtering, few of them are concerning crimes. Most information is considered as noise.
- According to the above examples, the searched keyword is ‘shoot'. The expected gun shooting data relates to crime scenes. However, the above post is related to a film shoot and photo shoot. So, it can consider it as a noisy post.
- To eliminate the noise and increase the accuracy, implemented a machine learning-based filter at this stage. Used Support Vector Machine (SVM) as the machine learning algorithm.
- The SVM is a state-of-the-art classification method that uses a learning algorithm based on structural risk minimization.
- The classifier can be used in many disciplines because of its high accuracy, ability to deal with high dimensions, and flexibility in modeling diverse sources of data.

Support Vector Machine:

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems.

In the SVM algorithm, plot each data item as a point in n-dimensional space (where n is the number of features you have) with the value of each feature being the

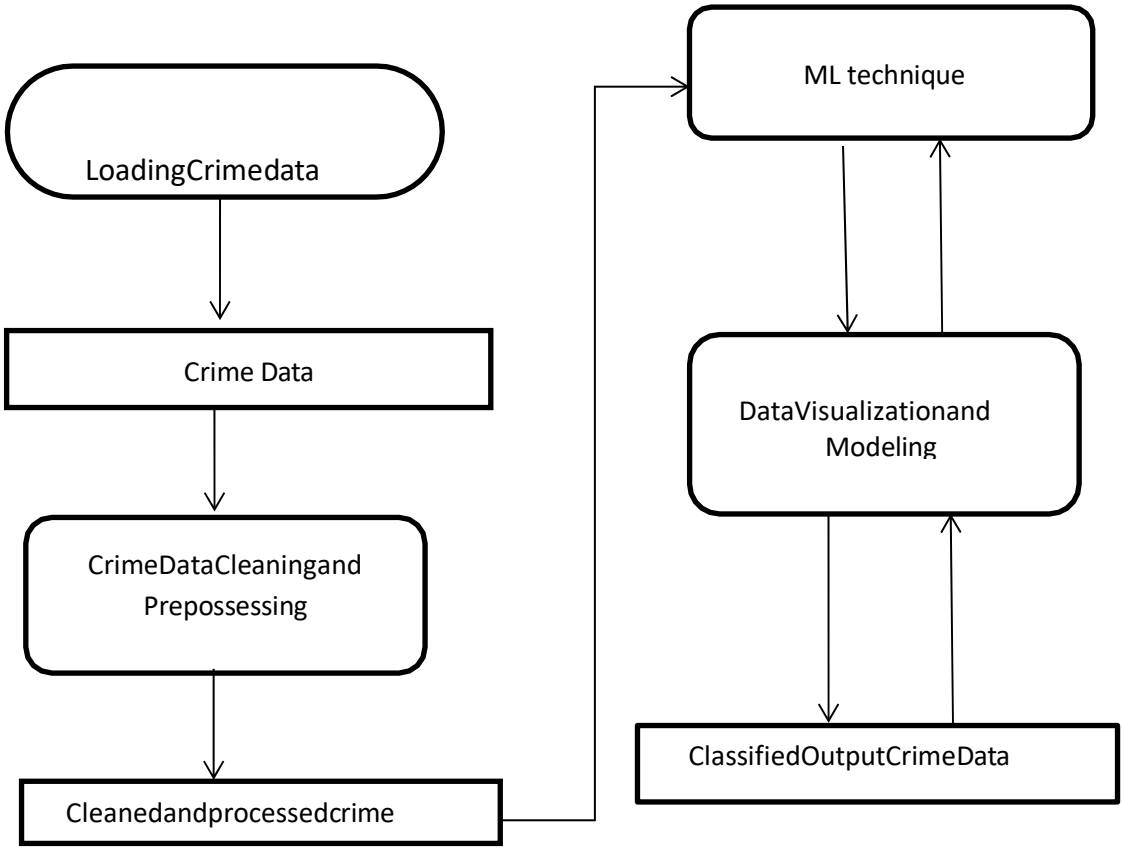
value of a particular coordinate. Then, perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontierwhich best segregates the two classes (hyperplane/ line). So looking at support vector machines and a few examples of its working are here. Above, got accustomed to the process of segregating the two classes with a hyper-plane.

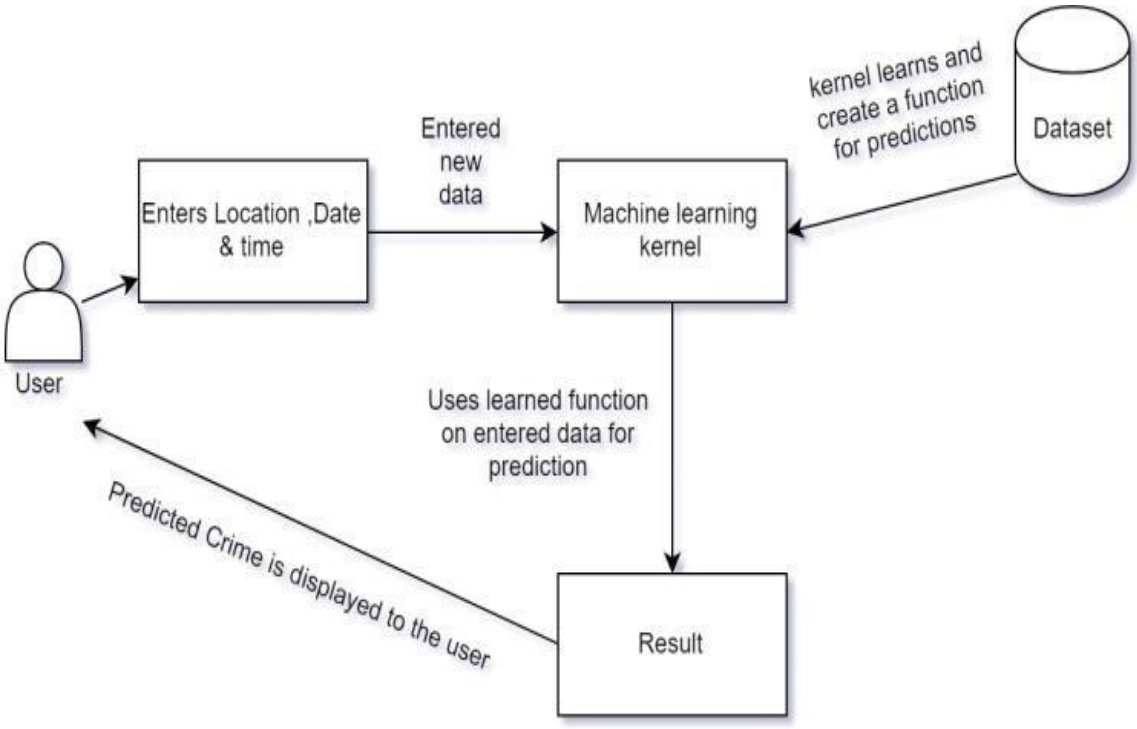
CHAPTER 5

SYSTEM DESIGN

5.1 DATAFLOW DIAGRAM:



5.2 ARCHITECTURAL DIAGRAM:



CHAPTER 6

TESTING

6.1 TESTING DESCRIPTION

Testing is a series of different tests that whose primary purpose is to fully exercise the computer based system. Although each test has a different purpose, all work should verify that all system element have been properly integrated and performed allocated function. Testing is the process of checking whether the developed system works according to the actual requirement and objectives of the system. The philosophy behind testing is to find the errors. A good test is one that has a high probability of finding an undiscovered error. A successful test is one that uncovers the undiscovered error. Test cases are devised with this purpose in mind. A test case is a set of data that the system will process as an input.

