**TABLE OF CONTENTS**

**TITLE PAGE I**

**CERTIFICATE II**

**DECLARATION III**

**ACKNOWLEDGEMENT IV**

**TABLE OF CONTENTS V**

**LIST OF FIGURES VI**

**LIST OF TABLES VII**

**LIST OF ABBREVIATIONS VIII**

**CHAPTER 1: INTRODUCTION ……………………………………………………………... 1**

1.1 Project Objectives ……………………………………………………...………... 1

1.2 Methodology ……………………………………………………………...….… 4

**CHAPTER 2: REVIEWS OF LITERATURE ………………………………………….....…. 6**

2.1 Literature Review ……………………………………………………………...… 6

**CHAPTER 3: SYSTEM DEVELOPMENT ………………………………………………….16**

3.1 Proposed Model ………………………………………………………...……….16

3.2 Design ……………………………………………………………………...…... 16

3.2.1 Predictive Analysis ……………………………………………………...…..17

3.2.2 Algorithms Used for Development ………………………………………….17

3.3.3 Testing Used for development ………………………………………….. .....18

3.3.4 Source Code …………………………………………………………………29

**CHAPTER 4: TEST PLAN AND PERFORMANCE ……………………………………...49**

4.1 Software Requirements and Tools Used …………………………………........49

4.2 Hardware Requirements ……………………………………………………..…..49

4.3 Limitations ………………………………………………………………………49

**CHAPTER 5: CONCLUSION ………………………………………………………..............50**

5.1 Application …………………………………………………………………....…52

5.2 Future Scope ………………………………………………………………….…52

**REFERENCES IX**

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Figure Name** | **Page No.** |
| 1. | Predictive Policing Approach | 02 |
| **2.** | Circle of Prediction and Prevention | 04 |
| 3. | Flow of Procedural Analysis | 05 |
| 4. | Flowchart | 06 |
| 5. | Predictive Analysis Benefits | 09 |
| 6. | Relation between claimed and proven benefits of predictive policing | 13 |
| **7.** | Identification of hotspots and profiling of offenders | 18 |
| 8. | ER Diagram | 48 |
| 9. | Predictive Policing Crime rate | 49 |
| 10. | Predictive Policing in the Context | 49 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Table Name** | **Page No.** |
| 1. | Examples of variables used in predictive policing | 08 |
| 2. | Methodology-Data Collection Processes | 18 |

**CHAPTER 1: INTRODUCTION**

In the past years, an increasing number of police forces around the world have adopted software that uses statistical data to guide their decision-making: predictive policing. This approach means that police departments analyze statistical historic data to predict in what geographic areas there is an increased chance of criminal activity. This type of information can be used by law enforcers to efficiently deploy their resources to prevent criminal behavior. Predictive policing does not replace conventional policing methods (e.g. problem-oriented policing, intelligence-led policing or hotspot policing) but enhances these traditional practices by applying advanced statistical models and algorithms.

The use of statistical models can be of immense value for reducing crime and ensuring the safety in cities. Indeed, some cases in the United States indicate that when predictive policing software is used, the crime rate decreases. For instance, with the use of historic data, Richmond’s police department tried to forecast where gun firing would occur on New Year’s Eve, in 2003, and adapted their surveillance routes to these predictions. It was deemed a success: the random gunfire decreased on this night with 47%, 246% more weapons were seized, while the police force became more efficient as $15.000 was saved.

There are, however, also indications that predictive policing may have important drawbacks. When predictive models are administered, crime-forecasting is not dependent on theory anymore, but takes the large amount of available data as a starting point. These models might result in possibly skewed depictions of society and criminal behavior as they tend to remove context. The risk here is that predictive policing could result in less effective and maybe even discriminatory police interventions.

In view of this debate about the benefits and risks of predictive policing, there is a need for a state-of-the-art overview of existing literature on the benefits and drawbacks of predictive policing. By conducting a PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis) study, we offer a systematic overview of the literature and illuminate how predictive policing is conceptualized, and to what extent the claimed benefits and drawbacks are empirically supported. Full information about this literature review can be obtained from the authors: we present the key findings of this extensive review in this article. In short, this paper gives police practitioners an overview of the claimed benefits and drawbacks of predictive policing and highlights that they need to realize that, for the moment, this innovative method lacks a clear evidence basis.

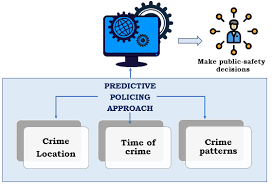
****

Fig. Predictive Policing Approach

**1.1 Project Objectives**

Predictive methods allow police to work more proactively with limited resources. The objective of these methods is to develop effective strategies that will prevent crime or make investigation efforts more effective. However, it must be understood at all levels that applying predictive policing methods is not equivalent to finding a crystal ball. For a policing strategy to be considered effective, it must produce tangible results. The objective of this research was to develop a reference guide for departments interested in predictive policing, providing assessments of both the most promising technical tools for making predictions and the most promising tactical approaches for acting on them. In many cases, we were able to illustrate how predictive technologies are being used to support police operations through a set of examples and case studies. Although some of the methods are promising and describe the current state of field, they are still more academic than practical.

In our assessment of predictive policing, we found that predictive methods can be divided into four broad categories:

1. Methods for predicting crimes: These are approaches used to forecast places and times with an increased risk of crime.

2. Methods for predicting offenders: These approaches identify individuals at risk of offending in the future.

3. Methods for predicting perpetrators’ identities: These techniques are used to create profiles that accurately match likely offenders with specific past crimes.

4. Methods for predicting victims of crimes: Similar to those methods that focus on offenders, crime locations, and times of heightened risk, these approaches are used to identify groups or, in some cases, individuals who are likely to become victims of crime.

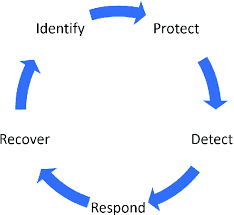


Fig. Circle of Prediction and Prevention

**1.2 Methodology**

Methods to identify individuals at high risk of offending in the future. The bulk of these methods relate to assessing individuals’ risk. Here, conventional methods rely on clinical techniques that add up the number of risk factors to create an overall risk score. The corresponding predictive analytics methods use regression and classification models to associate the presence of risk factors with a percent chance that a person will offend. Also of interest are methods that identify criminal groups (especially gangs) that are likely to carry out violent assaults on each other in the near future. Hence, these methods can also be used to assess the risk that an individual will become a victim of crime.

Methods used to identify likely perpetrators of past crimes. These approaches are essentially real-world versions of the board game Clue™: They use available information from crime scenes to link suspects to crimes, both directly and by processes of elimination. In conventional approaches, investigators and analysts of Law Enforcement Use of Predictive Technologies: Predicting Crimes Problem Conventional Crime Analysis (low to moderate data demand and complexity) Predictive Analytics (large data demand and high complexity) Identify areas at increased risk Using historical crime data Crime mapping (hot spot identification) Advanced hot spot identification models; risk terrain analysis Using a range of additional data (e.g., 911 call records, economics) Basic regression models created in a spreadsheet program Regression, classification, and clustering models Accounting for increased risk from a recent crime Assumption of increased risk in areas immediately surrounding a recent crime Near-repeat modeling Determine when areas will be most at risk of crime Graphing/mapping the frequency of crimes in a given area by time/date (or specific events) Spatiotemporal analysis methods Identify geographic features that increase the risk of crime Finding locations with the greatest frequency of crime incidents and drawing inferences Risk terrain analysis do this largely by tracing these links manually, with assistance from simple database queries (usually, the names, criminal records, and other information known about the suspects). Predictive analytics automate the linking, matching available “clues” to potential (and not previously identified) suspects across very large data sets.

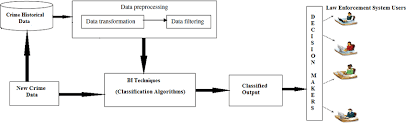
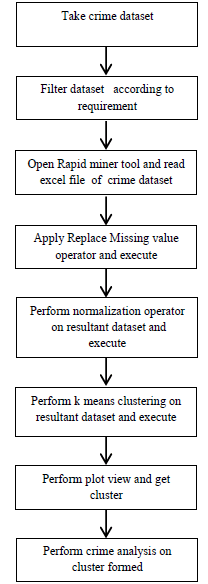


Fig. Flow of procedural Analysis

**FLOWCHART**

****

**CHAPTER 2: LITERATURE REVIEW**

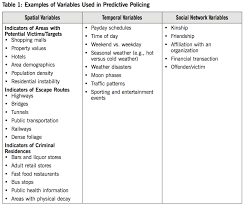
#### In the literature, a unanimous definition of predictive policing is missing but there is some consensus on its key features. Many of the articles indicate that predictive policing entails the application of quantitative techniques to forecast where criminal activities might occur in the (near) future. The predictions based on these analytic tools can guide the decision-making of law enforcement agencies, especially with the deployment of its personnel.

#### Analysts refers to a conceptualization of Craig Uchida of the National Institute of Justice that captures the essence of predictive policing: “Predictive policing is a concept that is built on the premise that it is possible to predict when and where crimes will occur again in the future by using sophisticated computer analysis of information about previously committed crimes” . Conceptualizations of predictive policing by Uchida are also used in other bodies of literature . They wield a slightly different conceptualization that not only focuses on place but also on the identification of individuals by these models: “Predictive policing is the application of analytical techniques- particularly quantitative techniques- to *identify* likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions” (p. xiii, own emphasis).

#### On the basis of these different but complementary notions about predictive policing in the literature, we developed the following definition: *Predictive policing is the collection and analysis of data about previous crimes for identification and statistical prediction of individuals or geospatial areas with an increased probability of criminal activity to help developing policing intervention and prevention strategies and tactics.*

#### The first key feature of predictive policing is the *usage of a broad variety of data*. There is general agreement that predictive policing is mainly occupied with descriptive analytics that have the aim to expose and understand crime trends by processing a wide variety of (un)unstructured data. Potentially this could help law enforcers in their strategic and tactical planning and how they can effectively deploy their resources. Indeed, it is argued that this type of policing strategy uses data mining methods to collect data that can help in the decision-making of law enforcement agencies. This indicates that all data is relevant, whereas the traditional policing methods only rely on criminal data.

#### A second key feature of predictive policing is the *connection with pre-emptive policing*, which is the notion that law enforcers act before criminal activities take place to prevent crime from happening. It is argued that predictive policing can be regarded as a form of pre-emptive policing that is mainly depended on statistical data. It is described that the police can prevent criminal behavior by engaging in “*upstream prevention”*. Herein, law enforcers should work with multiple actors in society to take away factors that cause criminal behavior: safety becomes co-production. This is a theoretical caveat, however, as little literature argue how predictive policing can use in a manner that prevents criminal activity by taking away factors that cause it, in the spirit of upstream prevention.



## Benefits of predictive policing: Potential benefits

#### In the conceptualization of predictive policing, general potential benefits are embedded: law enforcement agencies apply these methods to deploy their resources more efficiently and effectively. It is indicated in their paper that predictive policing can identify patterns in enormous data sets, which can be used for interference by police forces. We analyzed the literature to identify the specific claims.

#### A first specific claim of the benefits of predictive policing is that *resources can be deployed more accurately in place and time*. In respect to identifying areas at increased risk, predictive policing techniques are used that rely both on historic crime data and a wider range of data. For instance, advanced hot spot identification models and risk terrain analysis are used to forecast where criminal activity is most likely to occur. With this geospatial analysis, both criminal data and data that is retrieved through data mining are of importance: data that have no immediate relevance but can potentially help to prevent and predict crime from happening. In addition, these different types of data can also be used to determine *when* criminal activity is most likely to occur through spatiotemporal analysis. These models aim to forecast on what times the criminal activity is the highest in a specific geographic area. The assumption of near-repeat crimes, the theory that future crimes are more likely to take place near to the time and place of current crimes, is studied but also more specific patterns are presented in the literature. For instance, Dario, Morrow, Wooditch, and Vickovic tested with the use of criminal historic data from the Ventura Police Department whether good surfing conditions in California (i.e. weather conditions that attract surfers, locals, and tourists to surf spots), can be linked with a rise of criminal activity. They conclude that weather conditions indeed lead to more criminal activity in these areas, but only for a specific time-interval: between 2:30 pm and 5:29 pm . Haberman and Ratcliffe ([2012](https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)) found comparable results, as a chance of near-repeat events of armed robberies is increased in the first seven days, but hereinafter diminishes.

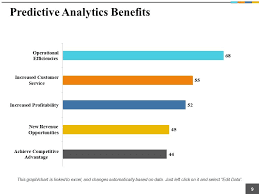
its

Fig. Predictive Analysis benefits

#### The analysis of time and space forms the basis for deployments of resources. Camacho-Collados and Liberatore ([2015](https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)) have developed in collaboration with the Spanish National Police Corps a Decision Support System (DSS) to efficiently distribute police officers in a geographic area. With their study, they tried to offer a solution of the Police Districting Problem (PDP), which is the challenge how police officers can optimally determine patrol sectors in which the chance of criminal behavior is the highest. The DSS-model proposed by the authors can help to better allocate police officers and determine the most optimal patrol routes. As part of their study, the authors tested their system and conclude that this method has the potential to distribute police forces more efficiently throughout the city. Although this is very promising, it remains a solution that needs to be implemented in practice to determine its actual value.

#### A second specific claim is that predictive policing techniques help to *identify individuals that potentially will be involved in an act of crime – either as victim of offender*. It is described that predictive algorithms can be used to identify members of criminal groups that show an elevated risk of a violent outbreak between them (e.g. gang shootings). Also, individuals can be identified that might become offenders in the future: inductive profiling . With these techniques, individuals that have attributes that correlate with a higher chance of displaying criminal behavior can already be monitored or targeted on the forehand. This profiling transcends only demographic characteristics of individuals, but can also consist of (social) behavioral patterns . It is illustrated with his research that sex crimes are most likely concentrated in activity spaces of the offender (i.e. locations that are frequently visited by individuals). On the basis of a more sophisticated analysis, It is demonstrated through social network analysis that the crime risks of individuals increase for a period of time if they are socially connected to a certain degree to an offender. Williams, Burnap, and Sloan draw similar conclusions in respect to social media and criminal activity, as they find an association “between aggregated open-source communications data and aggregated police recorded crime data in London”.

