Crime rate prediction

**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 require 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 robbery crime and predicting the percapita of robbery 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 Andhra pradesh crime data set. It consists of crime information like year ,population, robberies ,and robberies percapita. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtain will be high .we use multi-linear regression to predict.

**1.Introduction**

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 is 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.

**PYTHON:**

Python, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks. The simple syntax rules of the programming language further makes it easier for you to keep the code base readable and application maintainable. Main reasons to use python language is:

1. Readable and Maintainable Code

2. Multiple Programming Paradigms

3. Compatible with Major Platforms and Systems

4.  Robust Standard Library

5. Many Open Source Frameworks and Tools

6. Simplify Complex Software Development

7. Adopt Test Driven Development.

**Machine learning**

Machine learning is an subset 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. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Basically Machine Learning is of three types:

1.Supervised Learning: Learning from characterized data.

2.Unsupervised Learning: Learning from raw data.

3.Reinforcement Learning: Learning from self mistakes (or) self learning data.

Predictive modeling is the way of building a model that is capable of making predictions. The process includes a machine learning algorithm that learns certain properties from a training dataset in order to make those predictions. Predictive modeling can be divided further into two areas: Regression and pattern classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables. In contrast to regression models ,the task of pattern classification is to assign discrete class labels to particular data value as output of a prediction Here we used multiple linear regression algorithm to predict the future crime rate.

**Steps involved in Machine learning**

Steps involved in the machine learning is shown in the below figure

1.Data collection

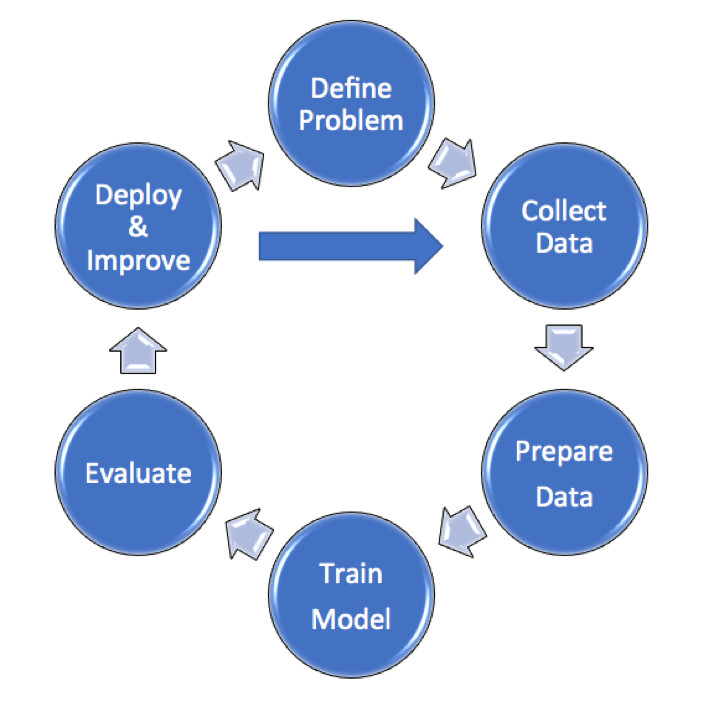
2.Data wrangling

3. Analyse the data

4.Train the algorithm

5.Test algorithm

6. Deployment



**Data collection:**

Here we collect the data from the police departments and crime departments and we use previous data and predict the future crime rate.

**Data Preprocessing:**

In this phase, the data is prepared for the analysis purpose which contains relevant information. Pre-processing and cleaning of data are one of the most important tasks that must be one before dataset can be used for machine learning. The real-world data is noisy, incomplete and inconsistent. So, it is required to be cleaned.

**Extraction of Feature Set/Training Data**

The feature sets and training set that has obtained by using any of the methods will be used for the implementation of machine learning algorithms.

**Implementation of Machine Learning Algorithm on Feature Set/Training Data**

**Regression:**

A Regression model is created when we want to find out a number – for example

how many days before a patient discharged from hospital with a chronic condition such as

diabetes will return.

**Classification:**

To determine a label or category – it is either one thing or another. We train the model using a set of labelled data.

**Testing of Data**

Testing of data is done based on training model which is classified using supervised learning algorithm. Evaluation of the total responses for every question and determine the polarity of feedback received in context of the given data.

**Multiple Linear Regression**

Multiple linear regression  is the most common form of linear regression analysis.  As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables.  The independent variables can be continuous or categorical (dummy coded as appropriate). The multiple linear regression algorithm is having one and more independent variables and one dependent variable. Here it consist of hyperplane i.e is all the straight lines joined together is called hyperplane. The iterations are performed upto a0=0. The equation of the multiple linear regression y=a0+a1\*x1+a2\*x2........

**1.2.Objective of research**

The above problem made me to go for a research about how can we reduce a crime for making easier. Through many documentation and cases, it came out that machine learning and data science can make the work easier and faster. The objective of this project is to analyze dataset which consist of robberies crimes per year and predicting the rate of crime which may happen in future depending upon dataset consisting of previous years data.

**1.3.Problem Statement**

Finding out the future prediction of crime rate for robberies by using Machine learning algorithm using python as core.

**1.4.Industry profile**

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 central bureau of investigation and crime departments to control and reduce the crime activities. Crime prediction is the the major problems to the police department and central bureau of investigation and crime departments as there are tremendous amount of crime data that exist.