Unit Testing

The first test in the development process is the unit test. The source code is normally divided into modules, which in turn are divided into smaller units called units. These units have specific behavior. The test done on these units of code is called unit test. Unit test depends upon the language on which the project is developed. Unit tests ensure that each unique path of the project performs accurately to the documented specifications and contains clearly defined inputs and expected results. Functional and reliability testing in an Engineering environment. Producing tests for the behavior of components (nodes and vertices) of a product to ensure their correct behavior prior to system integration.

Integration Testing

Testing in which modules are combined and tested as a group. Modules are typically code modules, individual applications, source and destination applications on a network, etc. Integration Testing follows unit testing and precedes system testing. Testing after the product is code complete. Betas are often widely distributed or even distributed to the public at large in hopes that they will buy the final product when it is

release.

.

System Testing:

System testing tests the integration of each module in the system. It also tests to find discrepancies between the system and it is original objective, current specification and system documentation. The primary concern is the compatibility of individual modules. Entire system is working properly or not will be tested here, and specified path ODBC connection will correct or not, and giving output or not are tested here these verifications and validations are done by giving input values to the system and by comparing with expected output. Top-down testing implementing here.

Acceptance Testing:

This testing is done to verify the readiness of the system for the implementation. Acceptance testing begins when the system is complete. Its purpose is to provide the end user with the confidence that the system is ready for use. It involves planning and execution of functional tests, performance tests and stress tests in order to demonstrate that the implemented system satisfies its requirements.
Tools to special importance during acceptance testing include:

Test coverage Analyzer – records the control paths followed for each test case.

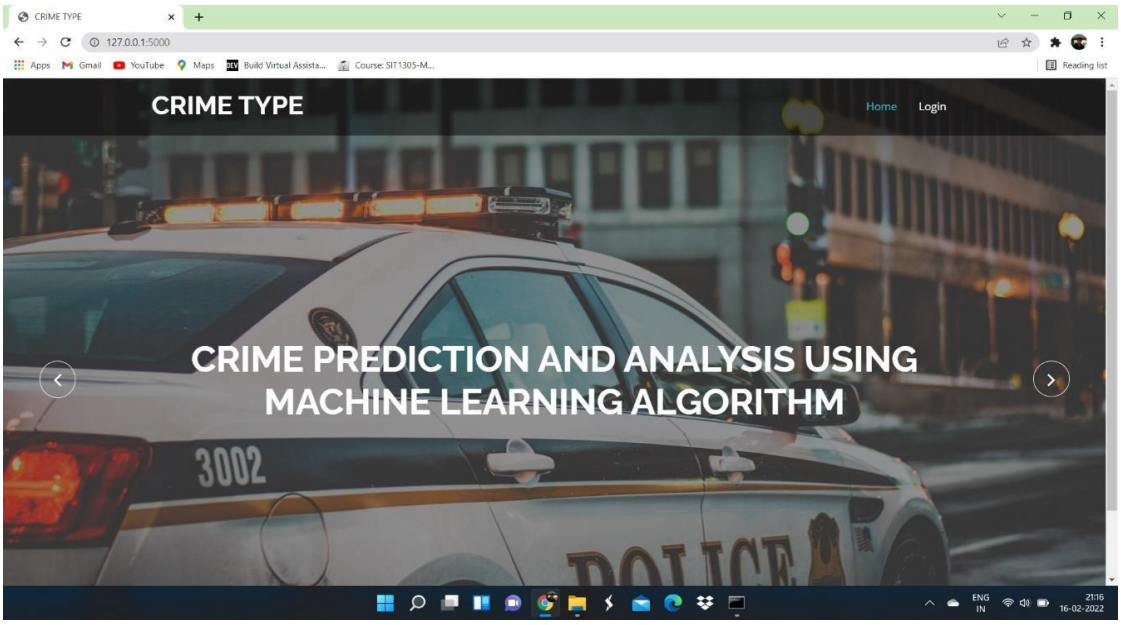
Timing Analyzer – also called a profiler, reports the time spent in various regions of the code are areas to concentrate on to improve system performance.

Coding standards – static analyzers and standard checkers are used to inspect code for deviations from standards and guidelines.