### Evidence for these benefits

#### Santos argued with his assessment of different policing techniques that there is little evidence regarding the effectiveness of predictive policing. Nevertheless, his article was written several years ago and this field of research is developing at high speed and several empirical studies have been conducted in the meantime. For this reason, we re-assessed the evidence-base for the benefits of predictive policing. We found that still only a limited number of studies in our corpus focused on the effectiveness of predictive policing methods in practice, but some recent studies tested whether the application of predictive policing techniques results in improvements of crime reduction.

#### It is evaluated that the utilization of predictive policing techniques by the New York Police Department (NYPD). They assessed the Domain Awareness system, which is a network of sensors, databases, devices, software and infrastructure that delivers tailored information and analytics to mobile devices and precinct desktops. The NYPD combined video analysis of cameras, environmental sensors, license plate readers, the 911-feed and an acoustic correlation processing of gun firing (i.e. ShotSpotter) to keep track of criminal activity in the city. The authors compared traditional hot spot policing with NYPD’s predictive policing software and the accuracy of predicting certain types of criminal behavior (i.e. burglary, felony assault, grand larceny, robbery & shootings). They did this by comparing traditional hot spot policing with this new predictive policing system in a 24-week cross-validation period and the results were striking: the accuracy of the predictions on the different types of criminal behavior have increased, especially for shootings. In addition, the efficiency of the officers was also improved as they could better respond on criminal activity and find suspects through the full network of sensors by which it is easier to find suspects or stolen vehicles through license plate recognition. Also, officers can respond faster on shootings through ShotSpotter (which registers the sounds of shootings). Overall, the overall crime index of New York decreased with 6% since the implementation of DAS. The authors recognize that this cannot be fully attributed to this system but still qualify the system as a success.

#### Analysts randomized controlled trials of predictive policing techniques in divisions of police departments of Los Angeles and Kent. They tested whether an epidemic-type aftershock sequence model (ETAS) that calculates the risk of criminal behavior in long-term hotspots and short-term near-repeat risks. In this experiment, every 24 h they randomly assigned police divisions to configure their patrol route with the use of either the ETAS-algorithm (treats of this study the tournament) or with the use of a traditional crime analyst (control). The main findings that the configuration of police patrols with ETAS forecasts resulted in a decrease in crime as a function of patrol time of 7,4%. In contrast, the forecasts made by analysts did not have any significant effect in terms of crime reduction. Thus, they conclude, predictive algorithms such as the ETAS algorithm can indeed help to reduce crime.

#### However, not all studies that tested the effectiveness of predictive policing methods found positive results. For instance, Hunt, Saunders, and Hollywood ([2014](https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)) evaluated an experiment conducted by the Shreveport police department, Louisiana, in 2012. In this study, a predictive policing programme is used to determine geographical areas with a higher risk of criminal activity: Predictive Intelligence Operational Targeting (PILOT). In addition, this programme also derives concrete plans for action. This experiment uses program theory – i.e., the determination of indicators that increases the likelihood that property crime occurs- to construct prevention models on how to react to these indicators. These models are distributed to the command staff (i.e. intermediate outcomes), and it assessed whether this resulted in a reduction of crime and an increase in the quality of arrests (i.e. final outcomes). The results are indecisive: there is no considerable evidence that the application of PILOT leads to a reduction in crime rates when compared to the control districts which used conventional crime mapping. Possible explanations for these no-results are the (1) statistical power which was too limited (because few districts were incorporated in the experiment), (2) the police departments did not implement the strategies from PILOT rightly, (3) or the possibility that the programme was inadequate.

#### One evaluation study focused specifically on profiling systems. Saunders, Hunt, and Hollywood tested whether a Strategic Subject List (SSL) which estimates the risk of individuals that might be involved in gun violence, either as offender or victim – can help to prevent criminal activity. In their research, they did not find any clues that individuals on this list have an increased chance of being a victim of gun violence. The authors found an increase in the chance of individuals on this list being arrested for shooting but this could result from the fact that officers used the Strategic Subject List as leads for cases that were unresolved or because of extra monitoring. Hence, it remains unclear how this specific predictive policing method should be used and whether it can contribute to existing policing practices.

#### Thus, existing evaluations and assessments produced mixed results. Whereas several studies show a positive significant effect for geospatial predictions, other studies have no significant results. Only one study focused on profiling and this study produced ambiguous outcomes. The mixed findings can be attributed to the type of evaluation, to the type of predictive policing or to the type of method that was used for predictive policing. A preliminary conclusion is that this approach has potential but not all types of crimes can be effectively reduced through predictive policing models and therefore the officers executing these strategies need to adequately use these models. Every individual predictive model that is applied by police departments should be individually evaluated to determine their effectiveness and efficiency. Below, an overview is given to what extent the two different types of predictions are empirically proven ([Figure 1](https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664#F0001)).

#### Figure 1. Relation between claimed and proven benefits of predictive policing.

#### 

## 

## Drawbacks of predictive policing

#### Although many police departments and academics are convinced of a bright future for predictive policing, several academics also raise some concerns regarding the usage of data mining and algorithms to forecast criminal behavior. These concerns will be discussed and determined to what extent these drawbacks are based on hypothetical assumptions or on empirical evidence. Many of these potential drawbacks concern both the spatial-temporal predictions and the profiling and therefore we only discuss this distinction if it is specifically highlighted.

#### A cautionary remark that is raised in the existing literature is that the algorithms cannot be fully comprehended by law enforcement agencies because of a lack of transparency of the predictive policing models. If the models are not fathomed by law enforcement agencies, it might become a challenge to determine how risky geospatial areas or individuals are: *riskier* is not the equivalent of *risky*. If law enforcers do not understand the factors that lead to an increased chance of crime, the effectiveness of their actions might be reduced. In addition, it is important for law enforcement agencies to make adequate inferences of the data and to make sure that it is properly understood to develop fitting strategies.

#### Most of the predictive *models are mainly data driven instead of theory driven*, which can also have major implications on how these models are used. The usage of big data and data-based approaches might have the consequence that there is too much emphasis on correlations, instead of causality. This could be problematic as predictions that are derived on algorithms are opaque and are hard to interpret. If existing models are not assessed and evaluated with the use of practical insights (e.g. tacit knowledge of police officers), the models will be outdated and present a skewed image of reality. The study by Saunders et al. ([2016](https://www.tandfonline.com/doi/full/10.1080/01900692.2019.1575664)), as already described, also indicates that predictive algorithms are possibly not self-explanatory. One of the reasons why they did not find significant results could be the fact that although the contact with potential offenders increased, the models do not provide any enough recommendations how to interact with these offenders or how the models should be used. This reinforces the assumption that the predictive models can never be used on itself without further instructions to police officers how to act in the streets, and hamper their effectiveness.

#### With a lack of transparency and understanding of predictive models, *accountability problems* might occur. Bennett Moses and Chan raise the potential consequence that law enforcers cannot fully understand and interpret the outcomes of the software and deem the outcomes as sufficient input for decision-making. This could lead to an accountability gap in which police officers are unable to understand the models and therefore cannot deduce biases in the models. In other words, it becomes unclear who is responsible for decision-making when there is full reliance on predictive algorithms.

#### As a consequence of the lack of transparency, use of a model of predictive policing for profiling may result in *stigmatizing individuals and groups* and thus forms of discrimination based on algorithms. Law enforcers may overlook and underestimate the effect when the predictive models are used inadequately as they can potentially lead to stigmatization of individuals. In their article, they provide a hypothetical example of how the resocialization of ex-convicts might be affected by the actions of law enforcement agencies. They make a compelling argument how the stigmatization of groups of people with a criminal record can lead to aversion, and eventually, relapse in criminal behavior as their reintegration in society is stagnated by these predictive algorithms and how they are treated by officers. Thus, the profiling of individuals can eventually be self-fulfilling as it drives individuals towards criminal behavior.

#### The administration of predictive policing techniques can also entail *unintended consequences*. Brannon compared two data-driven projects – one of which concerned with predictive policing – and came to a remarkable conclusion. He reviewed the Kansas City No Violence Alliance (KCNOVA), which uses network-analysis software to identify individuals that are most likely to be involved in criminal activity and a living lab in the downtown of Kansas that is aimed to improve the quality of live and stimulate capital investment in this part of the city. Brannon concludes that the application of predictive policing in a geographical area in the city also impacts this space and its inhabitants: when one area is monitoring criminal activity while the other area is flourishing as capital investments are encouraged, this leads to spatial inequality across racial and social classes.

#### Next, to the practical issues that are accompanying the administration of predictive algorithms, more fundamental *concerns regarding privacy and ethics* are raised. Edwards and Urquhart review whether the usage of open source and social media data by law enforcement agencies should be permitted and to what extent the digital identity of citizens is protected. The authors raise the question to what extent the digital footprint of citizens (e.g. what citizens share on social media and the data that can be collected such as our movements with public transport) is private and whether it can be used unconditionally. De Hert and Lammerant discuss tensions between the profiling of individuals in society and legal safeguards, as these are often loosened to resolve these tensions. Even tough jurisprudence on privacy is very clear in the legal limits of predictive profiling, there remain little cases which makes it hard to set precise boundaries what is eligible. This conclusion is underlined by Costanzo, D’Onofrio, and Friedl as they argue that legislation is important to retain trust between citizens and governments as there should be a balance between the utilization of big data and the privacy of citizens. If there are no clear boundaries citizens might develop a profound sense of mistrust towards governments as they are unaware whether, and to what extent, they are monitored.

### 

### 

### Evidence for these drawbacks

#### This literature review highlights the potential drawbacks of predictive policing have been discussed quite extensively, but empirical evidence for these drawbacks is lacking. The risk of predictive policing lacking transparency, with affiliated problems such as accountability issues, is plausible. In addition, if law enforcement agencies have limited boundaries or legislation they need to comply to, a wedge might develop between the government and its citizens since the mutual trust is reduced. However, in the academic literature, there is little empirical evidence how predictive policing methods lead to difficulties in practice. The focus in the (limited number of) empirical evaluation studies is on testing whether the desirable outcomes were realized and not whether this resulted in adverse effects. This is a gap that needs to be filled by empirical research, to show whether these claimed drawbacks actually occur in the implementation of predictive policing.

#### CHAPTER 3: SYSTEM DEVELOPMENT

#### 3.1 Proposed Model

The system is based on a decade of detailed academic research into the causes of crime pattern formation. That research successfully linked several key aspects of offender behavior to a mathematical structure that is used to predict how crime patterns will evolve from day-to-day, from moment-to-moment.

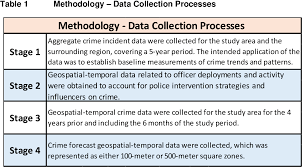
The mathematics looks complicated, but the behaviors upon which the math is based are very understandable. There are three aspects of offender behavior that make their way into our model.

1. **Repeat victimization**, which describes – taking burglary as an example – that if a house is broken into today, the risk that it is broken into tomorrow actually goes up. This is because it is “rational” for offenders to return to the places where they have been successful before. It makes less sense to go to some other unknown house where they don’t know if the house is empty of people, they don’t know how hard it is to break in, and they don’t know what there is to be stolen. The house they broke into two or three days ago is much less risky.
2. **Geographical search** ties it all together. We know that offenders rarely travel very far from their key activity points such as their home, work and play locations, meaning that crimes tend to cluster together.

**3.2 Design**

**3.2.1 Predictive Analysis**

* Predictive analytics uses algorithms to recognise data patterns and predict future outcomes
* Predictive analytics encompasses data mining, predictive modeling, machine learning, and forecasting
* The aim of such analysis is to identify relationships among variables that may not be immediately apparent using hypothesis-driven methods.

****

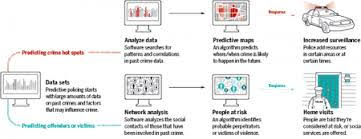
****

Fig. Identification of hotspots and profiling of offenders

**3.2.2 Algorithm Used for Development**

* **K-Means Algorithm**

K-means clustering is a simple unsupervised learning algorithm that is used to solve clustering problems. It follows a simple procedure of classifying a given data set into a number of clusters, defined by the letter "k," which is fixed beforehand. The clusters are then positioned as points and all observations or data points are associated with the nearest cluster, computed, adjusted and then the process starts over using the new adjustments until a desired result is reached.

* **KNN Algorithm**

K-Nearest Neighbors (KNN) is one of the algorithms used in [Machine Learning for regression](https://quantra.quantinsti.com/course/trading-with-machine-learning-regression) and classification problem. KNN algorithms use data and classify new data points based on similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase.

