A Project report
On

CRIME PREDICTION USING MACHINE LEARNING

Submitted in partial fulfillment of the requirement for the award of degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE & ENGINEERING

Submitted by

K.Koushik	**(16311A0525)**
Y.Gopi	**(16311A0554)**
B.Sai Harshank	**(16311A0560)**

Under the Guidance of
Dr. V. Satheesh Kumar
Professor
Department of CSE

Department of Computer Science & Engineering
SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY
(AUTONOMUS)
Yamnampet (V), Ghatkesar (M), Hyderabad – 501301, A.P.

2019-2020

Department of Computer Science & Engineering
SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY
(AUTONOMUS)
Yamnampet (V), Ghatkesar (M), Hyderabad – 501301, A.P.
2019-2020

CERTIFICATE

This is to certify that the project entitled "**CRIME PREDICTION USING MACHINE LEARNING**" is being submitted by, **K.KOUSHIK(16311A0525), Y.GOPI (16311A0554), B.SAI HARSHANK (16311A0560)** in partial fulfilment of the requirements for the award of **BACHELOR OF TECHNOLOGY** to **SNIST, Hyderabad**. This record is a Bonafide work carried out by them under my guidance and supervision. The result embodied in this project report has not been submitted to any other university or institute for the award of any degree of diploma.

Internal guide
Dr. V. Satheesh Kumar
Professor
Department of CSE

Project Co-ordinator
Mrs. P. Vasavi
Associate Professor
Department of CSE

Head of Department
Dr. Aruna Varanasi
Professor & HOD
Department of CSE

External Examiner

Date:

DECLARATION

We, **KASHETTY KOUSHIK (16311A0525), YADAPALLI GOPI (16311A0554), BACHU SAI HARSHANK (16311A0560),** students of **SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY, YAMNAMPET, GHATKESAR,** studying 4th year 2nd semester, **COMPUTER SCIENCE AND ENGINEERING** solemnly declare that the project work, titled **"CRIME PREDICTION USING MACHINE LEARNING"** is submitted to **SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY** for partial fulfillment for the award of degree of Bachelor of technology in **COMPUTER SCIENCE AND ENGINEERING**.

It is declared to the best of our knowledge that the work reported does not form part of any dissertation submitted to any other University or Institute for award of any degree.

ACKNOWLEDGEMENT

We would like to express our gratitude to all the people behind the screen who helped us to transform an idea into a real application.

We would like to express our heart-felt gratitude to our parents without whom we would not have been privileged to achieve and fulfill our dreams. We are grateful to our principal, **Dr. T. Ch. Siva Reddy,** who most ably run the institution and has had the major hand in enabling us to do my project.

We profoundly thank Dr**. ARUNA VARANASI**, Head of the Department of Computer Science & Engineering who has been an excellent guide and also a great source of inspiration to our work.
We would like to thank our project co-ordinator **Mrs. P. Vasavi**, for her guidance and encouragement.
We would like to thank our internal guide **Dr.V.Satheesh Kumar** for his technical guidance, constant encouragement and support in carrying out our project at college.
The satisfaction and euphoria that accompany the successful completion of the task would be great but incomplete without the mention of the people who made it possible with their constant guidance and encouragement crowns all the efforts with success. In this context, We would like thank all the other staff members, both teaching and non-teaching, who have extended their timely help and eased our task.

KASHETTY KOUSHIK (16311A0525)
YADAPALLI GOPI (16311A0554)
BACHU SAI HARSHANK (16311A0560)

ABSTRACT

crime detection and prevention is a standardized method for the identification and study of crime patter ns and trends. Our program can predict regions that are highly likely to occur in crime, and can imagin e areas vulnerable to crime. Crime data analysts will help law enforcement officers speed up the crime resolution process with the growing advent of computerized systems.using the machine learning princi ple we can derive valuable, previously unknown knowledge from unstructured data. Here we have an a pproach between computer science and criminal justice to build a method of machine learning which c an help to solve crimes more quickly. Rather than dwelling on the causes of crime such as criminal rec ord, political enmity, etc.

The challenge at the outset is to predict which crime category is most likely to occur, provided a time and place. We picked up various types of features that are seriously affecting the area of high crime, ba sed on different indicators using different algorithms.

INDEX

Introduction

We use python, open source software in this analysis to perform a comparative study between the com munities' violent crime trends and the kaggle generated crime dataset. We must implement the algorith ms Linear Regression, Multiple Linear Regression, Lasso Regression, Ridge Regression, SVR, Tree Regression Decision, Elastic Net Regression, and Random Forest Regression to solve the problem

Purpose of the Systems

Crime analysis involves exploring data about crimes provide necessity information and prevent crimes. The manual techniques for predicting crime takes more time and waste of time. Mining technologies, the use of information data and the growing capacity of intelligence technology h ave expanded the collection, storage and exploitation of data and are being used with the goal of discov ering similar patterns. Machine learning assists in predicting the crime using machine using certain algorithms.

It can therefore be described as "Data Mining is the process of finding new patterns from large data. Crime is an unhygienic, complex condition arising from social, cultural, and environmental influences. Crimes are the collective nuisance, and in many respects cost our society tremendously. Crime criteria and criminality matching are the two key elements that are typically used in the process of crime analysis

Those are the ones important for the cycle of investigating crime. Historically reported crime or unreso lved events are classified in the process of linking crime to historical data collection or crimes. The pri mary concern of Crime Analysis is a systematic approach to detecting, finding and predicting crime tre nds. The input of a crime analysis program consists of information and data obtained from the local Th e vast amount of crimerelated data in police forces and also the nature of the relationship between thes e types of data has made the conventional approach of investigating crime unreliable. Current approach es require much documentation and human effort to get data trends. Data mining turns the collected dat a into usable information for achieving the desired result.

After conversion of the data into useful information, various data mining techniques are applied to obta in data patterns

Existing System

So, many projects done on crime prediction for predicting the crime With the aid of machine learning, we can predict crime as a core of Python But, Main feature is the building the model. Model is the tells about the accuracy of the crime So many algorithms are used to construct the models.

Based on the models accuracy of the models will be dependent .Some of the models and their accuracy given below. Prediction of crime using k-means and their accuracy is 79%.Prediction of crime using deep learning and their accuracy is 83% .Prediction of crime using some algorithms and their accuracy is 85%.Based on models crime prediction accuracy is increased so, Constructing the best models with maximum accuracy.

Proposed System

Predictive modeling is the way to construct a computer that can make predictions.

The method involves a machine learning algorithm that learns those properties to make certain predictions from a training dataset.

Predictive modeling can be divided further into two areas:

1. Regression

2. pattern classification.

Models of regression are used to analyze the relation between data set variables

and the models are used to clean the data, if any missing values are in the dataset then these

models are used to make the clean of the particular of the values.

Construing the models using the training data. while constructing the models we specified two things time and place what time exactly crime occurred and where the mostly crime occurred these two things mainly necessity reduing the crime rate

SYSTEM REQUIREMENTS

The specifications include certain criteria about software and hardware.They are as follows:

Software Requirements:

- Programming language : Python
- Excel : For datasets(kaggle.com)
- Spyder: It is an integrated development environment(IDE) which is open source cross-platform for scientific programming in the python language .

Hardware Requirements:

The hardware requirements include the system specifications on what it runs. They include Processor and RAM.

- RAM : 4GB
- Processor : 3.3 GHz
- Internet connection

2.3. Modules:

1. Spyder

Spyder is an open source crossplatform integrated development ENVIRONMENT (IDE) for pytho nlanguage science programming. Spyder incorporates into the computational Python stack a range of p opular packages, including Numpy, SciPy, Matplotlib, Pandas, IPython, SymPy, Cython and other ope n source applicationsSpyder imports all the functions that are needed to perform on the dataset.
The spyder includes predefined methods, using the templates that we use for your project.

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Numpy and Matplotlib are the two libraries are used in the project.

2.Multiple Linear Regression:

Linear regression is the most common and easiest statistical methodology for predictive modelling. It essentially gives us an equation where we have our attributes as independent variables depending on our target variable [in our case sales].

And what does it look like? It looks like the linear regression equation:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here, we have Y as our dependent variable (Sales), X's are the independent variables and all thetas are the coefficients. Coefficients are basically the weights assigned to the features, based on their

3.Decison Tree Regression:

Decision tree is a type of supervised learning algorithm (which has a predefined target variable) often used in classification problems. It functions for the input and output variables both categorical and continuous.

In this technique, we divide the population or sample into two or more homogeneous sets (or sub-populations) based on the most important input variables splitter / differentiator.

4. Random Forest Regression:

Random forest is like an algorithm for bootstrapping with a Decision tree (CART) layout. Think, in th e whole population, we have 1000 observations, with 10 variables. Random Forest attempts to create s everal CART models with various samples and initial variables.

For example, to construct a CART model, it will take a random sample of 100 observations and 5 initial variables chosen at random.It will repeat the cycle (say) 10 times, and then make a final forecast for each statement. Final prediction is a feature of any forecast. This final prediction can be clearly the mean of any prediction

5.Ridge Regression:

Our concept to date was to essentially minimize the cost function, so that expected values are far closer to the desired outcome.

Now take a look back at ridge regression cost feature again.

$$\min\left(\left|\left|Y - X(\theta)\right|\right|_2^2 + \lambda||\theta||_2^2\right)$$

Here if you remember, we consider an extra term, called the penalty word. The π given here is actually denoted in the ridge function as an alpha parameter. So, we are essentially regulating the penalty word by changing the alpha values. The higher the alpha values, the greater is the penalty and thus the magn itude of the coefficients is decreased.

6. Lasso Regression:

Mathematics behind lasso regression is secretly identical to that of ridge differential only because inst ead of adding theta squares, we can add absolute meaning Θ.

$$\min\left(\left|\left|Y - X\theta\right|\right|_2^2 + \lambda||\theta||_1\right)$$

Here too, λ is the hypermeter, whose value is equal to the alpha in the Lasso function.

Elastic Net:

Elastic net is essentially a variation of the regularization of both L1 and L2. So if you know elastic net, the parameters can be tuned to enforce both Ridge and Lasso. So it uses both the penalty term L1 and the penalty term L2 and the equation looks like:

$$\min\left(\left|\left|Y - X\theta\right|\right|_2^2 + \lambda_1||\theta||_1 + \lambda_2||\theta||_2^2\right)$$

6. Gaussian Naive Bayes :

Gaussian Naive Bayes is a supervised classifier using the naive assumption that between two features t here is no dependency. Applying Bayesian Theorem implements this classifier. According to the theorem, class y and a vector of the dependent function consisting of x, .., has the following relation:

3.SYSTEM ANALYSIS

This Machine Design is closely correlated with the study of requirements. It is also "an specific system atic investigation undertaken to help someone (referred to as the decisionmaker) determine a better course of action and make a better decision than he would otherwise have made." This process includes breaking down the system into various pieces to evaluate the situation and examine project priorities, breaking down the structure.is what required to be created and attempting to engage users so that requirements can be defined.

Feasibility Study:

The System after careful analysis has been identified to be present with the following functional specifications.

Decision Tree :

Spyder module offers class for Decision Tree Classifier. In this case, two parameters are useful among other parameters of this classsplit samples show numberof splits to make at each stage of building a decision tree, and criterion indicates the function to calculate split quality. There are two types of function to calculate the quality of the split, as discussed above: information gain and impurity. Entropy is indicated in this class for knowledge gain, and gini for impurity.

Mulitple Linear Regression:

Linear regression is the most common and easiest statistical methodology for predictive modelling. It essentially gives us an equation where we have our attributes as independent variables depending on our target variable [in our case sales].

So what does the equation look like? Linear regression equation looks like this:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here we have Y as our dependent variable (Sales), X's are the independent variables and the coefficients areall Θ. Essentially, coefficients are the weights assigned to the functions, based on their equations.

Software Environment:

Reliability is calculated by application with the assistance of the data received. The definition of specifications plays an significant part in the design of a program. It will only display when properly defined requirement requirements, a device can be built that will fit into the necessary environment.It depends largely on the current system users who offer the specifics of the requirements as they are the people who use the system. It is because during the initial stages the specifications will be identified so that the framework can be set up according to those specification.Solving the problem once built is ver y difficult and, on the other hand, designing a system that does not fulfill the user's requirements is of no benefit.

The requirement specification for any system can be broadly stated as given below:

- The system should be able to interface with the existing system
- The system should be accurate

Selecting the right features for your model:

As we have a collection of high dimensional data, Usage of all the variables will be extremely inefficient,as some of them might provide redundant details. We will need to pick the right set of varia bles thatgive us a precise model and can describe the dependent variable well. Selecting the correct set ofvariables for the model has several forms.Company awareness and domain expertise will be amongst them first. For example, when forecasting sales we know marketing activities can have a positive effect on sales and are an essential feature of your model.

We should also be careful not to compare the variables we're choosing with one another.Instead of ma nually selecting the variables, using forward or backward selection, we can automate this process. In th e model, forward selection starts with the most important predictor and adds variable for every move. Backward elimination starts with all of the model's predictors and eliminates the least significant varia ble per step. Criteria can be chosen according to any statistical factor such as R-square, t-stat etc.

2.SYSTEM DESIGN

Architecture Design:

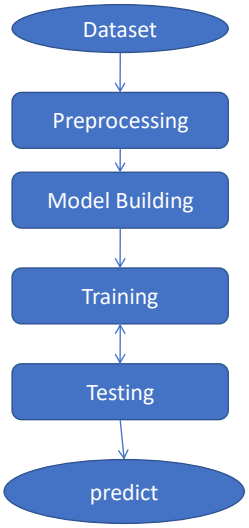


Fig 4.1 Architectural Design

4.2 UMLConcepts

UML stands for Unified Modeling Language which is used to visualize ,specify ,construct and document during software design of a project.

- Visualizing
- Specifying
- Constructing
- Documenting

4.2.1 Building blocks of UML:

UML have three building blocks. They are

- Things
- Relationships
- Diagrams

- Things are real world objects.
- Relationships will say the connection between these objects;
- Diagrams will describe the system as a whole.

Things:

There are four types of things:

- Structural things
- Behavioral things
- Grouping things
- An notational things

Structural things are also called as the physical parts of UML.

Different types of things are:

- Class
- Use case

- Node

Class is a collection of set of objects.“They are build based on OOPs concepts.

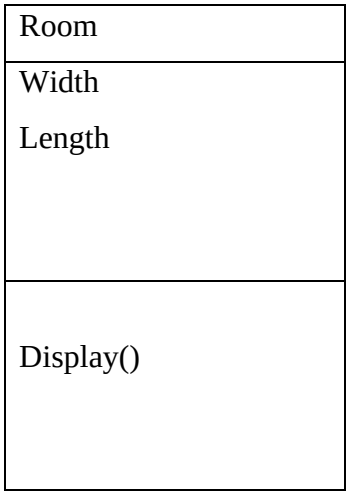


Fig: Example of Room class

Use case is a set of actions which are performed by a particular user.

Fig: Example use case

Node is a physical element which is present during runtime,it may even occupy some memory and can have processing capability. It is represented by a square box.

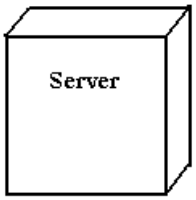


Fig: Node

Behavioral things are also called as the dynamic parts of UML models.