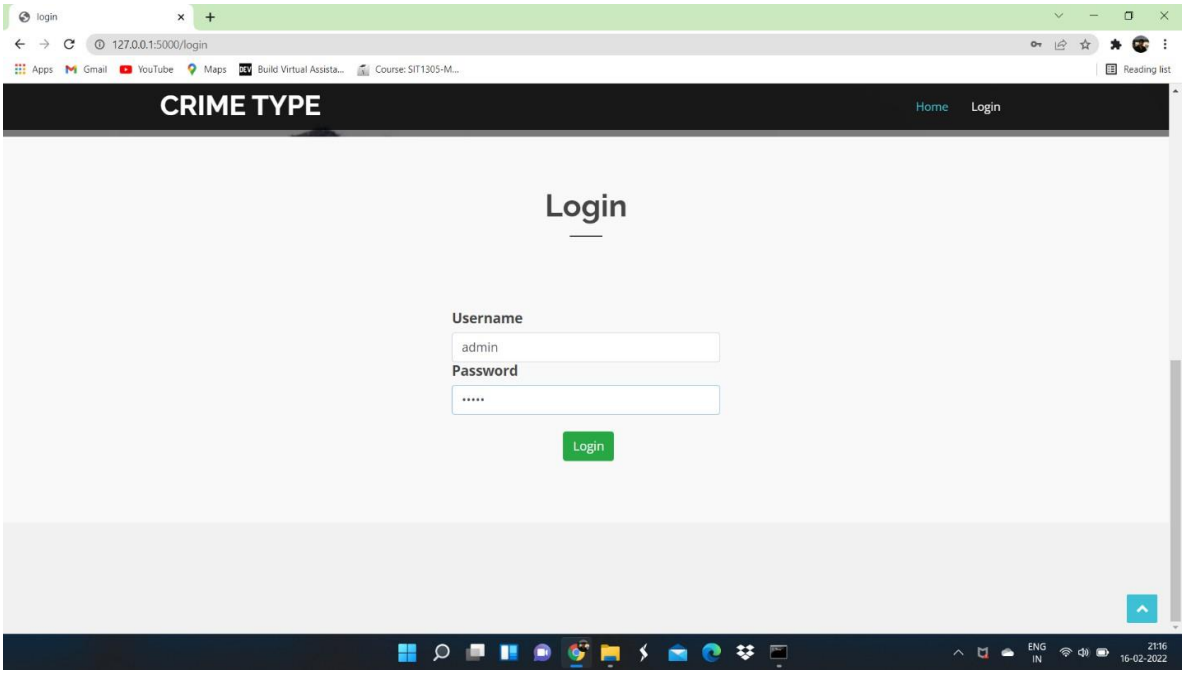
CHAPTER 7

SYSTEM IMPLEMENTATION

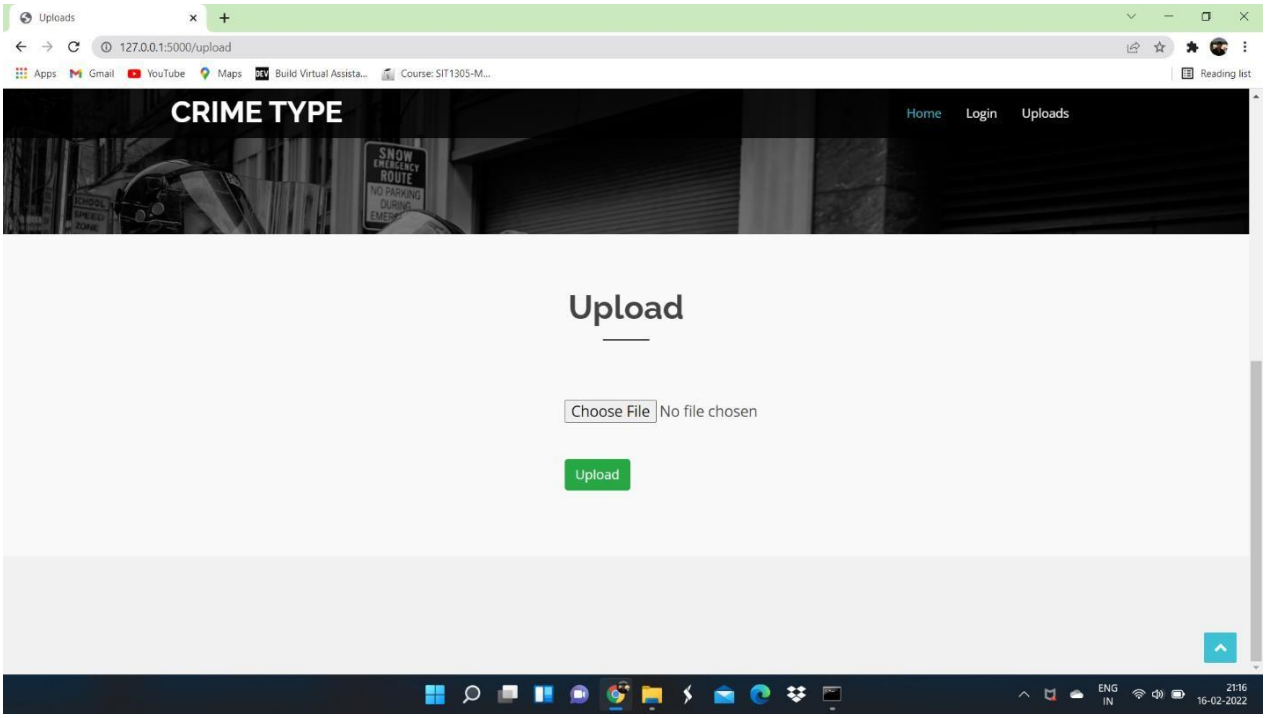
7.1 SCREEN LAYOUT DESIGN:



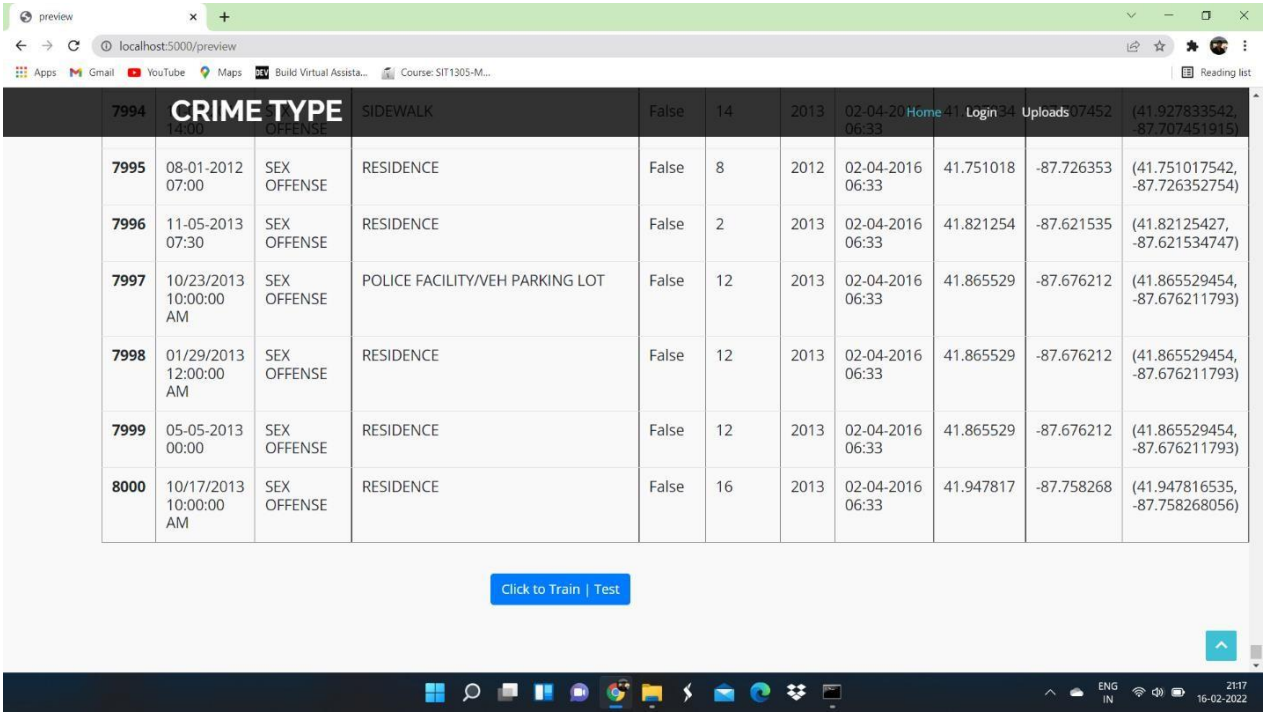
Shows the homepage ofour application of crime analysis and prediction.



Shows the login page for predicting the crime.