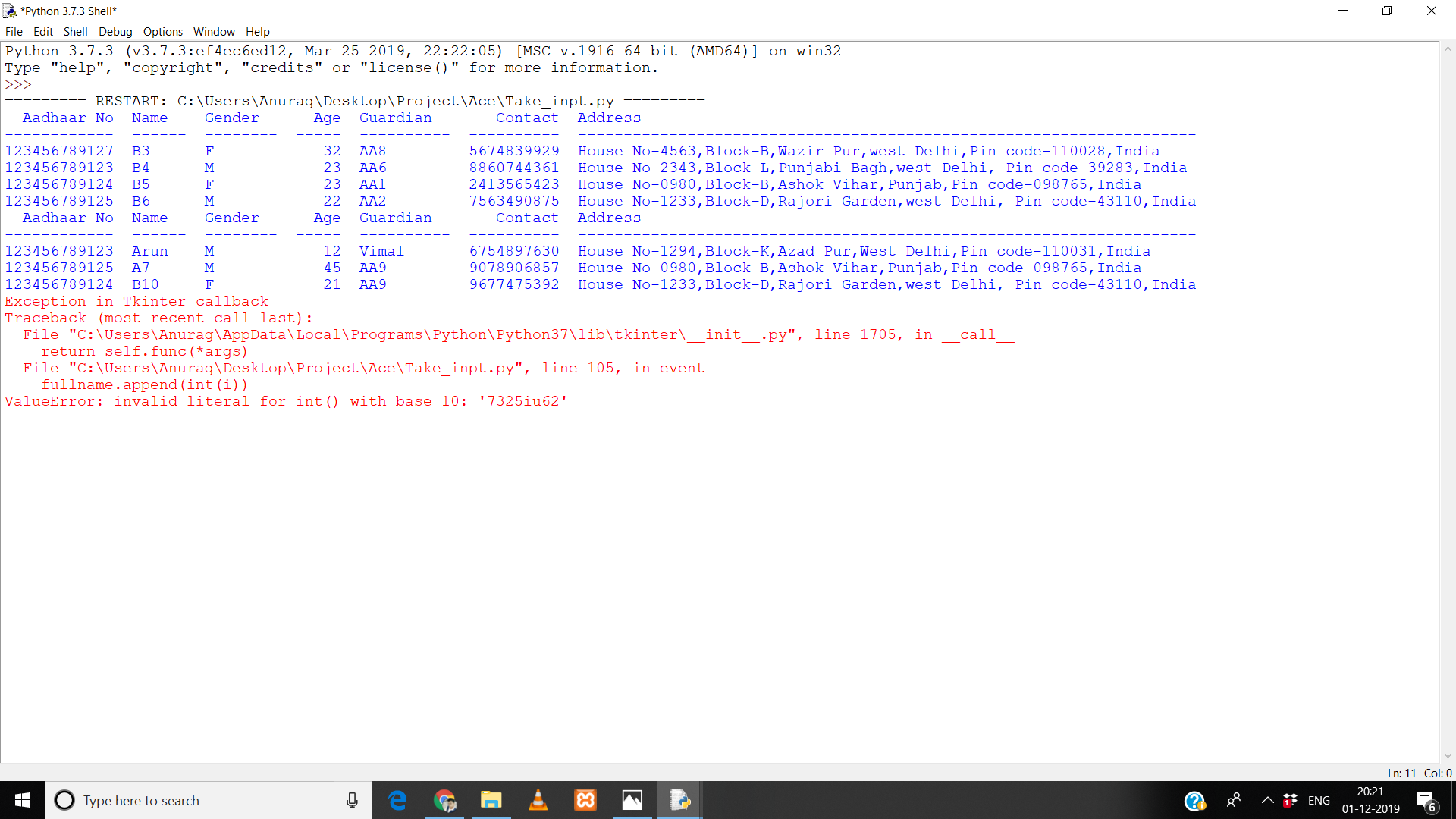
* **Logistic Regression**

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. The algorithm classifies incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classifications within data sets. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.

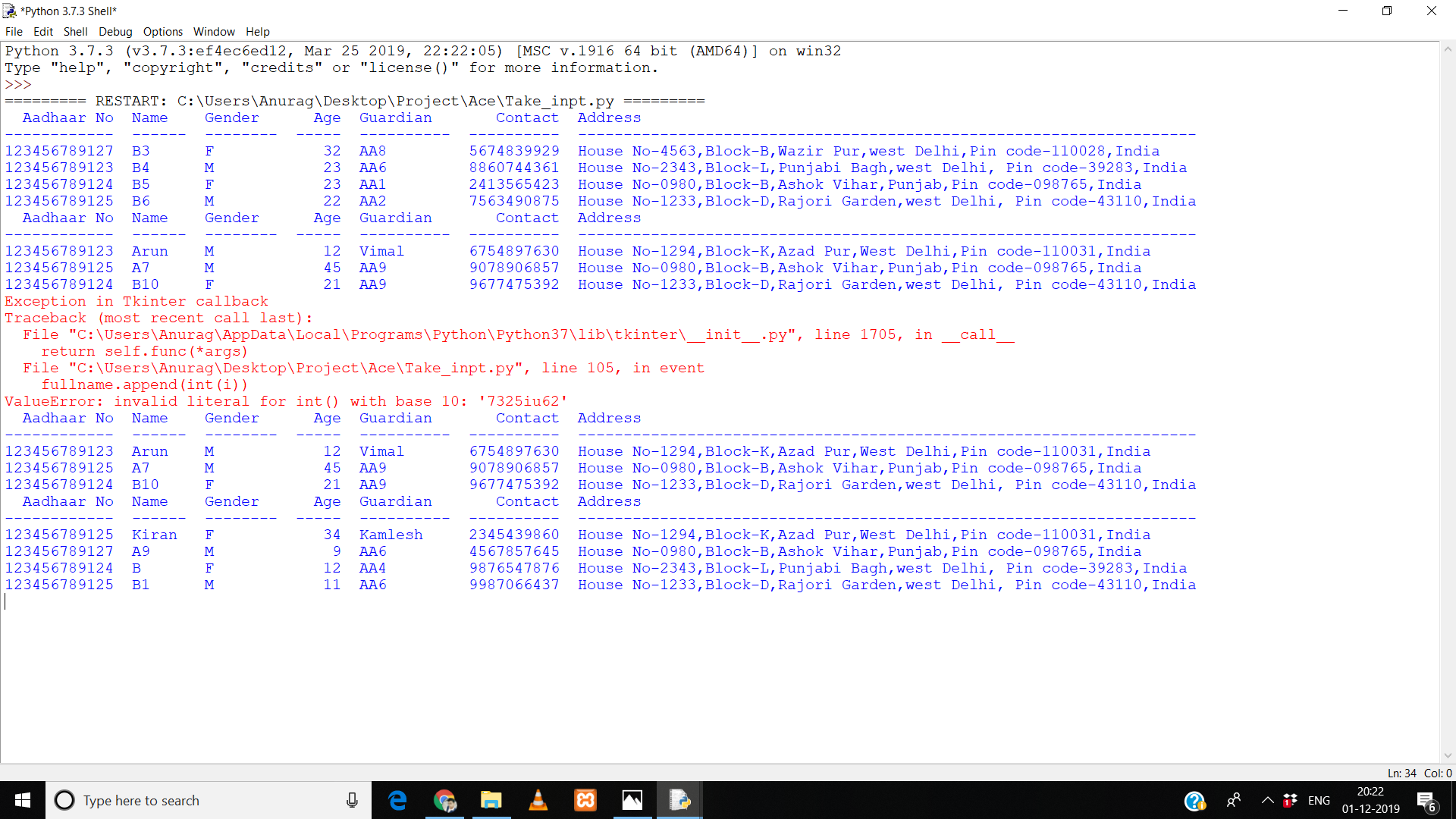
**3.2.3 Testing Used for Development**

1. **Unit Testing** - It is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output.
2. **Manual Testing** - Manual Testing is a type of Software Testing where Testers manually execute test cases without using any automation tools. Manual Testing is the most primitive of all testing types and helps find bugs in the software system. Any new application must be manually tested before its testing can be automated.
3. **Integration Testing** - It is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units. Test drivers and test stubs are used to assist in Integration Testing.

