JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

FBI CRIME DATA ANALYSIS USING MACHINE LEARNING

[1]*Anthoti Raju Assistant Professor,*
[2]*E Krishna Assistant Professor,*
krishna.cseit@gmail.com,
[3]*Dr. J Rajaram Associate Professor,*
drrajaram81@gmail.com,
[4]*D Navya Associate Professor,*
dubbaka.navya@gmail.com
Department of CSE Engineering,
Pallavi Engineering College,
Kuntloor(V),Hayathnagar(M),Hyderabad,R.R.Dist.-501505

ABSTRACT

Crime is one of the biggest and dominating problem in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, we will be using the technique of machine learning and data science for crime prediction of Chicago crime data set. For this supervised classification problem, Decision Tree, Gaussian Naive Bayes, k-NN, Logistic Regression. This approach involves predicting crimes classifying, pattern detection and visualization with effective tools and technologies. Use of past crime data trends helps us to correlate factors which might help understanding the future scope of crimes. In this work, various visualizing techniques and machine learning algorithms are adopted for predicting the crime distribution over an area. In the first step, the raw datasets were processed and visualized based on the need.

Keywords: crime analysis, prediction analysis, machine learning, decision trees, pattern detection.

INTRODUCTION

1.1 Introduction Of FBI Crime Data Analysis

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster. The objective would be to train a model for prediction. The training would be done using the training data set which will be validated using the test dataset. Building the model will be done using better algorithm depending upon the accuracy. The K-Nearest Neighbor (KNN) classification and other algorithm will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in the country.

This work helps the law enforcement agencies to predict and detect crimes in Chicago with improved accuracy and thus reduces the crime rate. There has been tremendous increase in machine learning algorithms that have made crime prediction feasible based on past data. The aim of this project is to perform analysis and prediction of crimes in states using machine learning models. It focuses on creating a model that can help to detect the number of crimes by its type in a particular state. In this project various machine learning models like K-NN, boosted decision trees will be used to predict crimes. Area Wise geographical analysis can be done to understand the pattern of crimes. Various visualization techniques and plots are used which can help law enforcement agencies to detect and predict crimes with higher accuracy. This will indirectly help reduce the rates of crimes and can help to improve securities in such required areas. Crimes can be predicted as the criminals are active and operate in their comfort zones. Once successful they try to replicate the crime under similar circumstances.

1.2 INTRODUCTION TO DOMAIN

Machine learning (**ML**)

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task.

Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

The name machine learning was coined in 1959 by Arthur Samuel. Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E. This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence, in which the question Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do In Turing's proposal the various characteristics that could be possessed by a thinking machine and the various implications in constructing one are exposed.

Machine learning uses data to detect various patterns in a given dataset.

1.It can learn from past data and improve automatically.

2.It is a data-driven technology.

3.Machine learning is much similar to data mining as it also deals with the huge amount of the data.

How does Machine Learning Work?

A Machine Learning system learns from historical data, builds the prediction models, and whenever it receives new data, predicts the output for it. The accuracy of predicted output depends upon the amount of data, as the huge amount of data helps to build a better model which predicts the output more accurately.

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample input doesn't have labels.

Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

In unsupervised learning, the algorithm builds a mathematical model from a set of data that contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of features, or inputs, in a set of data.

Active learning algorithms access the desired outputs (training labels) for a limited set of

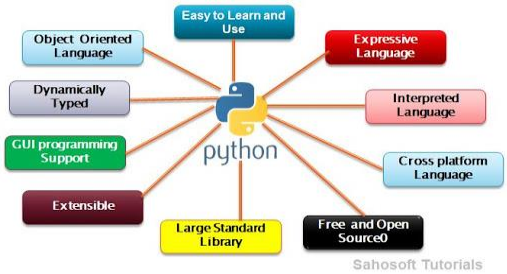
ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

inputs based on a budget and optimize the choice of inputs for which it will acquire training labels. When used interactively, these can be presented to a human user for labeling. Reinforcement learning algorithms are given feedback in the form of positive or negative reinforcement in a dynamic environment and are used in autonomous vehicles or in learning to play a game against a human opponent Other specialized algorithms in machine learning include topic modeling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Meta learning algorithms learn their own inductive bias based on previous experience. In developmental robotics, robot learning algorithms generate their own sequences of learning experiences, also known as a curriculum, to cumulatively acquire new skills through self-guided exploration and social interaction with humans. These robots use guidance mechanisms such as active learning, maturation, motor synergies, and imitation.

