A Novel Solution to Predict Crime Rate

V.Arvindh[1], K .Dhivagar[2], M.Sirija[3] ,Dr.P.Veeralakshmi[4]

[1,2]Student -Prince Shri Venkateshwara Padmavathy Engineering College.

[3,4]Faculty-Prince Shri Venkateshwara Padmavathy Engineering College.

ABSTRACT – Machine learning (ML) is that the revise of computer algorithms that look up involuntarily through skill and by the consumption of facts. It's seen as a vicinity of AI . Machine learning algorithms build a model bear sample data, referred to as "training data", so as to form predictions or decisions without being clearly programmed to try to do so. Crime is one of the largest and dominating crises in our civilization and its deterrence is an important task. On a daily basis there are enormous figures of crimes committed repeatedly. The current crisis faced are maintaining of proper dataset of sin and analyzing this data to aid in predicting and solving crimes in attitude. The purpose of this project is to evaluate dataset which consist of abundant crimes and predicting the category of crime which may occur in outlook depending upon various circumstances. So, in this project we will be using the method of machine learning and data science. Before preparation of the model data preprocessing will be done following this aspect choice and scaling will be done so that precision obtain will be greater. The K-Nearest Neighbor (KNN) sorting and various other algorithms will be tested for crime rate forecast and one with enhanced precision will be used for training. The soul function of this project is to give a finest proposal of how machine learning can be used by the commandment enforcement agencies to identify, envisage and resolve crimes at a much quicker rate and thus minimize the crime rate.

Keywords- Machine learning, crime, prediction, dataset, K-Nearest Neighbor, forecast, crime rate

1) INTRODUCTION:

DOMAIN:

Machine learning is a computer science subfield that is also known as predictive analytics or predictive modeling. Its purpose and application is to create new and/or exploit existing algorithms to learn from data to create generalizable models that make accurate predictions or to find trends, particularly with new and previously unseen data. Consider a dataset to be a table, with each row representing an observation (a.k.a. calculation, data point, etc.) and each column representing the features of that observation and their values. A dataset is normally divided into two or three subsets at the start of a machine learning project. The training and evaluation datasets, as well as an optional third validation dataset, are the bare minimum subsets. Machine learning, as previously mentioned, uses algorithms to automatically model and find patterns in data, to predict a specific performance or response. Statistics and mathematical optimization are extensively used in these algorithms. In a nutshell, machine learning involves using learning algorithms and optimization techniques to automatically learn a highly accurate predictive or classifier model, or to discover unknown patterns in results.

MACHINE LEARNING APPLICATIONS:

I. Image Recognition:

It is one of the most widely used applications of machine learning. You may identify an object as a digital image in several circumstances. The outputs of each pixel in a digital image are defined by the measurements as N2. Each pixel in a colored image is thought to provide three measurements of the intensities of three key colour components, i.e. RGB. So there are three N2 measurements in the N*N colored image.

II. Speech Recognition:

Speech recognition (SR) is the relocation of utter words into text. In speech recognition, a software approach concedes utter words. The quantifying in this Machine Learning suit might be a set of numbers that represent the speech

signal. We can chunk the signal into a hunk that contains distinct words or phonemes. In each chunk, we can represent the speech gesture by the potency or energy in different time-abundance bands.

2) RELATED WORKS:

