

A Thesis/Project/Dissertation Report

On
CRIME RATE PREDICTION USING MACHINE LEARNING

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

*Bachelor of Technology in Computer
Science and Engineering*



(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of

Dr. Harshvardhan Choudhary

Submitted By

Cyrus Justo

(19SCSE1010777)

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
OCTOBER,2022**



**SCHOOL OF COMPUTING SCIENCE AND
ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA**

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled
“CRIME RATE PREDICTION USING MACHINE LEARNING
” in partial fulfillment of the requirements for the award of the B.Tech submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Dr. Harshvardhan Choudhary, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Cyrus Justo,19SCSE1010777

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. Harshvardhan
Choudhary

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Cyrus Justo (19SCSE1010777) and has been held on _____ and their work is recommended for the award of B.Tech.

Signature of Examiner(s)

Signature of Supervisor(s)

Signature of Project Coordinator

Signature of Dean

Date:

Place: Greater Noida

Abstract

Crime investigation and prediction is a precise methodology for breaking down and recognizing various examples, connections and patterns in crime. The locales with high likelihood of event of crime is predicted by the framework. The framework created will assist with accelerating the most common way of settling crime for the law authorization offices. The utilization of AI and machine learning to identify crime by means of sound or cameras presently exists, is demonstrated to work, and expected to keep on growing. The utilization of AI/ML in anticipating crime or a singular's probability for committing a crime has guarantee yet is even a greater amount of an unknown. The current information from the police is utilized which subsequent to utilizing diverse prediction and clustering algorithms gives an understanding that will assist with foreseeing the probability of incidents, track crimes and help the law authorization specialists to convey assets and furthermore settle crime cases at a quicker rate. Upgrades in crime prevention innovation will probably prod expanded all out spending on this innovation. We likewise attempt to make our characterization task more significant by consolidating different classes into bigger classes. At last, we report and think about our outcomes with various classifiers, and well on roads for future work.

Table of Contents

Title		Page No.
Candidates Declaration		1-3
Acknowledgement		4
Abstract		5
CONTENTS		7-8
List of tables		9
List of Figures		10
Chapter 1	Introduction	
	1.1 Introduction	
	1.2 MACHINE LEARNING	
	1.3 Objectives of The Project	
	1.4 Problem Statement	
	1.5 Modules Description	
	1.5.1 Data set	
	1.5.2 Data Abstraction	
	1.5.3 Test Dataset	
	1.5.4 Result Evaluation	

Chapter 2 Literature Review

2.1 Machine Learning

2.1.1 Supervised Learning

2.1.2 Unsupervised Learning

2.1.3 Reinforcement Learning

2.2 Proposed Methodology

2.3 Software Description

2.3.1 Jupyter Notebook

Chapter 3 Project Design

3.1 Design Goals

3.1.1 Data Collection

3.1.2 Data Collection

3.1.3 Training Model

3.2 System Architecture

3.3 Use case Diagram

3.4 Data flow Diagram

Chapter 4

Results and Conclusion

4.1 Results

4.2 Conclusion

Reference

List of Figures

S.No.	Title
1	Supervised Learning
2	Unsupervised Learning
3	Reinforcement Learning
4	System Flow
5	System Architecture
6	Use Case Diagram
7	Data flow Diagram

CHAPTER 1

INTRODUCTION

1.1 INFORMATION

Crimes are the significant threat to the humankind. There are many crimes that happen in regular intervals 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.

Through many documentations and cases, it came out that machine learning and data science can make the work easier and faster. The aim of this project is to make crime prediction using the features present in the dataset. The dataset is extracted from the official sites. With the help of machine learning algorithm, using python as core we can predict the type of crime which will occur in a particular area with crime percapita.

The objective would be to train a model for prediction. The training would be done using Training data set which will be validated using the test dataset. The Multi Linear Regression (MLR) will be used for crime prediction. Visualization of dataset is done to analyze the crimes which may have occurred in a particular year and based on population and number of crimes. This work helps the law enforcement agencies to predict and detect the crime_per capita in an area and thus reduces the crime rate.

1.2 MACHINE LEARNING:

Machine Learning is a sub-area of artificial intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by recognizing patterns in databases. In other words: Machine Learning enables IT systems to recognize patterns on the basis of existing algorithms and data sets and to develop adequate solution concepts. Therefore, in Machine Learning, artificial knowledge is generated on the basis of experience.

