Crime Analysis & Prediction Using Machine Learning

Submitted as a partial fulfillment of Bachelor of Technology in Computer Science & Engineering
of
Maulana Abul Kalam Azad University of Technology
(Formerly known as West Bengal University of Technology)

Project Report

Submitted by

Name of Students	University Roll No.
Suprity Paul	**11600118015**
Anik Das Mondal	**11600118067**
Soumyadeb Dhar	**11600118021**
Indranil Sen	**11600118055**
Debjyoti Bose	**11600118058**

Under the supervision of

MR. SUROJIT BHATTACHARYYA

Assistant Professor, Computer Science & Engineering

Department of Computer Science & Engineering,
MCKV Institute of Engineering
243, G.T. Road(N)
Liluah, Howrah - 711204

Department of Computer Science & Engineering
MCKV Institute of Engineering
243, G. T. Road (N),
Liluah, Howrah-711204

CERTIFICATE OF RECOMMENDATION

I hereby recommend that the thesis prepared under my supervision by Suprity Paul, Anik Das Mondal, Soumyadeb Dhar, Indranil Sen & Debjyoti Bose entitled Crime Analysis & Prediction Using Machine Learning be accepted in partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science & Engineering Department.

Mr. Avijit Bose
Assistant Professor & Head of the Department,
Computer Science & Engineering Department.
MCKV Institute of Engineering,Howrah

Mr. Surojit Bhattacharyya
Assistant Professor,
Computer Science & Engineering Department.
MCKV Institute of Engineering, Howrah

MCKV Institute of Engineering
243, G. T. Road (N), Liluah
Howrah-711204

Affiliated to
Maulana Abul Kalam Azad University of Technology
(Formerly known as West Bengal University of Technology)

CERTIFICATE

This is to certify that the project entitled Crime Analysis & Prediction Using Machine Learning and submitted by

Name of students	University Roll No.
Suprity Paul	11600118015
Anik Das Mondal	11600118067
Soumyadeb Dhar	11600118021
Indranil Sen	11600118055
Debjyoti Bose	11600118058

has been carried out under the guidance of myself following the rules and regulations of the degree of Bachelor of Technology in Computer Science & Engineering of **Maulana Abul Kalam Azad University of Technology (**Formerly West Bengal University of Technology).

_____

(Signature of the project guide)

Mr. Surojit Bhattacharyya,
Assistant Professor,
Computer Science & Engineering Dept.

1. Suprity Paul
2. Anik Das Mondal.
3. Soumyadeb Dhar.
4. Indranil Sen
5. Debjyoti Bose.

MCKV Institute of Engineering
243, G. T. Road (N), Liluah
Howrah-711204
Affiliated to
Maulana Abul Kalam Azad University of Technology
(Formerly known as West Bengal University of Technology)

CERTIFICATE OF APPROVAL

(B.Tech Degree in Computer Science & Engineering)

This project report is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is to be understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed and conclusion drawn therein but approve the project report only for the purpose for which it has been submitted

COMMITTEE ON FINAL	1.
EXAMINATION FOR	2.
EVALUATION OF	3.
PROJECT REPORT	4.
	5.

Acknowledgement

Success of any project depends largely on the encouragement and guidelines of many others. We take this sincere opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project work.

We would like to show our greatest appreciation to Mr. Surojit Bhattacharyya. We always feel motivated and encouraged every time by his valuable advice and constant inspiration; without his encouragement and guidance this project would not have materialized.

Words are inadequate in offering our thanks to the other faculties for their encouragement and cooperation in carrying out this project work. The guidance and support received from all the members and who are contributing to this project, was vital for the success of this project.

Contents

Topics : **Page No.**

PROJECT ABSTRACT

The Main Objective of the Project is to analyze the Crimes occurred in the span of recent years in a city and Predict the Future Crimes based on the available records.

As per the plan, Machine Learning (ML) is to be used for analysis of available records from the dataset and further prediction of future crimes if any, is to be anticipated so that appropriate measures can be taken to prevent the same beforehand.

