**Crime Analysis using K mean Clustering**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER** | **TITLE** | **PAGE NO** |
|  | **ABSTRACT** |  |
|  | **INTRODUCTION** |  |
|  | * 1. General Introduction |  |
|  | * 1. Project Objectives |  |
|  | * 1. Problem Statement |  |
|  | **SYSTEM** **PROPOSAL** |  |
|  | * 1. Existing System |  |
|  | * + 1. Disadvantages |  |
|  | * 1. Proposed System |  |
|  | 2.2.1 Advantages |  |
|  | * 1. Literature Survey |  |
|  | **SYSTEM DIAGRAMS** |  |
|  | * 1. Architecture Diagram |  |
|  | * 1. Flow Diagram |  |
|  | * 1. UML Diagrams |  |
|  | **IMPLEMENTATION** |  |
|  | * 1. Modules |  |
|  | * 1. Modules Description |  |
|  | **SYSTEM** **REQUIREMENTS** |  |
|  | * 1. Hardware Requirements |  |
|  | * 1. Software Requirements |  |
|  | * 1. Software Description |  |
|  | * 1. Testing of Products |  |
|  | **CONCLUSION** |  |
| **7.** | **FUTURE** **ENHANCEMENT** |  |
| **8.** | **SAMPLE** **CODING** |  |
| **9.** | **SAMPLE** **SCREENSHOT** |  |
| **10.** | **REFERENCES** |  |

**ABSTRACT**

Criminal analysis is a methodical approach for identifying and analysing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, system can analyze previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In our system, the crime dataset of India that contains record of serious fraud of property in all states and we will apply k-means clustering .we can implement the machine learning algorithms. Finally, the experimental results shows that the accuracy, precision, recall and f1-score.Then, we can visualize the heat map and ROC curve.

**CHAPTER 1**

**INTRODUCTION**

* 1. **General Introduction:**

A crime rate has become a topic of major concern certainly to limit the development of good governance and increasing day by day. Crimes are neither systematic nor random otherwise crime cannot be analysis. When crimes like robbery, firebombing etc. have been decreased, crimes like murder, sex abuse, gang rape etc. have been increased. We cannot analyze the victims of crime but can analyze the place where crime occurred or happened. It is difficult to analyze the data to detect crime patterns or predict future crimes by intelligence agencies or local law enforcement agencies. So, there is a need of an effective analyzing tool which can analyze crime data efficiently and quickly to give some useful crime patterns. Predictive policing means, using analytical and predictive techniques to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day’s, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviors in crime pattern are difficult to explain. This paper illustrates how social development may lead to crime prevention. The aim is to provide a comprehensive review of theory and research with respect to the prevention of the crime in the society and to implement different data analysis algorithms which address the connections between crime and its pattern. Crime is one of the major issues is continuing to grow in intensity and complexity. In the recent years, crime is one of the social problems influencing the nature of life and economic development in a community. Crime can be divided into a few types such as crime against properties (theft, burglary, and robbery) and crime of aggression (homicides, assaults and rape). The availability of information technologies has enabled law enforcement to collect detailed information of crime data. With the increasing numbers of crimes nowadays, crime analysis is needed which comprises measure and procedure that intend to reduce the risk of crime. Crime analysis can be done through both quantitative and qualitative methods. Qualitative approaches in predicting crime such as scenario writing or environmental scanning are valuable in identifying the future of criminal activity. Meanwhile, quantitative method is used to predict the crime rates in future specifically. Moreover, crime analysis is a practical approach to analyze and identify the pattern of crimes. Crime analysis is part of crime prevention which has the tasks of discovering and detection of crimes and their relation with criminals.

* 1. **Objectives:**

The main objective of our project is,

* To analyze the crime behavior.
* To cluster the serious fraud of property.
* To implement the machine learning algorithms.
* To enhance the overall performance analysis.
  1. **Problem Statement:**

Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them.so, **to overcome this problem by using the clustering and machine learning algorithms**.