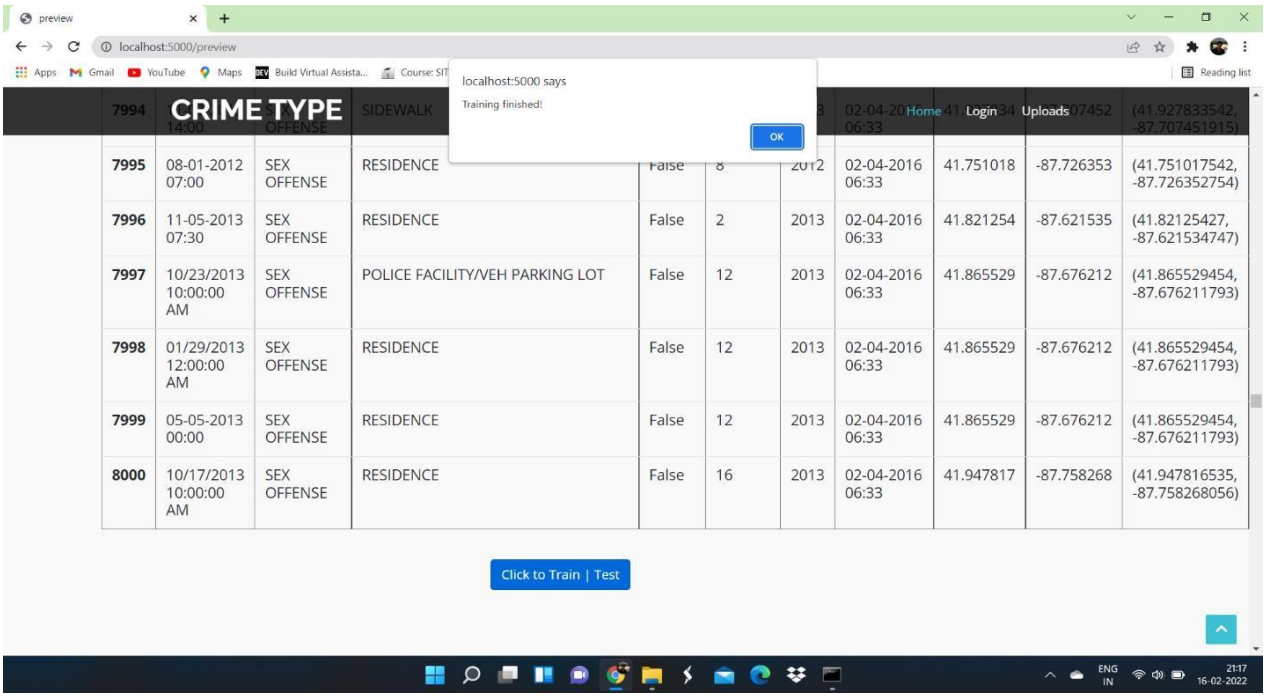


Shows the page to upload the dataset of crime prediction



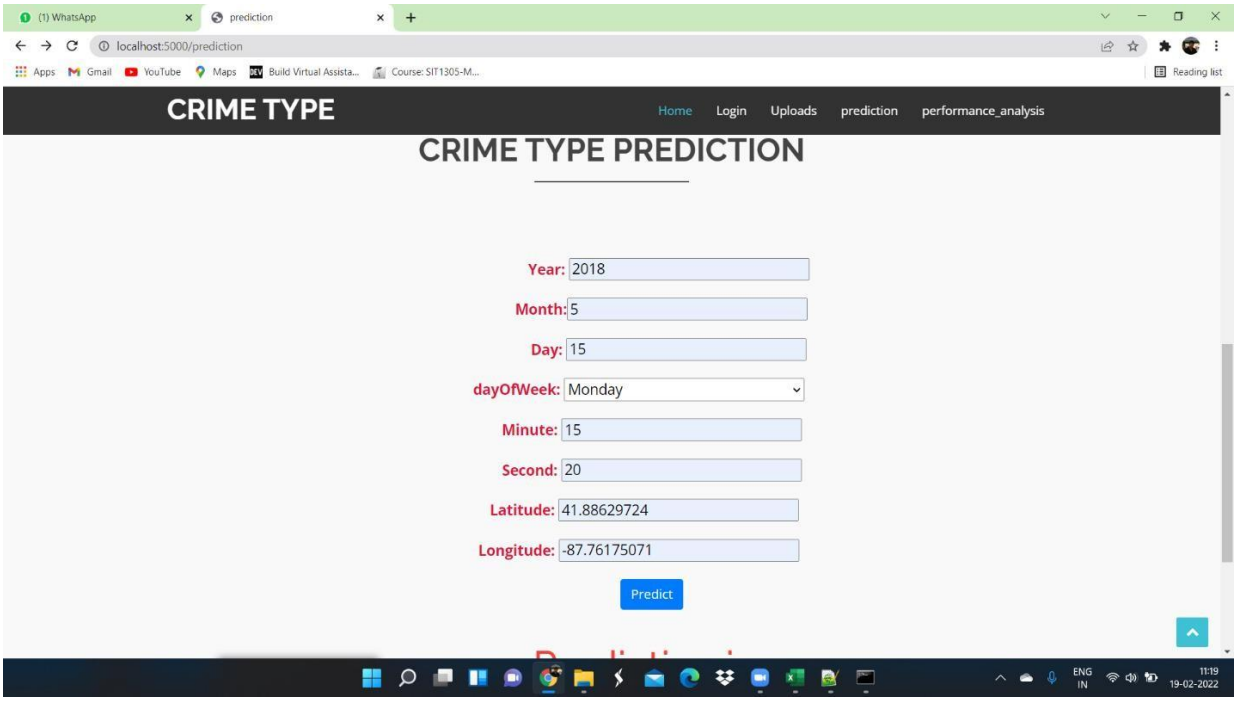
7994	14:00	OFFENSE	SIDEWALK	False	14	2013	02-04-2016 06:33	41.[illegible]34	[illegible]07452	(41.927833542, -87.707451915)
7995	08-01-2012 07:00	SEX OFFENSE	RESIDENCE	False	8	2012	02-04-2016 06:33	41.751018	-87.726353	(41.751017542, -87.726352754)
7996	11-05-2013 07:30	SEX OFFENSE	RESIDENCE	False	2	2013	02-04-2016 06:33	41.821254	-87.621535	(41.82125427, -87.621534747)
7997	10/23/2013 10:00:00 AM	SEX OFFENSE	POLICE FACILITY/VEH PARKING LOT	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
7998	01/29/2013 12:00:00 AM	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
7999	05-05-2013 00:00	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
8000	10/17/2013 10:00:00 AM	SEX OFFENSE	RESIDENCE	False	16	2013	02-04-2016 06:33	41.947817	-87.758268	(41.947816535, -87.758268056)

shows the dataset to train and test using machine learning algorithms.

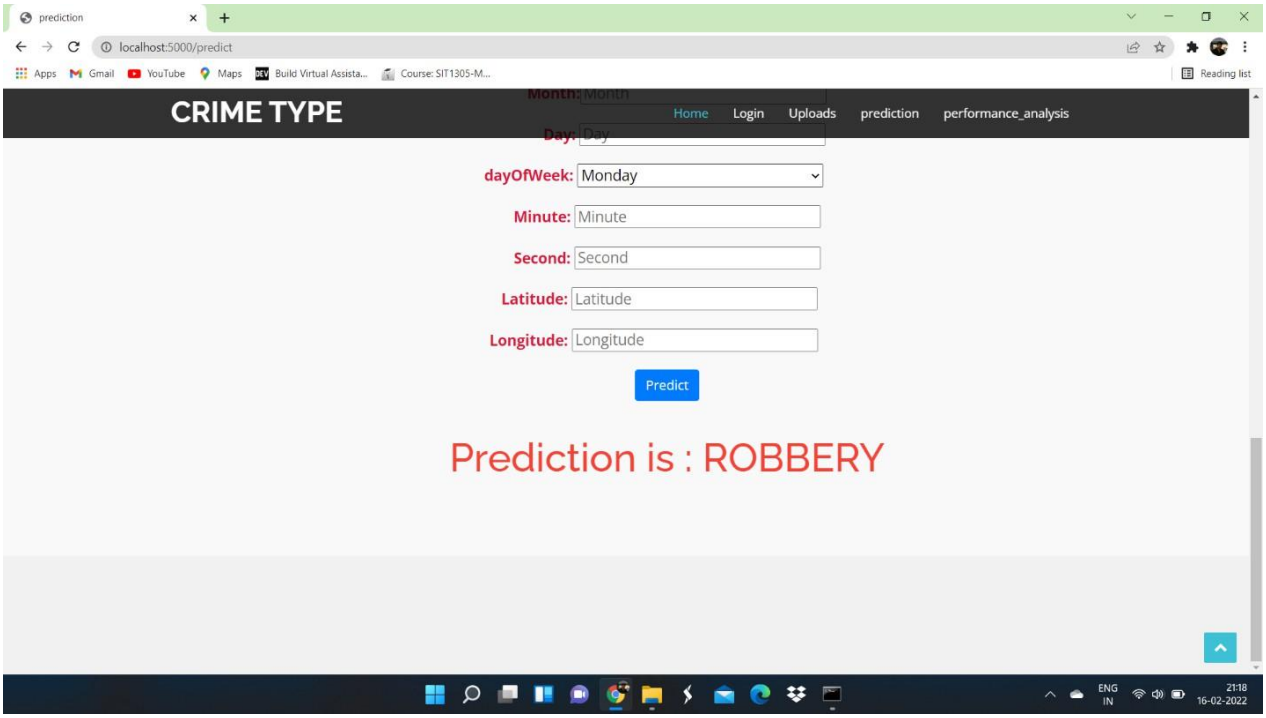


7994	14:00	OFFENSE	SIDEWALK				02-04-20 06:33	Home	Login	Uploads
7995	08-01-2012 07:00	SEX OFFENSE	RESIDENCE	False	8	2012	02-04-2016 06:33	41.751018	-87.726353	(41.751017542, -87.726352754)
7996	11-05-2013 07:30	SEX OFFENSE	RESIDENCE	False	2	2013	02-04-2016 06:33	41.821254	-87.621535	(41.82125427, -87.621534747)
7997	10/23/2013 10:00:00 AM	SEX OFFENSE	POLICE FACILITY/VEH PARKING LOT	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
7998	01/29/2013 12:00:00 AM	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
7999	05-05-2013 00:00	SEX OFFENSE	RESIDENCE	False	12	2013	02-04-2016 06:33	41.865529	-87.676212	(41.865529454, -87.676211793)
8000	10/17/2013 10:00:00 AM	SEX OFFENSE	RESIDENCE	False	16	2013	02-04-2016 06:33	41.947817	-87.758268	(41.947816535, -87.758268056)