LITRATURE STUDY

2.1 Features of Python Programming Languages

At some point in time, we had about as many programming language as we could count on our fingers. Today, there are so many, and all with their own specialties. But what make a language unique are its features. And ultimately, it is its featured that get it chosen or passed for project. So before beginning with deeper concepts of Python, lets take a look at the major features of python programming languages that give you reasons why you should learn Python as compared to R or other tool. Solets start with the features of Python Programming Language.



Easy to Code

Python is a very developer-friendly language which means that anyone and everyone can learn to code it in a couple of hours or days. As compared to other object-oriented programming languages like Java, C, C++, and C#, Python is one of the easiest to learn.

Open Source and Free

Python is an open-source programming language which means that anyone can create and contribute to its development. Python has an online forum where thousands of coders gather daily to improve this language further. Along with this Python is free to download and use in any operating system, be it Windows, Mac or Linux.

Support for GUI

GUI or Graphical User Interface is one of the key aspects of any programming language because it has the ability to add flair to code and make the results more visual. Python has support for a wide array of GUIs which can easily be imported to the interpreter, thus making this one of the most favorite languages for developers.

Object-Oriented Approach

One of the key aspects of Python is its object-oriented approach. This basically means that Python recognizes the concept of class and object encapsulation thus allowing programs to be efficient in the long run.

High-Level Language

Python has been designed to be a high-level programming language, which means that when you code in Python you don't need to be aware of the coding structure, architecture as well as memory management.

Integrated by Nature

Python is an integrated language by nature. This means that the python interpreter executes codes one line at a time. Unlike other object-oriented programming languages, we don't need to compile Python code thus making the debugging process much easier and efficient. Another advantage of this is, that upon execution the Python code is immediately converted into an intermediate form also known as byte-code which makes it easier to execute and also saves runtime in the long run.

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

Highly Portable

Suppose you are running Python on Windows and you need to shift the same to either a Mac or a Linux system, then you can easily achieve the same in Python without having to worry about changing the code. This is not possible in other programming languages, thus making Python one of the most portable languages available in the industry.

Highly Dynamic

As mentioned in an earlier paragraph, Python is one of the most dynamic languages available in the industry today. What this basically means is that the type of a variable is decided at the run time and not in advance. Due to the presence of this feature, we do not need to specify the type of the variable during coding, thus saving time and increasing efficiency.

Extensive Array of Library

Out of the box, Python comes inbuilt with a large number of libraries that can be imported at any instance and be used in a specific program. The presence of libraries also makes sure that you don't need to write all the code yourself and can import the same from those that already exist in the libraries.

Support for Other Languages

Being coded in C, Python by default supports the execution of code written in other programming languages such as Java, C, and C#, thus making it one of the versatile in the industry.

2.2 Python Classes and Objects

Python Classes/Objects

- Python is an object oriented programming language.
- Almost everything in Python is an object, with its properties and methods.
- A Class is like an object constructor, or a "blueprint" for creating objects.

2.3 Software Environment:

Python:

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms,
including structured (particularly, procedural,) object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released in 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released in 2008, was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3.

The Python 2 language, i.e. Python 2.7.x, was officially discontinued on 1 January 2020 (first planned for 2015) after which security patches and other improvements will not be released for it. With Python 2's end-of-life, only Python 3.5.xand later are supported.

Python interpreters are available for many operating systems. A global community of programmers develops and maintains CPython, an open sourcereference implementation. A non-profit organization, the Python Software Foundation, manages and directs resources for Python and CPython development.

Python is used for:

- web development (server-side),
- software development,
- mathematics,system scripting.