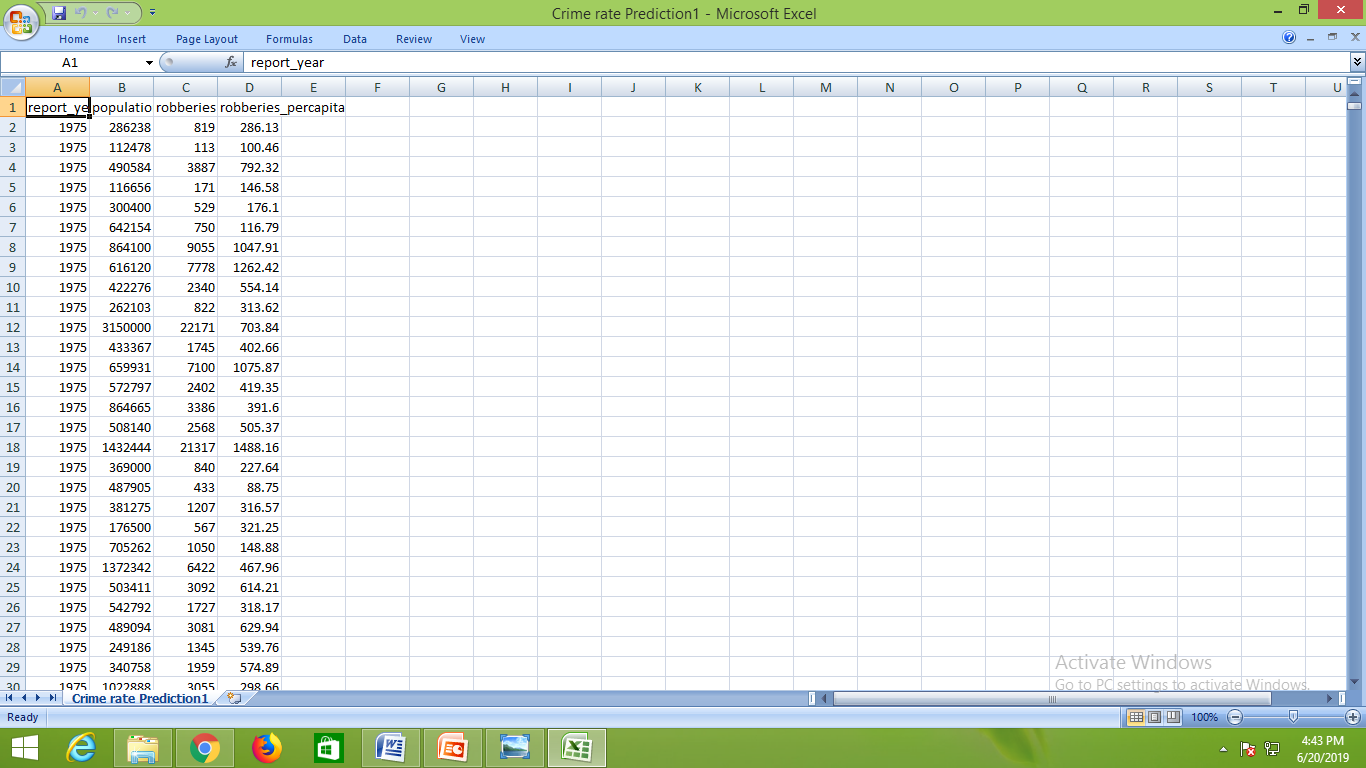
**2.Review of literature**

**www.kaggle.com**

The dataset used in this project is taken from Kaggle.com. The dataset obtained from kaggle is maintained and updated by the Andhra pradesh police department. Crime dataset from kaggle is used in CSV format.

**3.Data Collection:**

The data is collected and uploaded in the Jupiter notebook in Anaconda navigator. we use pandas , numpy and we declare pandas as pd and numpy as np. And we assign dataset to dataset variable by using "pd.read\_csv" following with dataset name and we print the dataset. And next we slice the data with independent variables as x and dependent variable as y. here the independent variables are year, population ,robberies and robberies percapita as dependent variable. Here we use "iloc" to slice and we convert the data set into arrays using ".values". we use corr() to find the co-relation of the dataset.



**4.Methodology**

**4.1 Statistical techniques :**

. . Multiple linear regression is the most common form of linear regression analysis.  As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable and two or more independent variables.  The independent variables can be continuous or categorical (dummy coded as appropriate)The multiple linear regression algorithm is having one and more independent variables and one dependent variable. Here it consist of hyperplane i.e is all the straight lines joined together is called hyperplane. The iterations are performed upto a0=0. The equation of the multiple linear regression y=a0+a1\*x1+a2\*x2........

From "sklearn.model\_selection" we import "train\_test\_split " ,and we declare variables x\_train, x\_test, y\_train,y\_test and these are assigned to train\_test\_split of x,y with test\_size =0.2, and we declare here random\_state as zero.

Random Sampling (Train and Test)

Training Sample: Model will be developed on this sample. 70% or 80% of the data goes here.   Test Sample: Model performances will be validated on this sample. 30% or 20% of the data goes here.

**4.2 Data modeling**

From sklearn.linear\_model we import the linear regression and we assign "lr" variable with linear regression and then we give the splitted values to the model with "fit" .to predict the values we use predict. Here we are giving inputs as year and population and robberies and we get robberies percapita .

Here we found out the by giving input as year 2025 and population 385625 and robberies as 437 and we get prediction as 330.95

**5.conclusion:**

With the help of machine learning technology, it has become easy to find out relation and patterns among various data’s. The work in this project mainly revolves around predicting the future crime rate. Using the concept of machine learning we have built a model. We generated the future crime rate value which helps the police departments and central bureau of investigation and crime departments to decrease the crime rate in that years.