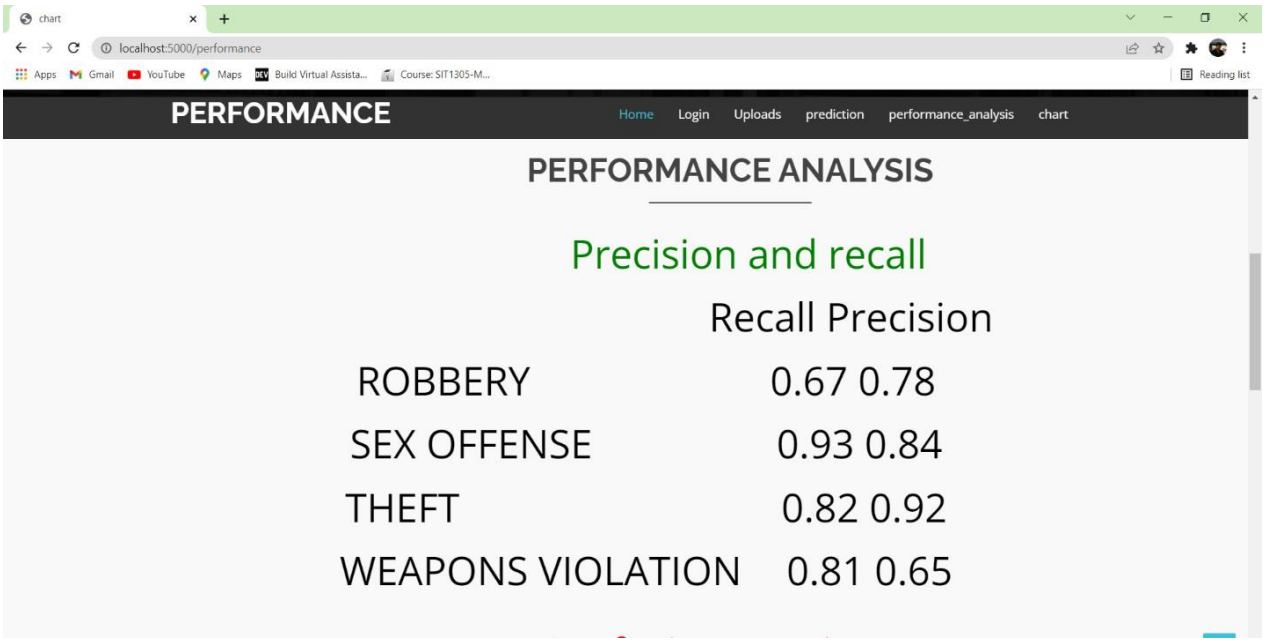
Shows the message to indicate the successful training using machine learning algorithms.



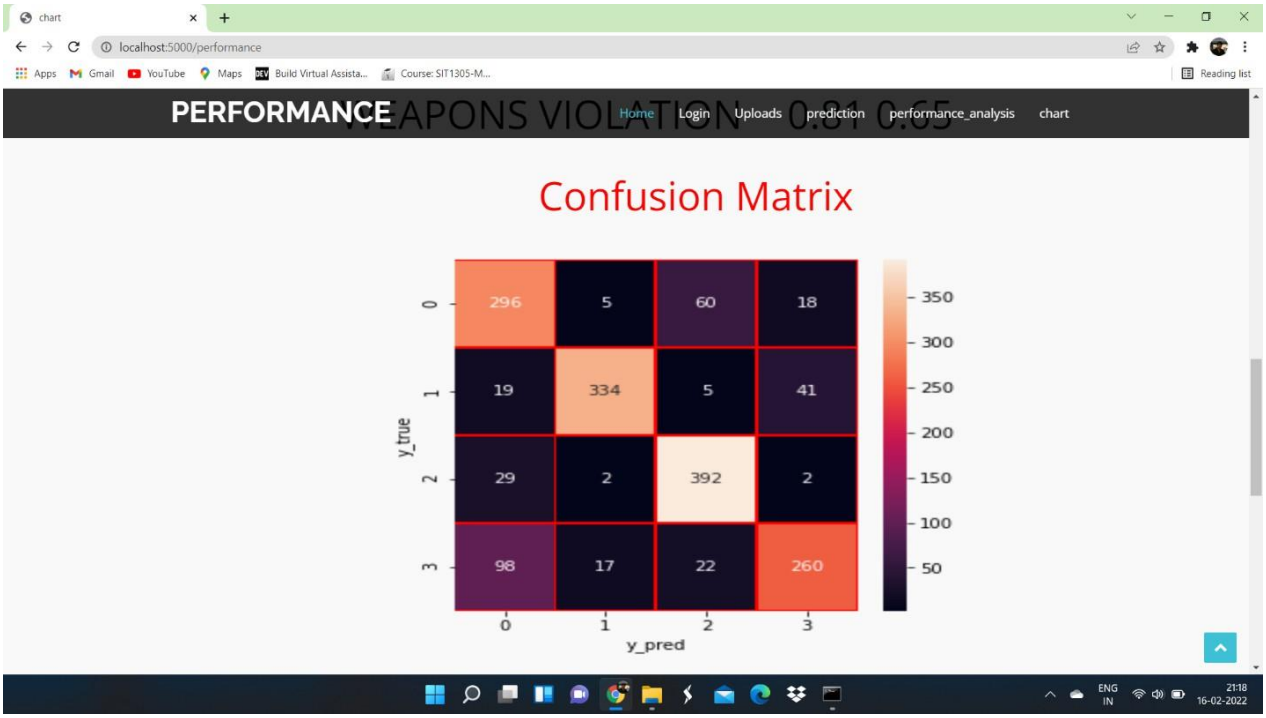
Shows the web page of filling the details about crime location, date and time to predict the crime.



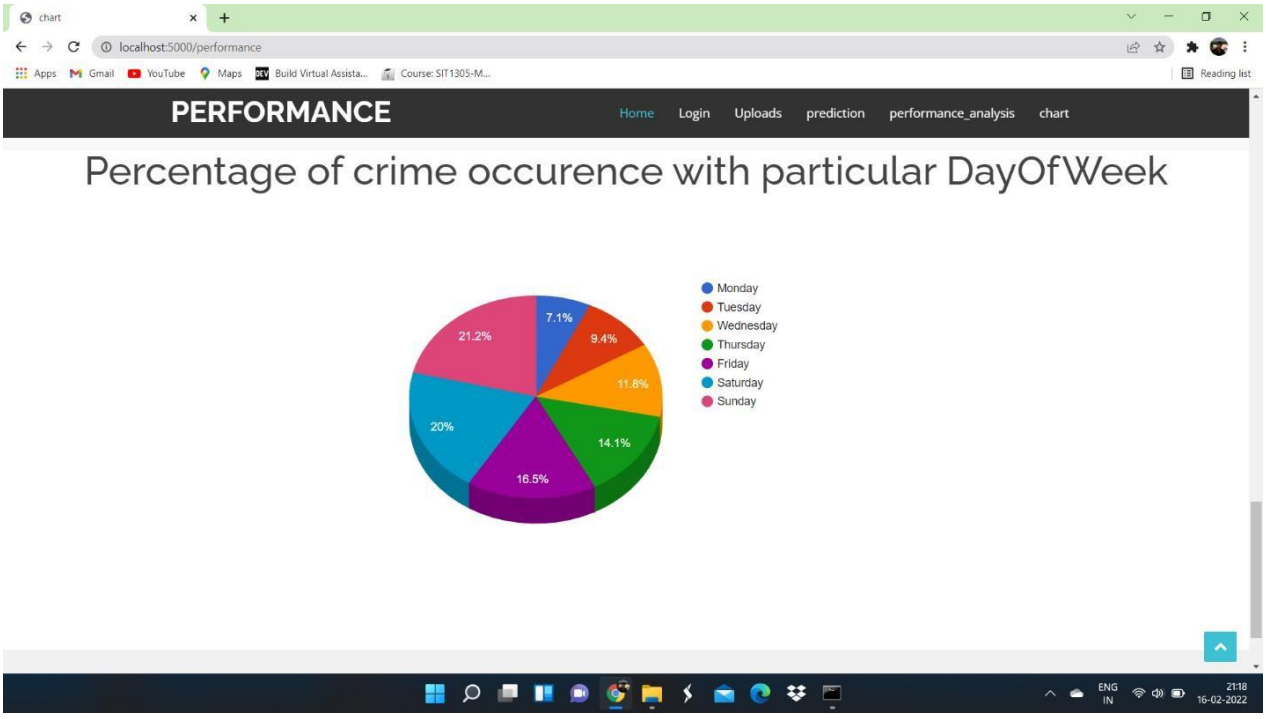
Shows the prediction of type of crime.