**CHAPTER 2**

**SYSTEM PROPOSAL**

**2.1 EXISTING SYSTEM:**

Crimes are increasing with a high frequency rates in this new era of world and hence it’s a devastating issue that everyone has been experiencing. For finding a pattern that can be used for prediction is necessary. The objective of this paper is to understand the concept of data mining and machine learning which can be used for finding criminal patterns and behaviours. The paper is further divided by providing basic differentiation of the clustering techniques used in unsupervised learning. And then after the crime dataset of India that contains record of serious fraud of property in all states and we will apply k-means clustering to find generic patterns. The main reason for this paper is to give a quip thought of how machine learning can be utilized by the law authorization to distinguish, anticipate and illuminate violations by a lot quicker rate.

**2.1.1 DISADVANTAGES:**

* The results is low when compared with proposed
* It doesn’t efficient for large volume of data’s
* Theoretical limits.
  1. **PROPOSED SYSTEM:**

In the recent past, crime analyses are required to reveal the complexities in the crime dataset. In this system, the crime in India dataset was taken as input. The input data was taken from the dataset repository. Then, we have to implement the data preprocessing step. In this step, we have to handle the missing values for avoid wrong prediction, to encode the label for input data. After that, we have to implement the unsupervised learning such as k-means clustering algorithm to find generic patterns. Then we have to implement the machine learning algorithms such as random forest, KNN and decision tree. Finally, the experimental results shows that the accuracy, precision, recall and f1-score. Then, we can visualize the clustered data and ROC curve.

**2.2.1 ADVANTAGES:**

* It is efficient for large number of datasets.
* The experimental result is high when compared with existing system.
* In addition, to implement the ROC curves for each algorithm.

**2.3 LITERATURE SURVEY:**

# **2.3.1Crime Pattern Detection, Analysis & Prediction using Machine Learning, 2020**

# ***Author:*** Rohit Patil1, Muzamil Kacchi2, Pranali Gavali3, Komal Pimparia

***Methodology*:**

Criminal analysis is a methodical approach for identifying and analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, system can analyze previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, system will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead of them. In this paper, the main focus is on the review of algorithms and techniques used for identify the criminals.

***Disadvantage*:**

* It however suffers from some disadvantages such as determination of the number of clusters by user, affectability from outlier data, high-dimensional data, and sensitivity toward centers for initial clusters and thus possibility of being trapped into local minimum may reduce efficiency of the K-means algorithm.

# **2.3.2Crime Pattern Detection, Analysis & Prediction, 2017**

# ***Author*:** Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav

***Methodology*:**

Crimes are a social irritation and cost our society deeply in several ways. Any research that can help in solving crimes quickly will pay for itself. About 10% of the criminals commit about 50% of the crimes. The system is trained by feeding previous year’s record of crimes taken from legitimate online portal of India listing various crimes such as murder, kidnapping and abduction, dacoits, robbery, burglary, rape and other such crimes. As per data of Indian statistics, which gives data of various crime of past 14 years (2001-2014) a regression model is created and the crime rate for the following years in various states can be predicted. The crime rates accelerate continuously and the crime patterns are constantly changing. As a result, the behaviours in crime pattern are difficult to explain. This paper illustrates how social development may lead to crime prevention. We have used supervised, semi-supervised and unsupervised learning technique on the crime records for knowledge discovery and to help in increasing the predictive accuracy of the crime.

***Advantage***:

* It will be helpful to the local police stations in crime suppression.
* To provide a comprehensive review of theory and research with respect to the prevention of the crime in the society and to implement different data analysis algorithms which address the connections between crime and its pattern.

# **2.3.3 An Overview on Crime Prediction Methods, 2017**

# ***Author*:** Nurul Hazwani Mohd Shamsuddin1 , Nor Azizah Ali2 , Razana Alwee

***Methodology*:**

In the recent past, crime analyses are required to reveal the complexities in the crime dataset. This process will help the parties that involve in law enforcement in arresting offenders and directing the crime prevention strategies. The ability to predict the future crimes based on the location, pattern and time can serve as a valuable source of knowledge for them either from strategic or tactical perspectives. Nevertheless, to predict future crime accurately with a better performance, it is a challenging task because of the increasing numbers of crime in present days. Therefore, crime prediction method is important to identify the future crime and reduces the numbers of crime. Currently, some researchers have been conducted a study to predict crime based on particular inputs. The performance of prediction models can be evaluated using a variety of different prediction methods such as support vector machine, multivariate time series and artificial neural network. However, there are still some limitations on their findings to provide an accurate prediction for the location of crimes. A large number of research papers on this topic have already been published previously. Thus, in this paper, we thoroughly review each of them and summarized the outcomes. Our objective is to identify current implementations of crime prediction method and the possibility to enhance it for future needs.