In order to enable the software to independently generate solutions, the prior action of people is necessary. For example, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the following tasks by Machine Learning:

1.3 OBJECTIVES OF THE PROJECT

- The main objective of the project is to predict the crime rate and analyse the crime rate to be happened in future. Based on this Information the officials can take charge and try to reduce the crime rate.
- The concept of Multi Linear Regression is used for predicting the graph between the Types of Crimes (Independent Variable) and the Year (Dependent Variable)
- The system will look at how to convert crime information into a regression problem, so that it will help detectives in solving crimes faster.
- Crime analysis based on available information to extract crime patterns. Using various multi linear regression techniques, frequency of occurring crime can be predicted based on territorial distribution of existing data and Crime recognition.

1.4 PROBLEM STATEMENT

The main problem is that day to day the population is going to be increased and by that the crimes are also going to be increased in different areas by this the crime rate can't be accurately predicted by the officials. The officials as they focus on many issues may not predict the crimes to be happened in the future. The officials/police officers although they tries to reduce the crime rate they may not reduce in full-fledged manner. The crime rate prediction in future may be difficult for them.

There has been countless of work done related to crimes. Large datasets have been reviewed, and information such as location and the type of crimes have been extracted to help people follow law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. There are several maps applications that show the exact crime location along with the crime type for any given city (see Figure 1). Even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future.

CHAPTER 2

LITERATURE REVIEW

Crime rate prediction is different in various applications, some of the studies are given below: C.P. Chaithanya, N. Manohar, Ajay Bazil Issac, describes Text detection is the method of locating areas in a picture wherever, text is present. Text detection and classification in natural pictures is very important for several computer vision applications like optical character recognition, distinguish between human and machine inputs and spam removal. Currently the challenge in text identifying is to detect the text in natural pictures due to many factors like, low- quality image, unclear words, typical font, image having a lot of color stroke than the background color, blurred pictures due to some natural problems like rain, sunny, snow, etc. The main aim of this work is to identify and classify the text in natural pictures. Here system detects the text and finds the connected regions, chain them together in their relative position. Uses a text classification engine to filter chains with low classification confidence scores.

Shiju Sathyadevan, Devan M.S, proposed that Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it .It is only within the last few decades that the technology made spatial data mining a practical solution for wide audiences of Law enforcement officials which is affordable and available. Since the availability of criminal data or records is limited we are collecting crime data from various sources like web sites, news sites, blogs, social media, RSS feeds etc. This huge data is used as a record for

creating a crime record database. So the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively.

PAC Duijn, V Kashirin, proposed, policymakers and law enforcement agencies across the globe struggle to find effective strategies to control criminal networks. The effectiveness of disruption strategies is known to depend on both network topology and network resilience. However, as these criminal networks operate in secrecy, data-driven knowledge concerning the effectiveness of different criminal network disruption strategies is very limited. By combining computational modeling and social network analysis with unique criminal network intelligence data from.

Nishat Shama proposed that Criminal activities are present in every region of the world affecting quality of life and socioeconomical 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 dataset from San- Francisco Open Data is used which contains the reported criminal activities in the neighbour hoods 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 Neighbour, 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.

Riya Rahul Shah proposed that Vancouver is most populated city in Canada. It is most ethnically diverse cities in Canada. Crime is one of the biggest and dominating problem in our society and its prevention is an important task. Even though Vancouver known to be the safest city it is observed that vehicle breakings and many more thefts is still a problem. 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.

2.1 MACHINE LEARNING

One of the finest word heard in today time is Machine Learning. Either it be at work or different places the machine learning has been an integral part of todays technology. Though its evolutionalizing and developing in a rapid rate and development and deployment of the same is still in progress. The machine learning itself had brought a random changes in today worlds because of which automation is in frame which was a mare existence in the past.

It's an aspiring term in todays time. One of the move that all the firm are interested into. It's a leading pillar for tomorrow leading the world to a better future of evolution where the customization and labor work can be reduce to half and the safety of the survival can be with

held to stand tall for the better utilization of human mind. Keeping that in picture it's been a hazard to many more in terms of irrespective field of interest. Since Machine is considered most efficient and the level of mistakes are kept at the minimum the level of work flow can be a work of hazard and further improvement on the same may create a thousands sitting idle in home creating a larger impact on unemployment and livelihood. Which in other way is a threat to the society too.