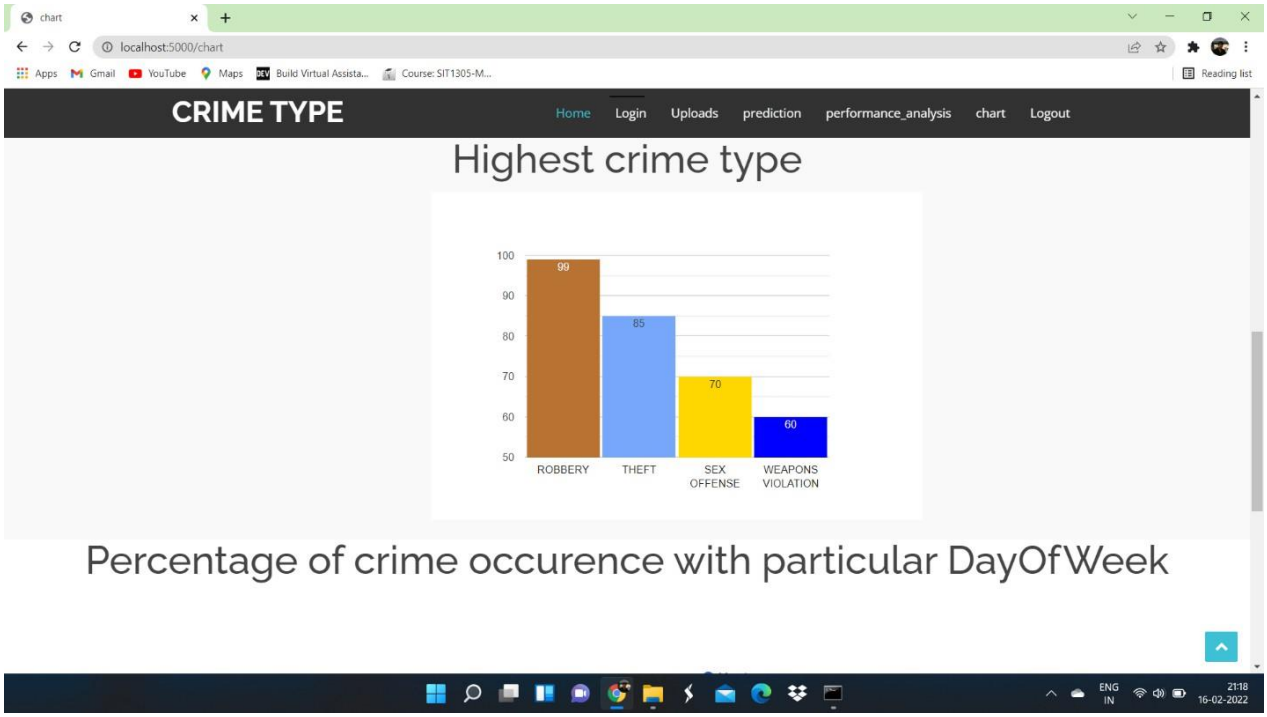
Shows the performance analysis for the type of crime



Shows the confusion matrix for the analysis of crime.



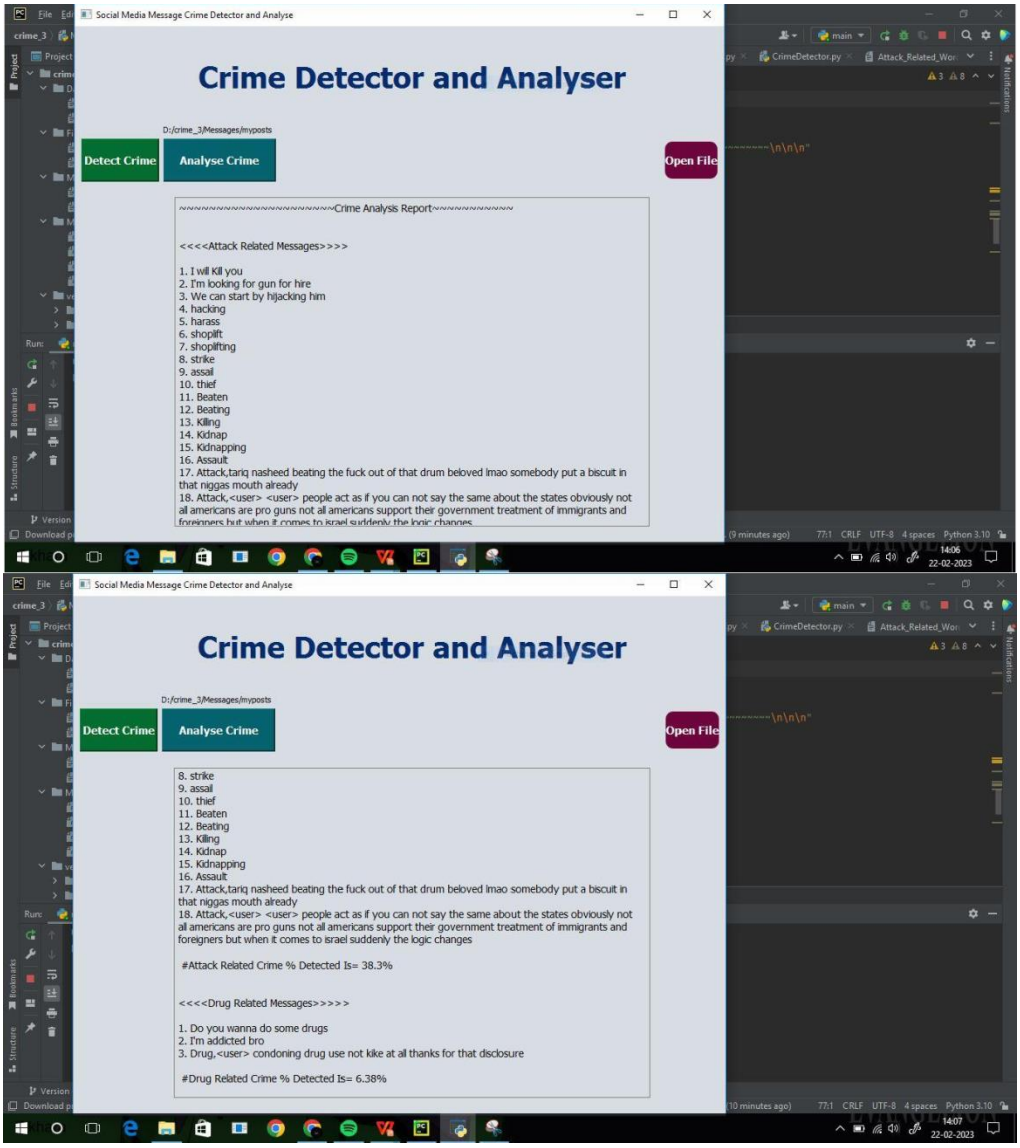
Shows the performance of crime occurrences with particular day of week.