Interaction:

An interaction is a behavior that consits of a collection of messages that interchangedbetween a set of objects. It is the connection between multiple objects. This is how they communicate with other objects.

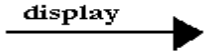


Fig: Messages

4.2.2.Relationships in the UML:

There are four types of relationships, they are:

- Dependency
- Association
- Generalization
- Realization

A **dependency** is a kind of relationship between two things where one thing is dependent on other thing.

Fig: Dependencies

An **association** is a simple connection or between different objects. Aggregation is a special kind of association, aggregation is used when there is a whole to part relationship.

Fig: Association Representation

A **generalization** is a generalization relationship in which objects are mapped to generalized elements or objects.

Fig: Generalization Representation

A **realization** is a relationship between classifiers, where there will contract system. One will give

contract and other will do it".

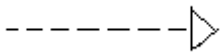


Fig: Realization

UML Diagrams

UML Diagrams for our application are as follows:

Use Case Diagrams

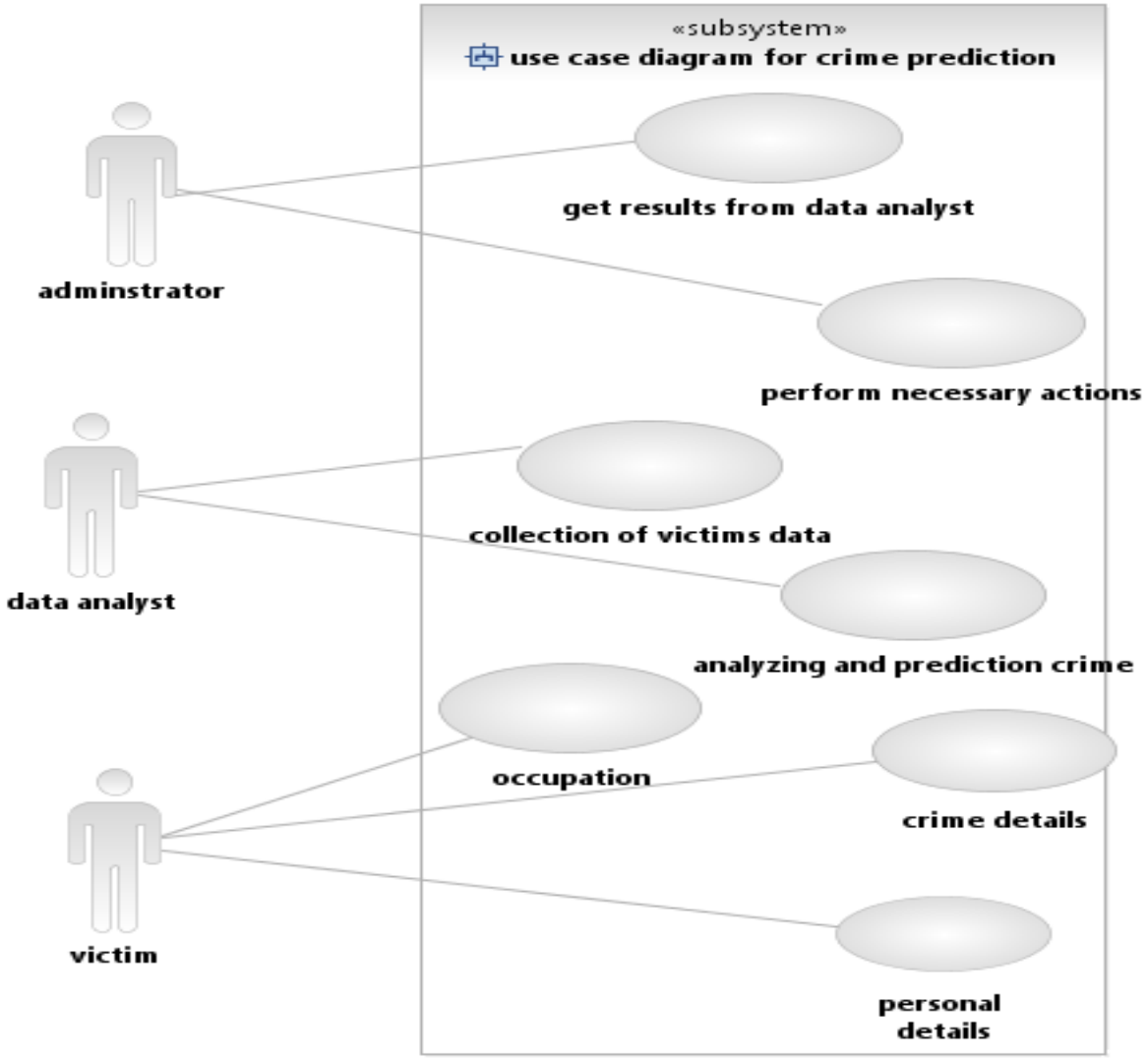


Fig 4.2.1.Usecase Diagram

The use case diagram to depicts the functionality of users

- Victim
- Data Analyst
- Adminstrator

Victim enter their details like occupation, crime details and personal details

Data Analyst take the victim data and prediction the crime results.

Adminstrator maintaining the dataset and taking necessary actions.

Class Diagram:

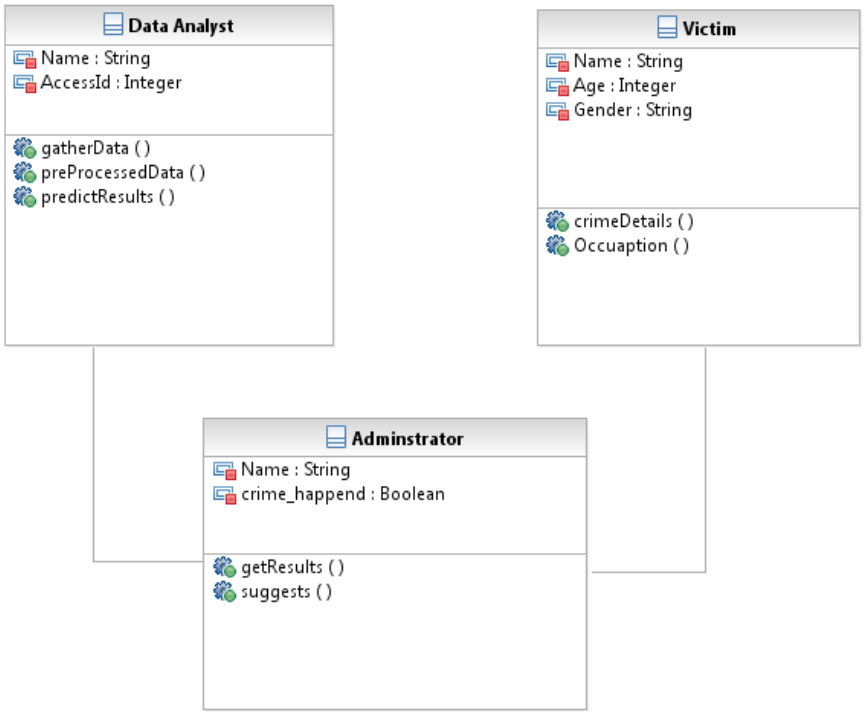


Fig 4.2.2 Class Diagram for Crime Prediction using Machine Learning

Here victim class contain some attributes and methods:

Attributes:

Name of the victim
Age of the victim
Gender of the victim

Methods:

Crime details()
Occupation();

Data Analyst contain some attributes and methods:

Attributes:

Name of the Data Analyst
Access Id

Methods:

gatherData()

preprocessedData()

predictResults()

Adminstrator contain some attributes and methods:

Attributes:

Name
Crime_happend

Methods:

Get_results()
Suggets()

The data set given to the Data Analyst and using the preprocessedData() method starts Clean the data and using the predictResults() method getting the output. Adminstrator Maintains the results.

4.2.1. Object Diagram

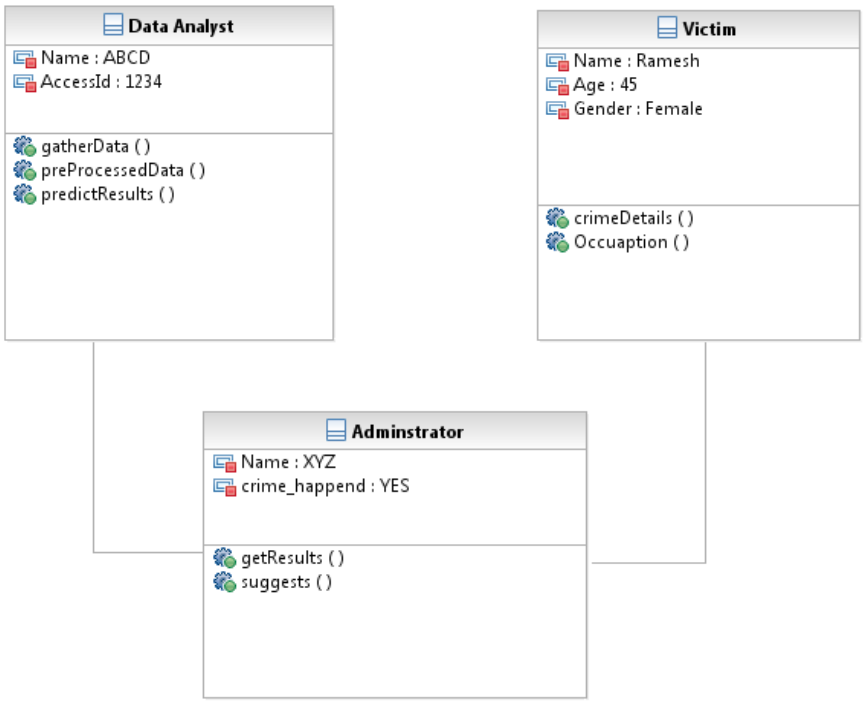


Fig 4.2.3 Object Diagram for Crime prediction

Victim contains Data :

Victim name: Ramesh
Victim age: 45
Victim gender: Male

Data Analyst data:

Name: ABCD
Access Id: 1234

Adminstrator Data:

Name: XYZ
Crime_happend: YES

Here object diagram provides data to all objects according to the attributes.

Interaction Diagram:

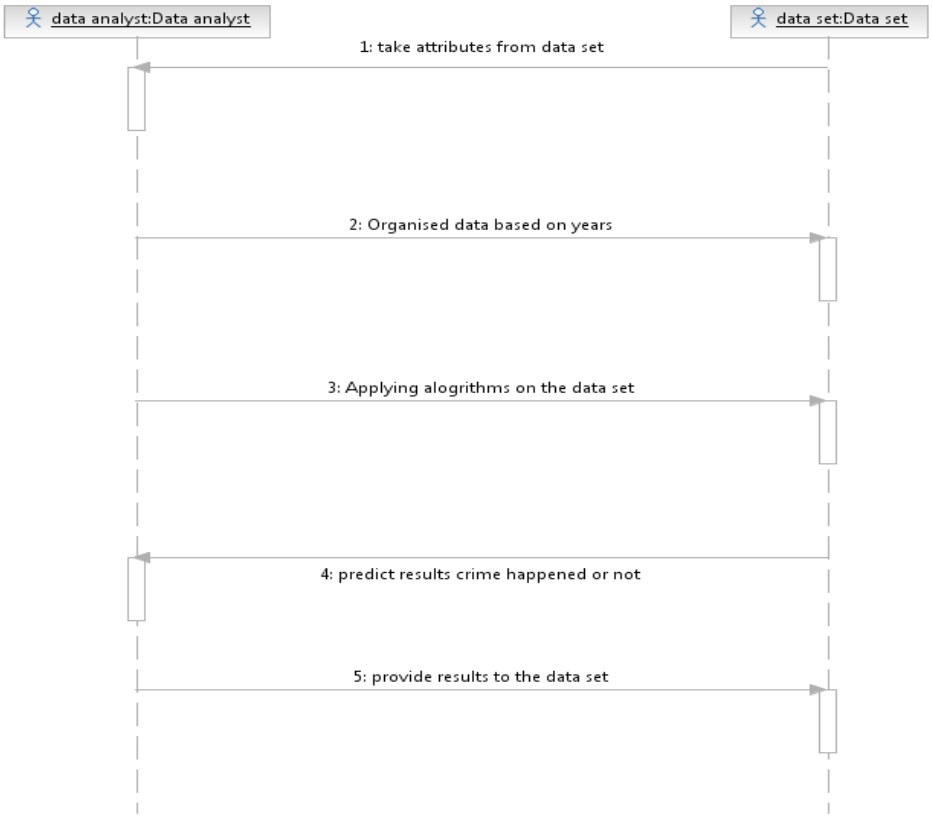


Fig 4.2.4 Activity Diagram

Activity diagram depicts the flow of the events in the whole system as shown above.

5. Coding

The implementation stage of any project is a true display of the defining moments that make a project a success or a failure. The implementation stage is defined as the system or system modifications should be installed and made operational in a production environment. The phase is initiated after the system has been tested and accepted by the user. This phase continues till the system is operating in production in accordance with the defined userrequirements.