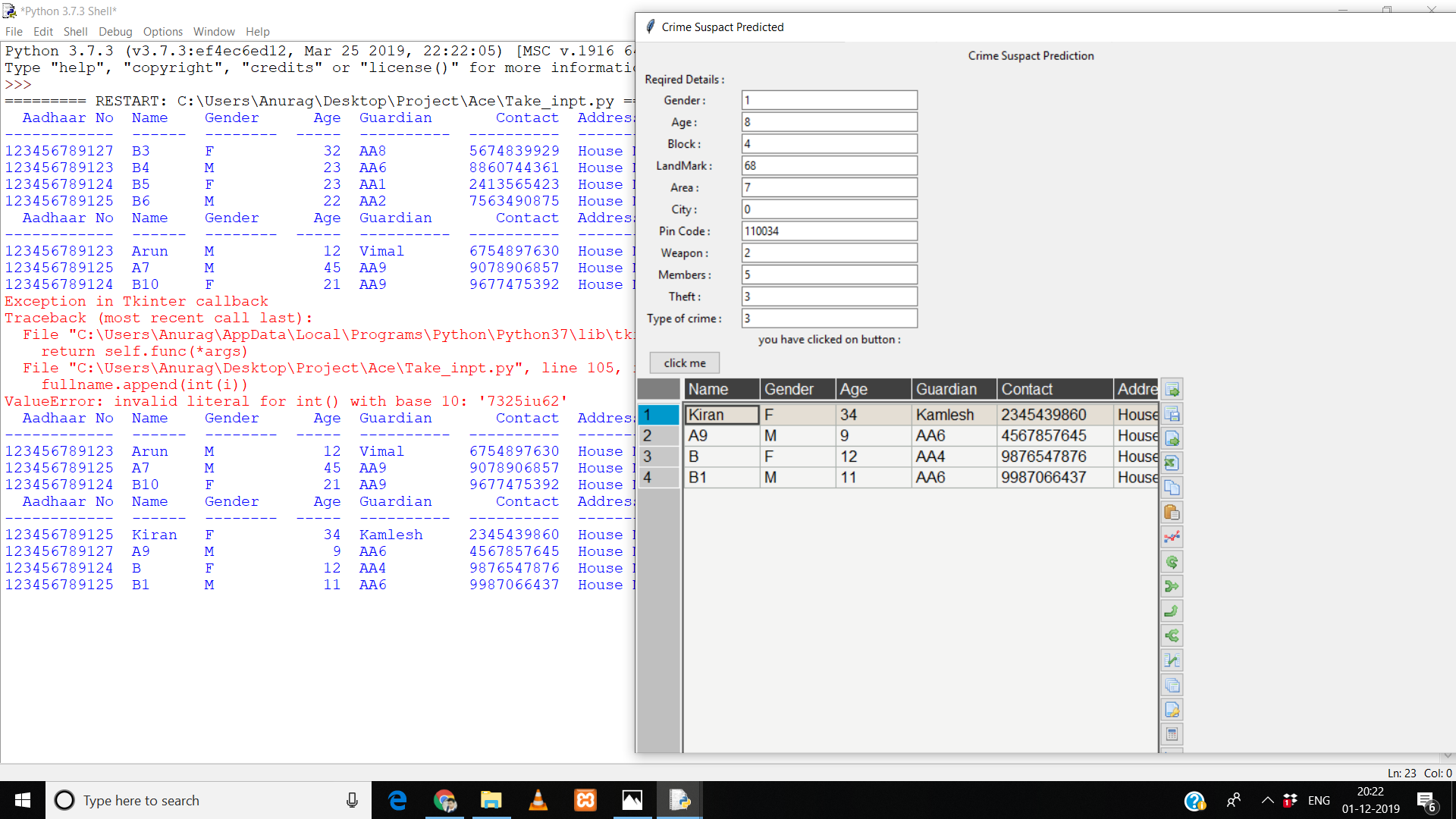
**Case 1: For testing**

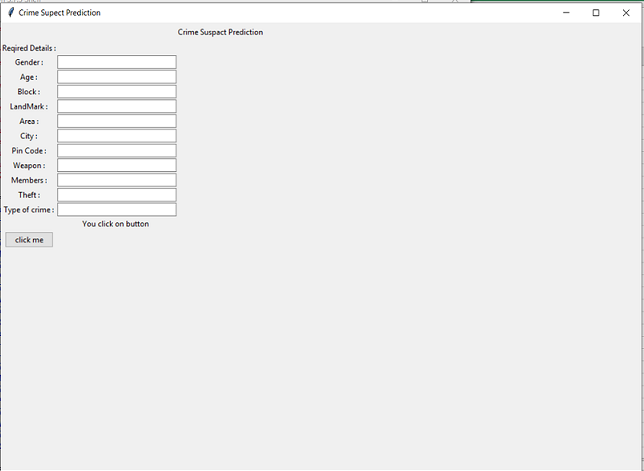


**Case 2:**

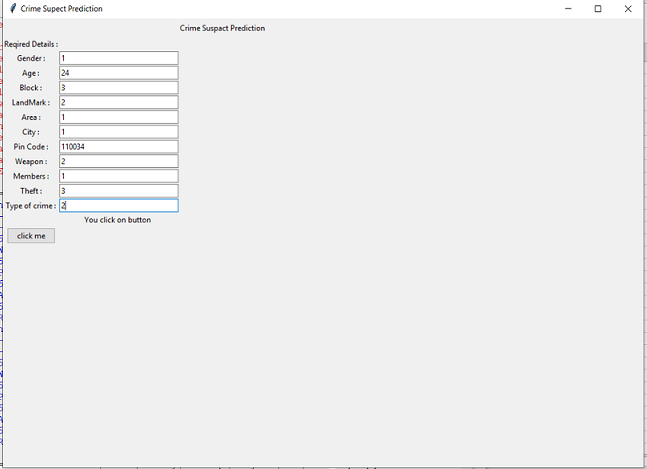


**Case 3:**

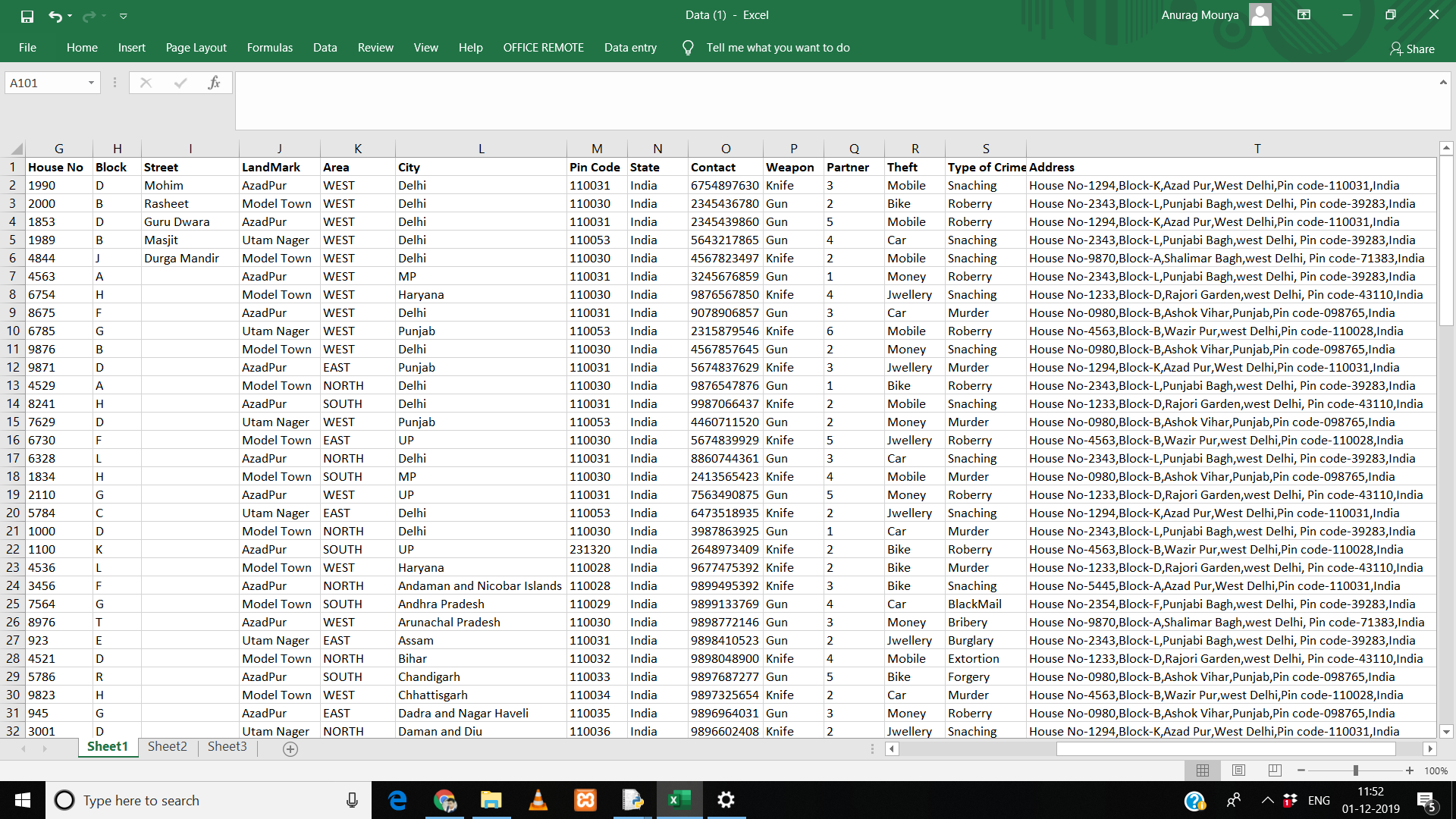
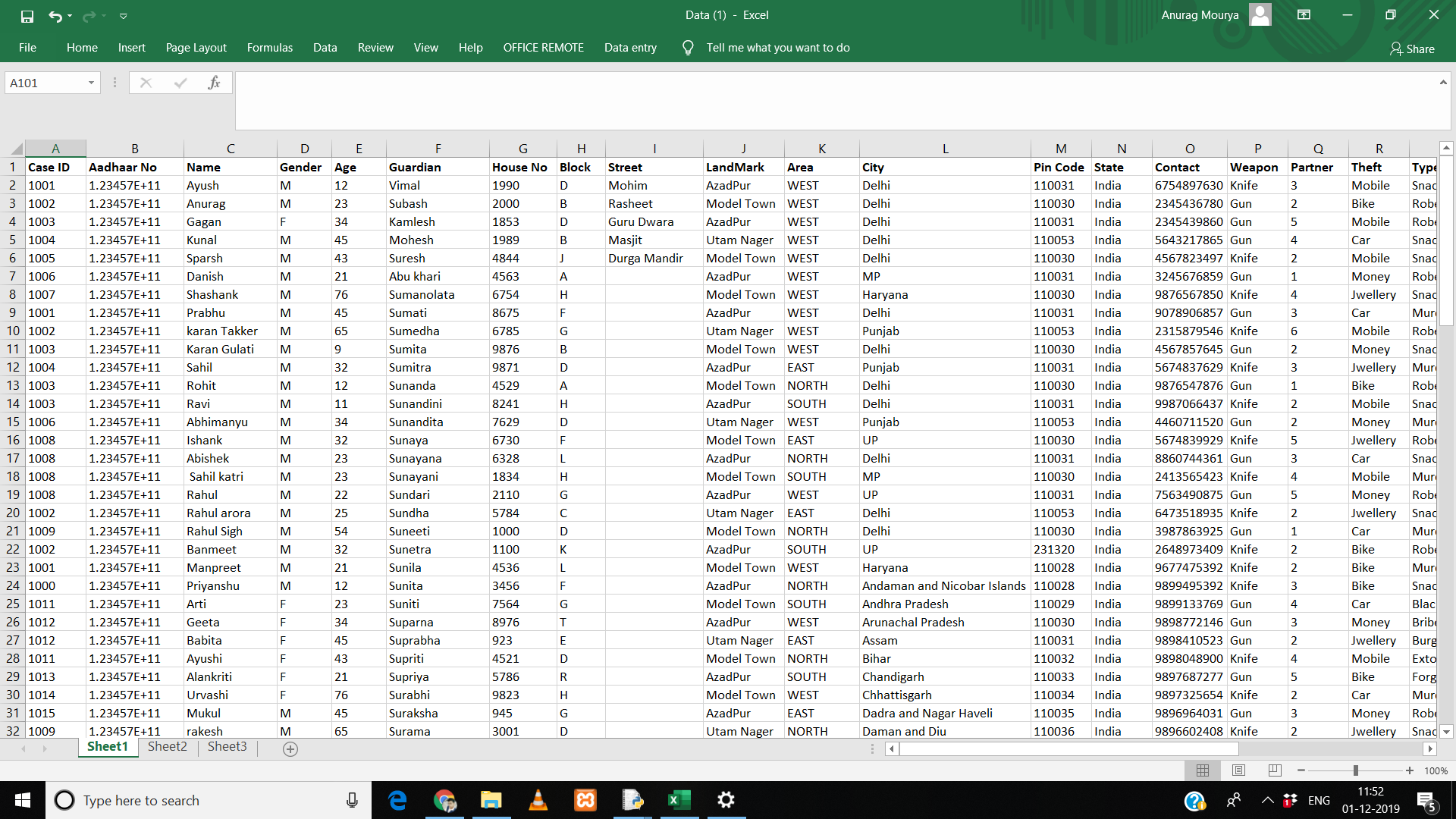


**Input Design**

Fields have been carefully devised so as to get optimised results while effectively using the database.