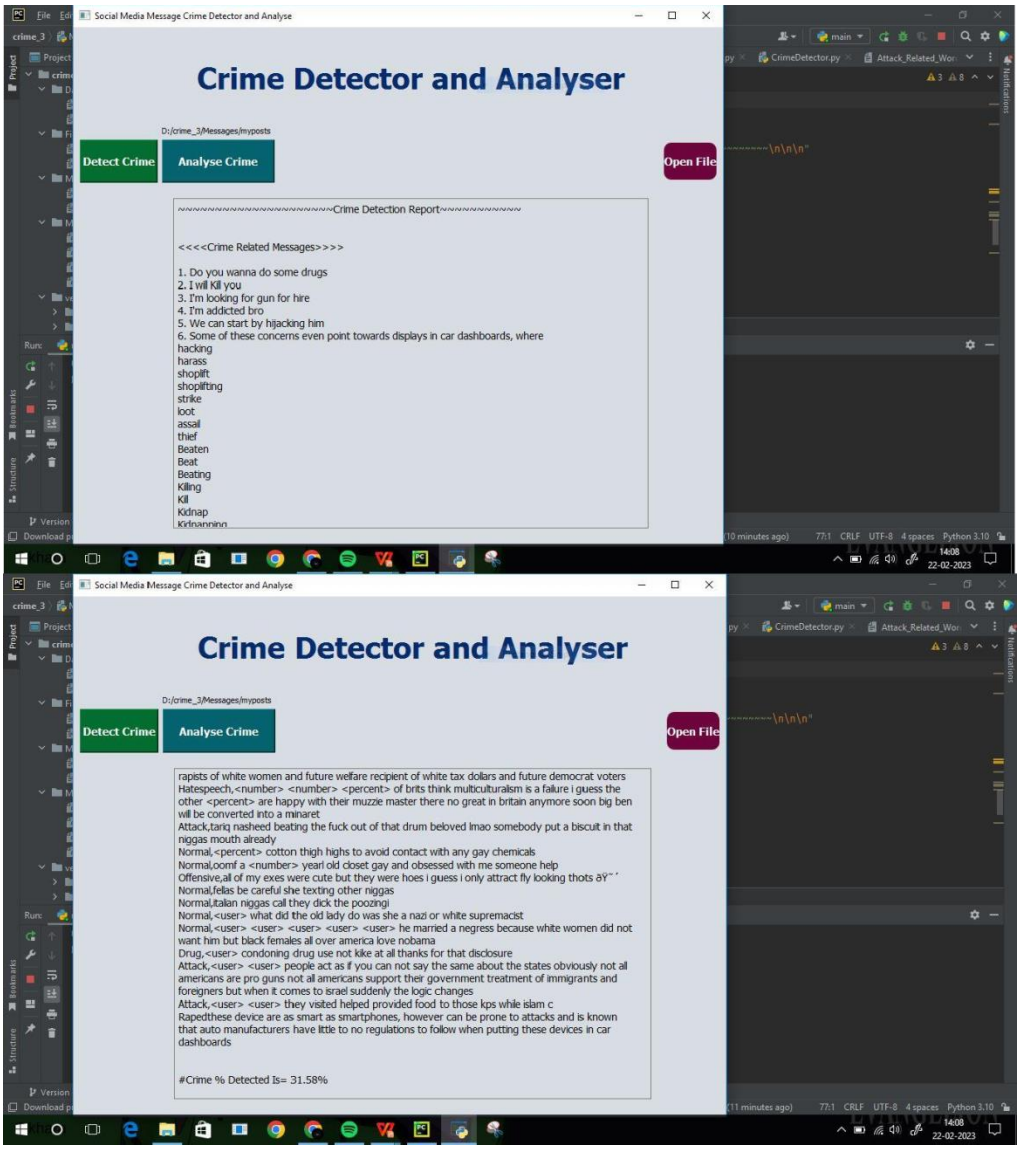
Shows the graph for percentage of crime occurrences with particular day of week.

CHAPTER 8

REPORTS



A method that uses SVM filtering techniques can be used to find postings from the social media data collections that are related to crimes. By testing and training the data we will analyze the type of the crime percentage.



By observing crime related messages through social media using SVM filter and random forest we conclude that type of the crime and percentage of the crime detected.

CHAPTER 9
SAMPLE SOURCE CODE

```
import numpy as np
importpandasaspd
fromflaskimportFlask,request,jsonify,render_template,redirect,flash,send_file from
sklearn.preprocessing import MinMaxScaler
fromsklearn.naive_bayesimportGaussianNB import
pickle
app=Flask(_name_)#InitializetheflaskApp

model=pickle.load(open('model.pkl','rb')) @app.route('/')

@app.route('/first') def
first():
        returnrender_template('first.html')



@app.route('/login') def
login():
        returnrender_template('login.html')
@app.route('/upload')
defupload():
    return render_template('upload.html')
@app.route('/preview',methods=["POST"]) def
preview():
```

```
    ifrequest.method =='POST':
        dataset= request.files['datasetfile']
        df=pd.read_csv(dataset,encoding='unicode_escape')
        df.set_index('Id', inplace=True)
        returnrender_template("preview.html",df_view=df)


#@app.route('/home')
#def home():
 #    return render_template('home.html')

@app.route('/prediction',methods=['GET','POST']) def
prediction():
    return render_template('prediction.html')


#@app.route('/upload')
#def upload_file():
#    return render_template('BatchPredict.html')



@app.route('/predict',methods=['POST'])
def predict():
            int_feature=[xforxinrequest.form.values()]
            print(int_feature)
```

```
  ifrequest.method =='POST':
      dataset= request.files['datasetfile']
      df=pd.read_csv(dataset,encoding='unicode_escape') df.set_index('Id',
      inplace=True)
      returnrender_template("preview.html",df_view=df)
@app.route('/login') def
login():
    returnrender_template('login.html') @app.route('/upload')
defupload():
   return render_template('upload.html')
@app.route('/preview',methods=["POST"]) def
preview():
   ifrequest.method=='POST':
       dataset=request.files['datasetfile']
       df=pd.read_csv(dataset,encoding='unicode_escape') df.set_index('Id',
       inplace=True)
       returnrender_template("preview.html",df_view=df)


#@app.route('/home') #def
home():
 #returnrender_template('home.html')

@app.route('/prediction',methods=['GET','POST']) def
prediction():
```

```
    returnrender_template('prediction.html')


#@app.route('/upload') #def
upload_file():
#return render_template('BatchPredict.html')



@app.route('/predict',methods=['POST']) def
predict():
    int_feature=[xforxinrequest.form.values()] print(int_feature)
    int_feature = [float(i) for i in int_feature] final_features
    = [np.array(int_feature)]
    prediction=model.predict(final_features)

    output=format(prediction[0])
    print(output)
    returnrender_template('prediction.html',prediction_text=output) @app.route('/chart')
defchart():
    returnrender_template('chart.html')
@app.route('/performance')
defperformance():
    returnrender_template('performance.html')
```

```
if_name_=="_main_":
    app.run(debug=True)
```