At first, Data with significant features are to be extracted from the dataset followed by preprocessing of the same. These processed data will further be used to train the proposed ML-Model using suitable ML-Algorithm to meet the criterion of the Model. After multiple times of training, the Model will be reached to an effective accuracy. Then it will be ready for testing and in the testing, phase predicted values will be generated.

1. INTRODUCTION

1.1 Background of the Project:

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively 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 infeasible to develop an algorithm of specific instructions for performing the task.

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 procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

The objective of our project is the analysis of crimes occurred over the years in Chicago and a prediction of future crimes based on that data. Also we will reach to a conclusion that whether the Chicago police dept. is doing a good job or not. We have used Machine Learning using Python and the method have undertaken is Logistic Regression so that we can easily classify and predict certain values. We can refer to the Crime Arrest Report graph and easily reach the necessary conclusion.

2. MACHINE LEARNING

2.1 What is machine learning?

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

2.2 Categories of Machine learning:

Machine learning tasks are classified into several broad categories.

- **Supervised learning** - 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.

- **Semi Supervised learning** - Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample input doesn't have labels.

- **Unsupervised Learning** - In unsupervised learning, the algorithm builds a mathematical model from a set of data which 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.

2.3 Some Common Applications of Machine learning:

- While using YouTube or Netflix, recommendations are received based on our interest, which is a basic example of machine learning implementation.
- Machine Learning can be used in Weather Forecasting and Stock Market Analysis.
- Machine Learning can be used to recognize user's voice and execute instructions given using the voice of a person, this technology is known as Voice Recognition.
- Machine Learning can be used for face detection and Image Recognition to automatically find the face of the person which matches it's database.
- Google Map gives us the route where traffic is less-application of machine learning.
- Google tracks your search and your search history and recommends ads based on your interest which is determined using the search history.
- In this project, detection is done weather any crime will take place or not, at a certain place at a certain time based on previous records with the help of machine learning.

2.4 Regression and Classification in Machine Learning:

Data scientists use many different kinds of machine learning algorithms to discover patterns in big data that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they "learn" about data to make predictions: supervised and unsupervised learning.

Supervised Machine Learning: The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output Y= f(X). The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

Techniques of Supervised Machine Learning algorithms include linear and logistic regression, multi-class classification, Decision Trees and support vector machines. Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. For example, a classification algorithm will learn to identify animals after being trained on a dataset of images that are properly labeled with the species of the animal and some identifying characteristics.

Linear Regression:

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple near regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables (or predictors) is assumed to be an affine function of those values: less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

Logistic Regression:

Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).

In logistic regression, the dependent variable is binary or dichotomous, ie. it only contains data coded as I (TRUE, success, pregnant, etc.) or 0 (FALSE, failure, non-pregnant, etc.). The goal of logistic is to find the best fitting (yet biologically reasonable) model to describe the relationship

between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a logit transformation of the probability of presence of the characteristic of interest:

$$\boldsymbol{logit(p)= b_0 + b_1X_1 + b_2X_2+ b_3X_3+ b_4X_4+\ldots\ldots\ldots+ b_kX_k}$$

where p is the probability of presence of the characteristic of interest. The logit transformation is defined as the logged odds:

odds = (p/1-p) = probability of presence of characteristic / probability of absence of characteristic

and

logit (p) = ln(p/1-p)

Rather than choosing parameters that minimize the sum of squared errors (like in ordinary regression), estimation in logistic regression chooses parameters that maximize the likelihood of observing the sample values.

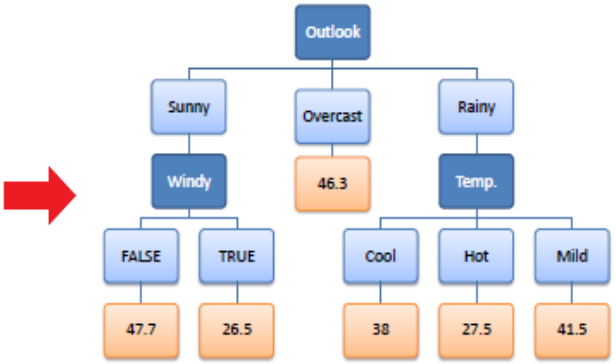
Advantages of Logistic Regression:

- Logistic regression is easier to implement, interpret, and very efficient to train.
- It makes no assumptions about distributions of classes in feature space.
- Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.