PROGRAM

Cp.py:

```
# -*- coding: utf-8 -*-
"""
Created on 2020

@author: kaushik
"""

# -*- coding: utf-8 -*-



# Multiple Linear Regression

# Importing the libraries
import numpy as np
```

```
import matplotlib.pyplot as plt
import pandas as pd
from pandas import Series, DataFrame
# Importin  the dataset
dataset = pd.read_csv('rape.csv')
X = dataset.iloc[:, [0,1,2,3]].values
y = dataset.iloc[:,[4,5,6,7,8,9,10]].values

# Encoding categorical data
 from sklearn.preprocessing import LabelEncoder, OneHotEncode
 labelencoder_X_0 =  LabelEncoder()
X[:, 0] = labelencoder_X_0.fit_transform(X[:, 0])
labelencoder_X_1 = LabelEncoder()
X[:, 2] = labelencoder_X_1.fit_transform(X[:, 2])

onehotencoder = OneHotEncoder(categorical_features = [0])
X =  onehotencoder.fit_transform(X).toarray()

 onehotencoder = OneHotEncoder(categorical_features = [2])
 X = onehotencoder.fit_transform(X).toarray()

#taking care of missing data
'"from  sklearn.preprocessing  import  Imputer
 imputer= Imputer(missing_values='NaN',strategy='mean',axis=0)
imputer= imputer.fit(X[:,[0,2]])
```

```
X[:,[0,2]]=  imputer.transform(X[:,[0,2]])
'''
 #Avoiding the Dummy Variable Trap
X = X[:, 1:]

# Splitting the dataset into the Training set and Test set
from sklearn.cross_validation import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 0)

# Feature Scaling
"""from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
sc_y = StandardScaler()
y_train = sc_y.fit_transform(y_train)"""

# Fitting Multiple Linear Regression to the Training set
from sklearn.linear model impor  LinearRegression
regressor = LinearRegression()
 regressor.fit(X_train, y_train)

# Predicting the Test set results with multiple linear regression
y_pred_MultiplelinearRegression = regressor.predict(X_test)
```

```
#calculating mse for linear regression

mse_linearRegression = np.mean((y_pred_MultiplelinearRegression - y_test)**2)


#calculating r-square for multiple linear regression

regressor.score(X_test,y_test)









#fitting ridge regression

 from sklearn.linear_model import Ridge


## training the model


ridgeReg = Ridge(alpha=0.05, normalize=True)


ridgeReg.fit(X_train,y_train)


y_pred_RidgeRegression = ridgeReg.predict(X_test)


#calculating mse for Ridge linear regression

mse_ridgeRegression = np.mean((y_pred_RidgeRegression - y_test)**2)


#calculating r-square for ridge regression
```

```
ridgeReg.score(X_test,y_test)

#fitting lasso regression
 from sklearn.linear_model import Lasso

lassoReg =  Lasso(alpha=0.3, normalize=True)

 lassoReg.fit(X_train,y_train)

y_pred_LassoRegression = lassoReg.predict(X_test)
#calculating mse for Lasso regression
mse_lassoRegression = np.mean((y_pred_LassoRegression - y_test)**2)

#calculating r-square for lasso regression
lassoReg.score(X_test,y_test)

#fitting elastic net regression
from sklearn.linear_model import ElasticNet

 ENreg = ElasticNet(alpha=1, l1_ratio=0.5, normalize=False)

 ENreg.fit(X_train,y_train)

y_pred_ElasticNet = ENreg.predict(X_test)
#calculating mse for ElasticNet linear regression
```

```
mse_ElasticNet = np.mean((y_pred_ElasticNet - y_test)**2)

#calculating r-square for elastic net regression

ENreg.score(X_test,y_test)
```

```
  # Fitting Decision Tree Regression to the dataset

from sklearn.tree import DecisionTreeRegressor

regressor_tree = DecisionTreeRegressor(random_state = 0)

 regressor_tree.fit(X_train, y_train)

# Predicting a new result

y_pred_Dtree = regressor_tree.predict(X_test)

#calculating mse for DecisionTree regression

mse_Dtree = np.mean((y_pred_Dtree - y_test)**2)

#calculating r-square for decisiontree regression

regressor_tree.score(X_test,y_test)

#random forest

 # Fitting Random Forest Regression to the dataset

from sklearn.ensemble import RandomForestRegressor
```

```
 regressor_randomForest = RandomForestRegressor(n_estimators = 15, random_state = 0)
regressor_randomForest.fit(X_train, y_train)

# Predicting a new result
y_pred_randomForest = regressor_randomForest.predict(X_test)
#calculating mse for RandomForest regression
mse_randomForest = np.mean((y_pred_randomForest - y_test)**2)

#calculating r-square for random forest regression
regressor_randomForest.score(X_test,y_test)

#for andhrapradesh

total_crime_ap = dataset[dataset['Subgroup'] == 'Total Rape Victims']
ap = total_crime_ap[total_crime_ap['Area_Name'] == 'Andhra Pradesh']
ap.set_index('Year')[['Victims_Above_50_Yrs','Victims_Between_10-14_Yrs', 'Victims_Between_14-
18_Yrs','Victims_Between_18-30_Yrs','Victims_Between_30-
50_Yrs','Victims_Upto_10_Yrs']].plot(kind = 'line', figsize = (15,15))
plt.xlabel('Years')
plt.ylabel('No. of Cases in Andhra Pradesh')
plt.title('Crime against Women in Andhra Pradesh')
plt.show()

#for odisha
total_crime_ap = dataset[dataset['Subgroup'] == 'Total Rape Victims']
```

```
ap = total_crime_ap[total_crime_ap['Area_Name'] == 'Odisha']
ap.set_index('Year')[['Victims_Above_50_Yrs','Victims_Between_10-14_Yrs', 'Victims_Between_14-18_Yrs','Victims_Between_18-30_Yrs','Victims_Between_30-50_Yrs','Victims_Upto_10_Yrs']].plot(kind = 'line', figsize = (15,15))
plt.xlabel('Years')
plt.ylabel('No. of Cases in odisha')
plt.title('Crime against Women in odisha')
#for year 2010
data24 = dataset[(dataset['Subgroup'] == 'Total Rape Victims') & (dataset['Year'] == 2010)]
allstates24 = data24[['Victims_Above_50_Yrs','Victims_Between_10-14_Yrs', 'Victims_Between_14-18_Yrs','Victims_Between_18-30_Yrs','Victims_Between_30-50_Yrs','Victims_Upto_10_Yrs']].plot(kind = 'barh', figsize = (20,20), width = 1)
allstates24.set_xlabel('No. of Cases in 2010')
allstates24.set_yticklabels(data24['Area_Name'])
plt.show()
```

ABOUT.PY:

```
# This file  is dual licensed under the terms of the Apache License,  Version
# 2.0,  and the BSD License.  See  he LICENSE file in the root of this repository
# for complete details.
 from __future__ import absolute_import, division, print_function
```

```
__all__ = [
    "__title__", "__summary__", "__uri__", "__version__", "__author__",
    "__email__", "__license__", "__copyright__",
]


 __title__ = "packaging"
  __summary__ = "Core utilities for Python packages"
  __uri__ = "https://github.com/pypa/packaging"


  __version__ = "17.1"


  __author__ = "Donald Stufft and individual contributors"
__email__ = "donald@stufft.io"


  __license__ = "BSD or Apache License, Version 2.0"
  __copyright__ = "Copyright 2014-2016 %s" % __author__
```

VIEW:

Workspace.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<project version="4">
  <component name="ChangeListManager">
    <list default="true"  id="0cecd330-6077-4cfe-87d1-9c2a35a3e3fa" name="Default Changelist"
comment="" />
    <option name="EXCLUDED_CONVERTED_TO_IGNORED"  value="true" />
    < option name ="SHOW_DIALOG"  value ="false" />
    < option name ="HIGHLIGHT_CONFLICTS"  value ="true" />
    < option name ="HIGHLIGHT_NON_ACTIVE_CHANGELIST"  value ="false" />
    < option name ="LAST_RESOLUTION"  value ="IGNORE" />
  </component>
  <component name="FUSProjectUsageTrigger">
    <session id="-1129816163">
      <usages-collectori d="statistics.lifecycle.project">
        < counts>
          <entry key="project.closed"  value="2" />
          < entry key ="project.open.time.17"  value ="1" />
          < entry key ="project.open.time.6"  value ="1" />
          < entry key ="project.opened"  value ="2" />
        </counts>
      </usages-collector>
        <usages-collector id="statistics.file.extensions.open">
          <counts>
           <entry key="py"  value="2" />
          </counts>
        </usages-collector>
        <usages-collector id="statistics.file.types.open">
```

```
      <counts>
        <entry key="Python"  value ="1" />
        < entry key ="Scratch  value ="1" />
      </ counts  >
    </usages-collector >
    <usages-collector id="statistics.file.extensions.edit">
      <counts>
        < entry key ="py"  value ="103" />
      </ counts  >
    </usages-collector >
    <usages-collector id="statistics.file.types.edit">
      <counts >
        < entry key ="Python" value ="103" />
      </counts >
    </usages-collector>
  </session>
 </component>
 <component name="FileEditorManager">
  <leaf SIDE_TABS_SIZE_LIMIT_KEY="300">
    <file pinned="false" current-in-tab="true">
      <entry file="file://$PROJECT_DIR$/myprogram.py">
        <provider selected="true" editor-type-id="text-editor">
          <state relative-caret-position="76">
            <caret line="28" column="18" selection-start-line="28" selection-start-column="18"
selection-end-line="28" selection-end-column="18" />
            <folding>
              <element signature="e#244#262#0" expanded="true" />
              </folding>
            </state>
          </provider>
        </entry>
      </file>
    </leaf>
```

```
</component>
<componentiname="FileTemplateManagerImpl">
  <option name="RECENT_TEMPLATES">
    <list>
      <option value="Python Script" />
     </list>
    </option>
  </component>
  <component name="FindInProjectRecents">
    <findStrings>
      <find>categorical_features</find>
     <find>cross_validation</find>
    </findStrings>
  </component>
<component name="IdeDocumentHistory">
  <option name="CHANGED_PATHS">
    <list>
      <option value="$PROJECT_DIR$/myprogram.py" />
    </list>
  </option>
</component>
<component name="ProjectFrameBounds" extendedState="6">
  <option name="x" value="-10" />
  <option name="y"  value="-10" />
  <option name="width"  value="979" />
  <option name="height"  value="1049" />
</component>
<component name="ProjectView">
  <navigator proportions="" version="1">
    <foldersAlwaysOnTop value="true" />
  </navigator>
  <panes>
    <pane id="Scope" />
```

```
    <pane id="ProjectPane">
      <subPane>
        <expand>
          <path>
            <item name="koushik"  type="b2602c69:ProjectViewProjectNode" />
            <item name="koushik"  type="462c0819:PsiDirectoryNode" />
          </path>
        </expand>
        <select />
      </subPane>
    </pane>
  </panes>
 </component >
 <component name="PropertiesComponent">
  <property name="com.intellij.ide.scratch.LRUPopupBuilder$1/New Scratch File" value="Python"
/>
  <property name="settings.editor.selected.configurable"
 value="com.jetbrains.python.configuration.PyActiveSdkModuleConfigurable" />
 </component>
 <component name="RunDashboard">
  <option name="ruleStates">
    <list>
      <RuleState>
        <option name="name" value="ConfigurationTypeDashboardGroupingRule" />
      </RuleState >
      <RuleState >
        <option name="name" value="StatusDashboardGroupingRule" />
      </RuleState >
    </list >
  </option>
 </component>
 <component name="RunManager">
```

```
    <configuration name="myprogram" type="PythonConfigurationType" factoryName="Python"
temporary="true">
      <module name="kodi_proj" />
      <option name="INTERPRETER_OPTIONS" value="" />
      <option name="PARENT_ENVS" value="true" />
      <envs>
        <env name="PYTHONUNBUFFERED" value="1" />
      </envs>
      <option name="SDK_HOME" value="" />
      <option name="WORKING_DIRECTORY"  value="$PROJECT_DIR$" />
      <option name="IS_MODULE_SDK"  value="true" />
      <option name="ADD_CONTENT_ROOTS"  value="true" />
      <option name="ADD_SOURCE_ROOTS"  value="true" />
      <option name="SCRIPT_NAME"  ="$PROJECT_DIR$/myprogram.py" />
      <option name="PARAMETERS" value="" />
      <option name="SHOW_COMMAND_LINE"  value="false" />
      <option name="EMULATE_TERMINAL"  value="false" />
      <option name="MODULE_MODE"  value="false" />
      <option name="REDIRECT_INPUT" value="false" />
      <option name="INPUT_FILE"  value="" />
      <method v="2" />
    </configuration>
    <recent temporary>
      <list>
        <item itemvalue="Python.myprogram" />
      </list>
    </recenttemporary>
  </component>
  <component name="SvnConfiguration">
    <configuration  />
  </component >
  <component name="TaskManager">
    <task active="true" id="Default" summary="Default task">
```

```
    <changelist id="0cecd330-6077-4cfe-87d1-9c2a35a3e3fa" name="Default Changelist"
comment="" />
    <created>1587475126856</created>
    <option name="number"  value="Default" />
    <option name="presentableId"  value="Default" />
    <updated>1587475126856</updated>
   </task>
   <servers />
  </component>
  <component name="ToolWindowManager">
   <frame x="-7" y="-7" width="1550" height="838" extended-state="6" />
   <layout>
    <window info content_ui="combo"  id="Project" order="0"  visible="true" weight="0.24983476"
/>
    <window info id="Structure"  order="1" side tool="true"  weight="0.25" />
    <window_inf  id="Favorites"  order="2" side tool="true" />
    <window info   anchor="bottom"  id="Message" order="0" />
    <window info   anchor="bottom"  id="Find" order="1" />
    <window info anchor="bottom"  id="Run"  order="2"  weight="0.3286119" />
    <window_info  anchor="bottom"  id="Debug"  order="3"  weight="0.4" />
    <window info  anchor="bottom"  id="Cvs"  order="4"  weight="0.25" />
    <window info  anchor="bottom"  id="Inspection"  order="5"  weight="0.4" />
    <window info anchor="bottom"   id="TODO"  order="6" />
    <window_info anchor="bottom"  id="Version Control"  order="7" show_stripe_button="false" />
    <window info anchor="bottom"  id="Terminal"  order="8" />
    <window info anchor="bottom"  id="Event Log"  order="9" side tool="true" />
    <window info anchor="bottom"  id="Python Console"  order="10" />
    <window info anchor="right  id="Commander" internal type="SLIDING"  order="0"
type="SLIDING"  weight="0.4" />
    <window info anchor="right" id="Ant Build"  order="1"  weight="0.25" />
    <window info anchor="right"  content_ui="combo"  id="Hierarchy"  order="2"  weight="0.25" />
   </layout>
  </component>
```

```
  <component name="VcsContentAnnotationSettings">
    <option name="myLimit"  value="2678400000" />
  </component>
  <component name="editorHistoryManager">
    <entry file="file://$APPLICATION_CONFIG_DIR$/scratches/scratch_2.py" />
    <entry file="file://$PROJECT_DIR$/myprogram.py">
      <provider selected="true"  editor-type-id="text-editor">
        <state relative-caret-position="76">
          <caret line="28"  olumn="18" selection-start-line="28"  selection-start-column="18" selection-
end-line="28" selection-end-column ="18" />
          <folding>
            <element signature="e#244#262#0"  expanded="true" />
          </folding >
        </state>
      </provider>
    </entry>
  </component>
```

Modules.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<project version="4">
  <component name="ProjectModuleManager">
    <modules>
      <module fileurl="file://$PROJECT_DIR$/.idea/kodi_proj.iml"
filepath="$PROJECT_DIR$/.idea/koushik.iml" />
    </modules>
  </component>
</project>
```

Msc.xml:

```
<?xml version="1.0"  encoding="UTF-8"?>
<project version="4">
  <component name="ProjectModuleManager">
    <modules>
      <module fileurl="file://$PROJECT_DIR$/.idea/kodi_proj.iml"
filepath="$PROJECT_DIR$/.idea/koushik_proj.iml" />
    </modules>
  </component>
</project>
```

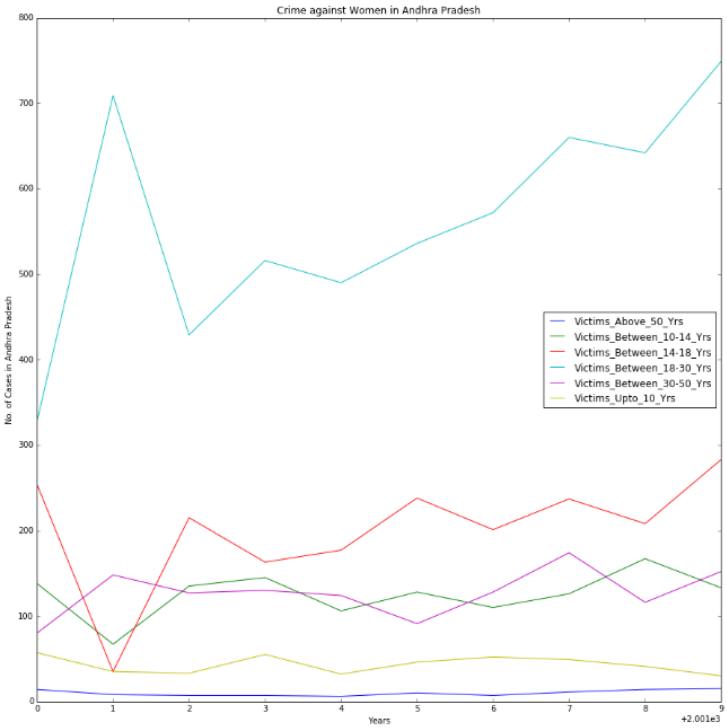
6. OUTPUTSCREENS

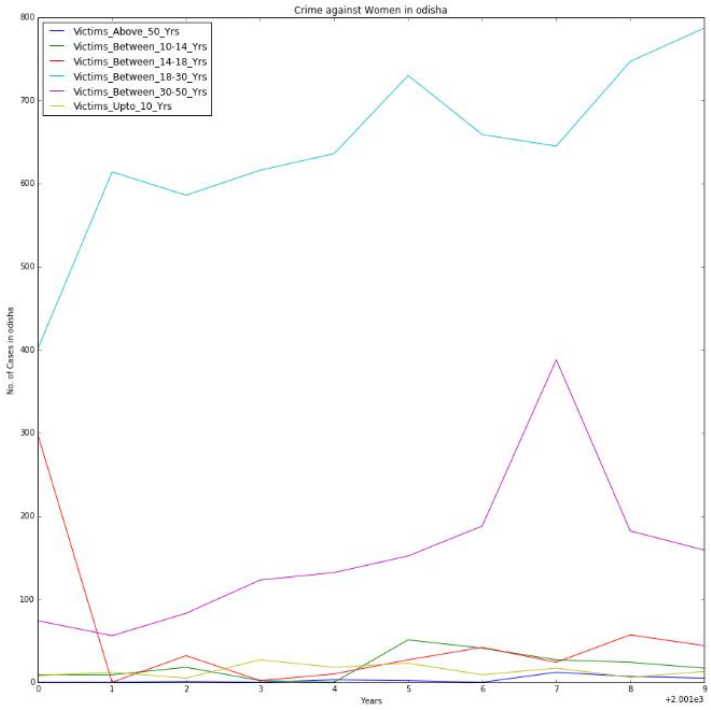
Output Screens of various functionalities in our application are shown over here along with the description.