Python installation procedure:

Windows Based

It is highly unlikely that your Windows system shipped with Python already installed. Windows systems typically do not.

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

Fortunately, installing does not involve much more than downloading the Python installer from the python.org website and running it. Let's take a look at how to install Python 3 on Windows:

Step 1: Download the Python 3 Installer

1. Open a browser window and navigate to the Download page for Windows at python.org.
2. Underneath the heading at the top that says **Python Releases for Windows**, click on the link for the **Latest Python 3 Release - Python 3.x.x**. (As of this writing, the latest is Python 3.6.5.)
3. Scroll to the bottom and select either **Windows x86-64 executable installer** for 64-bit or **Windows x86 executable installer** for 32-bit. (See below.)

Sidebar: 32-bit or 64-bit Python?
For Windows, you can choose either the 32-bit or 64-bit installer. Here's what the difference between the two comes down to:

- If your system has a 32-bit processor, then you should choose the 32-bit installer.
- On a 64-bit system, either installer will actually work for most purposes. The 32-bit version will generally use less memory, but the 64-bit version performs better for applications with intensive computation.
- If you're unsure which version to pick, go with the 64-bit version.

Note: Remember that if you get this choice "wrong" and would like to switch to another version of Python, you can just uninstall Python and then re-install it by downloading another installer from python.org.

Step 2: Run the Installer
Once you have chosen and downloaded an installer, simply run it by double-clicking on the downloaded file. A dialog should appear that looks something like this:

Important: You want to be sure to check the box that says **Add Python 3.x to PATH** as shown to ensure that the interpreter will be placed in your execution path.
Then just click **Install Now**. That should be all there is to it. A few minutes later you should have a working Python 3 installation on your system.

Mac OS based

While current versions of macOS (previously known as "Mac OS X") include a version of Python 2, it is likely out of date by a few months. Also, this tutorial series uses Python 3, so let's get you upgraded to that.
The best way we found to install Python 3 on macOS is through the Homebrew package manager. This approach is also recommended by community guides like The Hitchhiker's Guide to Python.

Step 1: Install Homebrew (Part 1)
To get started, you first want to install Homebrew:

1. Open a browser and navigate to http://brew.sh/. After the page has finished loading, **select the Homebrew bootstrap code under "Install Homebrew"**. Then hit cmd+c to copy it to the clipboard. Make sure you've captured the text of the complete command because otherwise the installation will fail.
2. Now you need to **open a Terminal app window, paste the Homebrew bootstrap code, and then hit** Enter. This will begin the Homebrew installation.
3. If you're doing this on a fresh install of macOS, you may get a pop up alert **asking you to install Apple's "command line developer tools"**. You'll need those to continue with the

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

installation, so please **confirm the dialog box by clicking on "Install"**.

At this point, you're likely waiting for the command line developer tools to finish installing, and that's going to take a few minutes. Time to grab a coffee or tea!

Step 2: Install Homebrew (Part 2)
You can continue installing Homebrew and then Python after the command line developer tools installation is complete:

1. Confirm the "The software was installed" dialog from the developer tools installer.
2. Back in the terminal, hit Enter to continue with the Homebrew installation.
3. Homebrew asks you to enter your password so it can finalize the installation. **Enter your user account password and hit** Enter to continue.
4. Depending on your internet connection, Homebrew will take a few minutes to download its required files. Once the installation is complete, you'll end up back at the command prompt in your terminal window.

Whew! Now that the Homebrew package manager is set up, let's continue on with installing Python 3 on your system.

Step 3: Install Python

Once Homebrew has finished installing, **return to your terminal and run the following command**:

$ brew install python3
Note: When you copy this command, be sure you don't include the $ character at the beginning. That's just an indicator that this is a console command.
This will download and install the latest version of Python. After the Homebrew brew install command finishes, Python 3 should be installed on your system.

You can make sure everything went correctly by testing if Python can be accessed from the terminal:

1. Open the terminal by launching **Terminal app**.
2. Type pip3 and hit Enter.
3. You should see the help text from Python's "Pip" package manager. If you get an error message running pip3, go through the Python install steps again to make sure you have a working Python installation.