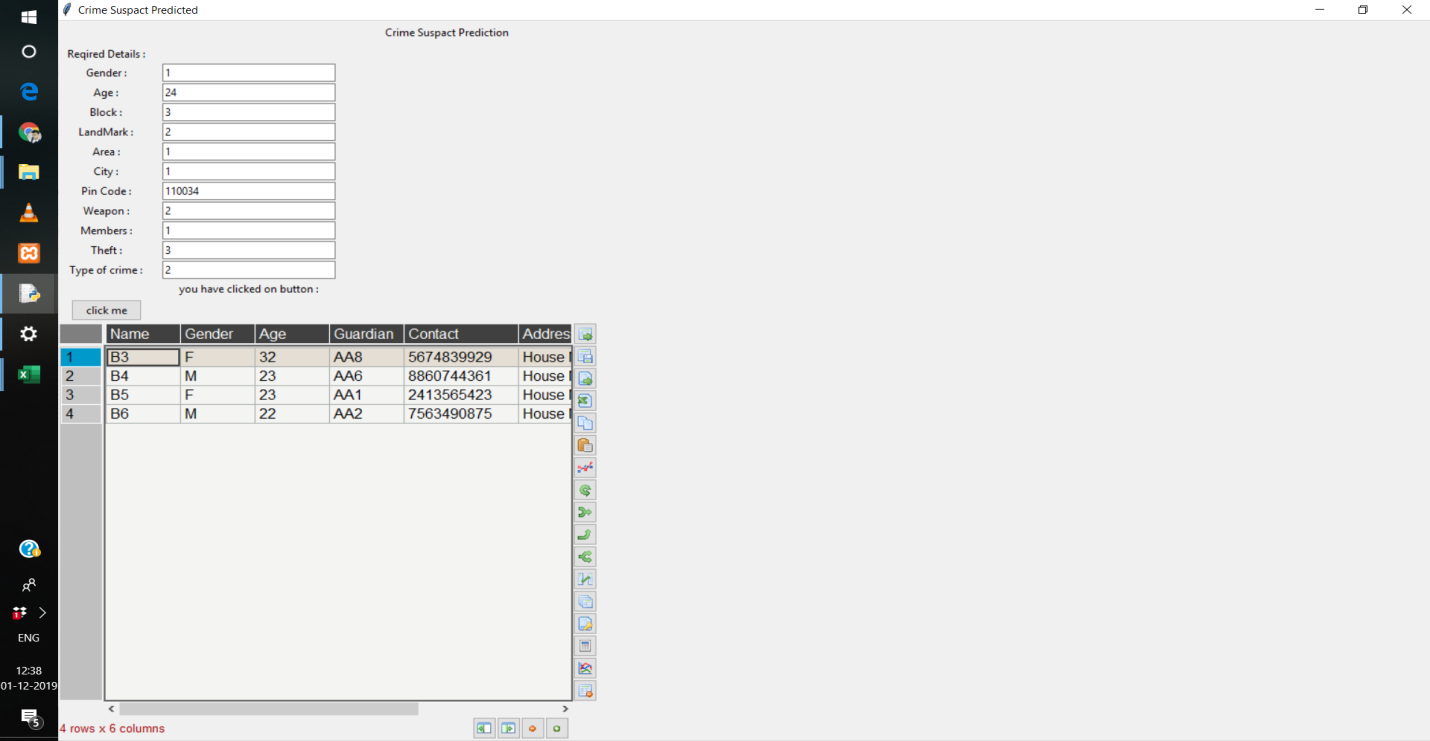
****

**Database Design**

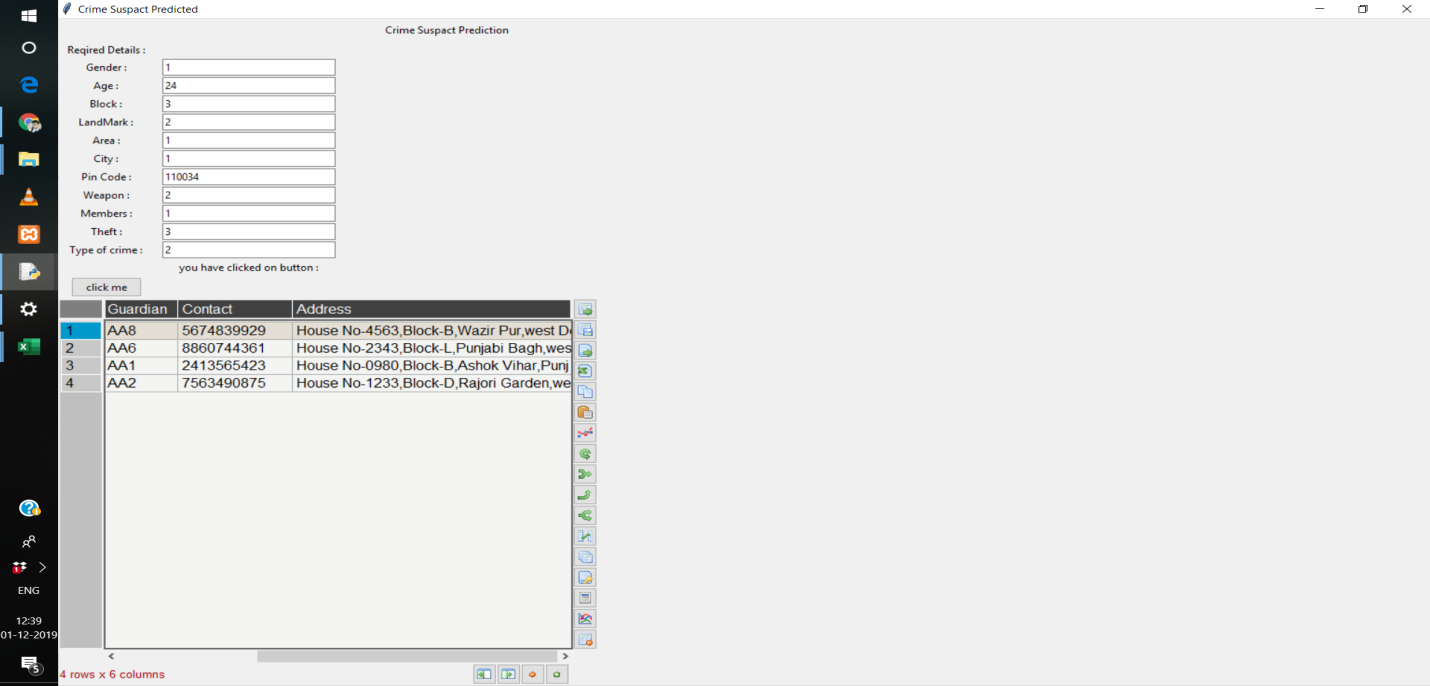
****

**Output Design**

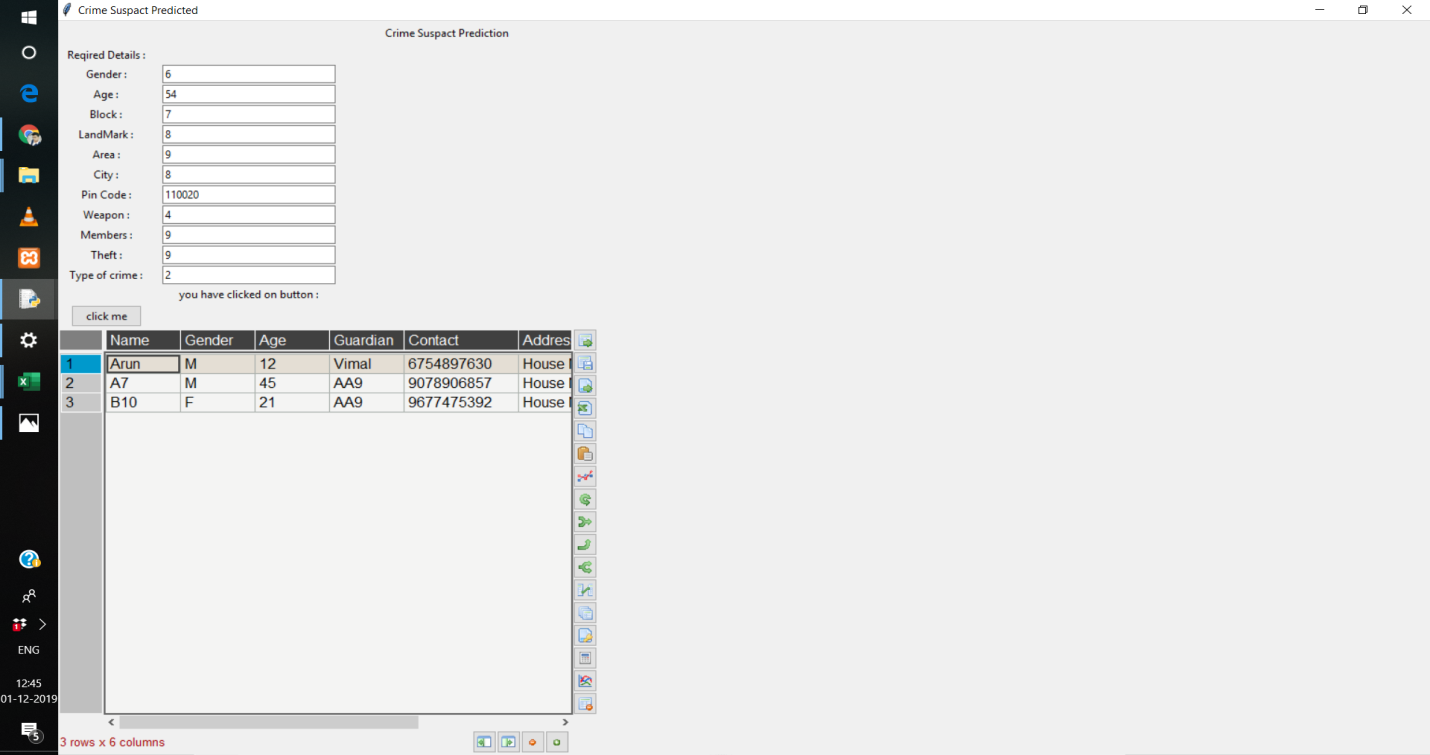
**Case 1:-**

****

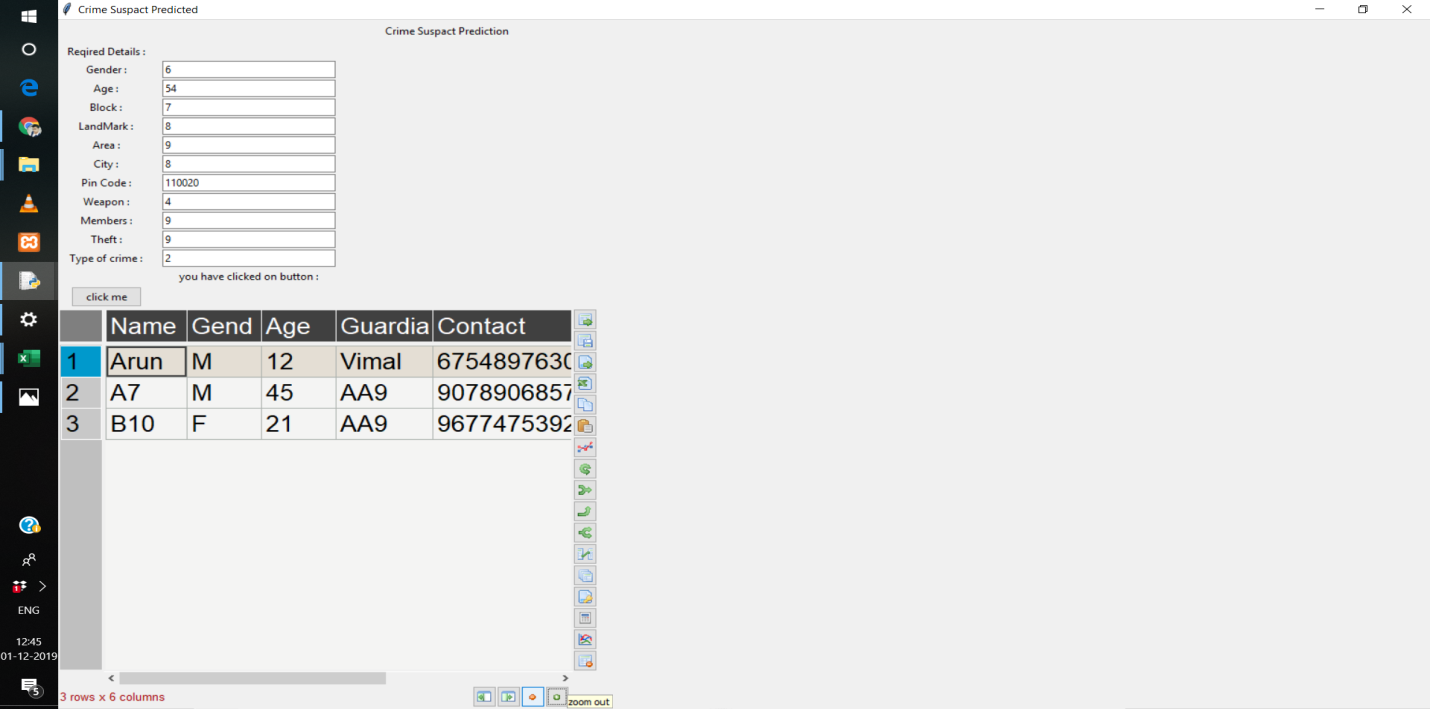
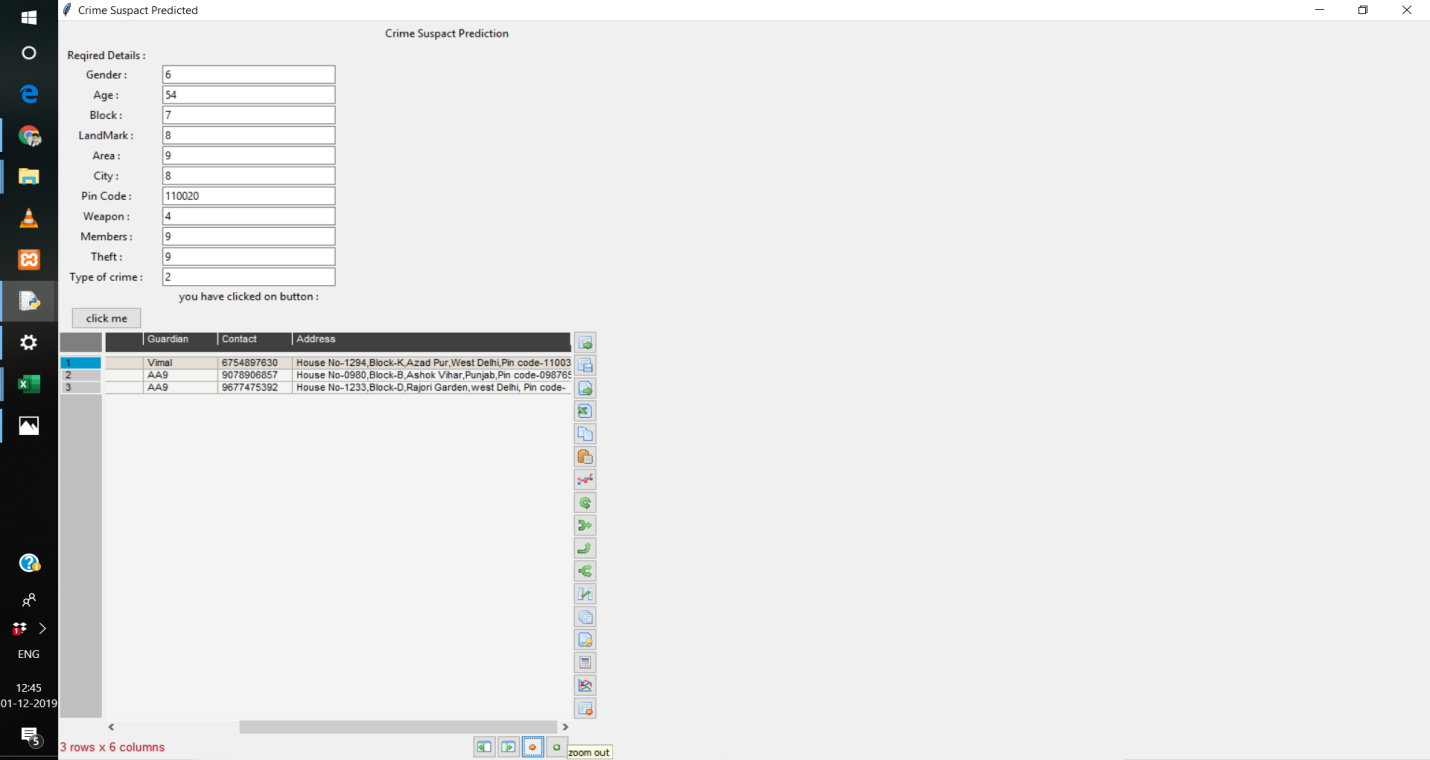
**Case 2:-**