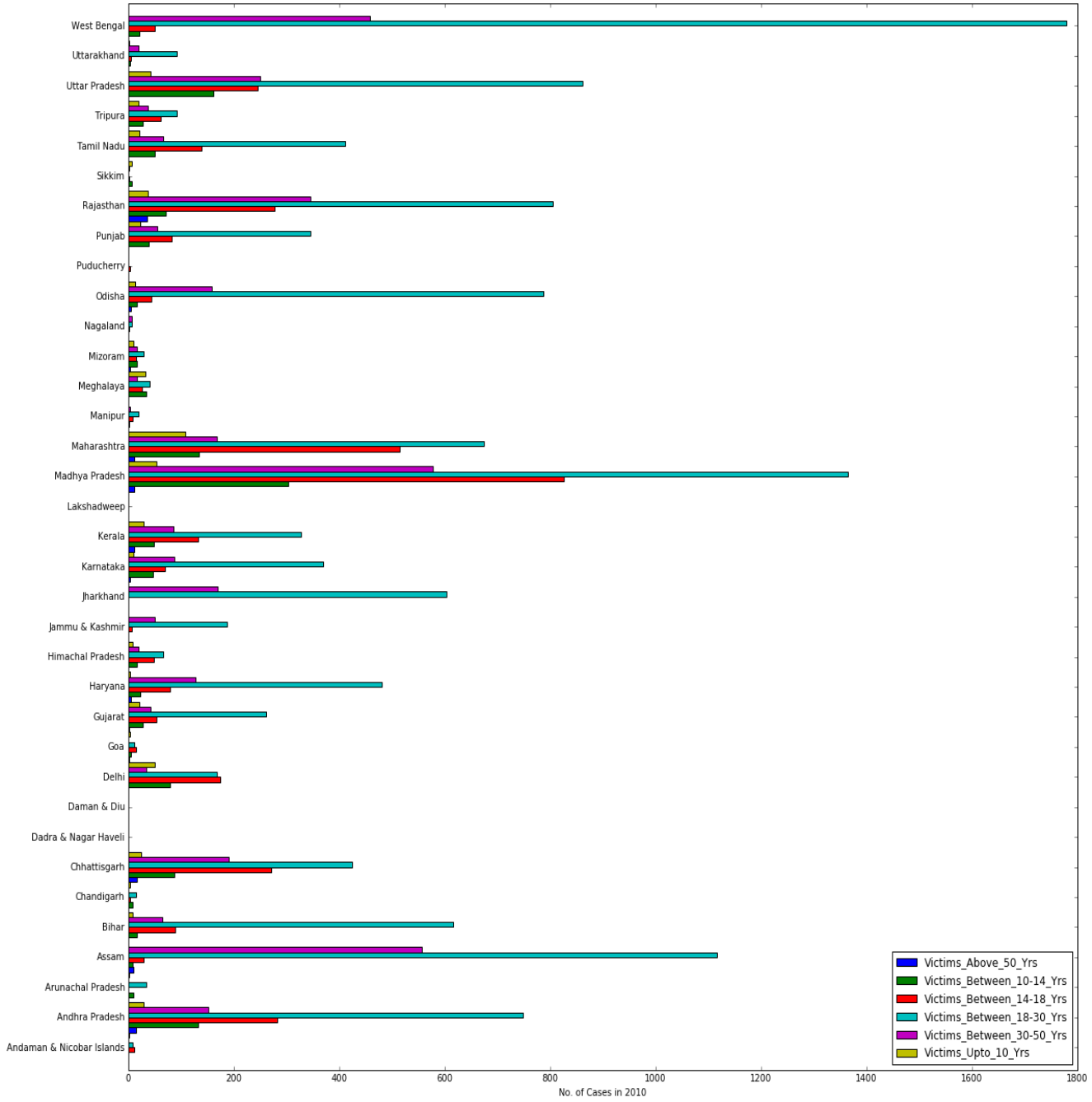
Initially the data was randomly distributed . We modified the data according to the sorted order of the years. Here is a snapshot of the sorted dataset**.**

	Area_Name	Year	Subgroup	Rape_Cases_Reported	Victims_Above_50_Yrs	Victims_Between_10-14_Yrs	Victims_Between_14-18_Yrs	Victims_Between_18-30_Yrs
2	Andaman & Nicobar Islands	2001	Total Rape Victims	3	0	0	3	0
3	Andaman & Nicobar Islands	2001	Victims of Incest Rape	1	0	0	1	0
4	Andaman & Nicobar Islands	2001	Victims of Other Rape	2	0	0	2	0
5	Andhra Pradesh	2001	Total Rape Victims	871	14	138	254	328
6	Andhra Pradesh	2001	Victims of Incest Rape	4	0	2	0	2
7	Andhra Pradesh	2001	Victims of Other Rape	867	14	136	254	326
8	Arunachal Pradesh	2001	Total Rape Victims	33	0	0	1	28

Later we segregated the input and output attributes and has split the data into training and test sets. The next step is preprocessing of the data. Later we applied the mentioned algorithms and found the accuracy with each of the algorithms.


Crime against Women in Andhra Pradesh
No. of Cases in Andhra Pradesh
Years
+2.001e3
Victims_Above_50_Yrs
Victims_Between_10-14_Yrs
Victims_Between_14-18_Yrs
Victims_Between_18-30_Yrs
Victims_Between_30-50_Yrs
Victims_Upto_10_Yrs


Crime against Women in odisha
Victims_Above_50_Yrs
Victims_Between_10-14_Yrs
Victims_Between_14-18_Yrs
Victims_Between_18-30_Yrs
Victims_Between_30-50_Yrs
Victims_Upto_10_Yrs
No. of Cases in odisha
Years
+2.001e3



Our main objective is to find out about and mitigate the mistake. But before that, let's talk about how t o tackle the first component, namely measuring the error. We already know that error represents the difference between the value we expect and the observed value. Let us consider only three ways by which error can be calculated:

Sum of residuals($\sum(Y - h(X))$) – can result in positive and negative errors being cancelled.

Sum of the absolute residual value($\sum|Y-h(X)|$) absolute value (will prevent error cancellation).

Sum of square of residuals ($\sum$ (Y-h(X))2) .

This is the method mostly used in practice since we penalize higher error values much more than a sma ller one, so that there is a significant difference between large errors and small errors, which makes it e asy to differentiate and and select the best fit line. Evaluation of the Sample R square and R-square changed How reliable do you think the sample is?

Do we have measurement metrics so we can check that? We have a number which is usually known as R-Square.

RSquare: Identifies how much of the overall Y variance (dependent variable) is explained by the X var iance (independent variable).Mathematically,it can be written as:

$$R - Square = 1 - \frac{\sum(Y_{actual} - Y_{predicted})\text{^}2}{\sum(Y_{actual} - Y_{mean})\text{^}2}$$

7. SYSTEM TESTING

Testing is the Error Detection process. Testing can play a very important role in maintaining quality co ntrol and assuring device reliability. The effects of the measurement performance are also used later during maintenance.

Psychology of Testing

The primary aim of testing is also to prove that a system works by ensuring it has no errors in it. The basic aim of phase testing is to identify the errors that may occur in the system.So the testing process should not be begun with the intention of showing that the software has no errors and the software works, but the testing person should have the intention of showing that a program is not work ig. Testing is the execution cycle of a program aimed at identifying errors in the output.

Testing Objectives

The main objective of testing is to uncover a host of errors, systematically and with minimum effort and time. Stating formally, we can say,

- Testing is a process of executing a program with the intent of finding anerror.
- A successful test is one that uncovers or not yet undiscoverederror.
- A good test case is one which has a high probability of finding error, if itexists.
- The tests are inadequate to detect possibly presenterrors
- The software more or less confirms to the quality and reliablestandards.

Types of testing the software:

The software testing types are based on the phases of the testing which are done in a phased manner. The testing is cut down into two types primarily:

- testing the software stability
- testing the software dynamically

Software stability:

We don't deal with the back end code running behind the program while checking whether the program is stable or not. Here we will only concentrate on the software's control flow and the specifications and behavior that the program needs to fulfill according to user needs.
Through this sort of research we only have to learn if we have a safe program for us. We might ignore t his phase in some of the software's development as it isn't too important.

We are actually reviewing the program specifications and their context entirely here and not concerned with checking specific code.

Testing the system dynamically:

This type of testing is done solely for testing the software's backend code to know if the software starts smoothly and does the things that are needed for the user. In any testing process the dynamic form of testing is very necessary and the tester should not neglect it. It also involves testing every functionality of the threads that are involved in the software.
Testing the functions in the software involves the following:

Testing each unit in software:

Here the code for the software system is broken down to small chunks of code called as software device. Each software unit is evaluated for conditions for initialization, loop tests, conditions for ending loops.

Every piece of code gets checked for its validity. Only those units of code that do useful work are to be submitted for further review and units that do not do any work are thrown out.
It is the first step of the testing of any program or software.

Testing while integrating code:

After all the code units have been successfully tested we need to incorporate these units into a specific action. It is very necessary that we do the testing of every unit first, and then go for the integration of code. While integrating the code, we need to take certain measures to avoid integrating non compatible code units that may result in program failure.

Testing the system as a whole:

By testing the system as a whole we will learn if any faults still occur between the interactions after the integration testing is finished. This testing is performed to ensure the software environment is as per th e design of the software.
It also includes checking the functions performed by each function in code and also tests if we have en ough resources to complete the test program

Test the software for acceptance:

A document is created similar to the srs doc to specify the requirements which the user needs. This type of document has all the measurements, user requirements, business needs. The tester will undergo the software search for all the details found in the text, and the acceptance test.

Testing for acceptance is required and necessary to know whether a product can be published and used or used without any complaints by a consumer for its needs.

There are two types of acceptance like internal and external. Internal test is performed by the product team and the external by the customers who give their feedback.

Test the software for stress and load:

When we give a lot of data as input to our system, we need to check whether our software can handle the large load of data at any particular time.

If we do not do this test then our software may crash and lead to losses for us. So it is important to know upto what amount of data our software can handle precisely.

Testing done by users:

After the software is released into the market and used by the customer, the feedback given by the customer who uses our software tool is really of a great help. This is because we can increase the product value of the software by inculcating the new features to be added to the software through the feedback of features recommended for the software.

It also helps to know the defects which are present through the customer complaints about the software which can be removed in the later versions of software as well.

So this type of testing is of a great help to not only the business but also for customer who can have his recommended features in the later releasing versions of software.

Goal of Testing:

We might make huge number of mistakes while building a software which would give rise to bugs, errors, crashes, shut downs in the system. The job of testing is to detect all these types of errors and eliminate as far as possible.

The following are some of the aims of testing:

- The software building costs and time reductions are a part of the testing goal. This is because if you can build a software which can do a lot of things for a user and building it lesser costs gives a great amount of profits.

- To remove any type of error, bug, shut-downs, crash in software. This makes the software to work properly in a timely manner.

- The customer satisfaction also is needed which can be done by testing the software and system to know if all the actions performed by it are as per customer's satisfaction.

- Build test cases in aoptimised way which would cover all the tests for software.

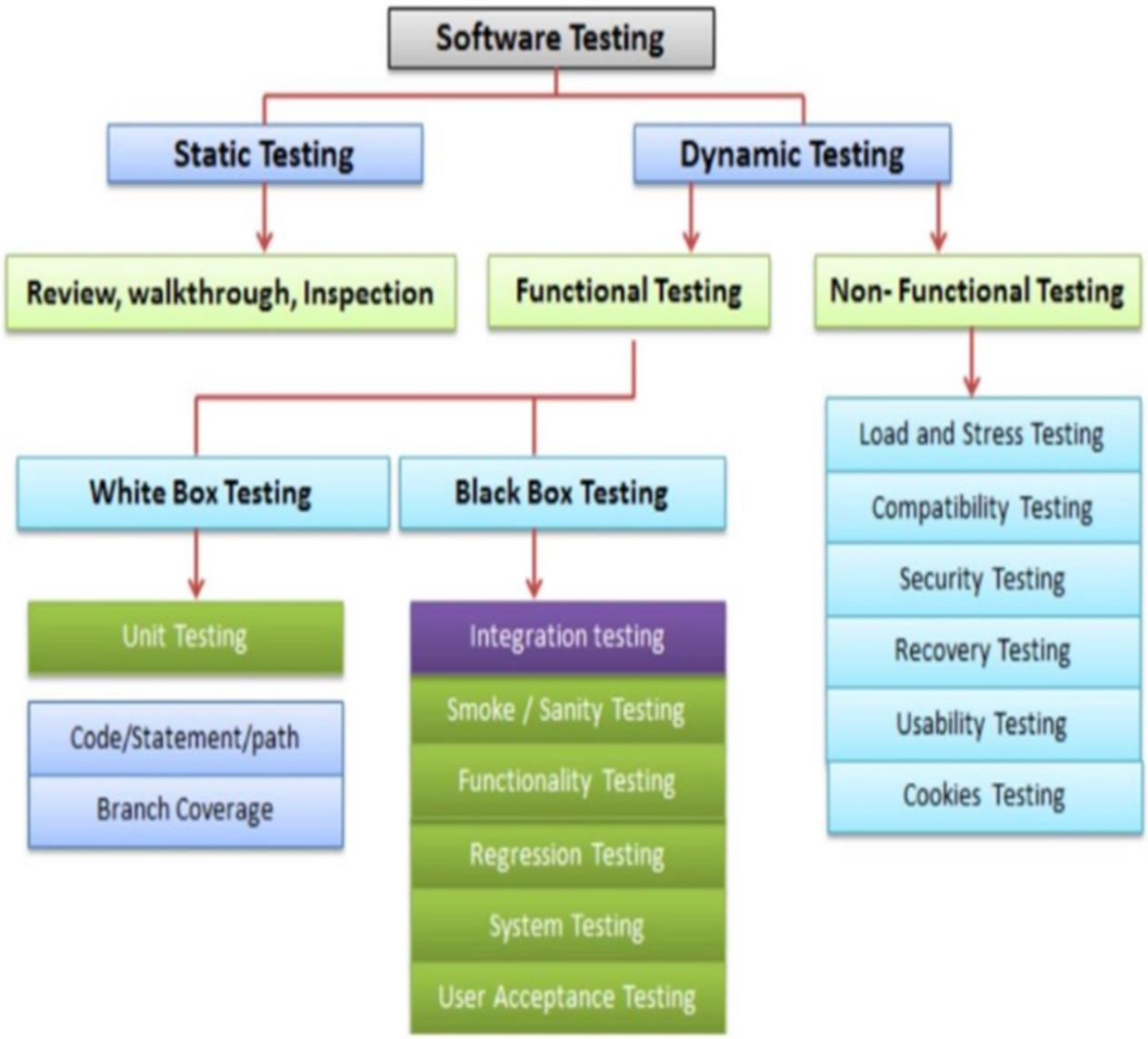
- The software quality standard need to be maintained while testing because it should be able to be released in market.