Decision Tree Algorithm:

Decision Tree is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

Predictors				Target
Outlook	Temp.	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	45
Sunny	Cool	Normal	False	52
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	46
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30



3. PYTHON

3.1 What is Python?

Python is a popular 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.

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

3.2 Applications of Python:

- Machine Learning, Artificial Intelligence and Data Science.
- Web Development
- Statistics and Advanced Mathematics.
- Software Development.
- Game Development.

3.3 Reasons for Using Python:

1. While writing a software application, you must focus on the quality of its source code to simplify maintenance and updates. The syntax rules of Python allow you to express concepts without writing additional code. At the same time, Python, unlike other programming languages, emphasizes on code readability, and allows you to use English keywords instead of punctuations.

2. Like other modern programming languages, Python also supports several programming paradigms. It supports object oriented and structured programming fully. Also, its language features support various concepts in functional and aspect-oriented programming.

3. Its large and robust standard library makes Python score over other programming languages. The standard library allows you to choose from a wide range of modules according to your precise needs. Each module further enables you to add functionality to the Python application without writing additional code.

4. As an open source programming language, Python helps you to curtail softwar development cost significantly. You can even use several open source Python frameworks, libraries and development tools to curtail development time without increasing development cost.

4. MACHINE LEARNING USING PYTHON

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this article, we'll see basics of Machine Learning, and implementation of a simple machine learning algorithm using python.

Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to computer whether a picture is cat or not. Then if we show the computer a new image. Then from the above training, computer should be able to tell whether this new image is cat or not.

4.1 Setting up the environment:

Python community has developed many modules to help programmers implement machine learning. In this article, we will be using NumPy, scipy and scikit-learn modules.

We can install them using cmd command:

pip install numpy scipy scikit-learn

A better option would be downloading miniconda or anaconda packages for python, which come prebundled with these packages. Follow the instructions given here to use anaconda.

4.2 Best Python Libraries for Machine Learning:

Here we will discuss about the best python libraries and its applications. We have used these libraries in our project.

- **NumPy** is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities.

- **Scikit-learn** is one of the most popular ML libraries for classical ML. algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with ML.

- **Pandas** is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

- **Matplotlib** is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc.

5. CRIME ANALYSIS AND PREDICTION USING ML

5.1 Objective:

The objective of our project is the analysis of crimes occurred over the years in Chicago and a prediction of future crimes based on that data. Also, we will reach to a conclusion that whether the Chicago police dept. is doing a good job or not.