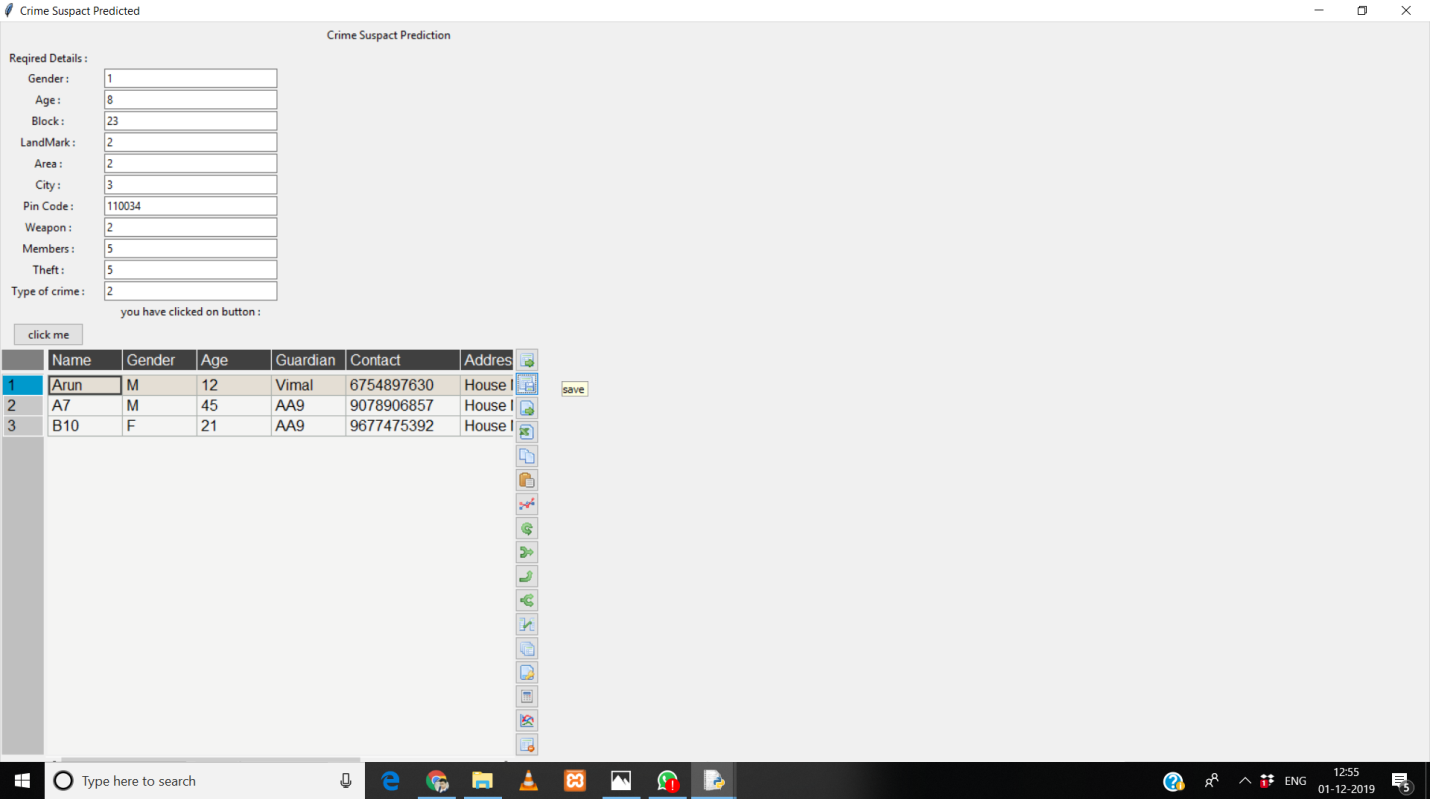
****

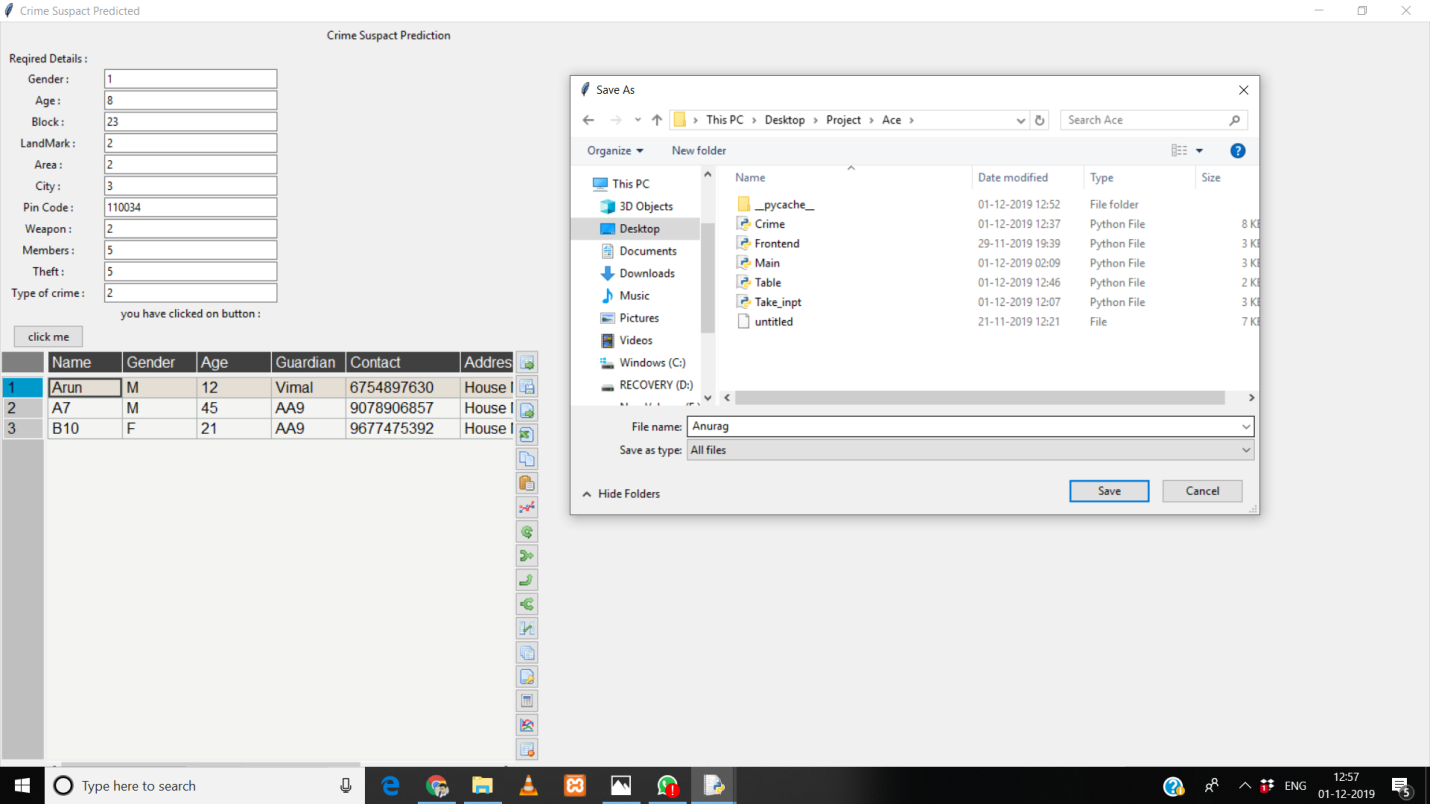
**Case 3:-**

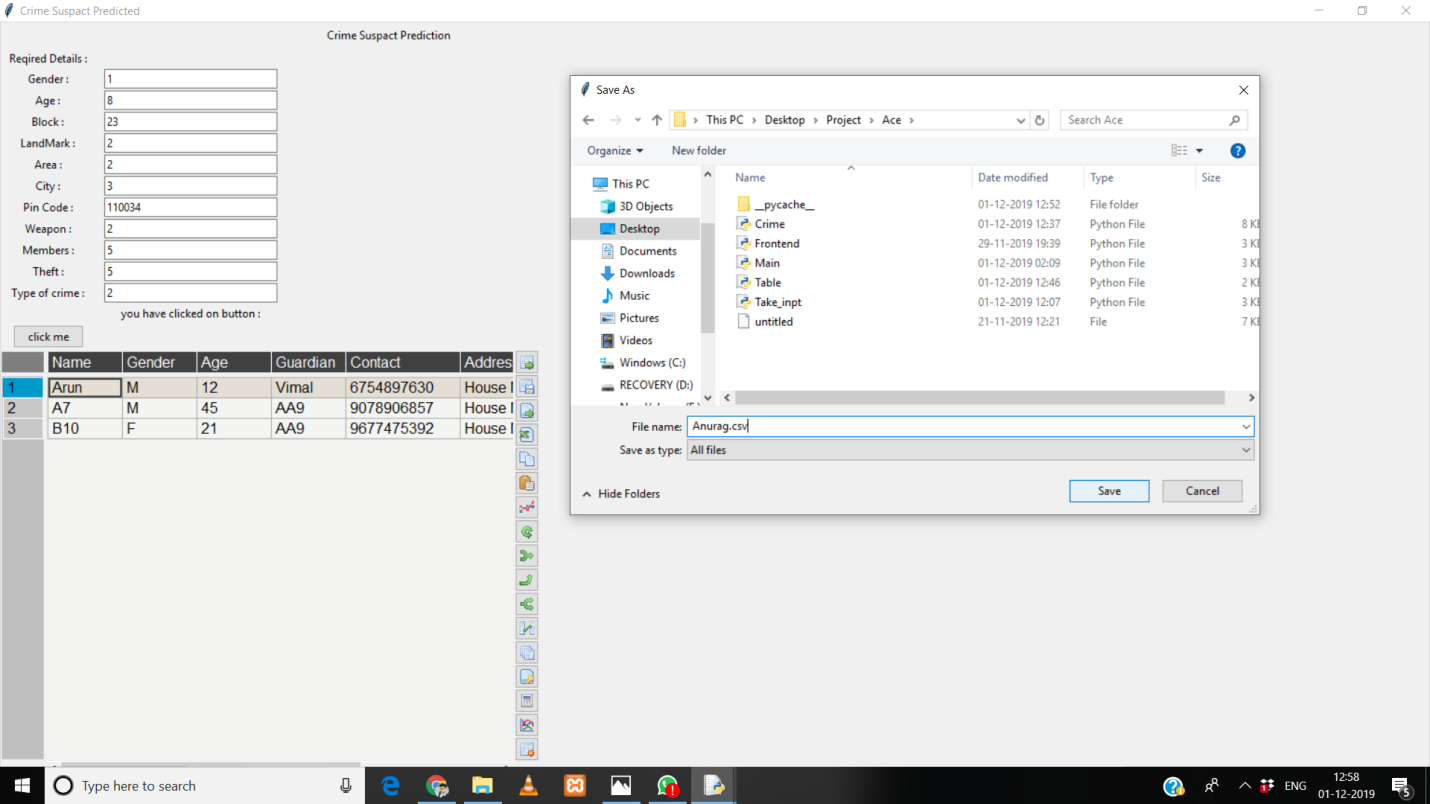
****

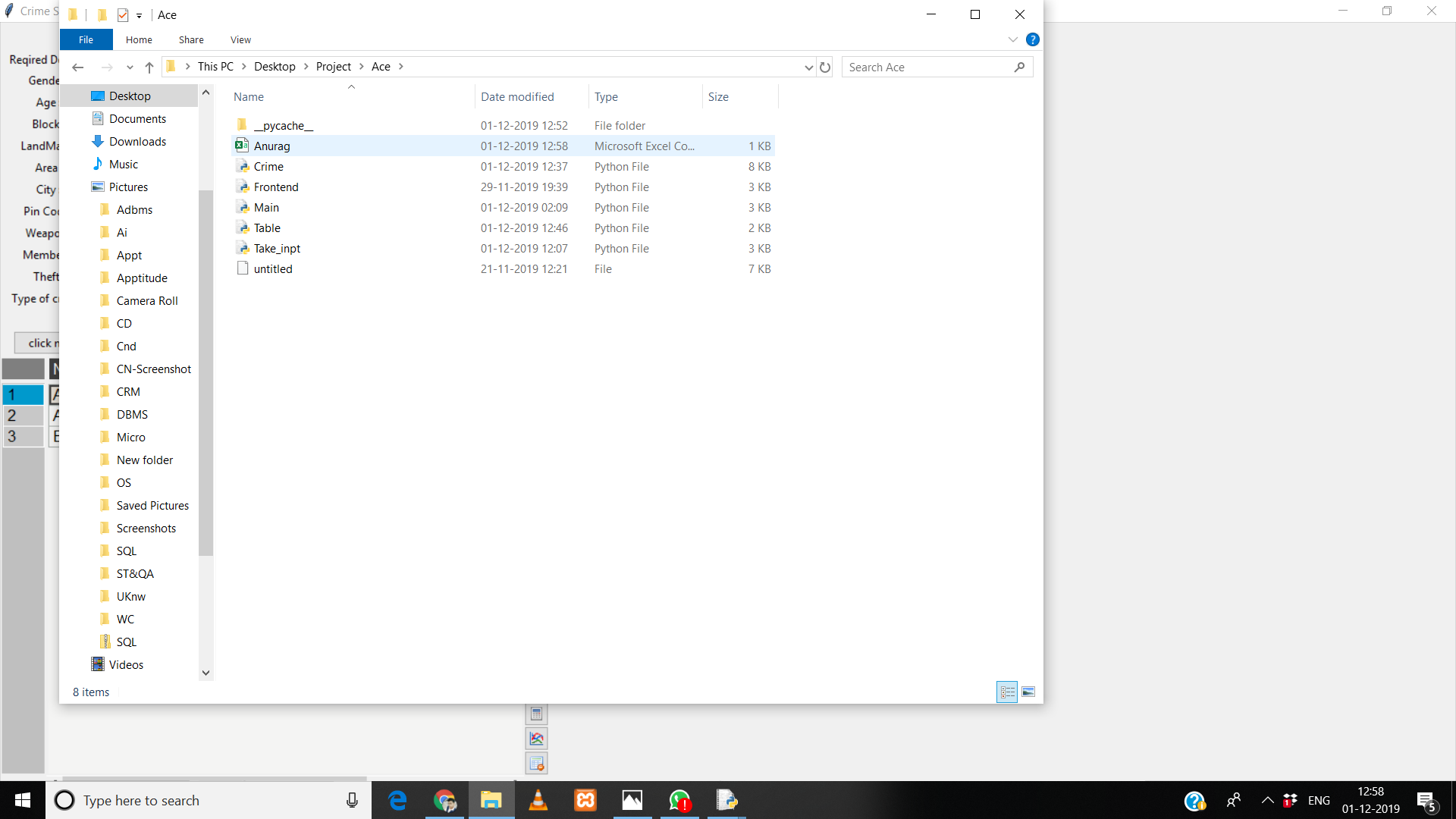
**Features of the Software Include**

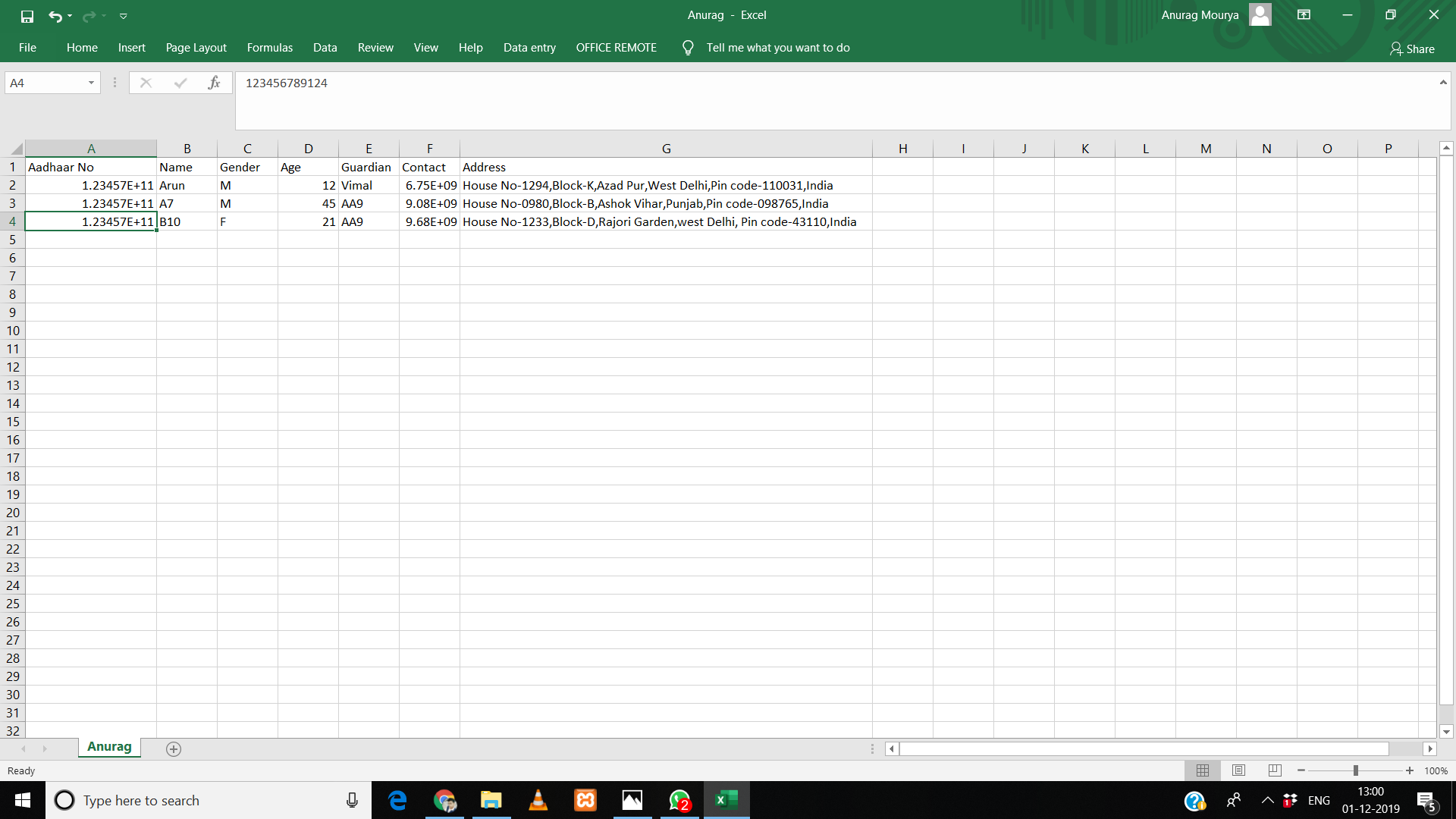
1. **Zoom In**
2. **Zoom Out**
3. **Save Table generated in CSV, Excel, etc. Format**