- It is insanely important that testing is done in right way on software by taking only those measures and testing tools which are fruitful to detect a error in software.

- We can build good test cases only if we can get some value from the test case to uncover an error from that particular test case.

- The primary goal of testing has always been to find the error in software and provide a method to remove them".

Types of Software Testing:



Code Testing:

This technique looks at the programme's reasoning. We created some test data to follow this method w hich resulted in the execution of Every program and module instruction i.e. every direction is checked Systems are neither planned as whole nor evaluated as single systems. Securing the coding is done correctly, two forms of checking are carried out or that the matter is carried out.

Unit Testing

Unit testingis a testing process that Focuses effort toward verificationon module likethe smallest device unit. Using the detailed design and the checking of process parameters is performed to verify all the errors within the boundary of that particular module. Until commencing integration checksprocess, all unit test errors must be accurate with the modules out.

Any service within this project can be regarded as a module. Each module has been tested to give speci fic set of inputs. When designing the module method as well as after the production is finished so that every module works without any mistake.The inputs are checked while the user accepts and takes the i nputs. The software units in a system are called as modules and the routines that are assembled and inc orporated to form a particular role in this application developer testing.

Unit testing is performed on modules first, to find the errors very quickly, independently of one another. This enables error detection. Initially prevented through this errors resulting in process interac tion between modules.

Unit testingis a testing process that Focuses on module verification effort including smallest computer unit.Using the detailed design and the checking of process parameters is performed to verify all the errors within the boundary of that particular module. Before starting the integration testing process, all modules must be successful with all unit test errors out.

Link Testing

Link testing does not deal with software testing but is about incorporating each module into the program. The key preoccupation of this test is the reliability of each element. The Programmer will check the modules that have different parameters, length, size, etc.

Integration Testing

We have to perform integration testing after the unit testing phase. This integration testingisto's main objective is to see if modules can be properly integrated, with the focus on testing interfaces between modules. This research activity can be regarded as research the design, and thus emphasizing the inter actions of the test module.

Integrating all the modules will form the main structure in this project. When integrating all the modules, you need to test if the integration effects work on any of the services by giving different type of combinations to the inputs with which the two services will operate perfectly before the Integration process.

System Testing

Here they will check the whole operating program. The reference document for this process will act as the specifications document for the entire process, and the operating system's purpose is to see whet her or not software meets its specifications.

Acceptance Testing

After the integration testing we have to undergo the acceptance test the main objective of the acceptance test is to conduct with practical customer data to prove that the customer program is working satisfactorily. Testing here focuses on the system's external behaviour; no attention is placed on the program's internal logic.

The test cases should be chosen in such a way that the maximum number of equivalence class attributes are exercised at once. The testing process is an important part of designing the program. This is the method of identifying errors and incomplete operations and it is also a full check to decide whether the goals are achieved and if the user specifications are achieved or not.

White Box Testing

White box testing is a unit testing approach where a sample can be taken from the whole code at a time and carefully tested to ensure that the most potential errors are not present at a statement point in it.I have checked each piece of code phase wise, we have to be careful that every sentence in the code should be executed at least once. The checking for the white box is often called Glass Box Testing.

I built up a list of test cases and sample data. At every level of the module we must test all possible co mbinations of execution paths through the code.

Black Box Testing

The approach of black box testing considers the whole module as a single unit and tests the device at the interface and contact with other modules at the level of the argument. Here the module should be treated as a block box and some input from that module will be taken and output generated. Output for a given set of input combinations and then forwarding the input to another modules.

Criteria Satisfied by Test Cases

1.The test cases that decreased by more than one percent, the amount of additional test cases intended to accomplish fair testing.

2.In this the test cases that tell us us about the existence or absence of groups of the test errors, rather t han an error that is only correlated with the particular test at hand.

Test Cases:

SNo	Test case ID	Input	Expected output	Actual Output	Test Case Pass/fail
1	100	Upload Data-Set for preprocessing	Data is cleaned	More number of attributes	Fail
2	245	Upload Data-Set for preprocessing	Data is cleaned	Data is cleaned	Pass
3	234	Using predicte_results function to predicte crime	Crime happened: YES	Crime happened :NO	Fail
4	235	Using Predicte results function to predicte	Crime happened:YES	Crime happened: YES	Pass

		crime			
5	235	Using Predicte results function to predicte crime	Crime happened:YES	Crime happened: YES	Pass

8. CONCLUSION

We have performed all the algorithms mentioned above . Among them we have found that decision tree and random forest algorithms are giving more accurate results.

```
In [6]: from sklearn.tree import DecisionTreeRegressor
   ...: regressor_tree = DecisionTreeRegressor(random_state = 0)
   ...: regressor_tree.fit(X_train, y_train)
   ...:
   ...: # Predicting a new result
   ...: y_pred_Dtree = regressor_tree.predict(X_test)
   ...:
   ...: #calculating mse for DecisionTree regression
   ...: mse_Dtree = np.mean((y_pred_Dtree - y_test)**2)
   ...:
   ...: #calculating r-square for decisiontree regression
   ...: regressor_tree.score(X_test,y_test)
Out[6]: 0.98961789897647257

In [7]: from sklearn.ensemble import RandomForestRegressor
   ...: regressor_randomForest = RandomForestRegressor(n_estimators = 15, random_state = 0)
   ...: regressor_randomForest.fit(X_train, y_train)
   ...:
   ...: # Predicting a new result
   ...: y_pred_randomForest = regressor_randomForest.predict(X_test)
   ...: #calculating mse for RandomForest regression
   ...: mse_randomForest = np.mean((y_pred_randomForest - y_test)**2)
   ...:
   ...: #calculating r-square for random forest regression
   ...: regressor_randomForest.score(X_test,y_test)
Out[7]: 0.98941134075220838
```

9. Future Scope

For this research, only crime data has been used, but as many researched have showed that a particular area's socio-economic standard is also a key indicator of possible criminal activity. This machine learning agent could incorporate those data and might perform better. This model can be also used for other geographic locations. This would also help to analyze crimes occurring in different locations and build a better understanding of different crimes and its relation with particular demography. 50 Also, there are many advanced machine learning approaches that can be explored. Deep Learning & Neural Networks can provide a more balanced understanding of criminal activities. As it has been seen on this research, imbalanced classes has been a major issue in dealing with the particular database. Advanced techniques to deal with imbalanced classes are also something that remains to be explored.

10 BIBLIOGRAPHY

[1] H. Adel, M. Salheen, and R. Mahmoud, "Crime in relation to urban design. Case study: the greater Cairo region," *Ain Shams Eng. J.*, vol. 7, no. 3, pp. 925-938, 2016.

[2] "Overall crime rate in Vancouver went down in 2017, VPD says," *CBC News*, Feb. 15, 2018. [Online] Available: https://www.cbc.ca/news/canada/british-columbia/crime-rate-vancouver-2017-1.4537831. [Accessed: 09- Aug- 2018].

[3] J. Kerr, "Vancouver police go high tech to predict and prevent crime before it happens," *Vancouver Courier*, July 23, 2017. [Online] Available: https://www.vancourier.com/news/vancouver-police-go-high-tech-to- predict-and-prevent-crime-before-it-happens-1.21295288. [Accessed: 09- Aug- 2018]

[4] J. Han, *Data mining: concepts and techniques*, Morgan Kaufmann, 2012.

[5] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," *Indian J. of Sci. and Technol.*, vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.

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

[7] T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," *Proc. of Artificial Intell. for Develop. (AID 2010)*, pp. 14-19, 2010.

[8] M. Al Boni and M. S. Gerber, "Area-specific crime prediction models," *15th IEEE Intl. Conf. on Mach. Learn. and Appl.*, Anaheim, CA, USA, Dec. 2016.

[9] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," *IEEE 20th Intl. Conf. on Comput. Supported Cooperative Work in Des. (CSCWD)*, Nanchang, China, May 2016.

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

[11] Y. L. Lin, L. C. Yu, and T. Y. Chen, "Using machine learning to assist crime prevention," *IEEE 6th Intl. Congr. on Advanced Appl. Inform. (IIAI- AAI)*, Hamamatsu, Japan, Jul. 2017.

[12] F. K. Bappee, A. S. Júnior, and S. Matwin, "Predicting crime using spatial features," *Can. AI 2018: Advances in Artificial Intel.-Lecture Notes in Comput. Sci.*, vol. 10832, pp. 367-373, Springer, Mar. 2018.

APPENDIX-A: Python Programming

About Python:

Python is an interpreted as a high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Python is a interpreted language. Interpreted languages do not need to be compiled at the end of the program to run.A program is called an interpreter runs Python code on almost any kind of computer.By help of this interpreter that a programmer can change the code and quickly see the results as soon ast the line is entered.This also means Python is slower than a compiled language like C, because it is not running machine code directly. Python is a good programming language for the beginners. It is a high-level language, which means a programmer can focus on what to do instead of focusing on how to do it. Writing programs in Python takes less time than writing the programs in some other languages.

Python drew inspiration from other programming languages like C, C++, Java, Perl, and Lisp. Python has a very easy-to-read syntax. Some of Python's syntax comes from C, because that is the language that Python was written inthe c language. But the Python uses whitespace to delimit code: spaces or tabs are used to organize the code into groups. This is different from C. In C, there is a semicolon at the end of each line and curly braces ({}) are used to group code but in python we don’t use braces and semicolon at the end of the line.we use whitespace to delimit code makes Python a very easy-to-read the language. Python is used by hundreds of thousands of programmers and is used in many places. Sometimes only Python code is used for a program, but most of the time it is used to do simple jobs while another programming

language is used to do more complicated tasks.Its standard library is made up of many functions that come with Python when it isinstalled.

On the Internet there are many other libraries available that make it possible for the Python language to do more things. These libraries make it a powerful language; it can do many different things.

Some things that Python is often used for are:

Keywords and reserved terms in Python:

Three basic types of Python statements are

• conditionals (such as an “if-then”statement),

• assignments ,and

• iteration (such as a for or whileloop).

Python has set aside many commands to help you create such statements. Python also protects you from accidentally over-writing these commands by “reserving” these commands. When you make an assignment in Python, such as a = 1, you add the name (or “identifier” or “variable”) a to the Python namespace. You can think of a namespace as a mapping from identifiers (i.e., a variable name such as a) to Python objects (e.g., an integer such as 1). A name can be

• “local” (such as a in a =1),

• “global” (such as the complex constant j representing $\sqrt{-1}$),

• “built-in” (such as abs, the absolute value function),or

• “reserved”, or a “keyword” (such as and - see the tablebelow).

Features of Python Programming:

A simple language which is easier to learn : Python has a very simple and elegant

syntax.It's much easier to read and write Python programs compared to other languages like: C++, Java.Python makes programming fun and allows you to focus on the solutionrather than syntax. If you are a newbie, it's a great choice to start your journey withPython.

1. commercial use. Not only can you use and distribute softwares written in it, you can even make changes to the Python's source code. Python has a large community constantly improving it in eachiteration.
2. **Portability:** You can move Python programs from one platform to another with any risk, and run it without any changes. It runs seamlessly on almost all platforms including Windows, Mac OS X andLinux.
3. **Extensible and Embeddable:** Suppose an application requires high performance.

   YoucaneasilycombinepiecesofC/C++orotherlanguageswithPythoncode.This

   will give your application high performance as well as scripting capabilities which other languages may not provide out of the box.

4. **A high-level, interpreted language:** Unlike C/C++, you don't have to worry about daunting tasks like memory management and garbage collection and so on. Likewise, when you run Python code, it automatically converts the code to the language which your computer understands. You don't need to worry about any lower-leveloperations.
5. **Large standard libraries to solve common tasks**: Python has a number of standard libraries which makes life of a programmer much easier so you don't have to write the code yourself. For example: Need to connect MySQL database on a Web server You have to use MySQLdb library by using import MySQLdb . Standard libraries inPython are well tested and used by hundreds of people. So you can be sure that it won't break yourapplication at any point.
6. **Object-oriented**: Everything in Python is an object. Object oriented programming (OOP) helps you solve a complex problem intuitively. With OOP, you have to divide these complex problems into smaller sets by creatingobjects.

Python modules:

NumPy is module for Python. The name is an acronym for "Numeric Python" or "Numerical Python". It is pronounced NUM-py or less often NUM- pee. It is an extension module for Python, mostly written in C. This makes sure that the precompiled mathematical and numerical functions and functionalities of Numpy guarantee great executionspeed.

Furthermore, NumPy enriches the programming language Python with powerful data structures, implementing multi-dimensional arrays and matrices. These data structures guarantee efficient calculations with matrices and arrays. The implementation is even aiming at huge matrices and arrays, better know under the heading of "big data". Besides that the module supplies a large library of high-level mathematical functions to operate on

these matrices and arrays.

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional arrayobject
- sophisticated(broadcasting)functions
- tools for integrating C/C++ and Fortancode
- It is very useful linear algebra, Fourier transform, and random numbercapabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

Pandas:

The Pandas library is one of the most preferred tools for data scientists to do data manipulation and analysis, next to matplotlib for data visualization the fundamental library for scientific computing in Python on which Pandas was built.
The fast, flexible, and expressive Pandas data structures are designed to make real-world data analysis significantly easier, but this might not be immediately the case for those who are just getting started with it. Exactly because there is so much functionality built into this package that the options are overwhelming.
That's where this Pandas cheat sheet might come in handy.
It's a quick guide through the basics of Pandas that you will need to get started on wrangling your data with Python.
As such, you can use it as a handy reference if you are just beginning their data science journey with Pandas or, for those of you who already haven't started yet, you can just use it as a guide to make it easier to learn about and use it.

The Pandas cheat sheet will guide you through the basics of the Pandas library, going from the data structures to I/O, selection, dropping indices or columns, sorting and ranking, retrieving basic information of the data structures you're working with to applying functions and data alignment.

Installation:

- **Mac** and **Linux** users can install NumPy via pip command: pip installnumpy
- **Windows** does not have any package manager analogous to that in linuxormac.