The basis and role of the device and Physical Sciences Research Council (EPSRC) are described. The EPSRC has launched a search program for Crime Prevention and Detection. The aim is to encourage engineers and physical scientists to develop subsequent generations of technologies to fend off and dig out crime. The Crime Technology program is stimulating scientists and engineers to believe in how their discipline and expertise can help combat crime and improve security [1]. Identity Crime is taken into account as crimes which involve masquerading one's identity and steal tip concerning the concerned person's identity. This paper mainly deals with identity crime associated with the MasterCard application, which nowadays is sort of prevalent and dear even. The prevailing non-data-mining techniques for eliminating fraud have some flaws and to combat them a replacement data-mining layer of defence has been proposed [2]. Given the fast-paced nature of recent police investigations, the event and use of advanced data processing tools for crime analysis can play a critical think in mitigating future harm and helping with crime prevention. This paper aims to unravel the matter of identifying potential serial offending patterns using previously underutilized attributes from police-recorded crime data [3]. Crimes are a social irritation and price our society deeply in several ways. Any research which will help in solving crimes quickly can pay for itself. About 10% of the sinner commits about 50% of the crimes. The system is trained by feeding previous year's record of crimes taken from licit online egress of India listing various crimes like homicide, snatch and abduction, harpy, larceny, burglary, rape, and other such crimes [4]. The present learn has been engaged up for detecting the hotspots of lane crimes and embryonic the crime manage strategies in the Faisalabad city of Pakistan. The spatial patterns of built-up lane crimes were analyzed. The crime gossip of 2012 was decoded and the crime maps were geared up in ArcGIS 10. The tactical crime scrutiny was done in a chain of meetings with law division and crime domineering strategies were built. The Comp stat model with some modifications was followed for the responsibility and act supervision of the police department [5]. The datasets contain aggregated counts of sin and crime-related proceedings categorized by the law division. The location and time of these actions is fixed in the data. Further spatial and sequential features are harvested from the raw statistics set. A group of data mining categorization techniques is in employment to execute the sin forecasting [6]. The modified e-learning method architecture is existing, which detects and responds to training filling according to the students' learning capabilities. The primary aim includes the innovation of finest settings, in which the learners can look up their learning capabilities. Moreover, the supervision can find vital hidden patterns to bring the useful reforms in the accessible system. The clustering methods K-Means, K-Medoids, Density-based Spatial Clustering of Applications with racket, Agglomerative Hierarchical group hierarchy and Clustering by express explore and ruling of concentration Peaks via warmth dissemination (CFSFDP-HD) are analyzed using didactic data mining [7].

3) PROBLEM DESCRIPTION:
To identify a crime that has a fair chance of occurring. For crime prediction, the K-Nearest Neighbor (KNN) classification and other algorithms will be used. The dataset is visualized to investigate any crimes that may have occurred in the region. One typical approach to investigating gang crime suspects is to identify the particular target of multiple suspects ahead of time, then manually track and gather information on those suspects to find any similar criminal suspects or criminal groups. For the identification of possible criminal suspects, data sharing among these parties becomes extremely essential to protect data privacy Personal and social data, such as criminal records and contact details, are vulnerable in the meantime

4) EXISITING SYSTEM:

- In the existing method most of crime prediction method using data mining process to find the prediction.

- Data mining process look much time for analyse the data.

- Accuracy of the process also poor compare to new technique.

- With the aid of machine learning algorithm, by means of python as core we can estimate the type of sin which will arise in a particular area.

- This exertion helps the commandment enforcement agencies to forecast and perceive crimes in our locale with enhanced exactness and thus reduces the crime rate.

DISADVANTAGES:

• The method's effectiveness is directly dependent on subjective processes of assessment and selection of main parameters, as well as the variety of their alternatives and, ultimately, the most useful combination of these parameters.

PROPOSED SYSTEM:

- In this proposed system we are collect the data sets (user generated content) in Internet.

- Here we using classification or regression algorithm based on the data sets our requirement.

- In this process we are going to use K-Nearest Neighbours (KNN) and also some classification algorithm.

- Based on the accuracy we are going to find the prediction rate

ADVANTAGES:

• To find the prediction degree, we're using a machine learning algorithm.

• We can only select the algorithm based on that accuracy.

• The accuracy of the result is dependent on the algorithm's accuracy.

Cumulate Dataset:

- Data Collection is one of the most important tasks in building a machine learning model.

- We collect the specific dataset based on requirements from internet.

- The dataset contains some unwanted data also. So first we need to pre-process the data and obtain perfect data set for algorithm.

Fore-treating:

- It is the congregation of chore related in order based on a number of targeted variables to investigate and create some precious ending.

- However, some of the data may be piercing, i.e. may contain imprecise ethics, deficient values or wrong values.

- Hence, it is must to practice the data prior to analysing it and approaching to the results.