****

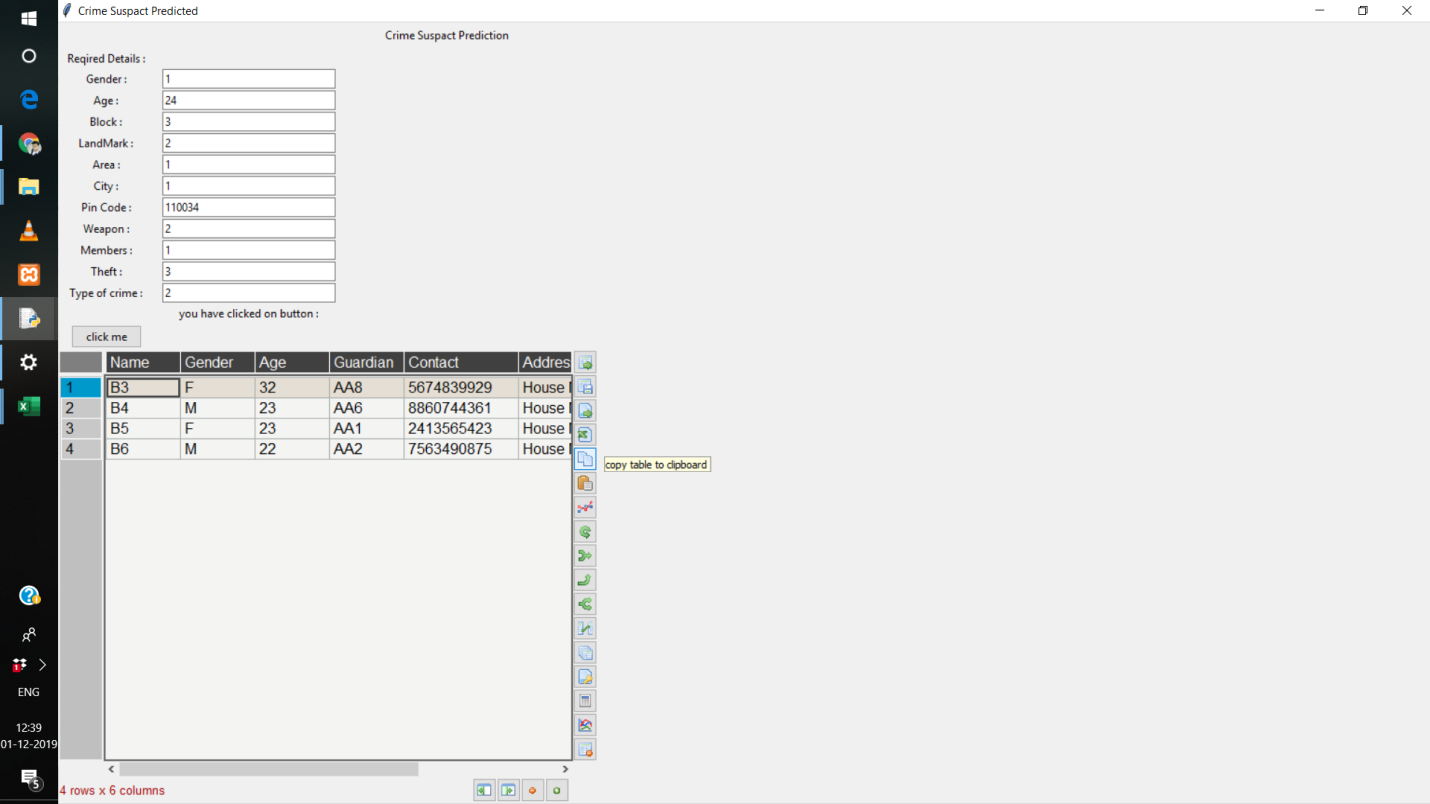
****

****

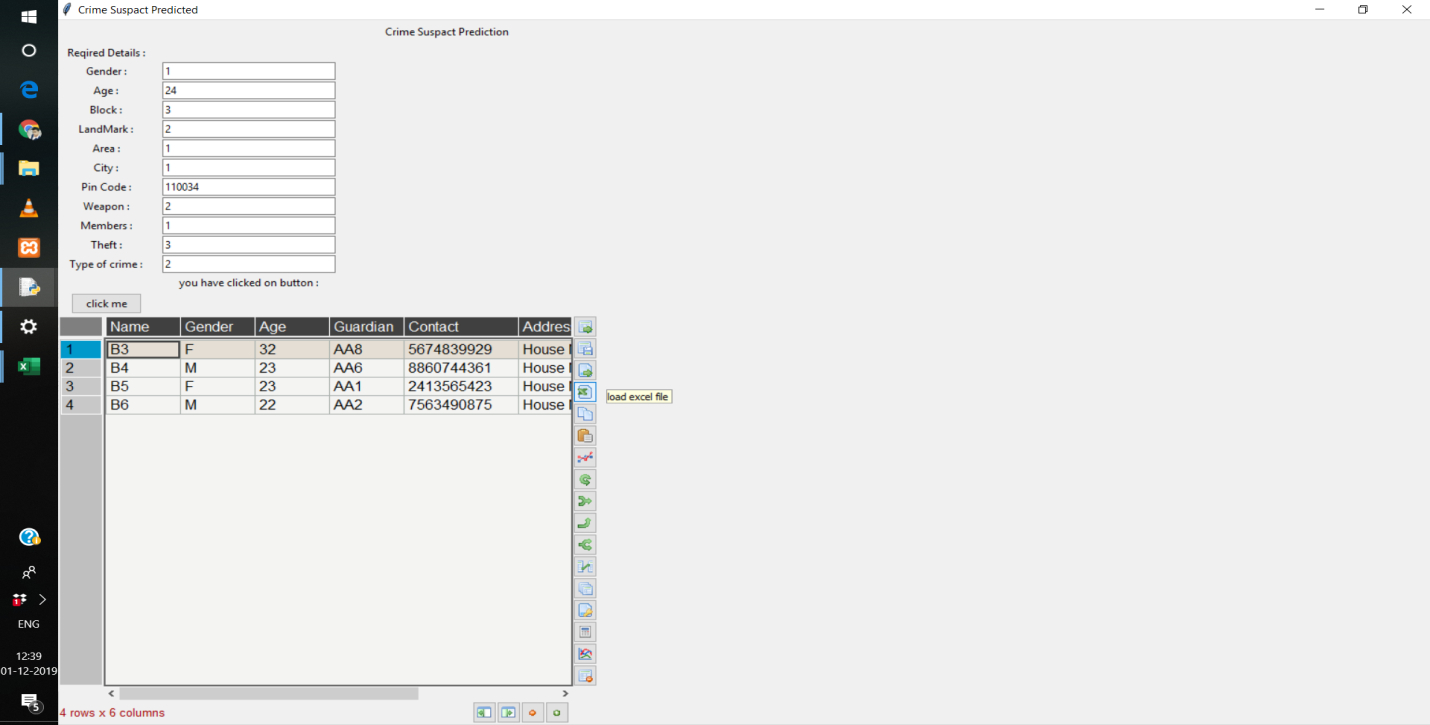
****

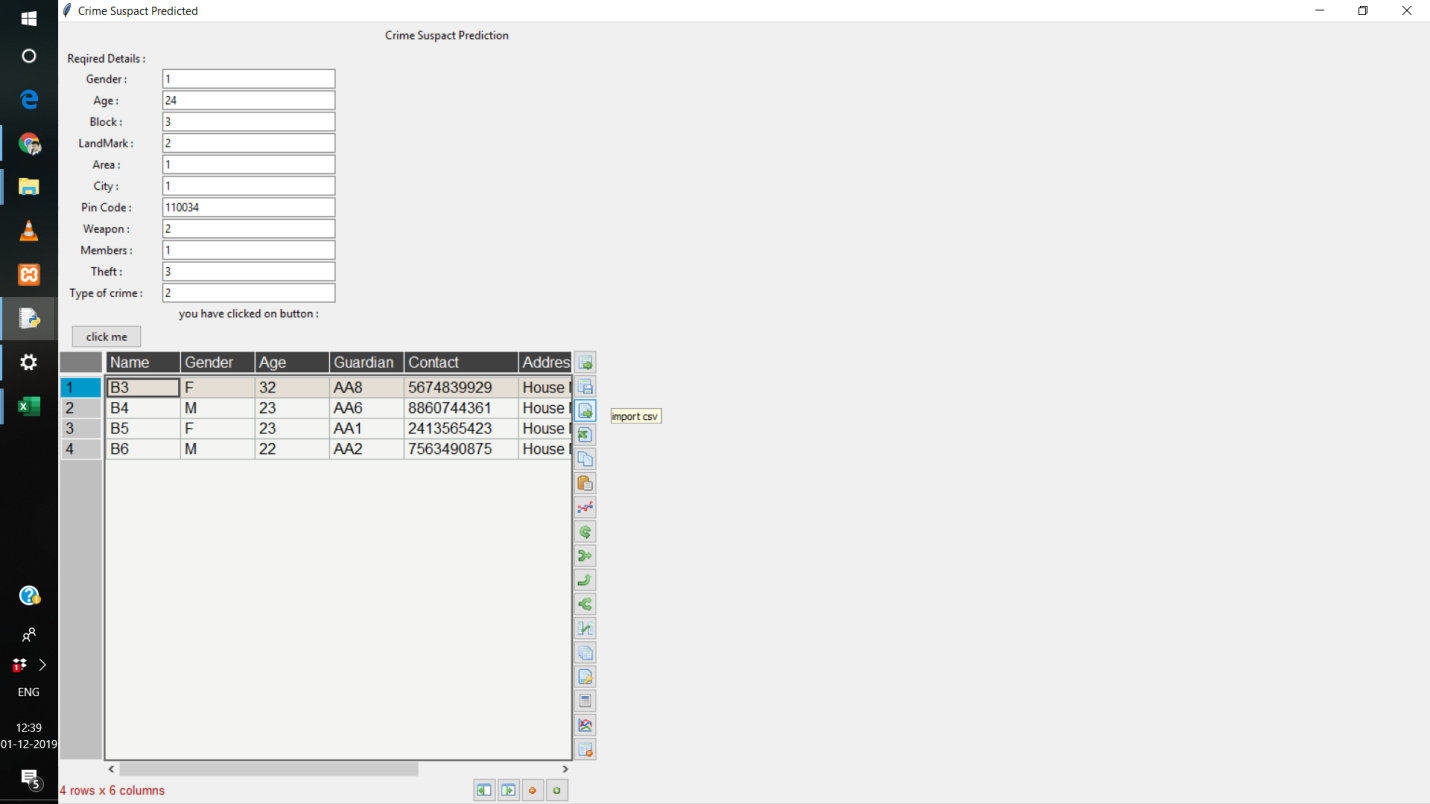
****

1. **Copy table to Clipboard**

****

1. **Load Excel File**

****

1. **Import csv**

**3.3.4 Source Code**

**BACKEND CODE**

################################# IMPORT LIBRARIES #####################

import numpy as np

import pandas as pd

from sklearn.preprocessing import LabelEncoder

from sklearn.pipeline import Pipeline

data= pd.read\_excel(r"C:\Users\Anurag\Desktop\Project\Data - Copy (2).xlsx")

data=pd.DataFrame(data)

#print(data)

##################################################################################################

############################### DATA SET ###################################

#data=data.head(23)

X1=data[["Gender","Block","LandMark","Area","City","State","Weapon","Theft","Type of Crime"]]

#print(X1)

class MultiColumnLabelEncoder:

def \_\_init\_\_(self,columns = None):

self.columns = columns # array of column names to encode

def fit(self,X,y=None):

return self # not relevant here

def transform(self,X):

'''

Transforms columns of X specified in self.columns using

LabelEncoder(). If no columns specified, transforms all

columns in X.

'''

output = X.copy()

if self.columns is not None:

for col in self.columns:

output[col] = LabelEncoder().fit\_transform(output[col])

else:

for colname,col in output.iteritems():

output[colname] = LabelEncoder().fit\_transform(col)

return output

def fit\_transform(self,X,y=None):

return self.fit(X,y).transform(X)

p=MultiColumnLabelEncoder(columns = X1).fit\_transform(data)

#print(p)

LM\_match=pd.DataFrame(p)

#####################################################################################################

import warnings

warnings.filterwarnings("ignore")

#################### MACHINE LEARNING USING ALGORITHM################

from pandas.plotting import scatter\_matrix

from sklearn.cluster import KMeans

import matplotlib.pyplot as plt

from sklearn import model\_selection

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import accuracy\_score

from sklearn.linear\_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

from sklearn.naive\_bayes import GaussianNB

from sklearn.svm import SVC

#split data

data1 = p #convert to dataframe to list/array

#get all numeric column

x\_train = data1[["Gender","Age","Block","LandMark","Area","City","Pin Code","Weapon","Partner","Theft","Type of Crime"]]

y\_train = data1[["Case ID"]]

#print(x)

#print(y)

"""

Gender\_x=int(input("Take gender M(0)and F (1) : " ))

Age\_x=input("Take age of criminal : " )

Block\_x=int(input("Block of crime spot (0-9): " ))

LandMark\_x=int(input("LandMark of crime spot(0-9) : " ))

Area\_x=int(input("Area of crime spot (0-9) : " ))

City\_x=int(input("City of crime spot (0-4) : " ))

PinCode\_x=int(input("PinCode of crime spot (6 digit) : " ))

Weapon\_x=int(input("Weapon used in crime (0-9) : " ))

Partner\_x=input("Member are involve in crime : " )

Theft\_x=int(input("Theft of crime (0-9) : " ))

Typeofcrime\_x=int(input("Type of crime (0-3) :" ))

"""

def alo(z):

cls= KNeighborsClassifier(n\_neighbors= 5,metric='minkowski',p=2)

cls.fit(x\_train,y\_train)

#z=[[1,48,2,1,1,1,110034,1,5,2,1]]

#z=[[Gender\_x,Age\_x,Block\_x,LandMark\_x,Area\_x,City\_x,PinCode\_x,Weapon\_x,Partner\_x,Theft\_x,Typeofcrime\_x]]

y\_pred=cls.predict(z)

#print(y\_pred)

y\_pred=int(y\_pred)

f=data[data["Case ID"]==y\_pred ]

#u=f[["Case ID","Aadhaar No","Name","Gender","Age","Guardian","Contact","Address"]]

u=f[["Aadhaar No","Name","Gender","Age","Guardian","Contact","Address"]]

ind(u)

#print(u)

#split in train tand test

#x\_train,x\_validation,y\_train,y\_validation = model\_selection.train\_test\_split(x, y, test\_size=.10, random\_state=7)

#print(x\_train)

#print(x\_validation)

#print(y\_train)

#print(y\_validation)

"""

L=LogisticRegression()

L.fit(x\_train,y\_train)

L\_pred=L.predict(z)

#print(L\_pred)

D=DecisionTreeClassifier()

D.fit(x\_train,y\_train)

D\_pred=D.predict(z)

#print(D\_pred)

S=SVC(gamma='auto')

S.fit(x\_train,y\_train)

S\_pred=S.predict(z)

#print(S\_pred)

G=GaussianNB()

G.fit(x\_train,y\_train)

G\_pred=G.predict(z)

#print(G\_pred)

LD=LinearDiscriminantAnalysis()

LD.fit(x\_train,y\_train)

LD\_pred=LD.predict(y\_pred)

#print(LD\_pred)

"""

'''

kmeans = KMeans(n\_clusters=2).fit(x\_train,y\_train)

centroids = kmeans.cluster\_centers\_

print(centroids)

print(kmeans)

plt.scatter(x\_train,x\_validation, c= kmeans.labels\_.astype(float), s=50, alpha=0.5)

plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=50)

plt.show()