Arrays in NumPy: NumPy’s main object is that homogeneous multidimensional array.

- It is a processofelements , all of the same type, indexed by a number of tuple for thepositiveintegers.

- In NumPy dimensions are known as axes. The number of axes isknownas a rank.
- In numPy array class is called as **ndarray**. It is also known by the alias**array**.

Pandasitis an open source, BSD-licensed library providing high-performance, easy-to-use in the data structures and data analysis tools for the Python programming language Pandas is a NumFOCUS sponsored project. This will help and ensure the success of development of pandas as a world-class open-source project, and makes it possible to donate to the project. Python has long been great for data munging and preparation, but less so for data analysis and modelling.

Pandas helps fill this gap, enabling you to carry out your entire data analysis workflow in Python without having a switch to a more domain specific language like R. Combined with the excellent IPython toolkit and other the libraries, the environment for doing data analysis in Python excels in performance, productivity, and the ability to collaborate the Pandas does not implement significant modeling for the functionality outside of linear and the panel regression; for this, look at statsmodels and scikit-learn. More work is still needed to make the Python a first class statistical modeling environment, but we are well on our way toward that goal**.**

SciPy (Scientific Python)is often mentioned in the same breath with NumPy. SciPy needs Numpy, as it is based on the data structures of Numpy and furthermore its basic creationand manipulation functions. It extends the capabilities of NumPy with the further useful functions for minimization the regression, Fourier-transformation

and many others transformation. Both NumPy and SciPy are not part of the basic Python installation. They have to be get installed after the Python installation. NumPy has to be installed before installing SciPy. (Comment: The diagram of the image on the right side is the graphical visualisation of a matrix with 14 rows and 20 columns. It's a so-called Hinton diagram. The size of a square within this diagram corresponds to the size of the value of the depicted matrix. The colour determines, if the value is positive or negative. In our example: the colour red denotes negative values and the colour green denotes positive values.)

NumPy is based on two earlier Python modules dealing with arrays. One of these is Numeric. Numeric is like NumPy a Python module for high-performance, numeric computing, but it is obsolete nowadays. Another predecessor of NumPy is Numarray,

which is a complete rewrite of Numeric but is deprecated as well. NumPy is a merger of those two, i.e. it is build on the code of Numeric and the features of Numarray.

NLTK Pythonmodule:

The NLTK module is a massive tool kit, aimed at helping you with the entire Natural Language Processing (NLP) methodology. NLTK will aid you with everything from splitting sentences from paragraphs, splitting up words, recognizing the part of speech of those words, highlighting the main subjects, and then even with helping your machine to understand what the text is all about. In this series, we're going to tackle the field of opinion mining, or sentimentanalysis.

In our path to learning how to do sentiment analysis with NLTK, we're going to learnthe following:

- Tokenizing - Splitting sentences and words from the body oftext.
- Part of Speechtagging
- Machine Learning with the Naive Bayesclassifier
- How to tie in Scikit-learn (sklearn) withNLTK

- Training classifiers withdatasets

- Performing live, streaming, sentiment analysis withTwitter.

This will give you all of the tokenizers, chunkers, other algorithms, and all of the corpora. If space is an issue, you can select to selectively download everything manually. The NLTK module will take up about 7MB, and the entire nltk_data directory will take up about 1.8GB, which includes your chunkers, parsers, and the corpora.

If you are operating headless, like on a VPS, you can install everything by running Python and doing:

import
nltknltk.do
wnload() d
(for
download)

all (for download everything)

That will download everything for you headlessly.

Now that you have all the things that you need, let's knock out some quick vocabulary:

- **Corpus** - Body of text, singular. Corpora is the plural of this. Example: A collection of medicaljournals.
- **Lexicon** - Words and their meanings. Example: English dictionary. Consider, however, that various fields will have different lexicons. For example: To a financial investor, the first meaning for the word "Bull"issomeone who is confident about the market, as compared to the common English lexicon, where the first meaning for the word"Bull"is an animal. As such, there is a special lexicon for financial investors, doctors, children, mechanics, and soon.
- **Token** - Each "entity" that is a part of whatever was split up based on rules. For examples, each word is a token when a sentence is "tokenized"

into words. Each sentence can also be a token, if you tokenized the sentences out of aparagraph.

These are the words you will at most commonly hear upon entering the Natural Language Processing (NLP) space, but they there are many more that we will be covering in time. With that, let's show an example of how one might actually tokenize something into tokens with the NLTKmodule.

Pickle python module:

It is used for serializing and de-serializing a Python object structure. Any object in python can be pickled so that it can be saved on disk. What pickle does is that it "serialises" the object first before writing it to file. Pickling is a way to convert apython object (list, dict, etc.) into a character stream. The are the idea is that this character stream contains all the information necessary to reconstruct the object in another pythonscript.

The **pickle** module has an optimized that cousin called the **cPickle**module. As its name implies, **cPickle**is written in C, so it can be up to the 1000 times faster than **pickle**. However it does not support subclassing of the **Pickler()** and **Unpickler()** classes, because in **cPickle**

these are functions, not classes. Most applications have not need to be no need for this functionality, and can benefit from the improved performance of **cPickle**. Other than that, the interfaces of the two modules are nearly identical; the common and the most interface is described in this manual and differences are pointed out where necessary. In the following discussions, we use the term "pickle" to collectively describe the **pickle** and **cPickle**modules.

The data streams the two modules produce are guaranteed to be interchangeable.

Python has a more primitive serialization module called **marshal**, but in general **pickle** should always be the preferred way to serialize Python objects. **marshal** exists primarily to support Python's .pyc files. The data format used by **pickle** is Python-specific. This has the advantage that there are no restrictions imposed by external

standards such as XDR (which can't represent pointer sharing); however it means that non-Python programs may not be ableto reconstruct pickled Pythonobjects.

By default, the **pickle** data format uses a printable ASCII representation. This is slightly more voluminous than a binary representation. The big advantage of using printable ASCII (and of some other characteristics of **pickle**'s representation) is that for debugging or recovery purposes it is possible for a human to read the pickled file with a standard text editor.

There are currently 3 different protocols which can be used for pickling.

1. Protocol version 0 is the original ASCII protocol and is backwards compatible with earlier versions ofPython.
2. Protocol version 1 is the old binary format which is also compatible with earlier versions ofPython.
3. Protocol version 2 was introduced in Python 2.3. It provides much more efficient pickling of new-styleclasses.

HTML

Hypertext MarkupLanguage(HTML), the languages of the world wide web(WWW), allows users to produces web pages that included text, graphics and pointer to other web pages(Hyperlinks).

HTML is not a programming language but it is an application of ISO Standard 8879,SGML(Standard Generalized Markup Language),but specialized to hypertext and adapted to the Web. We can navigate through the information based on out interest and preference. A markup language is simply a series of items enclosed within the elements should bedisplayed.

Hyperlinks are underlined or emphasized works that load to other documents or some portions of the same document.Html can be used to display any type of document on the host computer, which can be geographically at a different location. It is a versatile language and can be used on any platform ordesktop.

HTML provides tags(special codes) to make the document look attractive.HTML provides are not case-sensitive. Using graphics,fonts,different sizes, color, etc.. can enhance the presentation of the document. Anything that is not a tag is part of the document itself.

Basic Html Tags:

<!-- -->	SpecificComments.
<A>……</A>	Creates Hypertextlinks.
<B>……</B>	Text is changed tobold
<Body>…….</Body>	Contains all tags and text in theHtml-document
<DD>……..</DD>	Definition of aterm.
<TABLE>……</TABLE>	Createstable
<Td>……..</Td>	Indicates table data in atable.
<Tr>……..</Tr>	Designates a tablerow
<Th>…….</Th>	Creates a heading in atable.

Advantages:

- ✓ A HTML document is small and hence easy to send over the net.Itis small

because it does not include formattedinformation.

- ✓ HTML is platformindependent.

- ✓ HTML tags are notcase-sensitive.

Advantages of CSS

1. **CSS saves time** : You can write CSS once and then reuse the same sheet in multiple HTML pages. You can define a style for each HTML element and apply it to as many web pages as youwant.
2. **Pages load faster**: If you are using CSS, you do not need to write HTML tag attributes every time. Just write one CSS rule of a tag and apply it to all the occurrences of that tag. So, less code means faster downloadtimes.
3. **Easy maintenance** :Tomake a global change, simply change the style, and all the elements in all the web pages will be updated automatically.
4. **Superior styles to HTML** :CSS has a much wider array of attributes than HTML, so you can give a far better look to your HTML page in comparison to HTML attributes.
5. **Multiple Device Compatibility**: Style sheets allow content to be optimized for more than one type of device. By using the same HTML document, different versions of a website can be presented for handheld devices such as PDAs and cellphones or forprinting.
6. **Global web standards**:Now HTML attributes are being deprecated and it is being recommended to use CSS. Soit's a good idea to start using CSS in all the HTML pages to make them compatible with futurebrowsers.

APPENDIX-B: MODULES

Spyder:

Spyder is an open source cross-platform integrated development envrinoment (IDE) for scientific programming in the python language. Spyder integrates with a number of prominent packages in the scientific Python stack, including Numpy , SciPy , Matplotlib, pandas, IPython, SymPy and Cython as well as other open source software

Spyder is importing all the functions to perform on the dataset. the spyder contain predefined methods, using the methods we find the models for your project.

Numpy and Matplotlib are the two libraries are used in the project.

CSV:

A CSV is a comma-separated values file, which allows data to be saved in a tabular format. CSVs look like a garden-variety spreadsheet but with a .csv extension.

CSV files can be used with most any spreadsheet program, such as Microsoft Excel or Google Spreadsheets. They differ from other spreadsheet file types because you can only have a single sheet in a file, they can not save cell, column, or row. Also, you cannot not save formulas in this format.

These files serve a number of different business purposes. They help companies export a high volume of data to a more concentrated database, for instance. They also serve two other primary business functions: CSV files are plain-text files, making them easier for the website developer to create

Since they're plain text, they're easier to import into a spreadsheet or another storage database, regardless of the specific software you're using To better organize large amounts of data

Saving CSV files is relatively easy, you just need to know where to change the file type.Under the "File name" section in the "Save As" tab, you can select "Save as type" and change it to "CSV (Comma delimited) (*.csv). Once that option is selected, you are on your way to quicker and easier data organization. This should be the same for both Apple and Microsoft operating systems.

In the world of online commerce, one of your main objectives is to reach a large number of clients. Since CSV files are easy to organize, ecommerce business owners can manipulate these files in many different ways. CSV files are mostly used for importing

and exporting important information, such as customer or order data, to and from your database.

A more practical example of this would be an ecommerce business that buys customer data from a social media website. The online network would likely send the consumer information to your database in CSV format, making it quick and easy to exchange the data. If formatted correctly, CSV files are simple to convert to other file types. CSV files are also not hierarchical or object-oriented, meaning they have a ubiquitous structure, which is another factor that makes them easy to import, export and convert.

Cv2 Module:

Installation and Usage

1.If you have previous/other manually installed version of OpenCV installed, remove it before installation to avoid errors.

2.Start installing

3.Select the correct package which is suitable for your environment:

There are four packages among them select only one of them. Dont install multiple packages in the same environment. There is no available plugin architecture as all the packages use the same namespace. If you have installed multiple packages in the same environment, uninstall them all with pip uninstall and reinstall only one required one with specific version of particular package.

a. Packages for standard desktop environments such as Windows, macOS, almost any GNU/Linux distribution:

•run pip install opencv

b. Packages for server environments

These packages do not contain any GUI functionality. These are smaller and suitable for restricted environments.

•run pip install opencv-python

•run pip install opencv- contrib-python

4. Import the package:

import cv2

All packages contain files. cv2.data.haarcascades which can be used as a shortcut to the data

folder.

For example:

cv2.CascadeClassifier(cv2.data.haarcascades + "haarcascade_frontalface_default.xml")

Tensor flow Module:

• Open a new Anaconda (or) Command Prompt window and activate the tensorflow_cpu environment

• Once open, type the following commands:

>>pip install tensorflow==1.9

• Wait for the installation to finish

Test your Installation

Open a new Anaconda (or) Command Prompt window and activate the tensorflow_cpu environment y)

• Start typing a command to open a new Python interpreter session

>>python

•Once the interpreter opens up, type:

>>import tensorflow as tf

• If the above code pops out an error, then make sure that you have activated the tensorflow_cpu environment and that tensorflow_cpu was successfully installed or not

Then run the following:

>>hello = tf.constant('Hello, TensorFlow!')

>>ses = tf.Session()

Once the above command runs, if you see a similar to the one below, it means that you could benefit from installing TensorFlow by building the sources that correspond to you specific CPU.

• Your CPU supports instructions.

•Finally, for testing as described by TensorFlow themselves (see here), let's run the following:

• print(ses.run(hello))

output: Hello, TensorFlow

APPENDIX-C: UNIFIED MODELING LANGUAGE

Anaconda is a free and open source distribution of Python for scientific computing like as data science, machine learning applications, large-scale processing, predictive analytics, etc., that helps to simplify package management. The Anaconda distribution includes data-science packages suitable for many operating systems such as Windows, Linux, and MacOS.

Anaconda distribution comes with nearly 1,500 packages. It also includes a Graphical User Interface(GUI), Anaconda Navigator, as a graphical alternative to the instruction interface (CLI).

The big difference between conda and therefore the pip package manager is in how package dependencies are managed, which may be a significant challenge for Python data science and therefore the reason conda exists.

When pip installs a package, it automatically installs any dependent Python packages on faith if these conflict with previously installed packages. Due to this, a user with a working installation, can find that it stops working having used pip to put in a special package that needs a special version of the dependent numpy library than the one employed by Tensorflow. In some cases, the package may appear to figure but produce different leads to detail.

In contrast, conda analyses the present environment including everything currently installed, and, along side any version limitations specified, works out a way to install compatible set of dependencies, warning if this can't be done.

Open source packages are often individually installed from the Anaconda repository, Anaconda Cloud (anaconda.org), or your own private repository or mirror, using the conda install command. Anaconda compiles and builds all the packages with in the Anaconda repository itself, and provides binaries for different operating systems such as Windows 32/64 bit, Linux 64 bit and MacOS 64-bit. Anything available on PyPI could also be installed into a conda environment using pip, and conda will keep track of what it's installed itself and what pip has installed.

The default installation of Anaconda2 is version Python 2.7 and Anaconda3 is version Python 3.7. However, it is possible to make new environments that include any version of Python packaged with conda. Anaconda Navigator is a desktop graphical user interface (GUI) which includes Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using any command line commands. Navigator can be able to search for packages in Anaconda Cloud or in a local Anaconda Repository, and it will install them in an environment, run the packages and update them. It is available for various operating systems such as Windows, macOS and

Linux.

The following apps are available by default in navigator such as:

- JupyterLab
- Jupyter Notebook
- QtConsole
- Spyder
- Glueviz
- Orange
- RStudio
- Visual Studio Code

CONDA

Conda is open source, cross platform package manager and environment management system that installs, runs, and updates packages and their available dependencies. It is created for Python programs, but it can create package and distribute software for other languages including multi-language projects. The conda package and environment manager is included in all versions of Anaconda, such as Miniconda, and Anaconda Repository.

