iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3331

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

Crime Rate Analysis and Prediction Based on Spatial Data Mining

Jerrin Simla A[1], Sivashankari M[2], Kavitha E[3], Gopigha S[4]
[1]*Assistant Professor*
[2]*Student, IV Year*

Abstract: ***Crime analysis is one amongst the foremost necessary activities of the bulk of the intelligent and enforcement organizations everywhere the globe. Generally, they collect domestic Associate in foreign crime connected knowledge (intelligence) to stop future attacks and utilize a restricted range of enforcement resources in an optimum manner. A serious challenge moon-faced by most of the law enforcement and intelligence organizations is expeditiously and accurately analyzing the growing volumes of crime patterns have created the analyzing and recording of crime knowledge tougher. Data processing could be a powerful tool which will be used effectively for analyzing massive databases and derivation necessary analytical results. This paper presents Associate in Nursing intelligent crime analysis system that is meant to beat the on top of mentioned issues. Within the planned system crime happened is found and PySpark is employed to store great amount of information for crime analysis. The planned system consists of a chic and simplified setting which will be used effectively for processes of crime analysis.***

I. INTRODUCTION

Crime analysis has become one in all the foremost important activities of the fashionable world because of thanks to attributable to the high magnitude of crimes that could be a result of technological advancements and also the increase. Enforcement organizations and also the intelligence gathering organizations all round the world typically collect great amount of domestic and foreign crime information (intelligence) to stop future attacks. As this involves an outsized quantity of knowledge, manual techniques of analyzing like information with a huge variation have resulted in lower productivity and ineffective utilization of hands. This is often one all the dominant issues in several enforcement and intelligence organizations.

There are many important reasons for crime analysis like to spot general and specific crime trends, patterns associate degreed series in an on-going, timely manner, to maximize the usage of restricted law enforcements resources, to access crime issues domestically, regionally, across the nation inside and between enforcement agencies, to be proactive in sleuthing and preventing crimes and to satisfy the enforcement desires of the dynamical society. There are numerous crime data processing techniques offered like bunch techniques, association rule mining, sequential pattern mining, and classification and string comparison.

Several internet based mostly crime mapping systems are offered on the net like Narcotic network in city local department, however majority of them are custom created for legislative authorities in numerous countries and people systems aren't accessible to parties outside that exact enforcement or legislative authorities.

II. EXISTING SYSTEM

The system could be a web-based system that includes of crime analysis techniques like hotspot detection, crime comparison and crime pattern image. System can't be directly valid victimization records of the department of local government as a result of police records embrace each major and minor crime incidents. The system is predicated on newspaper articles therefore it includes therefore it includes solely a set of total crime incidents. Therefore individual parts of the system live evaluated and results of that analysis square measure wont to measure the effectiveness.

Crime data processing is that the application of information mining techniques for crime analysis. Numerous researches are allotted during this domain Crimes are often divided into subcategories supported totally different criteria. In eight crime classes square measure given. They're traffic violations, sex crimes, theft, fraud, arson, drug offenses, cybercrimes and violent crimes.

There were several efforts to analyze sorts differing types differing kinds of crimes victimization machine-driven techniques however there's no unified framework describing the way to apply those techniques to totally different crime types. In they need used a framework which incorporates a relationship between the crime data processing technique and crime sort characteristics.

There square measure many existing systems that use crime data processing framework for crime pattern identification and narcotics network in metropolis department of local government.