***Advantage:***

* The scanning algorithm based on geographical crime incidents used to identify clusters with relatively high level of crime hotspots.

# **2.3.4 Crime Pattern Detection Using Data Mining, 2006**

# ***Author***: Shyam Varan Nath

***Methodology*:**

Crimes are a social nuisance and cost our society dearly in several ways. Any research that can help in solving crimes faster will pay for itself. Here we look at use of clustering algorithm for a data mining approach to help detect the crimes patterns and speed up the process of solving crime. We will look at k-means clustering with some enhancements to aid in the process of identification of crime patterns. We will apply these techniques to real crime data from a sheriff’s office and validate our results. We also use semi-supervised learning technique here for knowledge discovery from the crime records and to help increase the predictive accuracy. We also developed a weighting scheme for attributes here to deal with limitations of various out of the box clustering tools and techniques. This easy to implement machine learning framework works with the geo-spatial plot of crime and helps to improve the productivity of the detectives and other law enforcement officers. It can also be applied for counter terrorism for homeland security.

***Advantage***:

* The automated detection of crime patterns, allows the detectives to focus on crime sprees first and solving one of these crimes results in solving the whole “spree”.

# **2.3.5 Using machine learning algorithms to analyze crime data, 2015**

***Author*:** Alejandro Guerra-Manzanares a, Jorge Medina-Galindo, Hayretdin Bahsi b and Sven Nomm

***Methodology***:

Data mining and machine learning have become a vital part of crime detection and prevention. In this research, we use WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. We implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the linear regression algorithm performed the best among the three selected algorithms. The scope of this project is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns. We implemented the Linear Regression, Additive Regression, and Decision Stump algorithms using the same finite set of features, on the communities and crime un normalized dataset to conduct a comparative study between the violent crime patterns from this particular dataset and actual crime statistical data.

***Advantage***:

* Finding and visually presenting groups of facts previously unknown or left unnoticed.
* Heterogeneous data is segmented into a number of homogenous clusters.

**CHAPTER 3**

**SYSTEM DIAGRAMS**

**3.1 SYSTEM ARCHITECTURE:**

Input data

Preprocessing

Clustering

Classification

Performance metrics

Data set ( Crime in India)

Handle missing value

Label Encoding

RF

DT

K-means

KNN

Accuracy.precision, recall

Heap map, ROC curve

FIGURE 3.1: SYSTEM ARCHITECTURE

**3.2 FLOW DIAGRAM**

Input Data

Preprocessing

Clustering

Classification

Performance analysis

FIGURE 3.2: FLOW DIAGRAM

**3.3 UML DIAGRAMS:**

**3.3.1 USE CASE DIAGRAM:**

System

User

FIGURE 3.3.1: USE CASE DIAGRAM

**3.3.2 ACTIVITY DIAGRAM:**

Input Data

Preprocessing

Clustering

Performance analysis

Classification

FIGURE 3.3.2: ACTIVITY DIAGRAM

**3.3.3 SEQUENCE DIAGRAM:**

Input Data

Preprocessing

Clustering

Classification

Select data

Missing value

K-means

Load data

Data splitting

Accuracy

FIGURE 3.3.3: SEQUENCE DIAGRAM

**3.3.4 ER DIAGRAM:**

Data selection

Preprocessing

Clustering

Classification

FIGURE 3.3.4: ER DIAGRAM

**3.3.6 CLASS DIAGRAM:**

Select data ()

Load data ()

View data ()

INPUT

K-means ()

Clustering

Accuracy ()

ROC ()

Performance analysis

Preprocessing

Missing values ()

Label encode ()

RF ()

DT ()