ML is the abbreviation for Machine Learning. In other word it is making a human mind fitting inside a machine which uses the same to perform the task of thousands. Machine Learning deals with the higher aspects of learning techniques and algorithm which are highly aligned to make the work flow seamlessly effortless with the human tendency of doing work.

Algorithm of such are improvising in nature which learns by themselves and fit themselves in the world of impairment by getting the required data and adjusting with the same giving the effective results out of the same. ML is a subsidiary or the subset of an AI(Artificial Intelligence). It is a mathematical model where computation of the testcases plays the major role in driving of the results.

A wide level of machine learning architecture are implemented today to turn on the yield factor and make people life more efficient in terms of livelihood. Various use of such in Message Filtering like spams, Trash automation are automated and carried out by the same. Since the efficiency is way more than a human tendency. Multi-tasking and processing is also initiated by the same giving a dual output which a human can never ever possibly be able to.

Statistics is the major key role in driving the machine learning in figure. It deals with computation of statistics in a wide range view and processing the same to give an data driven output causing it more sensible and resourcesable. Not only to the same it optimizes the resources and the efficiency is unbitable and reliable in terms of any means.

Though its being evolutionary in nature but it has integrated itself well with the terms of computational and digitalization. Various computational fields like Data Mining, Statistical Analysis, Optimization of resources, Automation are a major part of it. Here the machine has

the capacity to process the result on its own as same as the human bring. This process can be initiator as well as the derivable. The statistical flow is mainly reasonable with data driven pattern even the unstructured or the semi-structured data can be processed and approximate answer to the same can be derived.

All the equations are derived and the closest value to it's aligned field is found and the proximity is determined.

The classification of the same can be listed as follows:

2.1.1 SUPERVISED LEARNING

Supervised Learning deals with the supervision of the machine to derive the necessity input required. It's a mathematical model where the inputs and output of the same is already known and its passed to the machine to get the expected output so that the efficiency is determined and this is the learning phase for the machine. Here the feeding and derivation of the same is measured.

Here the machines filters the inputs learns from the functional unit. Compute it and stores it into its memory for further process and if found a matching pattern it uses the same and learns from it and plot a result out of the same.

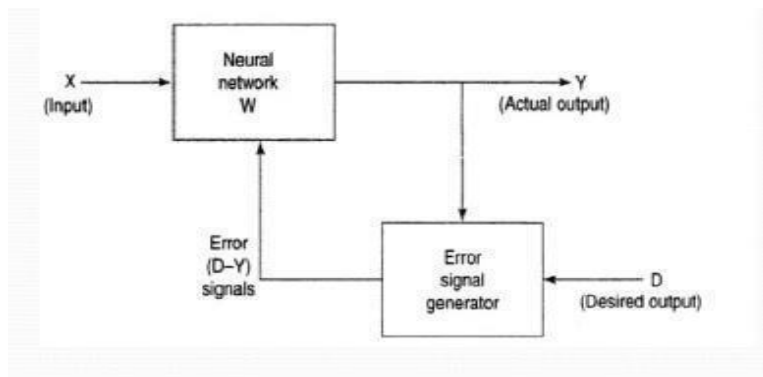


Fig 1.1: Supervised learning

This is a dependent process. The machine totally depends on the user who has to feed the inputs and has to check the efficiency of the same and correct it with the flow of iteration. It's an ANN network. During the training phase vectors are taken into consideration.

Up in the above figure There's an input vector and the output vector. The input vector derives and gives an output flow of the output vector. If the error signal is generated then the iteration is undergone where as lacking of the same means the output field is derived and the output result is accurate and no modification needs to be undergone for same.

2.1.2 UNSUPERVISED LEARNING

Unsupervised learning deals with learning by itself. It is also known as self learning algorithm. Here only the input vector is known and passed. So the variance of the result deals with the input factors. Here the input factors are grouped and clustered. Cluster is the main essence of this technique.

Test Data are passed and with the iteration of the same it learns from it derives itself more closer to the conclusion part. Labelled is missed in the data set and classification and categorization of the same had to be done by the machine itself.