Model.ipynb:

```
importpandasas pd
importmatplotlib.pyplotasplt import
seaborn as sns
df=pd.read_csv('new.csv') df
df = df.dropna()
df.info()
df['Hour'].unique()
df['dayOfWeek'].unique()
plt.rcParams["figure.figsize"]=(150,100) x = df
[['Year','Month','Day','dayOfWeek','Minute','Second','Latitude','Longitude']]X
y=df['Primary Type']
fromsklearn.model_selectionimporttrain_test_split #pip install
imblearn
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=0.2, random_state = 42)
fromsklearn.ensembleimportRandomForestClassifier from
sklearn import metrics
rf=RandomForestClassifier(n_estimators=100,random_state=60,max_depth= 50)
rf=rf.fit(X_train,y_train)
```

```
rf.score(X_train,y_train)
fromsklearn.metricsimportaccuracy_score y_pred =
rf.predict(X_test ) accuracy_score(y_pred,y_test)
import sklearn.metrics
print(sklearn.metrics.classification_report(y_test,y_pred)) y_pred =
rf.predict(X_test )
y_true=y_test


fromsklearn.metricsimportconfusion_matrix
cm=confusion_matrix(y_true,y_pred)
cm
importseabornassns
importmatplotlib.pyplotasplt

f, ax=plt.subplots(figsize=(5,5))
sns.heatmap(cm,annot=True,linewidths=0.5,linecolor="red",fmt=".0f",ax=ax) plt.xlabel("y_pred")
plt.ylabel("y_true")
plt.show()

importpickle
pickle.dump(rf, open('model.pkl', 'wb'))
model=pickle.load(open('model.pkl','rb')) Style.css
```

```
@charset"UTF-8";
/*----------------------------------------
# General
----------------------------------------*/
html, body {
  height:100%;
}

.floatleft {
  float:left;
}

.floatright {
  float:right;
}

.alignleft {
  float:left;
  margin-right: 15px;
  margin-bottom:15px;
}

.alignright {
  float:right;
  margin-left: 15px;
  margin-bottom:15px;
```

```
}

.aligncenter {
  display:block;
  margin:0auto15px;
}

a:focus{
  outline:0pxsolid;
}

img{
  max-width:100%;
  height: auto;
}

.fix{
  overflow:hidden;
}

p{
  margin:00 15px;
  color:#444;
}

h1,h2,h3,h4,h5,h6{
```

```
  font-family:'Raleway',sans-serif; margin:
  0 0 15px;
  color:#444;
  font-weight:500;
}

h1{
  font-size: 48px;
  line-height:50px;
}

h2{
  font-size: 38px;
  line-height:40px;
}

h3{
  font-size: 30px;
  line-height:32px;
}

h4{
  font-size: 24px;
  line-height:26px;
}
```

```
h5{
  font-size: 20px;
  line-height:22px;
}

h6{
  font-size: 16px;
  line-height:20px;
}

a{
  transition:all0.3sease0s; text-
  decoration: none;
}

a:hover{
  color:#3EC1D5;
  text-decoration:none;
}

a:active,a:hover{
  outline: 0 none;
}

body{
  background:#fffnonerepeatscroll00;
```

```
  color:#444;
  font-family:'OpenSans',sans-
  serif; font-size: 18px;
  text-align:
  left;
  overflow-
  x:hidden;
  line-height:
  22px;
}

/*Backtotopbutton*/
.back-to-top
  {
  position:fix
  ed; display:
  none;
  background:#3
  EC1D5; color:
  #fff;
  padding:6px12px9px1
  2px; font-size: 16px;
  border-
  radius:2px;
  right: 15px;
  bottom:
  15px;
  transition:background0.5s;
}

@media(max-width:768px){
```

```
    .back-to-top
     {
     bottom:15
     px;


#@app.route('/home')
#def home():
 #    return render_template('home.html')

@app.route('/prediction',methods=['GET','POST']) def
prediction():
    return render_template('prediction.html')


#@app.route('/upload')
#def upload_file():
#    return render_template('BatchPredict.html')



@app.route('/predict',methods=['POST']) def
predict():
            int_feature=[xforxinrequest.form.values()]
            print(int_feature)

            int_feature = [float(i) for i in int_feature]
            final_features = [np.array(int_feature)]
```

```
        prediction=model.predict(final_features)


        output=format(prediction[0])
        print(output)
        returnrender_template('prediction.html',prediction_text=output) @app.route('/chart')
defchart():
        returnrender_template('chart.html')
@app.route('/performance')
defperformance():
        returnrender_template('performance.html')


if _name_ == "_main_": app.run(debug=True)
```

CHAPTER 10
CONCLUSION

The project improves the accuracy of the detection of crime related posts from social media text messages. Here, First we applied a keyword based filter, and then to remove the noise, we applied the SVM based filter and random forest classification. According to the existing works, SVM has the best accuracy among the classifiers. So, The SVM is used for the research.

In conclusion, utilizing multi-model data and employing various machine learning algorithms can enhance the accuracy and effectiveness of crime prediction models. However, continuous refinement and evaluation of the models are necessary to ensure their reliability and validity in real-world applications.

CHAPTER 11
FUTURE ENHANCEMENT

Future enhancements of this research work on training bots to predict the crime prone areas by using machine learning techniques. method for enhancing the detection of posts about crimes in social media text messages. Here, a keywordbased filter is initially used, and then an SVM-based filter and a random forest classification are utilized to get rid of the noise. SVM has the highest accuracy of all the classifiers, according to the currently available research. Therefore, was employed in the studies.

CHAPTER 12
REFERENCES

12.1 BIBLIOGRAPHY:

[1] Ginger Saltos and Ella Haig, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology& Decision Making, 2017.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conference on networks& soft computing (IEEE) 2014.

[3] Khushabu A. Bokde, Tisksha P. Kakade, Dnyaneshwari S. Tumasare, Chetan G. Wadhai B.E Student, Crime Detection Techniques Using DataMining and KMeans, International Journal of Engineering Research & technology (IJERT) ,2018.

[4] H. Benjamin Fredrick David and A. Suruliandi, Survey on crime analysis and prediction using data mining techniques, ICTACT Journal on Soft computing, 2017.

[5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization and prediction Using Data Mining, Indian Journal of Computer Science and Engineering (IJCSE), 2015.

[6] RajKumar. S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends in engineering &research, 2019.

[7] Chhaya Chauhan, Smriti Sehgal, Crime analysis using data mining techniques and algorithms, International Conference on Computing, Communication and Automation, 2017.

[8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

[9] Dr .M.Sreedevi, A.Harha Vardhan Reddy, ch.Venkata Sai Krishna Reddy,Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology, 2018.

[10] K.S.N .Murthy, A.V.S.Pavan kumar, Gangu Dharmaraju, internationaljournal of engineering, Science and mathematics, 2017.

[11] Deepiika k.K, Smitha Vinod, Crime analysis in india using data mining techniques, International journal of Enginnering and technology, 2018. 40

[12] Hitesh Kumar Reddy ToppyiReddy, Bhavana Saini, Ginika mahajan, Crime Prediction &Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS 2018).

[13] Johnson, S.D., Bowers, K.J., Birks, D.J. et al. Predictive Mapping of Crime by Proactive Police Deployment. J Quant Criminol 24, 237–256 (2008). DOI: 10.1007/s10940-008-9051-2

[14] Mohler, G., Short, M., Brantingham, P. et al. Randomized Controlled Field Trials of Predictive Policing. J Exp Criminol 9, 247–274 (2013). DOI: 10.1007/s11292-013- 9184-4

[15] Ratcliffe, J.H., Taniguchi, T., Groff, E.R. The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots. Criminol Public Policy 16, 589–618 (2017). DOI: 10.1111/1745- 9133.12289

[16] Gerber, M.S., Johnson, B.D., Buerger, M.E. Predicting Gun Violence in Chicago. In: Gerber M.S., Johnson B.D. (eds) Policing and Crime Control. Springer, Cham (2016). DOI: 10.1007/978-3-319-28676-4_4.