Classification

KNN ()

FIGURE 3.3.5: CLASS DIAGRAM

**CHAPTER 4**

**IMPLEMENTATION**

**4.1 MODULES:**

* Data selection
* Data preprocessing
* Clustering
* Classification
* Result Generation

**4.2 MODULES DESCRIPTION:**

**4.2.1: DATA SELECTION:**

* The input data was collected from dataset repository.
* In our process, the crime in India dataset is used.
* The data selection is the process of analyzing the crime.
* This dataset contains complete information about various aspects of crimes happened in India from 2001.
* There are many factors that can be analysed from this dataset.
* In our process, we have to take the serious fraud dataset.

**4.2.2: DATA PREPROCESSING:**

* Data pre-processing is the process of removing the unwanted data from the dataset.
* Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning.
* This step also includes cleaning the dataset by removing irrelevant or corrupted data that can affect the accuracy of the dataset, which makes it more efficient.
* Missing data removal
* Encoding Categorical data
* Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0.
* Missing and duplicate values were removed and data was cleaned of any abnormalities.
* Encoding Categorical data: That categorical data is defined as variables with a finite set of label values.
* That most machine learning algorithms require numerical input and output variables.

**4.2.3: CLUSTERING:**

* **K-means** clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
* K-means clustering is **a type of unsupervised learning**, which is used when you have unlabeled data (i.e., data without defined categories or groups).
* The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K.
* Data points are clustered based on feature similarity.

**4.2.4: CLASSIFICATION:**

**Machine learning** is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

* **Random forest** algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting.
* It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result**.**
* **Decision Tree** Simple to understand and to interpret. Trees can be visualised. Requires little data preparation.
* Other techniques often require data normalisation, dummy variables need to be created and blank values to be removed.
* Note however that this module does not support missing values.
* The **k-nearest neighbors (KNN)** algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems.
* It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.

**4.2.5: RESULT GENERATION:**

The Final Result will get generated based on the overall classification and prediction. The performance of this proposed approach is evaluated using some measures like,

* **Accuracy**

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

AC= (TP+TN)/ (TP+TN+FP+FN)

* **Precision**

Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

Precision=TP/ (TP+FP)

* **Recall**

Recall is the number of correct results divided by the number of results that should have been returned. In binary classification, recall is called sensitivity. It can be viewed as the probability that a relevant document is retrieved by the query.

Recall=TP/ (TP+FN)

* **Confusion matrix**

A confusion matrix is a tabular summary of the number of correct and incorrect predictions made by a classifier. It can be used to evaluate the performance of a classification model through the calculation of performance metrics like accuracy, precision, recall, and F1-score.

* **ROC curve**

Area under the ROC curve is one of the most useful metrics to evaluate a supervised classification model. This metric is commonly referred to as ROC-AUC. Here, the ROC stands for Receiver Operating Characteristic and AUC stands for Area under the Curve.

**CHAPTER 5**

**SYSTEM REQUIREMENTS**

**5.1 HARDWARE REQUIREMENTS:**

* System : Pentium IV 2.4 GHz
* Hard Disk : 200 GB
* Mouse : Logitech.
* Keyboard : 110 keys enhanced
* Ram : 4GB

**5.2 SOFTWARE REQUIREMENTS:**

* O/S : Windows 7.
* Language : Python
* Front End : Anaconda Navigator – Spyder

**5.3 SOFTWARE DESCRIPTION:**

**5.3.1 Python**

Python is one of those rare languages which can claim to be both *simple* and powerful. You will find yourself pleasantly surprised to see how easy it is to concentrate on the solution to the problem rather than the syntax and structure of the language you are programming in. The official introduction to Python is Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms. I will discuss most of these features in more detail in the next section.

## **5.3.2 Features of Python**

### **Simple**

Python is a simple and minimalistic language. Reading a good Python program feels almost like reading English, although very strict English! This pseudo-code nature of Python is one of its greatest strengths. It allows you to concentrate on the solution to the problem rather than the language itself.

### **Easy to Learn**

As you will see, Python is extremely easy to get started with. Python has an extraordinarily simple syntax, as already mentioned.