Assuming everything went well and you saw the output from Pip in your command prompt window…congratulations! You just installed Python on your system, and you're all set to continue with the next section in this tutorial.

Packages need for python based programming:

- **Numpy**
  NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc.
- **Pandas**
  Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.
- **Keras**
  Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. Use Keras if you need a deep learning library that: Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).
- **Sklearn**
  Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.
- **Scipy**
  SciPy is an open-source Python library which is used to solve scientific and mathematical problems. It is built on the NumPy extension and allows the user to manipulate and visualize data with a wide range of high-level commands.
- **Tensorflow**

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow.

- **Django**

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

- **Pyodbc**

pyodbc is an open source Python module that makes accessing ODBC databases simple. It implements the DB API 2.0 specification but is packed with even more Pythonic convenience. Precompiled binary wheels are provided for most Python versions on Windows and macOS. On other operating systems this will build from source.

- **Matplotlib**

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.

3.1 SYSTEM SPECIFICATION:

3.1.1 HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 14' Colour Monitor.
- Mouse : Optical Mouse.
- Ram : 512 Mb.

3.1.2 SOFTWARE REQUIREMENTS:

- Operating system : Windows 7 Ultimate.
- Coding Language : Python.
- Front-End : Python.
- Designing : Html,css,javascript.
- Data Base : MySQL.

3.2.1 EXISTING SYSTEM

- The dataset used for this is real and authentic. The dataset is acquired from UCI machine learning repository website.
- The title of the dataset is 'Crime and Communities'. It is prepared using real data from socio-economic data from 1990 US Census, law enforcement data from the 1990 US LEMAS survey and crime data from the 1995 FBI UCR.
- This dataset contains a total number of 147 attributes and 2216 instances.

3.2.2 PROPOSED SYSTEM :

- From a large list of attributes, only eighteen attributes are chosen for Exploratory Data Analysis. The chosen attributes are namely state, HousVacant,PctHouseOccup, PctHouseOwnCC, PctVacantBoarded, PctVacMore6Mos, PctUnemployed, PctEmploy, murdperPop, rapesperPop, robbperPop, assaultperPop, burglperPop, larcperPop, autoTheftperPop, arsonsperPop, nonviolperpop and violentcrimesperpop.
- Regression Analysis is limited to the following predictor and response variable predictor variables: Housevacant,

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

PctHouseOccup, PctHouseownCC, PctVacantBoarded, PctVacmore6Mos, PctUnemployed, PctEmploy Response variables: Violentcrimesperpop.

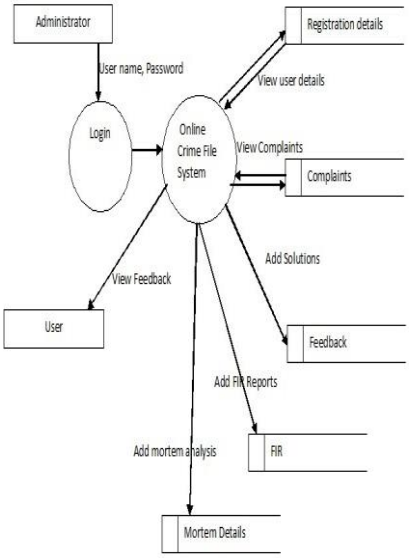
❖Solving the imbalanced class problem, the machine learning agent was able to categorize crimes with 81% accuracy.