5.2 About the project:

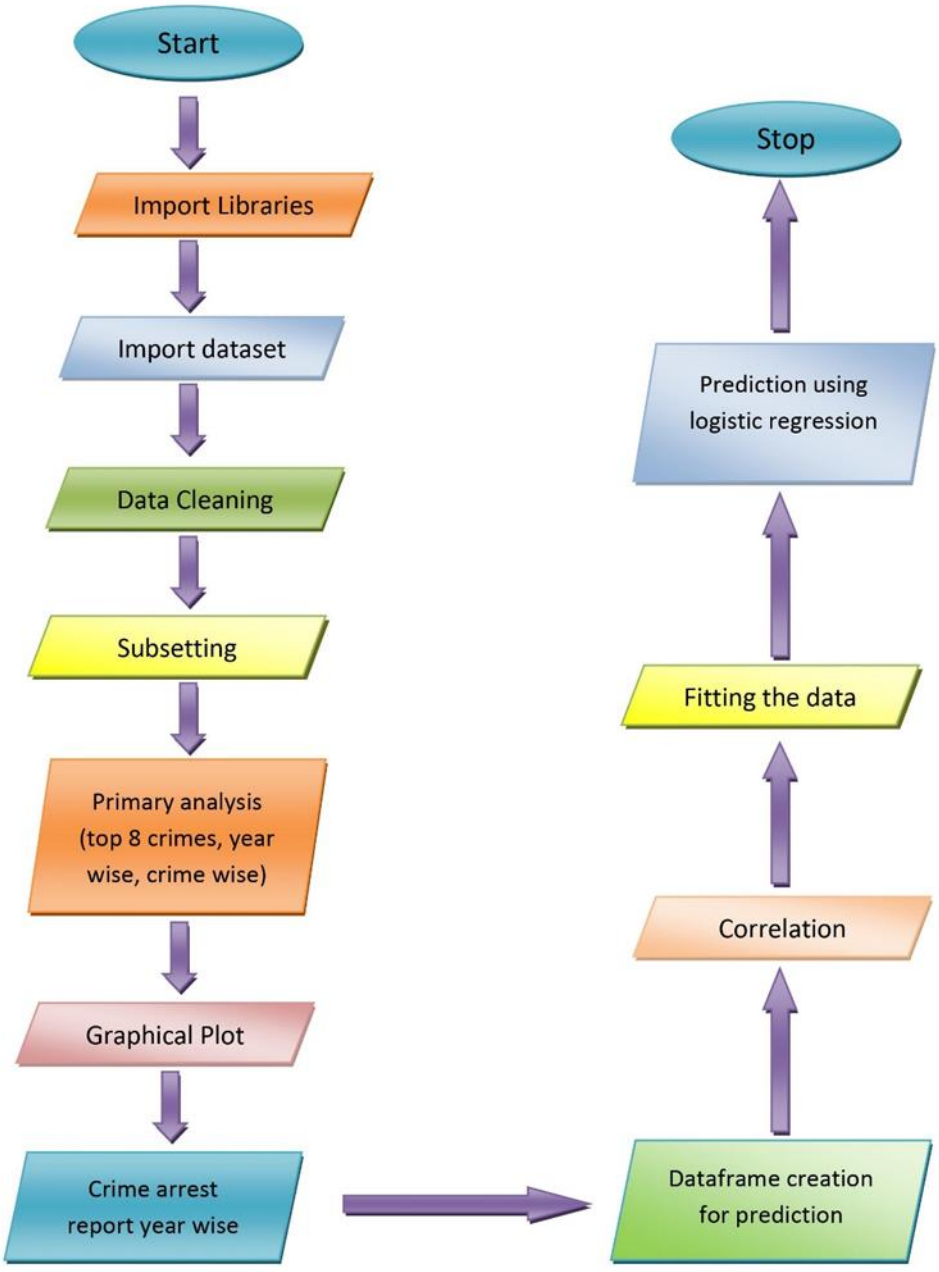
We have used Machine Learning using Python and the method we have undertaken is Logistic Regression- so that we can easily classify and predict certain values. After importing all the necessary libraries needed to mend this project into a success, we fetch the data of Chicago Crime. We can see various columns like- ID, Case Number, Date. Block, IUCR, Primary Type, Description, Arrest, Domestic, Beat, District, FBI Code. Year and Updated on.

Here ID refers to a particular crime. Case number is the unique identification number to pin point an individual crime of a certain type. The Date basically tells us the Date and Time of the occurence of the crime. Block gives away the exact location (Street no., PIN etc.) where the crime took place. IUCR stands for Illinois Uniform Crime Report-Governmental police & the number suggests active police force present at the time of the crime. Under the primary type's umbrella comes a miscellaneous offense, offense involving Children, Theft, Murder, Assault etc.

Under the description column the type of the primary crime is properly explained. The Arrest column tells us the wrong doer was caught or not- if yes, true. If no, false. The domestic column suggests if the type of crime can be labelled as domestic or not. The Beat value is a district code for different police stations. The FBI (Federal Buracu of Investigation) has a unique code to identify the crimes. The year column obviously tells us the year when the crime occurred. Updated on tells us when the crime data was updated again- to be specific-the time and date.