Spyder, Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda. It provides editing, along with interactive testing, debugging, and introspection features.It is initially created and developed by "Pierre Raybaut" in the year 2009, since 2012 Spyder it is maintained and continuously improved by a team of Python developers and the community.

Spyder is an open-source cross platform integrated development environment (IDE) for programming in the Python language. Spyder combines with many prominent packages in the Python stack, including NumPy, SciPy, Matplotlib, pandas, IPython, SymPy and Cython, as well as other open source software. It is licensed under the MIT.

Spyder may be a powerful scientific setting written in Python, for Python, and designed by and for scientists, engineers and information analysts. It offers a singular combination of the advanced written material, analysis, debugging, and identification practicality of a comprehensive development tool with the information exploration, interactive execution, deep review, and exquisite visual image capabilities of a scientific package.

After installing Anaconda, start Spyder on Windows, macOS, or Linux by running the command spyder.

Spyder is already installed in Anaconda Navigator, which is included in Anaconda.

Spyder is designed to use either of the PyQt or PySide Python bindings. QtPy, a thin abstraction layer developed by the Spyder project and later adopted by multiple other packages, provides the flexibility to use either backend.

Features include:

• An editor with syntax highlighter, introspection, code completion.

• Support for multiple Python consoles

• It has the ability to explore and edit variables from a GUI

• A facilitate pane ready to retrieve and render made text documentation on functions, categories and strategies mechanically or on-demand

• A programme coupled to IPdb, for in small stages execution

• Static code analysis, powered by Pylint

• A run-time Profiler, to benchmark code

• Project support, permitting work on multiple development efforts at the same time

• A integral file human, for interacting with the filesystem and managing comes

• A "Find in Files" feature, permitting full regular expression search over a such scope

• An on-line facilitate browser, permitting users to go looking and consider Python and package documentation within the IDE

• A history log, recording each user command entered in every console

• An internal console, allows for introspection and management over Spyder's own operation

The Unified Modeling Language (UML) is a general-purpose visual modeling language that is used to specify and visualize, construct, and document the artifacts of a software system. It captures decisions and understanding about the systems that must be constructed. It is used to understand design browse configure maintain and control information about such systems. It is intended for use with all development methods lifecycle stages, application domains and the media. The modeling language is intended to unify past experience about the modeling techniques and to incorporate the current software best practices into a standard approach. UML includes semantic concepts and

notation, and guidelines. It has static dynamic environmental, and organizational parts. It is intended to be supported by interactive visual modeling tools that have code generators and the report writers. The UML specification does not define a standard process but is intended to be useful with an iterative development process. It is intended to support most existing object oriented development processes.

The UML captures information about the static structure and dynamic behavior of the system. A system is modeled as a collection of discrete objects and that interact to perform work that ultimately benefits an outside user. The static structure defines the kinds of objects important to a system and to its implementation, as well as the relationships among the objects. The dynamic behavior defines the history of objects over the time and the communications among objects to accomplish the goals.

Modeling a system from several and separate but related viewpoints permits it to be understood for different purposes.

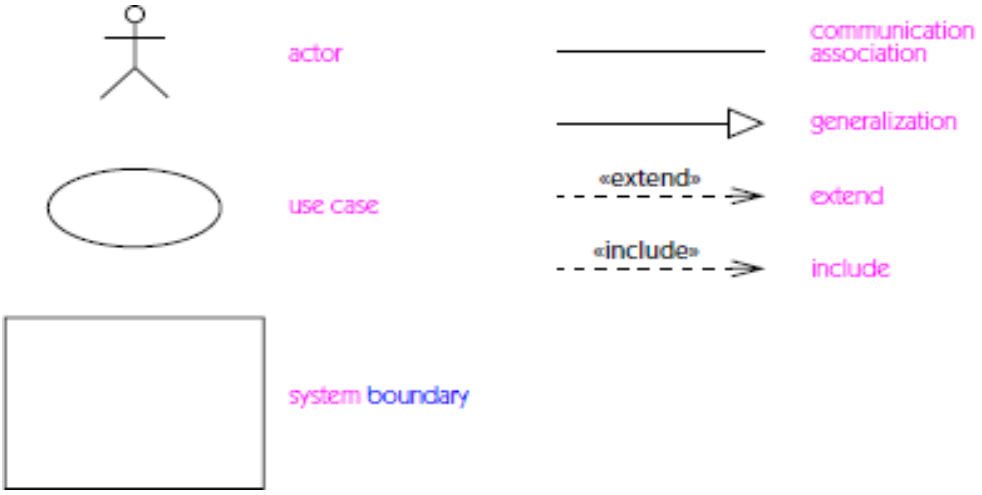
The UML also contains organizational constructs for arranging models into packages that permit software teams to partition large systems into workable pieces, to understand and control dependencies among the packages and to manage the versioning of model units in a complex development environment. It contains constructs for representing the implementation decisions and for organizing run-time elements into the components.

UML is not a programming language. Tools can provide code generators from UML into a variety of programming languages, as well as construct reverse engineered models from existing programs. The UML is not a highly formal language intended for theorem proving.

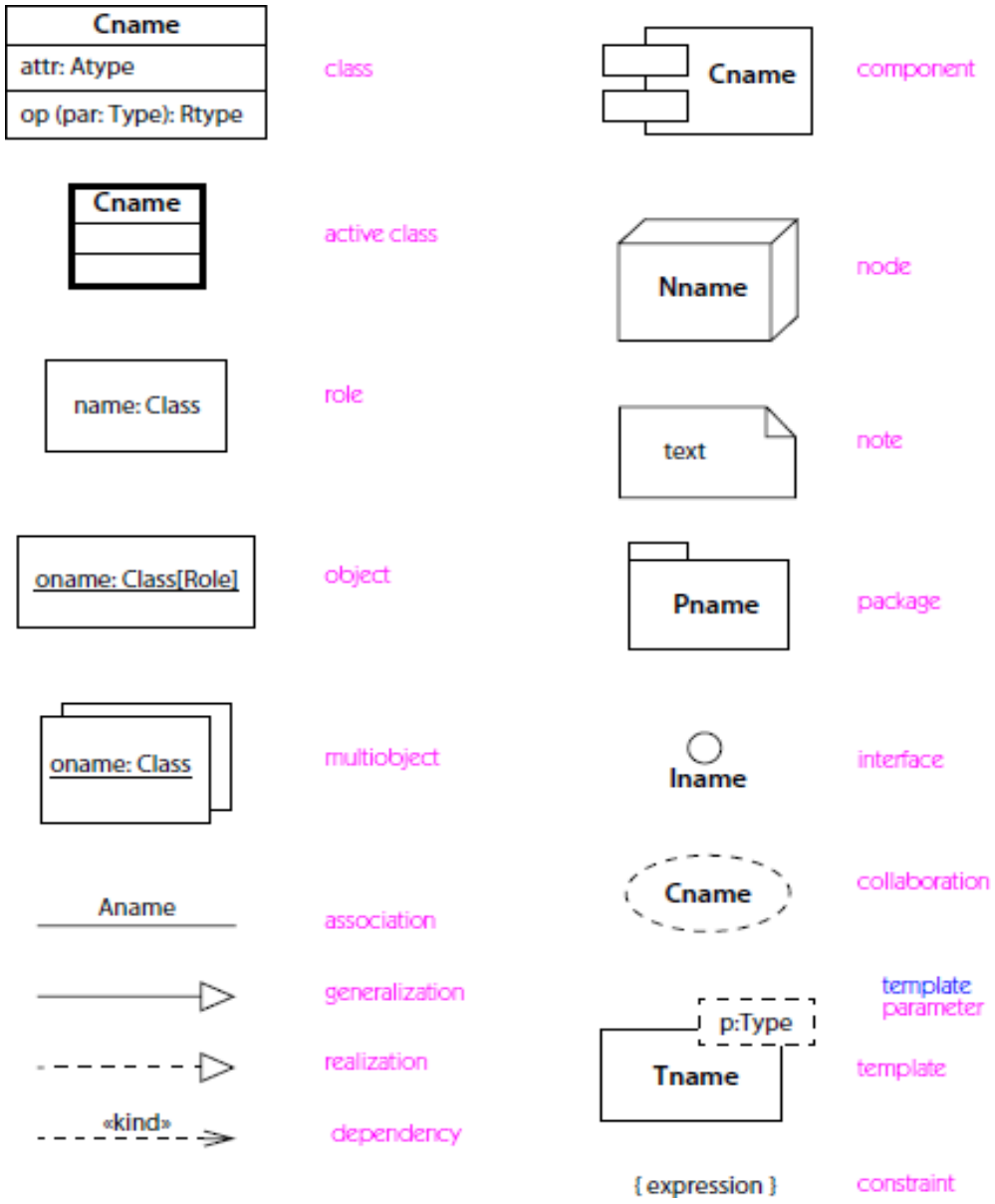
There are a number of such type of languages, but they are not easy to understand or to use for most of the purposes. The UML is a general-purpose modeling language. For only specialized domains, such as GUI layout, VLSI circuit design, or rule-based

artificial intelligence, a more specialized tool with a special language that might be appropriate. UML is a discrete modeling language.

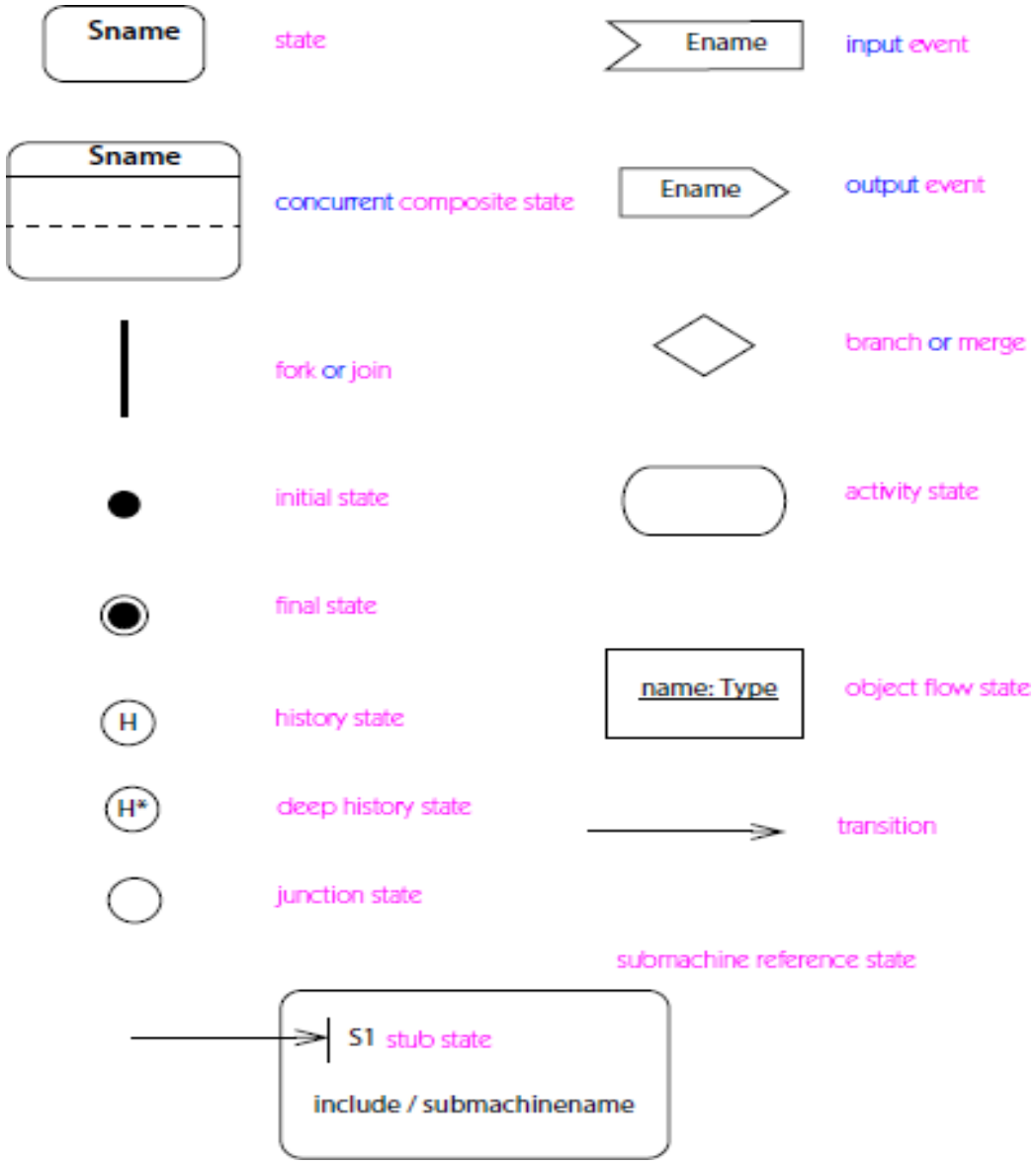
It is not intended to model continuous systems such as those found in engineering and physics. UML is intended to be a universal general-purpose modeling language for discrete systems such as those made of software, firmware, or digital logic.

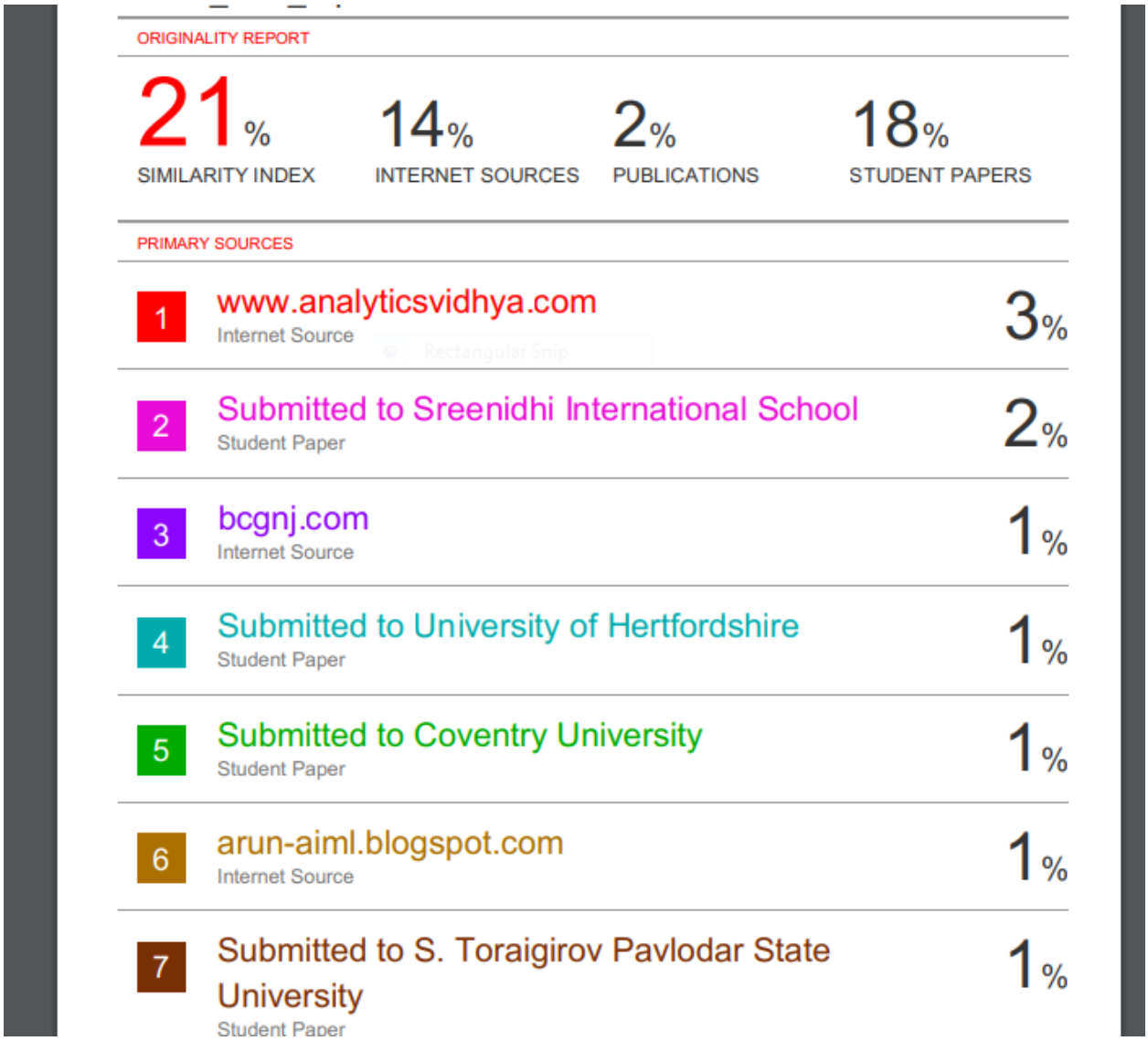


Icons on class, component, deployment, and collaboration diagrams



Icons on statechart and activity diagrams


Sname
state
Ename
input event
Sname
concurrent composite state
Ename
output event
fork or join
branch or merge
initial state
activity state
final state
name: Type
object flow state
H
history state
H*
deep history state
transition
junction state
submachine reference state
S1 stub state
include / submachinename