SYSTEM DESIGN

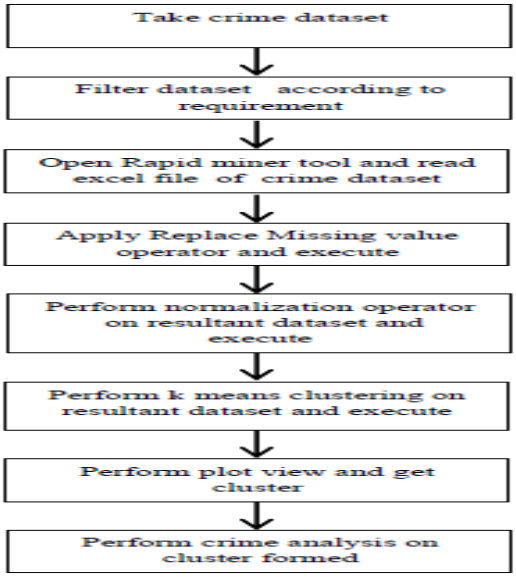
4.1 Introduction

Criminal activities are present in every region of the world affecting quality of life and socio-economical development. As such, it is a major concern of many governments who are using different advanced technology to tackle such issues. Crime Analysis, a sub branch of criminology, studies the behavioral pattern of criminal activities and tries to identify the indicators of such events. Machine learning agents work with data and employ different techniques to find patterns in data making it very useful for predictive analysis. Law enforcement agencies use different patrolling strategies based on the information they get to keep an area secure. A machine learning agent can learn and analyze the pattern of occurrence of a crime based on the reports of previous criminal activities and can find hotspots based on time, type or any other factor. This technique is known as classification and it allows to predict nominal class labels. Classification has been used on many different domains such as financial market, business intelligence, healthcare, weather forecasting etc. In this research, a data set. from San-Francisco Open Data[8] is used which contains the reported criminal activities in the neighborhoods of the city San Francisco for a duration of 12 years. I used different classification techniques like Decision Tree, Naive Bayesian, Logistic Regression, k-Nearest Neighbor, Ensemble Methods to find hotspots of criminal activities based on the time of day. Results of different algorithms have been compared and most the effective approach has also been documented.

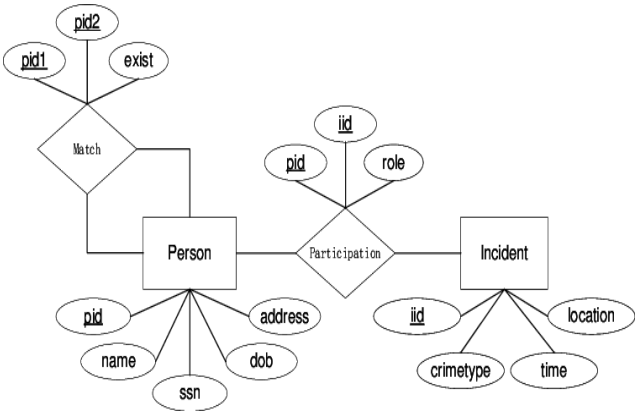
4.2 Data Flow Diagrams:



4.3 Flow Chart:

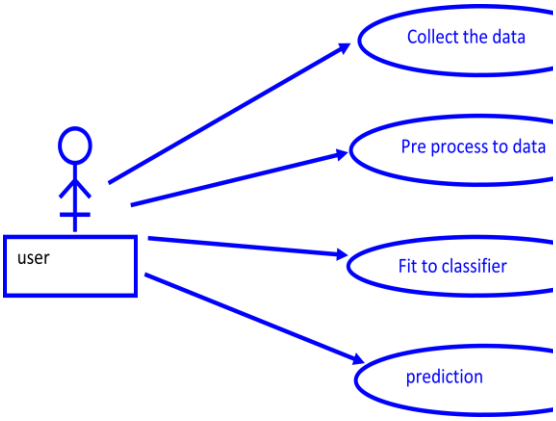


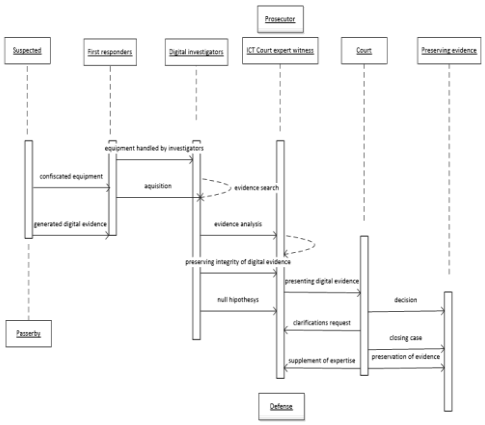
4.4 E-R Diagram



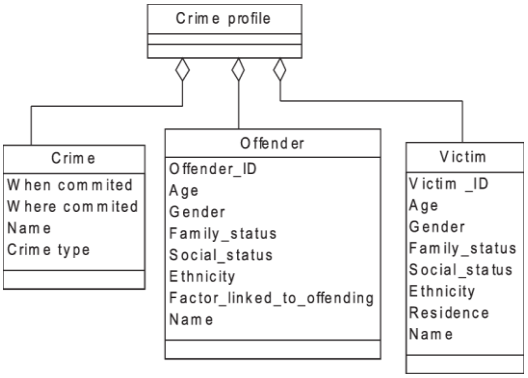
4.5 uml diagram

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

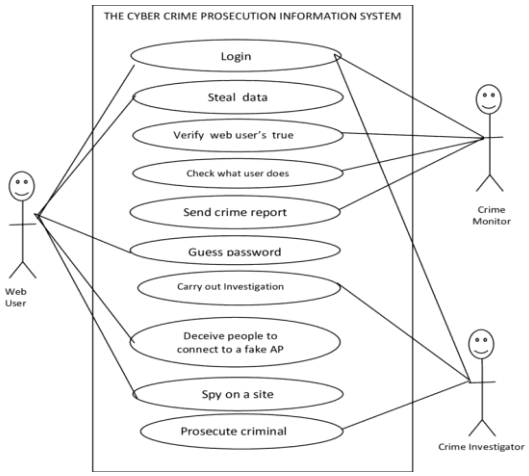




4.5.1 Class Diagram



4.5.2 Use Case Diagram:



4.5.3 Sequence diagram:

4.6 Input and Output Designs:

4.6.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1.Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3.When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

4.6.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2.Select methods for presenting information.

3.Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS :

Unit testing :
Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing :
Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test : Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

Valid Input: identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test :

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing :

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

Test strategy and approach Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

6.2 Integration Testing :

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

6.3 Acceptance Testing

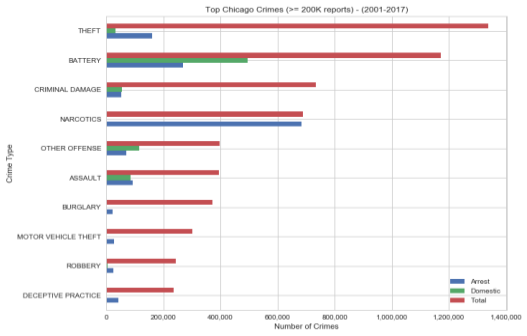
User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