'''

# Program to find most frequent

# element in a list

def most\_frequent(case\_id):

counter = 0

num = case\_id[0]

for i in case\_id:

curr\_frequency = case\_id.count(i)

if(curr\_frequency> counter):

counter = curr\_frequency

num = i

return num

#case\_id=np.concatenate((L\_pred, D\_pred, S\_pred, G\_pred, LD\_pred),axis=0)

#case\_id=np.concatenate(( y\_pred),axis=0)

#case\_id=list(y\_pred)

#case\_id=y\_pred

#print(case\_id)

#caseid=most\_frequent(case\_id)

#print(caseid)

#print(LM\_match["LandMark"])

#print(x\_validation["LandMark"])

#f=data[data["Case ID"]==caseid ]

#print(f)

#v=f[["Case ID","Aadhaar No","Name","Gender","Age","Guardian","Contact","Address"]]

#v=f[["Aadhaar No","Name","Gender","Age","Guardian","Contact"]]

from tabulate import tabulate

def ind(u):

k=["Aadhaar No","Name","Gender","Age","Guardian","Contact","Address"]

u.set\_index('Aadhaar No',inplace=True)

print(tabulate(u,headers=k))

import Table as t

t.app1(u)

#app.mainloop()

#print(type(u))

def usqt(y):

z=[y]

#print("pass",z)

alo(z)

def cooz():

return u

def tooz():

return v

def main():

return cooz()

return tooz()

"""

from IPython.display import HTML

s=HTML(u.to\_html(classes= 'table table-striped'))

print(s)

"""

"""

import Frontend

def Amu():

Frontend.Lt1()

class Entr:

def \_\_init\_\_(self):

print('Created Client')

def connected(self):

print('Connected')

def send\_message(self,data2):

print("clienrt sent '{} '".format(data2))

Fd=pd.DataFrame(f)

n=int(x\_validation["LandMark"].values)

print(n)

M=LM\_match[LM\_match["LandMark"]==n]

Md=pd.DataFrame(M)

if Fd["Name"]==Md["Name"]:

print(Fd)

"""

'''

#Build Model

# Spot Check Algorithms

models = []

models.append(('LR', LogisticRegression()))

models.append(('LDA', LinearDiscriminantAnalysis()))

models.append(('KNN', KNeighborsClassifier()))

models.append(('CART', DecisionTreeClassifier()))

models.append(('NB', GaussianNB()))

models.append(('SVM', SVC(gamma='auto')))

results =[]

names = []

for name,model in models:

kfold = model\_selection.KFold(n\_splits=10, random\_state=7)

cv\_results = model\_selection.cross\_val\_score(model, x\_train, y\_train, cv=kfold, scoring='accuracy')

results.append(cv\_results)

names.append(name)

msg = "%s: %f (%f)" % (name, cv\_results.mean(), cv\_results.std())

print(msg)

'''

**TABLE CODE**

from tkinter import \*

from pandastable import Table, TableModel

import tkinter as tk

from tkinter import ttk

def app1(u):

#print("U :",u)

y=u

#print("Y :",y)

sm(y)

def sm(y):

#print("Passsss :",y)

class TestApp(ttk.Frame):

#import Crime

#y=Crime.cooz()

'''Basic test frame for the table'''

def \_\_init\_\_(self, parent=None):

self.parent = parent

Frame.\_\_init\_\_(self)

self.main = self.master

#width\_value=self.main.winfo\_screenwidth()

#height\_value=self.main.winfo\_screenheight()

self.main.geometry('1000x750+200+100')

self.main.title('Crime Suspect Predicted')

f = ttk.Frame(self.main)

f.grid(row=15,columnspan=10,sticky=E+W)

df = TableModel.getSampleData()

self.table = pt = Table(f, dataframe=y,

showtoolbar=True, showstatusbar=True)

pt.show()

return

app=TestApp()

app.mainloop()

#app = TestApp()

#launch the app

#app.mainloop()

**TAKE INPUT CODE**

from tkinter import \*

import tkinter as tk

from tkinter import ttk

o = tk.Tk()

o.title('Crime Supect Prediction')

o.geometry("1000x700")

Tn = ttk.Label(o,text='Crime Suspact Prediction')

Tn.grid(row=0,column=4,pady=5)

Tn = ttk.Label(o,text='Reqired Details :')

Tn.grid(row=1,column=0,pady=1)

ln = ttk.Label(o,text='Gender :')

ln.grid(row=2,column=0,pady=1)

tln = ttk.Entry(o,width=30)

tln.grid(row=2,column=1,pady=1)

ln1 = ttk.Label(text='Age :')

ln1.grid(row=3,column=0,pady=1)

tln1 = ttk.Entry(o,width=30)

tln1.grid(row=3,column=1,pady=1)

ln2 = ttk.Label(text='Block :')

ln2.grid(row=4,column=0,pady=1)

tln2 = ttk.Entry(o,width=30)

tln2.grid(row=4,column=1,pady=1)

ln3 = ttk.Label(text='LandMark :')

ln3.grid(row=5,column=0,pady=1)

tln3 = ttk.Entry(o,width=30)

tln3.grid(row=5,column=1,pady=1)

ln4 = ttk.Label(text='Area :')

ln4.grid(row=6,column=0,pady=1)

tln4 = ttk.Entry(o,width=30)

tln4.grid(row=6,column=1,pady=1)

ln5 = ttk.Label(text='City :')

ln5.grid(row=7,column=0,pady=1)

tln5 = ttk.Entry(o,width=30)

tln5.grid(row=7,column=1,pady=1)

ln6 = ttk.Label(text='Pin Code :')

ln6.grid(row=8,column=0,pady=1)

tln6 = ttk.Entry(o,width=30)

tln6.grid(row=8,column=1,pady=1)

ln7 = ttk.Label(text='Weapon :')

ln7.grid(row=9,column=0,pady=1)

tln7 = ttk. Entry(o,width=30)

tln7.grid(row=9,column=1,pady=1)

ln8 = ttk.Label(text='Members :')

ln8.grid(row=10,column=0,pady=1)

tln8 = ttk.Entry(o,width=30)

tln8.grid(row=10,column=1,pady=1)

ln9 = ttk.Label(text='Theft :')

ln9.grid(row=11,column=0,pady=1)

tln9 = ttk.Entry(o,width=30)

tln9.grid(row=11,column=1,pady=1)

ln10 = ttk.Label(text='Type of crime :')

ln10.grid(row=12,column=0,pady=1)

tln10 = ttk.Entry(o,width=30)

tln10.grid(row=12,column=1,pady=1)

msg = Label(text='You click on button ')

msg.grid(row=13,column=1)

def inputbyuser(y):

import Crime

Crime.usqt(y)

def event():

#print('you have clicked on button')

ln = tln.get()

ln1 = tln1.get()

ln2 = tln2.get()

ln3 = tln3.get()

ln4 = tln4.get()

ln5 = tln5.get()

ln6 = tln6.get()

ln7 = tln7.get()

ln8= tln8.get()

ln9 = tln9.get()

ln10= tln10.get()

fullname1 = [ln,ln1,ln2,ln3,ln4,ln5,ln6,ln7,ln8,ln9,ln10]

fullname=[]

for i in fullname1:

fullname.append(int(i))

#print('name is ',fullname)

#print(type(fullname))

msg.configure(text="you have clicked on button :")

inputbyuser(fullname)

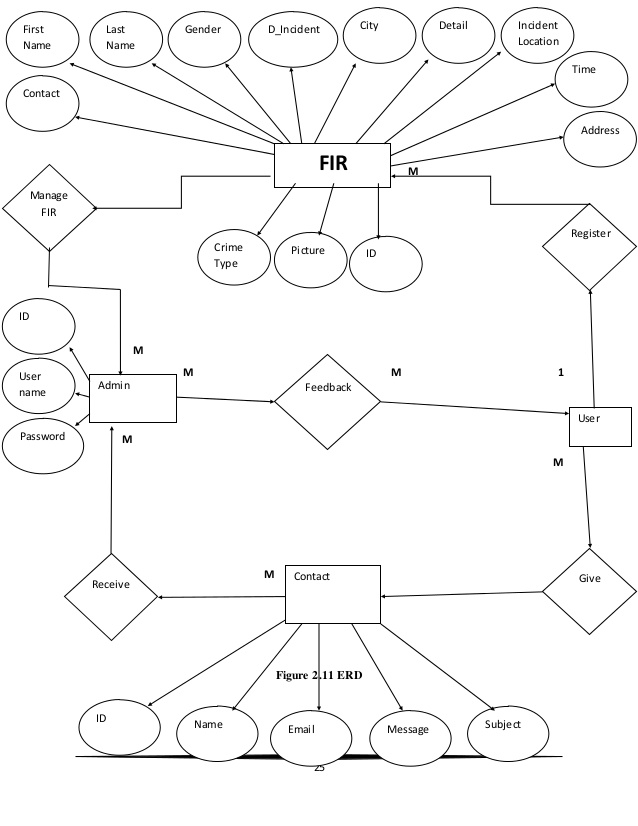
return fullname

b = ttk.Button(o,text='click me',command=event)

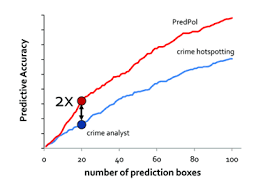
b.grid(row=14,column=0,pady=2)

o.mainloop()

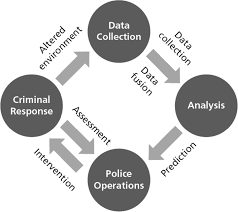
**E-R DIAGRAM**

****

**Predictive Policing Crime rate**

****

**Predictive Policing in the context**

****

**CHAPTER 4: PERFORMANCE ANALYSIS**

**4.1 Software Requirements :**

**Tools Used:**

* + - Python IDLE.
    - Interpreters for scripts.
    - Python version - minimum 3.6v or PyCharm
    - Python Pip install with variable environment support

**4.2 Hardware Requirements:**

* + - Windows Desktop
    - RAM - minimum 4GB
    - ROM - minimum 64GB
    - Screen - 1000x750

**4.3 Limitations:**

* + Real-time guidance
* Portability
* Power limitations
* Privacy preservation
* Continuous availability

**CHAPTER 5: CONCLUSION**

This literature review attempted to give a state-of-the-art overview of the scholarly attention to predictive policing. The purpose of this study is threefold as it assessed (1) how predictive policing is conceptualized, (2) what the potential and proven benefits are, and (3) what evidence there is for these claimed and proven drawbacks. We will summarize the outcomes in these conclusions and highlight the relevance for police practitioners.

In the current literature, a unanimous definition of predictive policing is absent. Nevertheless, most of the literature operationalize predictive policing as a method that applies quantitative techniques to predict in what geographical areas there is an increased chance of criminal behavior, but also which individuals and groups – through predictive profiling – are more likely to be involved in criminal activities. These models help to configure an optimal deployment of resources (e.g. patrol routes of officers) to reduce crime most efficiently and effectively. On the basis of our review of the literature, we developed the following definition that combines the geospatial focus and profiling: *Predictive policing is the collection and analysis of data about previous crimes for identification and statistical prediction of individuals or geospatial areas with an increased probability of criminal activity to help developing policing intervention and prevention strategies and tactics.*

With respect to the benefits of predictive policing, there are mismatches identified in the literature. There are many prospects described by predictive models, as it aims to reduce crime through more efficient and effective policing strategies. Nonetheless, actual evaluations of the usage of these models in practice lead to mixed results. Of the three existing studies that empirically tested whether geographic areas are better targeted with predictive software, only two show a positive correlation. In addition, a study that evaluated whether the profiling of potential victims and offenders of criminal activities neither showed a significant result. This implies that not all predictive policing models effectively reduce each form of crime and that that geospatial predicting and profiling are both very different variations of predictive policing. We conclude that the usage of every individual model should be thoroughly evaluated before any effectiveness-claims can be made.

The concerns surrounding predictive policing are mainly directed towards the lack of transparency of the predictive models. This has consequences for both the effectiveness and accountability of these models. If police officers do not comprehend why the predictive algorithms derive certain outcomes or how their patrol routes are configured, they might not be aware of how they should respond in certain situations or how to act. This might hamper the effectiveness of the geospatial predictions of predictive software. Besides, when the predictive models are not transparent, police departments potentially cannot legitimize their decision-making anymore. There are also glimpses identified that the administration of predictive policing software in certain areas can lead to inequality between social groups. Lastly, the ethical question regarding the protection of privacy is brought up. When the profiling of individuals is gained a more prominent role in the practices of law enforcers, it is important to revise the rights of the citizens in relation to their digital and online privacy, as legislation and jurisprudence are often vague and unclear. Arguably, this could impair the relationship between citizens and the government because of unclear civil rights. However, there is no empirical evidence to strengthen any of these assumptions. Hence, academics should further elaborate how the predictive models work out in practice and whether we actually see that the drawbacks of lack of transparency and stigmatizing of individuals and groups actually occurs.

In sum, this study has provided an overview of what predictive policing is and what the claimed benefits and drawbacks are. At the same time, the overview highlights that there is a need for a stronger empirical assessment of these approaches to understand the relation between features of the approaches and success in reducing certain forms of crime. When there is more evidence available to back-up the claimed benefits and drawbacks of predictive policing, it can be objectively determined how effective predictive policing methods are and how they can contribute to the traditional policing methods. Therefore, scholars are urged to evaluate different predictive policing models to increase our understanding of what type of predictive methods seem fruitful and under which conditions. To evaluate the claimed drawbacks of lack of transparency or accountability, it should be studied how predictive models are used in practice. If strategies that are derived from predictive algorithms are not executed properly or valued by officers, this undermines their effectiveness. Furthermore,

**5.1 Application**

This project is proposed for the betterment of society. It can be beneficial to evaluate to what extent there is too much focus on correlations instead of casualties by law enforcement agencies. Finally, it will be worthwhile to investigate how these predictive policing models can reduce crime through prevention instead of the controlling of geospatial areas and individuals. It should be evaluated how these predictive models can be used to resolve underlying factors that lead to an increased risk of criminal activity.

**5.2 Future Scope**

The use of this software could drastically increase the speed of how crimes are dealt with in the country. The software runs through major data sets and thus provides prediction for probable suspects which will be highly effective for law enforcers thereby helping society make a safe place to live.