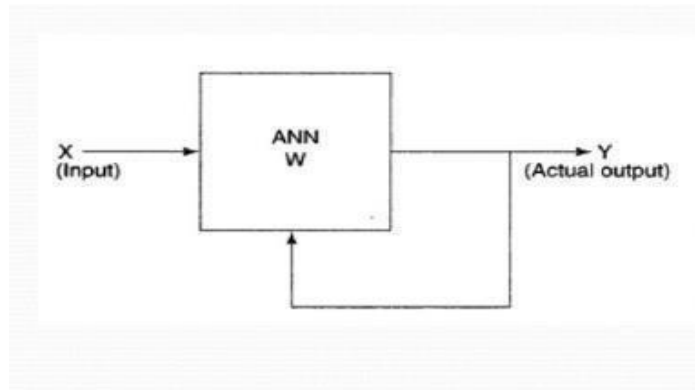


Fig 1.2: Unsupervised Learning

As described in the above figure, In this ANN network when the input is processed by the function the output had to be self derived and to be matched with the cluster set to provide the result. If the result lacks the interpretation then it undergoes the iteration. All the data sets are formed and combined in a cluster set for the effective uses of the same in further cases.

Feedbacks are not reciprocated incase of such it responds to commonalities. If the commonalities is found between the dataset then it applies the previous functionalities and derive the data. If not set then it learns and identifies for the others.

2.1.3 REINFORCEMENT LEARNING

In this type of learning a reinforced strategy is used. It deals with blooming of the knowledge. It's neither Supervised nor Unsupervised form of learning. They use dynamic techniques for letting the user know the output and the derivation of the same.

In these sort of algorithm set they don't assume the environmental set. These are even used in higher and complex mechanism finding likes genetic algorithm. They are widely in progress and implemented most in automation for the better efficiency of the establishment. These algorithms are used in Games and Automation of the vehicle resources.

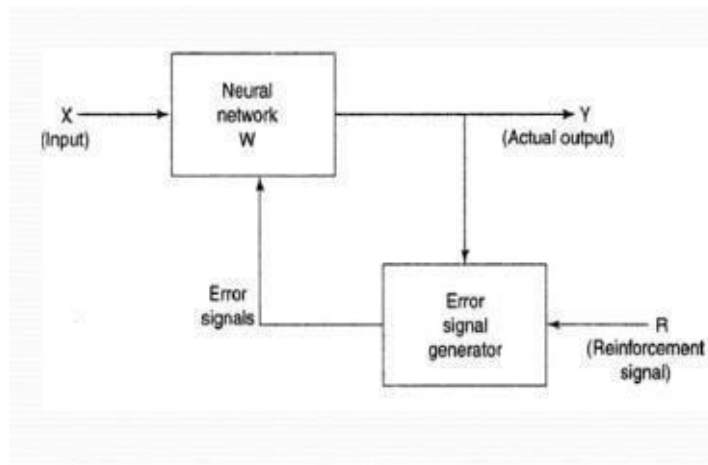


Fig 1.3: Reinforcement Learning

As described in the figure the input vector is passed to a ANN model where the functionalities of the same are stored. If the accurate output is derived then a reward is given to the user making it go to the next level for further task of completion. If not

then the Error signal is generated for the same. The accuracy level is calculated and passed down to the user stating the same.

The user sees the percentage of match and pass down and tries other keys of iteration to get the most out of it and complete the task to carry on the ladder of success. This is the same with the machine. Machine iterates the same and to the error signal an add on of reinforced signal is passed which the machine learn and iterates on the same to get closer to the actual results.

2.2 PROPOSED METHODOLOGY

METHODOLOGY:

This project has undergone the following process:

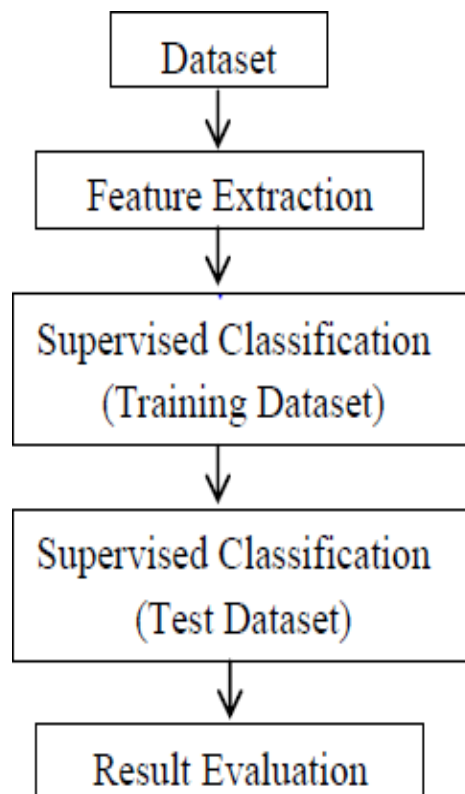


Fig 2.1: System Flow

Initially, the data set is prepared manually. After identifying the relationships and visualizing the data, we create a regression model for forecasting the percapita. For this model, we have used Multi Linear regression model. Other models such as the Linear Regression and Logistic Regression models were also tested, but the Multi Linear regression produced the minimal error while training the model. This regression model predicts the percapita of Crime rate that is going to be happen in future by taking different parameters.

Logistic Regression: In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1".

KNN: In pattern recognition, the k -nearest neighbors algorithm (k -NN) is a nonparametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k -NN is used for classification or regression. In k -NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. In k -NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors.

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the

explanatory (independent) variables and response (dependent) variable. So far, we have seen the concept of simple linear regression where a single predictor variable X was used to model the response variable Y . In many applications, there is more than one factor that influences the response. Multiple regression models thus describe how a single response variable Y depends linearly on a number of predictor variables.

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.

2.1 SOFTWARE DESCRIPTION

2.1.1 JUPYTER NOTEBOOK

Jupyter Notebook or so called IPython Notebook is an interactive web based computational mean for starting with Jupyter Notebook documents. The term notebook itself is a huge entity to represent the integration with different entity sets. JSON is the main document form from the same for the execution which follows the brief on the schema and the input and output means. It has high integration with several language set and has various flexibilities with the choices. The extension used for the same is “.ipynb” which runs in this platform. It’s an open-source software package with interactive communication means. It has it’s open standards for the same. It’s an open community best for budding programmers. The flexibility of the same is phenomenon and splendidly done the configuration and integration of the same is simplest and easy on hold so that no prior distortion is generated and the efficiency of the same is measured through out any system of choice. It’s the best software sets that been used across cross for designing and developing of the products and support wide help support. Not only to that, it provides scalability in the code and the deployment of the same. Various Language can be changed and the project can be undertaken on the same. The created notebook files can be shared and stored in various means for further utilization. It supports cultivated and interactive output sets. Easily crossed over for graphing, plotting and visualizing of the elements.

Data Integration of the same is to it's best. The integration of big data and it can process chunks of values in an approx. time which gives a better performance and the higher computational means. Various works on data like cleaning, cleansing, transforming modeling and visualizing can be done by the same.

CHAPTER 3

PROJECT DESIGN

3.1 DESIGN GOALS

To make the project runs smoothly it's required that we make plan and design some accepts like flowcharts and system architecture which are defined below.

3.1.1 Data Collection

Data collection is one of the important and basic things in our project. The right dataset must be provided to get robust results. Our data mainly consists of previous year or weeks stock prices. We will be taking and analyzing data from Kaggle. After that seeing the accuracy, we will use the data in our model.

3.1.2 Data Preprocessing

Human can understand any type of data but machine can't our model will also learn from scratch so it's better to make the data more machine readable. Raw data is usually inconsistent or incomplete. Data preprocessing involves checking missing values, splitting the dataset and training the machine etc.

3.1.2 Training Model

Similar to feeding somethings, machine/model should also learn by feeding and learning on data. The data set extracted from Kaggle will be used to train the model. The training model uses a raw set of data as the undefined dataset which is collected from the previous fiscal year and from the same dataset a refine view is presented which is seen as the desired output. For the refining of the dataset various algorithms are implemented to show the desired output.

3.2 SYSTEM ARCHITECTURE

The dataset we use for the proposed project is been taken from Kaggle. But this data set is in raw format. The data set is a collection of valuation of stock market information about some companies. The initial step is to convert raw data into processed data. Which is done by feature extraction, since the raw data collected have multiple attributes but only some of those attributes are needed for the prediction. Feature extraction is a reduction process.

The structure, behavior and views of a system is given by structural model.