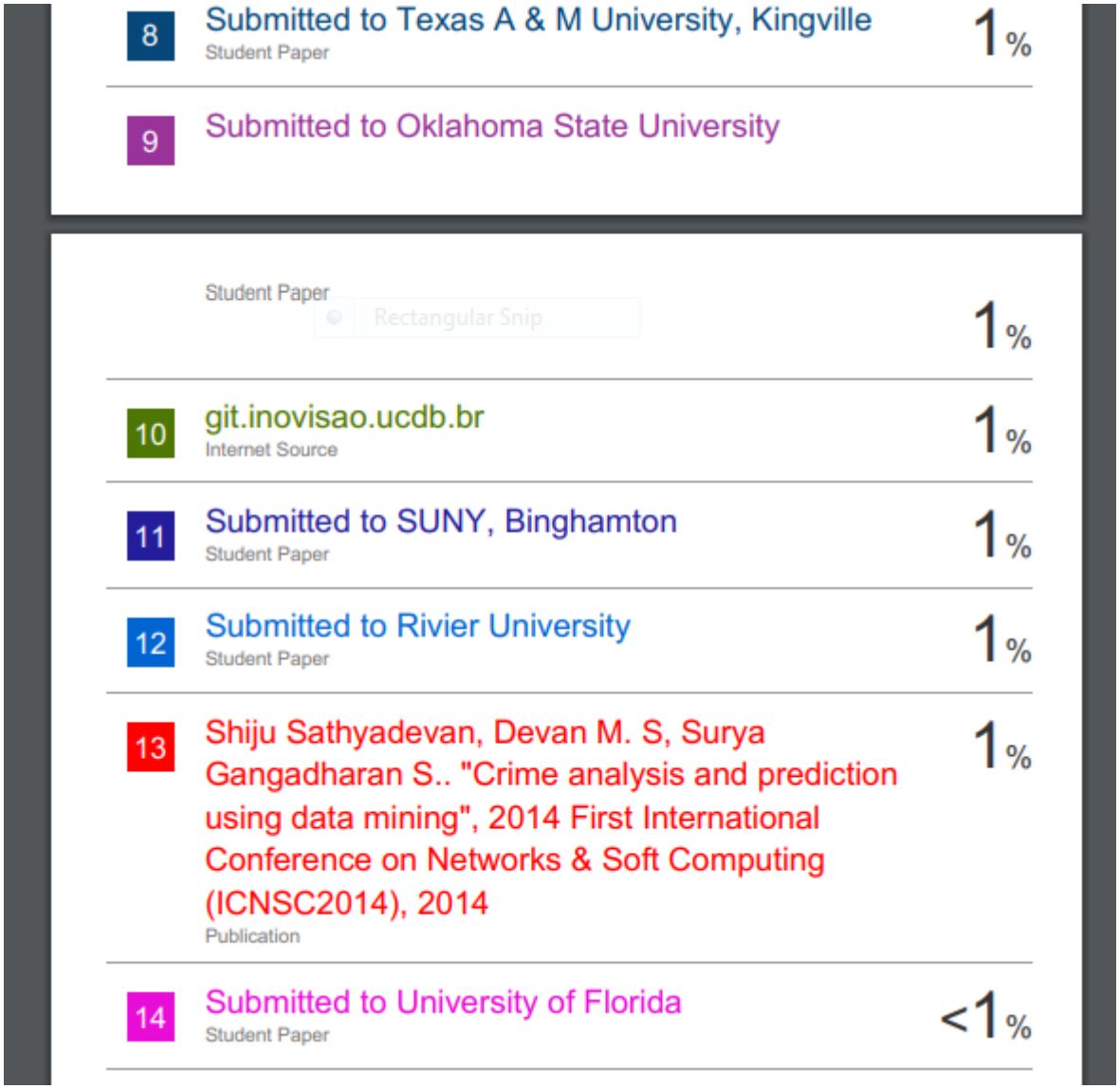


ORIGINALITY REPORT

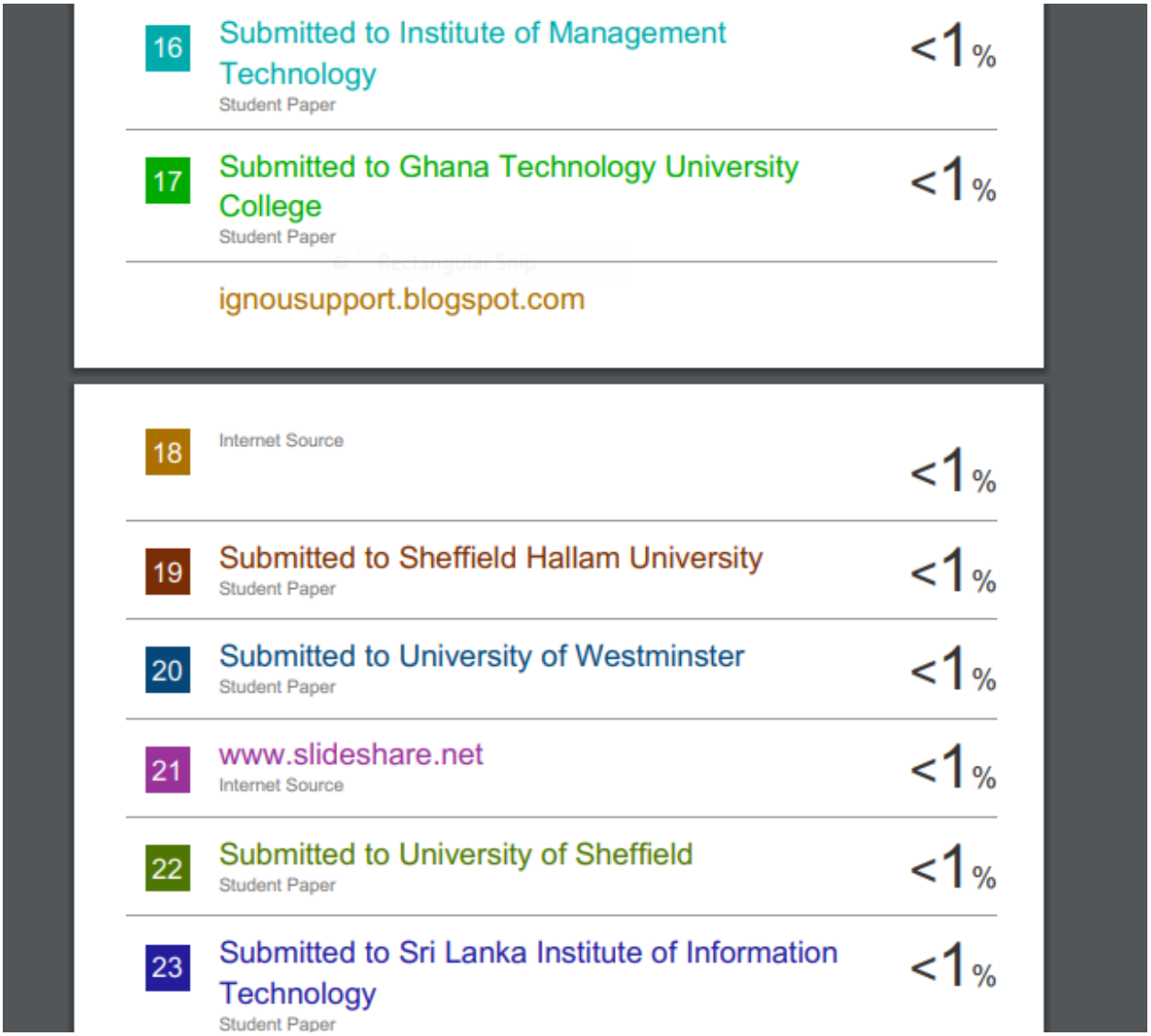
21%	14%	2%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

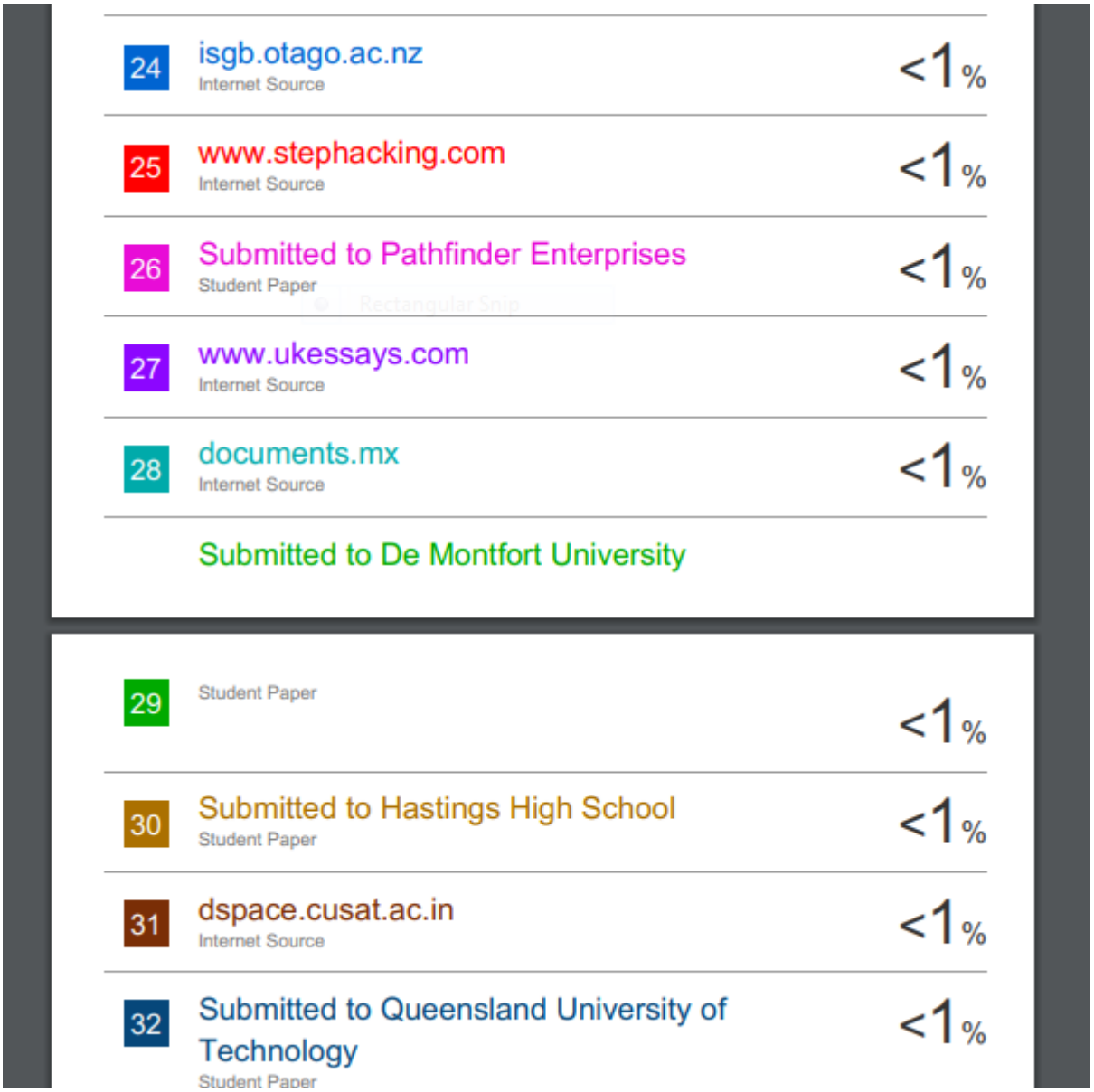
PRIMARY SOURCES

1	www.analyticsvidhya.com Internet Source	3%
2	Submitted to Sreenidhi International School Student Paper	2%
3	bcgnj.com Internet Source	1%
4	Submitted to University of Hertfordshire Student Paper	1%
5	Submitted to Coventry University Student Paper	1%
6	arun-aiml.blogspot.com Internet Source	1%
7	Submitted to S. Toraigirov Pavlodar State University Student Paper	1%



8	Submitted to Texas A & M University, Kingville Student Paper	1%
9	Submitted to Oklahoma State University	
	Student Paper	1%
10	git.inovisao.ucdb.br Internet Source	1%
11	Submitted to SUNY, Binghamton Student Paper	1%
12	Submitted to Rivier University Student Paper	1%
13	Shiju Sathyadevan, Devan M. S, Surya Gangadharan S.. "Crime analysis and prediction using data mining", 2014 First International Conference on Networks & Soft Computing (ICNSC2014), 2014 Publication	1%
14	Submitted to University of Florida Student Paper	<1%


16
Submitted to Institute of Management Technology
Student Paper
<1%
17
Submitted to Ghana Technology University College
Student Paper
<1%
ignousupport.blogspot.com
18
Internet Source
<1%
19
Submitted to Sheffield Hallam University
Student Paper
<1%
20
Submitted to University of Westminster
Student Paper
<1%
21
www.slideshare.net
Internet Source
<1%
22
Submitted to University of Sheffield
Student Paper
<1%
23
Submitted to Sri Lanka Institute of Information Technology
Student Paper
<1%


24 isgb.otago.ac.nz Internet Source <1%
25 www.stephacking.com Internet Source <1%
26 Submitted to Pathfinder Enterprises Student Paper <1%
27 www.ukessays.com Internet Source <1%
28 documents.mx Internet Source <1%
Submitted to De Montfort University
29 Student Paper <1%
30 Submitted to Hastings High School Student Paper <1%
31 dspace.cusat.ac.in Internet Source <1%
32 Submitted to Queensland University of Technology Student Paper <1%