- Data Fore-treating can be completed by data cleaning, data transformation, data selection.

- Data cleaning includes Fill in lost values, even loud data, classify or eliminate outliers, and resolve inconsistency.

- Data transformation may include smooth, aggregation, simplification, revolution which improves the worth of the data.

- Data selection includes some methods or functions which tolerate us to choice the valuable data for our organization.

Guidance dataset:

The aim is to produce a skilled (fitted) mock-up that generalizes well to innovative, unidentified facts. The built-in mock-up is evaluate using "fresh" example from the held-out datasets (validation and test datasets) to estimate the model's exactness in classify new data. To diminish the threat of issue such as more proper, the example in the corroboration and investigation datasets should not be used to instruct the mock-up.

Sturdy dataset:

An analysis dataset is a dataset that is autonomous of the preparation dataset, but that follow the identical prospect sharing as the preparation dataset. A test locate is consequently a set of example used just to review the recital (i.e. generalization fig 1) of a wholly precise classifier.
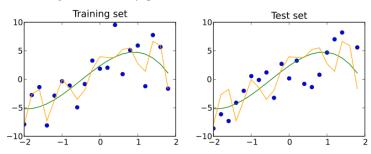

Figure1:- predictive models

Algorithm:

KNN Algorithm:

- o **Step-1:** Select the numeral K of the neighbours

- o **Step-2:** analyse the Euclidean space of K numeral of neighbours

- o **Step-3:** get the K nearest neighbours as per the calculated Euclidean distance.

- o **Step-4:** Among these k neighbours, reckon the figure of the statistics point in each sort.

- o **Step-5:** allocate the original statistics point to that grouping for which the figure of the neighbour is highest.

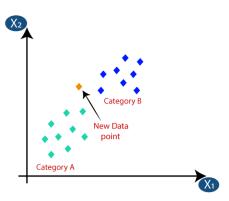- o **Step-6:** Our mock-up is complete.(fig 2)

Figure 2:- knn graph

o   Initially, we will opt the number of neighbours, so we will opt the k=5.

o   Next, we will investigate the Euclidean distance between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as :(fig 3)
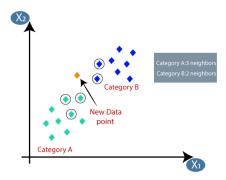

Figure 3:-Euclidean model

Linear decay Algorithm:

With easy linear decay when we have a solitary effort, we can use data to approximation the coefficients.

This requires that you analyze arithmetical property from the data such as resources, standard deviations, correlation and covariance. All of the data must be accessible to pass through and compute data.

Support Vector Machine:

"Support Vector Machine" (SVM) is a control mechanism knowledge algorithm which can be used for together classification and corrosion defy. However, it is mostly used in categorization troubles. In the SVM algorithm, we plan each statistics article as a spot in n-dimensional freedom (where n is number of features you have) with the assessment of each trait being the cost of a scrupulous synchronize. Then, we execute categorization by result the hyper-plane that differentiate the two course very fit (look at the below snapshot fig 4).
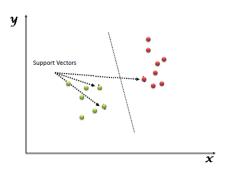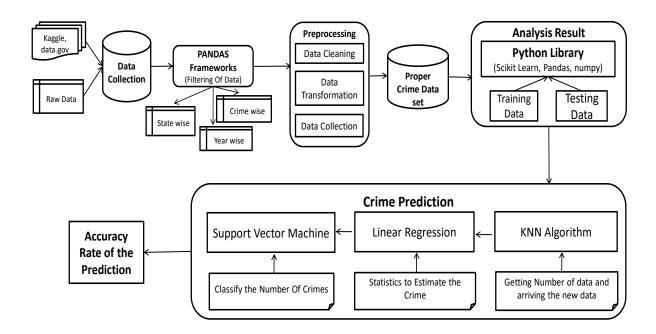
Figure 4:-svm prediction

Support Vectors are merely the co-ordinates of entity observation. The SVM classifier is a limit which finest segregate the two modules (hyper-plane/ line).