### **Free and Open Source**

Python is an example of a FLOSS (Free/Libré and Open Source Software). In simple terms, you can freely distribute copies of this software, read its source code, make changes to it, and use pieces of it in new free programs. FLOSS is based on the concept of a community which shares knowledge. This is one of the reasons why Python is so good - it has been created and is constantly improved by a community who just want to see a better Python.

### **High-level Language**

When you write programs in Python, you never need to bother about the low-level details such as managing the memory used by your program, etc.

### **Portable**

Due to its open-source nature, Python has been ported to (i.e. changed to make it work on) many platforms. All your Python programs can work on any of these platforms without requiring any changes at all if you are careful enough to avoid any system-dependent features.

You can use Python on GNU/Linux, Windows, FreeBSD, Macintosh, Solaris, OS/2, Amiga, AROS, AS/400, BeOS, OS/390, z/OS, Palm OS, QNX, VMS, Psion, Acorn RISC OS, VxWorks, PlayStation, Sharp Zaurus, Windows CE and PocketPC!

You can even use a platform like [Kivy](http://kivy.org) to create games for your computer and for iPhone, iPad, and Android.

### **Interpreted**

This requires a bit of explanation.

A program written in a compiled language like C or C++ is converted from the source language i.e. C or C++ into a language that is spoken by your computer (binary code i.e. 0s and 1s) using a compiler with various flags and options. When you run the program, the linker/loader software copies the program from hard disk to memory and starts running it.

Python, on the other hand, does not need compilation to binary. You just run the program directly from the source code. Internally, Python converts the source code into an intermediate form called bytecodes and then translates this into the native language of your computer and then runs it. All this, actually, makes using Python much easier since you don't have to worry about compiling the program, making sure that the proper libraries are linked and loaded, etc. This also makes your Python programs much more portable, since you can just copy your Python program onto another computer and it just works!

### **Object Oriented**

Python supports procedure-oriented programming as well as object-oriented programming. In procedure-oriented languages, the program is built around procedures or functions which are nothing but reusable pieces of programs. In object-oriented languages, the program is built around objects which combine data and functionality. Python has a very powerful but simplistic way of doing OOP, especially when compared to big languages like C++ or Java.

### **Extensible**

If you need a critical piece of code to run very fast or want to have some piece of algorithm not to be open, you can code that part of your program in C or C++ and then use it from your Python program.

### **Embeddable**

You can embed Python within your C/C++ programs to give scripting capabilities for your program's users.

### **Extensive Libraries**

The Python Standard Library is huge indeed. It can help you do various things involving regular expressions, documentation generation, unit testing, threading, databases, web browsers, CGI, FTP, email, XML, XML-RPC, HTML, WAV files, cryptography, GUI (graphical user interfaces), and other system-dependent stuff. Remember, all this is always available wherever Python is installed. This is called the Batteries Included philosophy of Python.

Besides the standard library, there are various other high-quality libraries which you can find at the Python Package Index.

**5.4 TESTING PRODUCTS:**

System testing is the stage of implementation, which aimed at ensuring that system works accurately and efficiently before the live operation commence. Testing is the process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an error. A successful test is one that answers a yet undiscovered error.

Testing is vital to the success of the system. System testing makes a logical assumption that if all parts of the system are correct, the goal will be successfully achieved. . A series of tests are performed before the system is ready for the user acceptance testing. Any engineered product can be tested in one of the following ways. Knowing the specified function that a product has been designed to from, test can be conducted to demonstrate each function is fully operational. Knowing the internal working of a product, tests can be conducted to ensure that “al gears mesh”, that is the internal operation of the product performs according to the specification and all internal components have been adequately exercised.

**5.4.1 UNIT TESTING:**

Unit testing is the testing of each module and the integration of the overall system is done. Unit testing becomes verification efforts on the smallest unit of software design in the module. This is also known as ‘module testing’.

The modules of the system are tested separately. This testing is carried out during the programming itself. In this testing step, each model is found to be working satisfactorily as regard to the expected output from the module. There are some validation checks for the fields. For example, the validation check is done for verifying the data given by the user where both format and validity of the data entered is included. It is very easy to find error and debug the system.