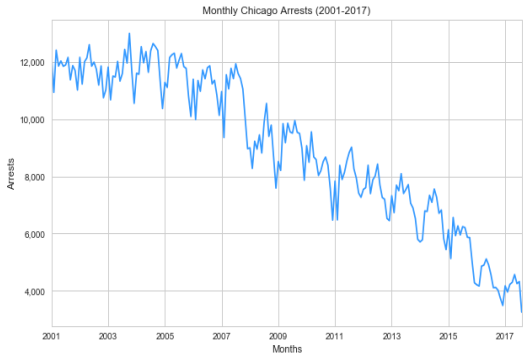
SCREENSHORTS

7.1 Chicago Crimes Graph

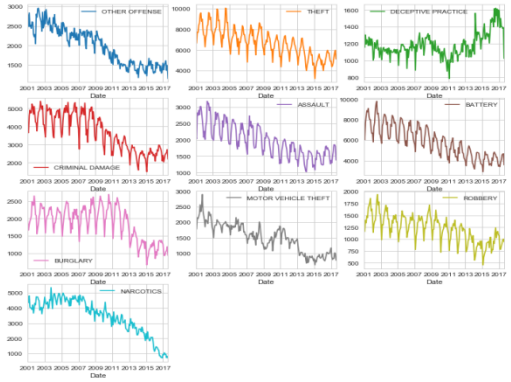
ISSN- 2394-5125 VOL 6, ISSUE 07, 2019



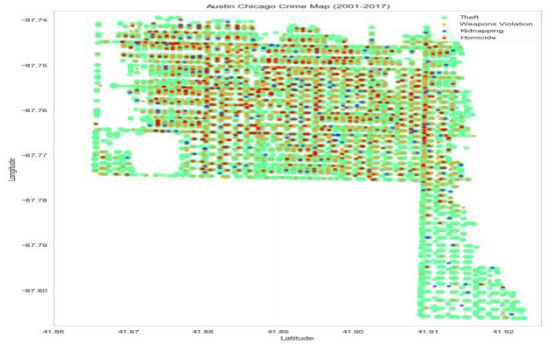
7.2 Monthly Chicago Arrests



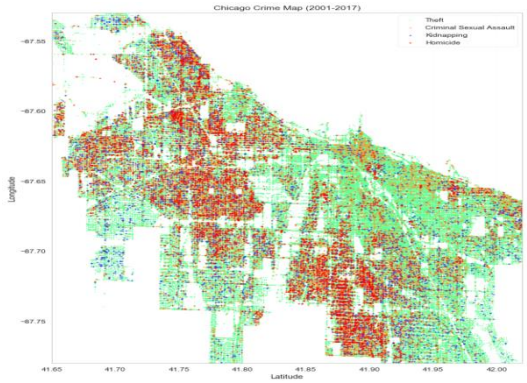
7.3 Various Crime Graphs



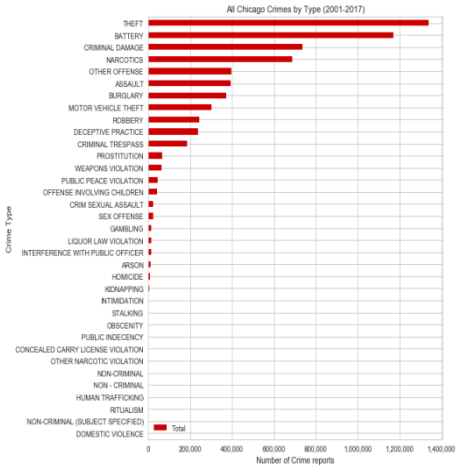
7.4 Austin Chicago Crime Map



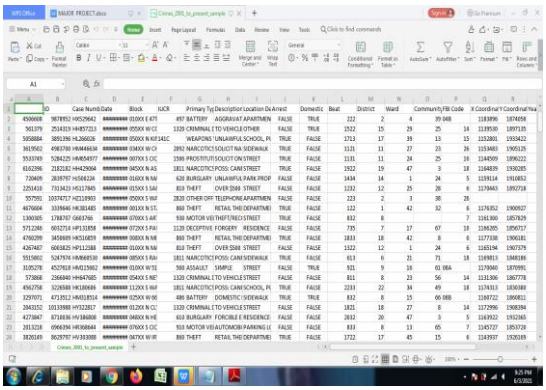
7.5 Chicago Crime Map



7.6 All Types Of Chicago Crimes

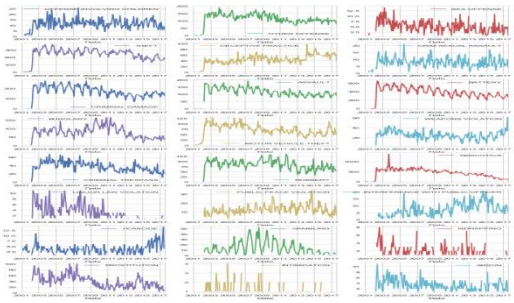


7.7 Crime Document



7.8 Previous Years Crime Graphs

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019



IMPLEMENTATION

8.1 Modules

1. **Decision Tree Classifier:** Decision tree classification model forms a tree structure from dataset. Decision tree is built by dividing a dataset into smaller pieces. At each step in the algorithm, a decision tree node is splitted into two or more branches until it reaches leaf nodes. Leaf nodes indicates the class labels or result. At each step, decision tree chooses a feature that best splits the data with the help of two functions: Gini Impurity and Information Gain. Gini Impurity measures the probability of classifying a random sample incorrectly if the label is picked randomly according to the distribution in a branch.
2. **Gaussian Naive Bayes:** Gaussian Naive Bayes is a supervised classifier that uses naive assumption that there is no dependency between two features. This classifier is implemented by applying Bayesian Theorem.
3. **Logistic Regression:** Logistic regression uses linear boundaries to classify data into different categories. Logistic regression can work on both binary and multiclass problems. For multiclass dataset, one vs the rest scheme is used. In this method, logistic regression trains separate binary classifiers for each class. Meaning, each class is classified against all other classes, by assuming that all other classes is one category.
4. **K-Nearest Neighbor:** Nearest Neighbors method is used in both supervised and unsupervised learning. While testing with new data, KNN looks at k data points in training dataset which are closest to the test data point. k indicates the number of neighbors voting to classify a datapoint. The distance can be measured with various metrics. Euclidean distance is the most common choice.
5. **Ensemble Methods:** Ensemble learning is a method of combining multiple learning algorithm together to achieve better performance over a single algorithm. Ensemble methods can be divided into two categories: averaging methods and boosting methods.

In this paper, two ensemble methods are used: Random Forest, which follows the principle of averaging method and Adaboost which is a boosting model.

- **Random Forest:** In this ensemble model several decision trees are built using samples drawn with replacement from the training set. The splitting of each node of a tree is not based on the best 19 split of all features, rather the best split among a random set of features.