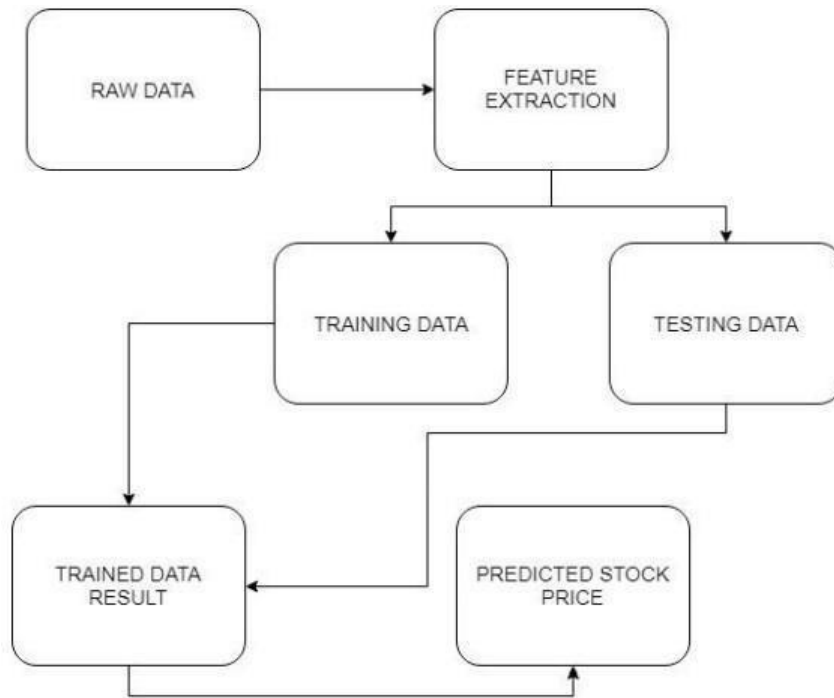


Fig 3.1: System Architecture

The above figure 3 .1 gives the demonstration on the dataset extraction and refining the raw dataset by categorizing into two phases of training and testing data

3.3 Use case Diagram

A dynamic and behavioral diagram in UML is use case diagram. Use cases are basically set of actions, services which are used by system. To visualize the functionality requirement of the system this use case diagram are used. The internal and external events or party that may influence the system are also picturized. Use case diagram specify how the system acts on any action without worrying to know about the details how that functionality is achieved.

For the project we have created the below mentioned use case diagram.

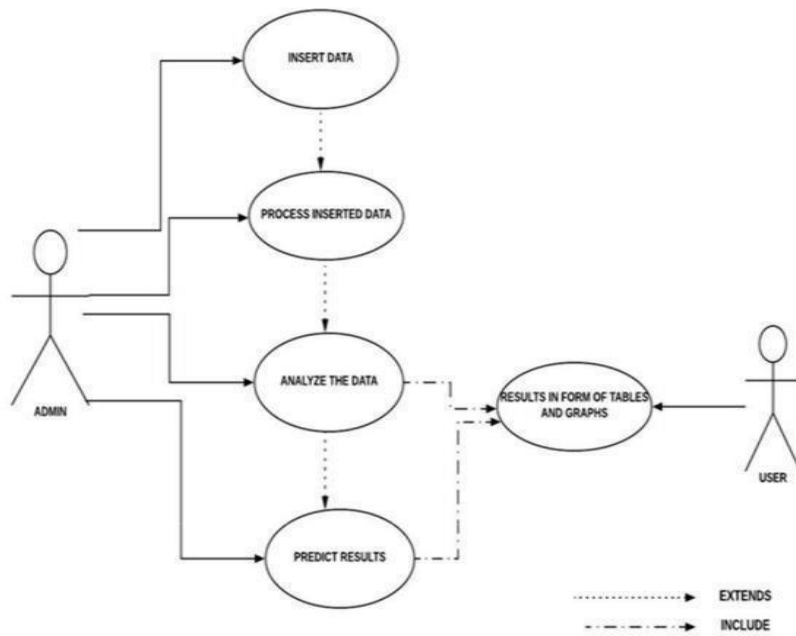


Fig 3.2: Use Case Diagram

The above figure 3.2 shows the use-case diagram of the entitled project and its flow. From the diagram it's seen that the user gives the raw dataset as input and with the flow of the input in the system.

The system evaluates and process the dataset train itself with the provided dataset and extract the meaningful dataset to process and refine the cluster data and from the given cluster of the data, the plotting of the data values are shown and with the given range the system plots the data gives a figurative output as prediction and display the same as the refined output in the display screen.