**5.4.2 INTEGRATION TESTING:**

Data can be lost across an interface, one module can have an adverse effect on the other sub function, when combined, may not produce the desired major function. Integrated testing is systematic testing that can be done with sample data. The need for the integrated test is to find the overall system performance. There are two types of integration testing. They are:

i) Top-down integration testing. ii) Bottom-up integration testing.

**5.4.3 TESTING TECHNIQUES/STRATEGIES:**

* **WHITE BOX TESTING:**

White Box testing is a test case design method that uses the control structure of the procedural design to drive cases. Using the white box testing methods, we

Derived test cases that guarantee that all independent paths within a module have been exercised at least once.

* **BLACK BOX TESTING:**

1. Black box testing is done to find incorrect or missing function
2. Interface error
3. Errors in external database access
4. Performance errors.
5. Initialization and termination errors

In ‘functional testing’, is performed to validate an application conforms to its specifications of correctly performs all its required functions. So this testing is also called ‘black box testing’. It tests the external behaviour of the system. Here the engineered product can be tested knowing the specified function that a product has been designed to perform, tests can be conducted to demonstrate that each function is fully operational.

**5.4.4 SOFTWARE TESTING STRATEGIES**

**VALIDATION TESTING:**

After the culmination of black box testing, software is completed assembly as a package, interfacing errors have been uncovered and corrected and final series of software validation tests begin validation testing can be defined as many,

But a single definition is that validation succeeds when the software functions in a manner that can be reasonably expected by the customer

**USER ACCEPTANCE TESTING:**

User acceptance of the system is the key factor for the success of the system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system at the time of developing changes whenever required.

**OUTPUT TESTING**:

After performing the validation testing, the next step is output asking the user about the format required testing of the proposed system, since no system could be useful if it does not produce the required output in the specific format. The output displayed or generated by the system under consideration. Here the output format is considered in two ways. One is screen and the other is printed format. The output format on the screen is found to be correct as the format was designed in the system phase according to the user needs. For the hard copy also output comes out as the specified requirements by the user. Hence the output testing does not result in any connection in the system.

**CHAPTER 6**

**CONCLUSION**

With the advancement in technologies that are coming recently in data science and especially in machine learning, it becomes easy and efficient to discover patterns and information which might get useful for future prediction in crime analysis and behaviour segmentation. Clustering is the process of grouping similarities in a dataset so that it can get useful for analysis, discovering patterns and prediction.

**CHAPTER 7**

**FUTURE ENHANCEMENT**

As a future work, it would be interesting to evaluate the performance of some unsupervised algorithms. Furthermore, we applied various deep and machine learning algorithms independently from each other. In the future, we should like to combine different machine learning and deep learning algorithms as a multi-layered model to improve the detection performance.

**CHAPTER 8**

**SAMPLE CODING**

#import libraries-------------------------------------------------

import pandas as pd

import warnings

warnings.filterwarnings("ignore")

from knn import \*

from DT import \*

from RF import \*

##1.data slection---------------------------------------------------

#def main():

dataframe=pd.read\_csv("Dataset.csv")

print("---------------------------------------------")

print()

print("Data Selection")

print("Samples of our input data")

print(dataframe.head(10))

print("----------------------------------------------")

print()

#2.pre processing--------------------------------------------------

#checking missing values

print("---------------------------------------------")

print()

print("Before Handling Missing Values")

print()

print(dataframe.isnull().sum())

print("----------------------------------------------")

print()

print("-----------------------------------------------")

print("After handling missing values")

print()

dataframe\_2=dataframe.fillna(0)

print(dataframe\_2.isnull().sum())

print()

print("-----------------------------------------------")

#label encoding

from sklearn import preprocessing

label\_encoder = preprocessing.LabelEncoder()

print("--------------------------------------------------")

print("Before Label Handling ")

print()

print(dataframe\_2.head(10))

print("--------------------------------------------------")

print()

dataframe\_2['Area\_Name']= label\_encoder.fit\_transform(dataframe\_2['Area\_Name'])

dataframe\_2['Group\_Name']= label\_encoder.fit\_transform(dataframe\_2['Group\_Name'])

dataframe\_2['Sub\_Group\_Name']= label\_encoder.fit\_transform(dataframe\_2['Group\_Name'])

print("-------------------------------------------------------")

print("After Label Encoding")

print()

print(dataframe\_2.head(10))

print("---------------------------------------------------------")

print()