- **Adaboost:** Adaboost or Adaptive Boosting is a boosting algorithm. Adaboost combines several weak learners to produce a stronger model. The final output is obtained from the weighted sum of the weak models. As it is a sequential process, in each step a weak learner is changed in favor of misclassified data points in previous classifiers.

CONCLUSION

Throughout the research it has been evident that basic details of a criminal activities in an area contains indicators that can be used by machine learning agents to classify a criminal activity given a location and date. Even though the learning agent suffers from imbalanced categories of the dataset, it was able to overcome the difficulty by oversampling and undersampling the dataset. Through the experiments, it can be seen the imbalanced dataset was benefitted by using ENN undersampling. Using the undersampled data, Adaboost decision tree successfully classified criminal activities based on the time and location. With a accuracy of 81.93%, it was able to outperform other machine learning algorithms. Imbalanced classes are one of the main hurdles to achieve a better result. Though the machine learning agent was able to predictive model out of simply crime data, a demographic dataset would probably help to further improve the result and solidify it.

Referencess

[1]Lakshman Narayana Vejendla and A Peda Gopi, (2019)," Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology", Revue d'IntelligenceArtificielle , Vol. 33, No. 1, 2019,pp.45-48.

ISSN- 2394-5125 VOL 6, ISSUE 07, 2019

[2]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), "Classification of tweets data

based on polarity using improved RBF kernel of SVM" . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409-

[3]. A Peda Gopi and Lakshman Narayana Vejendla, (2019)," Certified Node Frequency in Social Network Using Parallel Diffusion Methods", Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

[4]. Lakshman Narayana Vejendla and Bharathi C R ,(2018),"Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs", Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13- 1921-1_63 DOI: 10.1007/978-981-13-1921- 1_63

[5]. Lakshman Narayana Vejendla and Bharathi C R, (2018), "Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS", Modelling, Measurement and Control A, Vol.91,Issue.2, pp.73-76. DOI:10.18280/mmc_a.910207

[6]. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018)," Different

techniques for hiding the text information using text steganography techniques: A survey", Ingénierie des Systèmesd'Information, Vol.23, Issue.6,pp.115- 125.DOI: 10.3166/ISI.23.6.115-125

[7]. A Peda Gopi and Lakshman Narayana Vejendla (2018), "Dynamic load balancing for client server assignment in distributed system using genetic algorithm", Ingénierie des Systèmesd'Information, Vol.23, Issue.6, pp. 87-98. DOI: 10.3166/ISI.23.6.87-98