Moving on, we can see we have made a primary analysis and overview in the shape of a bar chart of top 8 crimes over the different years that even a layman can easily and effortlessly understand. Like for example, we can see within 2001-2004 ARSON' was the crime that occurred most in no-2,12,361. Thus we can easily visualize the data. The name of the bar chart to be referred is namely Crime Scenario. Then we have done a crime analysis of different crimes- for example we can see how the number of crime ARSON has varied over the years. Then we have done a year wise analysis of the top 8 crimes happening on that said year. Like for example, ARSON was the biggest crime in Chicago in the year 2001. The value being a whopping 98761. Then other than bar chart, we also have graphically plotted the trend of a particular crime over the years, like we can see ARSON Crime Rate graph over here. After that we have done a minute comparative analysis of 19 years of crime and made a crime arrest report year wise. Now finally we have used an efficient logistic recursion algorithm to find the total fraction/values of arrested or not arrested to judge whether the Chicago Police Dept. is doing a good job or not and if crime rate is going to decrease in Chicago in the coming years or not. We can refer to the Crime Arrest Report graph and easily reach the necessary conclusion.

5.3 Data Flow Diagram:



5.4 Model Deployment Pathway:

Let us take a sample raw dataset given below:

A	B	C	D	E	F	G	H	I	J	K	L	M
ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	District	Ward
12584710	JF103542	12/31/2021 11:59:00 PM	006XX N STATE ST	2250	LIQUOR LAW VIOLATION	LIQUOR LICENSE VIOLATION	HOTEL / MOTEL	TRUE	FALSE	1832	18	42
12581979	JF100042	12/31/2021 11:58:00 PM	082XX S MARQUETTE AVE	650	BURGLARY	HOME INVASION	APARTMENT	FALSE	TRUE	423	4	7
12584234	JF102969	12/31/2021 11:55:00 PM	087XX S WOOD ST	1310	CRIMINAL DAMAGE	TO PROPERTY	RESIDENCE	FALSE	FALSE	2221	22	21
12581975	JF100045	12/31/2021 11:50:00 PM	068XX S CHAMPLAIN AVE	1365	CRIMINAL TRESPASS	TO RESIDENCE	APARTMENT	FALSE	FALSE	321	3	20
12582027	JF100292	12/31/2021 11:50:00 PM	051XX S SPAULDING AVE	560	ASSAULT	SIMPLE	APARTMENT	FALSE	FALSE	822	8	14

M	N	O	P	Q	R	S	T	U	V
Ward	Commun	FBI Code	X Coordir	Y Coordir	Year	Updated On	Latitude	Longitude	Location
42	8	22	1176247	1904872	2021	2022 Jan 08 03:45:56 PM	41.89432785	-87.62814321	(41.894327846°, -87.62814321°)
7	46	5	1195601	1850868	2021	2022 Jan 07 03:47:10 PM	41.74568005	-87.5588512	(41.745680051°, -87.558851198°)
21	71	14	1165870	1846697	2021	2022 Jan 07 03:47:10 PM	41.73491689	-87.66790905	(41.734916889°, -87.66790905°)
20	42	26	1181741	1859870	2021	2022 Jan 07 03:47:10 PM	41.77071358	-87.60935864	(41.770713582°, -87.609358643°)
14	63	08A	1155247	1870367	2021	2022 Jan 07 03:47:10 PM	41.80008984	-87.70619529	(41.800089835°, -87.706195287°)

This data is retrieved from https://data.cityofchicago.org/ which demonstrates the data of crimes occurred in Chicago recently. After further data cleaning we will utilize the processed dataset to predict and analyze the crime rates using our ML algorithms.

According to the data flow diagram the steps are as follows:

- **Import Libraries**: We have to import the libraries like Pandas, Matplotlib, Skit-Learn, NumPy etc.
- **Import Dataset**: We have to import the dataset using ‘read_csv()’ method under panda library
- **Data Cleaning**: Here above, in the given raw dataset there are certain columns which we are not required while predicting and analyzing the crime using ML. The columns are as follows:

  Case Number, Domestic, Beat, District, Ward, Community, FBI Code, X and Y Coordinate, Year, Updated on, Latitude, Longitude, Location.