#3.Data splitting---------------------------------------------------

x=dataframe\_2.drop('Loss\_of\_Property\_Above\_100\_Crores',axis=1)

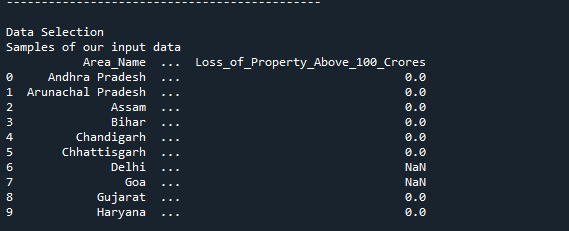
y=dataframe\_2.Loss\_of\_Property\_Above\_100\_Crores

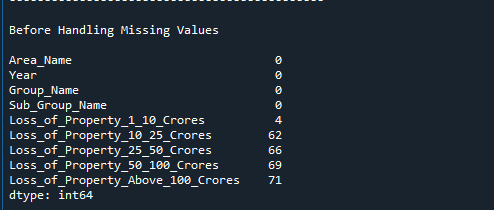
from sklearn.model\_selection import train\_test\_split

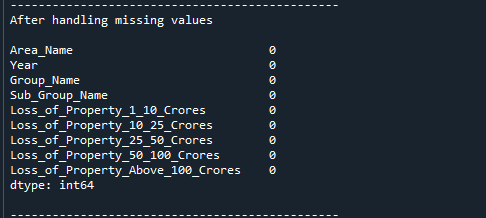
x\_train,x\_test,y\_train,y\_test = train\_test\_split(x,y,test\_size = 0.25,random\_state = 42)