o   Figure 5:-architecture diagram of crime prediction

Here raw data are collected from the web and required data are filtered using pandas framework supported state wise, crime wise, and year wise. After the info is filtered its preprocessed using 3 processes .after all the method is completed proper crime dataset is provided. Using python library we are getting to train and test the info of the right crime data set. The crime prediction is completed supported by the algorithms to classify and supply the statistics and getting the info and arriving at replacement data. Finally, accurate rate of prediction is produced result

5) RESULTS:

Based on that dataset we will get the result used from a machine-learning algorithm to predict the result .the data are pre-processed and filtered .the data is trained and tested using the python library. Here we will also determine the accuracy rate of the prediction. It'll help locate spring water prediction.

6) CONCLUSION AND FUTURE ENHANCEMENTS:

On the idea of that dataset, we will predict the result using our arbitrary timberland estimate. we will also determine how accurate the forecast is during this section. it might be beneficial in determining rate Prediction. Predictive models using machine learning. The accuracy of crime prediction was found to be between 39 and 44 percent using KNN and boosted decision trees. Algorithm precision, complexity, and training time varied slightly counting on the tactic and algorithm. Both the algorithm and therefore the data are often fine-tuned for particular applications to enhance prediction accuracy. The results of the experiments show that our method can produce good research results with an inexpensive overhead. we would like to expand we add the longer term to assist CO Offline.

7) REFERNCES:

1) Amit Kumar Manjhvar (Asst.prof) and Nidhi Tomar (Research Scholar) "An Improved Optimized Clustering Technique For Crime Detection "2016 IEEE Symposium on Colossal Data Analysis and Networking (CDAN)

2) Khushboo Sukhija,Shailendra Narayan Singh and Jitendra Kumar"Spatial Visualization Approach for Detecting Criminal Hotspots: An Analysis of Total Cognizable Crimes in the State of Haryana" 2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India

3) Julio Borges, Daniel Ziehr, Michael Beigl,Nelio Cacho, Allan Martins and Simon Sudrich, Samuel Abt,"Feature Engineering for Crime Hotspot Detection" 2017 IEEE

4) Julio Borges and Adelson Araujo Jr.,Cacho "Towards a Crime Hotspot Detection Framework for Patrol Planning "2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th Intl. Conference on Data Science and Systems

5) Shoaib KhalidI3*, liechen Wangl2, Muhammd Shakeel3, Xia Nan I "Spatio-temporal Analysis of the Street Crime Hotspots in Faisalabad City of Pakistan "

6) Chung-Hsien Yu1 , Max W. Ward1 , Melissa Morabito2 , and Wei Ding"Crime Forecasting Using Data Mining Techniques "2011 11th IEEE International Conference on Data Mining Workshops

7) Samina Kausar Xu Huahu, Iftikhar Hussain , Zhu Wenhao , And Misha Zahid "Integration of Data Mining Clustering Approach in the Personalized E-Learning System "Received October 26, 2018, accepted November 13, 2018, date of publication November 20, 2018, date of current version December 19, 2018.

8) X.Alphonse Inbaraj and A. Seshagiri Rao " Hybrid Clustering Algorithms for Crime Pattern Analysis" Department of Computer Science and Engineering PACE Institute of Technology and Sciences Ongole 2018 IEEE International

Conference on Current Trends toward Converging Technologies, Coimbatore, India .

9) Clifton Phua, Member, IEEE, Kate Smith-Miles, Senior Member, IEEE, Vincent Lee, and Ross Gayler, "Resilient Identity Crime Detection ",   IEEE transactions on knowledge and data engineering ,year 2012 .

10) Peng Chen and Justin Kurlan"Time, Place, and Modus Operandi: A Simple Apriori Algorithm Experiment for Crime Pattern Detection " New Zealand Institute for Security and Crime Science University of Waikato Hamilton,

11) Shayam Varan Nath,"Crime pattern detection using data mining   ", IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 09, pp. 41-44, 2010.

12) R. Adderley and P. B. Musgrove." Data mining case study: Modeling the behavior of offenders who commit serious sexual assaults." In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'01), pages 215.220, New York, 2001.