After performing data cleaning the processed dataset will be:

A	B	C	D	E	F	G	H	I
ID	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic
12584710	12/31/2021 11:59:00 PM	006XX N STATE ST	2250	LIQUOR LAW VIOLATION	LIQUOR LICENSE VIOLATION	HOTEL / MOTEL	TRUE	FALSE
12581979	12/31/2021 11:58:00 PM	082XX S MARQUETTE AVE	650	BURGLARY	HOME INVASION	APARTMENT	FALSE	TRUE
12584234	12/31/2021 11:55:00 PM	087XX S WOOD ST	1310	CRIMINAL DAMAGE	TO PROPERTY	RESIDENCE	FALSE	FALSE
12581975	12/31/2021 11:50:00 PM	068XX S CHAMPLAIN AVE	1365	CRIMINAL TRESPASS	TO RESIDENCE	APARTMENT	FALSE	FALSE
12582027	12/31/2021 11:50:00 PM	051XX S SPAULDING AVE	560	ASSAULT	SIMPLE	APARTMENT	FALSE	FALSE

- **Subsetting**: We will create subsets of the dataset for training, validating, and testing our model. A subset allows us to reduce the size of the dataset, and to split the examples into disjoint sets for the purpose of training, validation, and testing.
- **Primary Analysis**: We are planning to do next analysis based on the top 8 crimes. Three kinds of primary analysis will be done---count of crimes in a given year range, year-wise and crime wise.
- Based on the primary analysis we will plot the graphs and create the corresponding dataframes for prediction.
- Next, we will see how one or more variables are correlated to each other and plot the map accordingly.
- Then we will have to fit the data and use logistic regression model for prediction.

Thus, this is how we are planning to prepare our project and we will keep updating our solutions and techniques for the best outcome.

6. CONCLUSION

Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to computer whether a picture is cat or not. Then if we show the computer a new image, then from the above training, computer should be able to tell whether this new image is cat or not.

In our project, we will be using an efficient logistic regression algorithm to find the total fraction/values of arrested or not arrested to judge whether the Chicago Police Dept. is doing a good job or not and if crime rate is going to decrease in Chicago in the upcoming years or not. We can refer to the Crime Arrest Report graph and easily reach the necessary conclusion.

7. FUTURE SCOPE OF THE PROJECT

This project presented the techniques and methods that can be used to predict crime and help law agencies. The scope of using different methods for crime prediction and prevention can change the scenario of law enforcement agencies. Using ML, we can substantially impact the overall functionality of law enforcement agencies. In the near future, by combining ML and computer vision, along with security equipment such as surveillance cameras and spotting scopes, a machine can learn the pattern of previous crimes, understand what crime actually is, and predict future crimes accurately without human intervention. A possible automation would be to create a system that can predict and anticipate the zones of crime hotspots in a city. Law enforcement agencies can be warned and prevent crime from occurring by implementing more surveillance within the prediction zone. This complete automation can overcome the drawbacks of the current system, and law enforcement agencies can depend more on these techniques in the near future. Designing a machine to anticipate and identify patterns of such crimes will be the starting point of our future study. Although the current systems have a large impact on crime prevention, this could be the next big approach and bring about a revolutionary change in the crime rate, prediction, detection, and prevention, i.e., a “universal police officer”.

8. REFERENCES

1. Chung-Hsien Yu, Max W. Ward, Melissa Morabito, and Wei Ding : “Crime Forecasting using Data Mining Techniques”, IEEE, 2011

2. S.Sivaranjani, Dr. S. Sivakumari, Aasha M. : “Crime Detection and Forecasting in Tamil Nadu using Machine Learning Techniques ”, IEEE, 2016

3. Masila Abdul Jalil, Chia Pui Ling, Noor Maizura Mohamad, Noor Fatihah Mohd : “Knowledge Representation Model for Crime Analysis”, Elsevier, 2017

4. Al Boni, M., & Gerber, M. S. (2016, December) : “AreaSpecific Crime Prediction Models. In Machine Learning and Applications (ICMLA)”, 2016 15th IEEE International Conference on (pp. 671-676). IEEE

5. Zhao, X., & Tang, J. (2017, November) : “Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW)” , 2017 IEEE International Conference on (pp. 1158-1159). IEEE