3.3 Data Flow Diagram

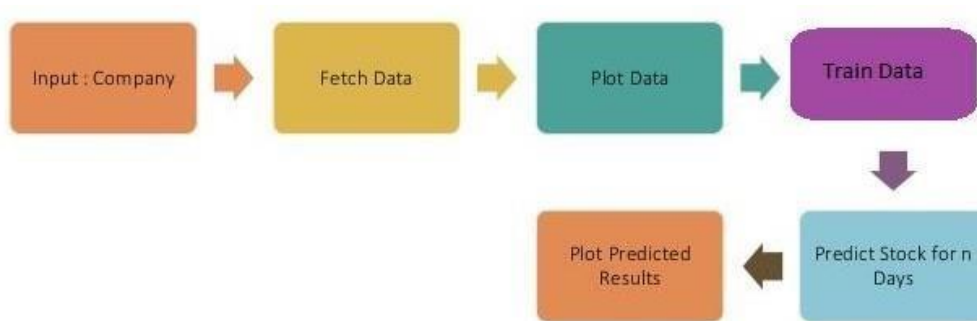


Fig 3.3 Data Flow Diagram

In the above fig 3.3 we are taking a company fetching the data of the company from the panda's data-reader library then we are plotting the data, then we train the data to predict the stock for certain number of days. In this way data is flowing in our syste

CHAPTER 4

RESULTS AND CONCLUSION

4.1 RESULTS

The Support vector Regression model results in classifying the test data. It predicts the stock crime rate value that is going to be happening in future. It uses train set to train and fit the model. The test data is transformed and predicts the accurate result.

By Using Django and Aws we prepared the UserInterface.

The below figure describes that home page takes the year, murder, attempts, cheating etc., values as input in text boxes.



A screenshot of a web form with a light blue background. The form contains nine input fields stacked vertically, each with a label to its left. The labels are: "Age", "Homelessness Rate", "Poverty Rate", "Arrests per person with intent to murder", "Arrests per individual of women", "Arrests by individual of education", "Representation of girls from through schools", and "Security threat by neighborhood". At the bottom center of the form is a "SUBMIT" button.

The below figure describes that new page which displays the predicted crime rate of the inputs produced by our algorithms.

A screenshot of a web page titled "CRIME RATE PREDICTOR". Below the title is a dropdown menu labeled "District" with "ADILABA" selected. Below the dropdown is a vertical list of ten input fields, each containing a number: 2001, 101, 60, 17, 50, 30, 16, 5, 8, and 47.

Crime Rate Prediction

4.2 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 type of crime and crime percapita which may happen in future. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation using Multi Linear Regression Algorithm. The model predicts the type of crime and Data visualization helps in analysis of data set and prediction of crimes. The graphs include bar, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding different crime datasets that can help in capturing the factors that can help in keeping society safe.

REFERENCES

- 1) Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
- 2) Machine Learning in Production: Developing and Optimizing Data Science Workflows and Applications (Addison-Wesley Data & Analytics)
- 3) [Python: The Complete Reference - Martin C. Brown
- 4) <https://docs.anaconda.com/anaconda/install/windows/>
- 5) A.U.S.S Pradeep, SorenGoyal, J. A. Bloom, I. J. Cox, and M. Miller, —Detection of statistical arbitrage using machine learning techniques in Indian Stock market,|| IIT Kanpur, April 15, 2013.
- 6) Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil—Parameters for Stock Market Prediction,|| Prashant S Chavan et al, Int.J.Computer Technology & Applications, Vol 4 (2),337-340.
- 7) NeelimaBudhani, Dr. C. K. Jha, Sandeep K. Budhani—Prediction of Stock Market Using Artificial Neural Network,|| International Conference on Soft Computing Techniques for Engineering and Technology (ICSTET)- 2014.
- 8) SharvilKatariya, Saurabh Jain—Stock Price Trend Forecasting using Supervised Learning Methods.
- 9) DhirajMundada, GauravChhapparwal, SachinChaudhari, and TruptiBhamare— Stock Value Prediction System,|| International Journal on Recent and Innovation Trends in Computing and Communication, April 2015
- 10) Hochreiter, Sepp, and Jürgen Schmidhuber—Long short-term memory,|| Neural computation 9.8 (1997): 1735-1780