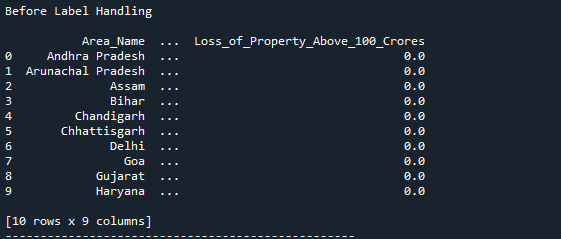
**CHAPTER 9**

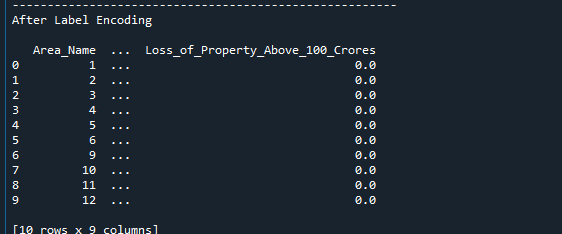
**SAMPLE SCREENSHOTS**

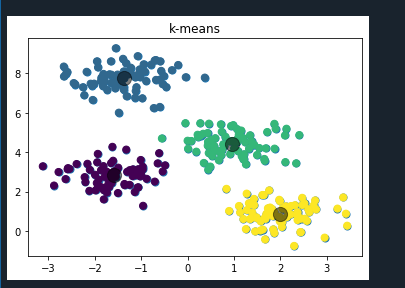


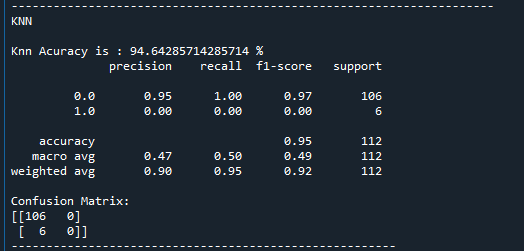


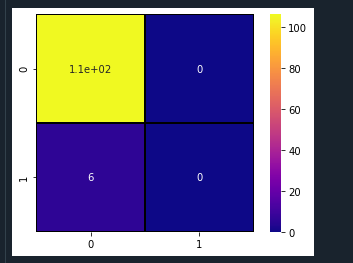


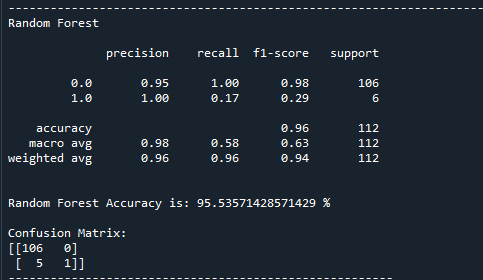


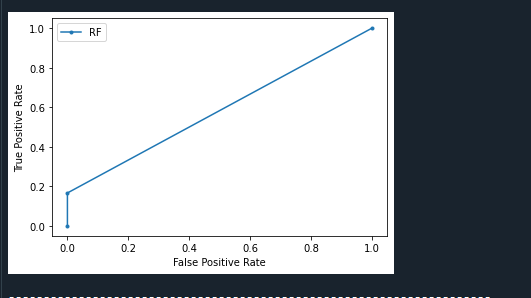


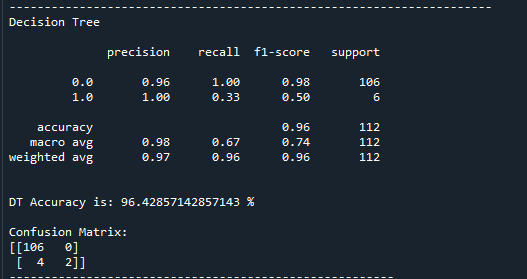












**CHAPTER 10**

**REFERENCES**

1. Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225-230). IEEE.
2. Suhong Kim ; Param Joshi ; Parminder Singh Kalsi ; Pooya Taheri. Crime Analysis Through Machine Learning, 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON).
3. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
4. Nath, S. V. (2006, December). Crime pattern detection using data mining. In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on (pp. 41-44). IEEE.
5. Suhong Kim, Param Joshi, Parminder Singh Kalsi, and Pooya Taheri Fraser. Crime Analysis Through Machine Learning, November 2018, International College, Simon Fraser University.
6. Alkesh Bharati, Dr Sarvanaguru RA.K, Crime Prediction and Analysis Using Machine Learning, International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 09 | Sep 2018.
7. Arpita Nagpal, Aman Jatain, Deepti Gaur. Review based on Data Clustering Algorithms, Proceedings of 2013 IEEE International Conference on Information and Communication Technologies (ICT 2013).
8. Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi. Analysis and Prediction of Crimes by Clustering and Classification, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, vol 4, No.8, 2015.
9. Lawrence McClendon and Natarajan Meghanathan\*. USING MACHINE LEARNING ALGORITHMS TO ANALYZE CRIME DATA, Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015.
10. Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.
11. N. Tyagi, A. Rana, “Fuel your growth with integration: Hybrid cloud computing”, in International Journal of Applied Engineering Research, Vol. 10, Isuue 13, pp 32761-32762 (2015).
12. A. Singh, A. Rana, J. Ranjan, “Proposed analytical customer centric model for an automobile industry”, in International Journal of Data Mining, Modelling and Management, Vol. 7, Isuue 4, pp 314-330 (2015).
13. A. Singh, A. Rana, J. Ranjan, “Data mining techniques and its effect in customer relationship management”, in International Journal of Data Analysis Techniques and Strategies, Vol. 7, Isuue 4, pp 406-427 (2015).
14. G. Dubey, A. Rana, N. K. Shukla, “ User reviews data analysis using opinion mining on web”, in 2015 1st International Conference on Futuristic Trends in Computational Analysis and Knowledge Management, ABLAZE 2015, pp 603- 612 (2015).
15. A. Rana, S. P. Singh, R. Soni, A. Jolly, “Incremental software version release planning for adaptive maintenance”, in International Journal of Software Engineering and its Applications, Vol. 9, Isuue 8, pp 217-228 (2015).
16. A. Singh, A. Rana, “Generate frequent queries for views in a data warehouse using data mining techniques”, in ACM International Conference Proceeding Series (2014).