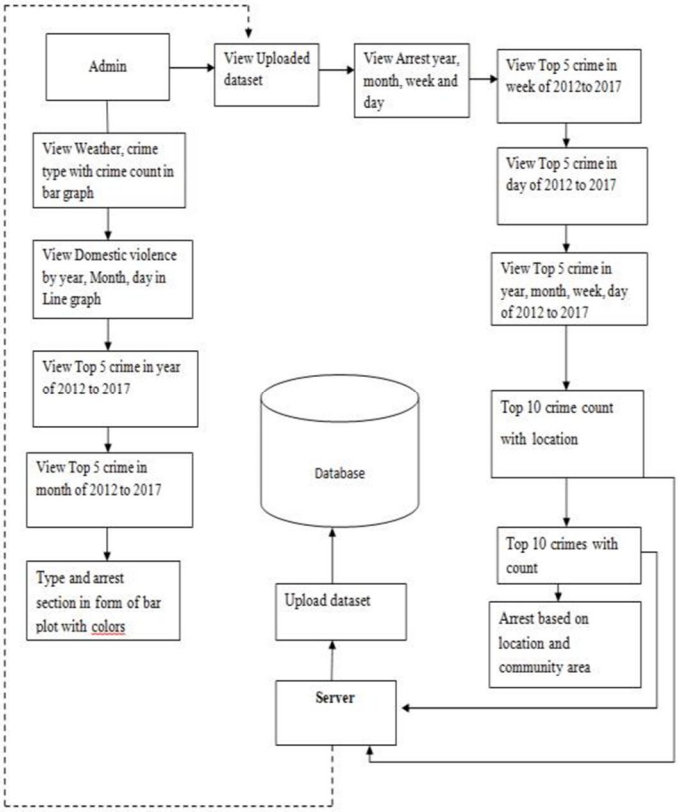
An intelligent crime identification system is represented within which are often to predict doable suspects for given crime. They used 5 styles of agents specifically, message house agent, entranceway agent, unfortunate person agent, criminal agent and proof agent.

A. Disadvantage

1) Crime analyzed here is by types of crimes happened. Rime has not been analyzed with any aspects.

2) Storing of dataset is less by compared to proposed system accuracy, which will not be perfect while prediction analysis. Here number of crimes per year, per month and per day has been calculated on basis of Chicago data from 2012 to 2017.

III. ARCHITECTURE



IV. PROPOSED SYSTEM

The proposed system consists of an upscale and simplified setting which will be used effectively for processes of crime analysis. We have a tendency to planned Pyspark for classification and datasets area unit keep into sparksql storing and retrieval of knowledge are going to be quicker.

1) The prediction of the arrest supported year, month, week, day and prime five crimes in 2012, 2013, 2014, 2015, 2016, 2017 in an exceedingly month, week and day wise is completed. Through location analysis we have a tendency to foreseen placement at that additional range of crimes has occurred. We have a tendency to analyze the quantity of crime and arrest has been premeditated for additional accuracy. Here, the entities area unit premeditated by primary sort and arrest section is pictured within the sort of bar chart to search out the results of criminal activity.

2) In the graph, red color refers to “Arrest” and different color refers to “Not arrested” that helps to point out effective operating.

3) Top ten crimes in mere location.

4) Top ten overall crimes with count.

5) Predicting arrest supported location with latitude, meridian and community space.

In this planned system approaches used are

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

A. Svm Pre-Processing

A support vector machine is additionally referred to as a support vector network. Import operate for pandas has been accustomed save dataset and read_csv has been created as variables. We have a tendency to print the sort crimes to point out information frame alongside import functions. We have a tendency to shall read transfer dataset by removing initial 3 columns by victimization iloc functions and heading the remaining columns with date as main header.

B. Getting Column Temperature, Crime Types And Events

1) Group By Columns With Count: Getting column temperature (Fahrenheit), Event, crime type (Primary type) and count by using group by function. We remove duplicate columns and view top 15 information from the cri_count type.

C. Logistic Regression Accuracy Calculation

Logistic regression is employed to predict outcomes or responses. Supplying regression is conceptually kind of like statistical regression, wherever statistical regression estimates the target variable. Rather than predicting values as within the statistical regression, supplying regression would estimate the percentages of sure event occurring.

Logistic regression gets yearly arrest values, monthly arrest values, weekly arrest values and daily arrest values to supply line map and predict the event of arrest. Plotting a graph with yearly, weekly and day may resample predefined operate accustomed classify year, week and day arrest. A is employed for year and W is employed for Week and M is employed for Month and D is employed for day. Line is employed to plot graph.

D. Finding Top 5 Crimes

Creating type from 2012 to 2017 with Crime type (primary type) column, heading the column and plotting with year, month, week and day with top 5 crimes.

E. K–Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN)Classification is employed for crime prediction. The projected system will predict the regions that have high likelihood for rate and might forecast crime prone areas, rather than specializing in causes of crime incidence like criminal background of wrong doer.

1. Obtaining Location Description and first sorts with Count.
2. Location Description with Arrest Analysis alongside Latitude.

3) Victimization Location specially Community space.

4) Total rate.

V. RESULT ANALYSIS

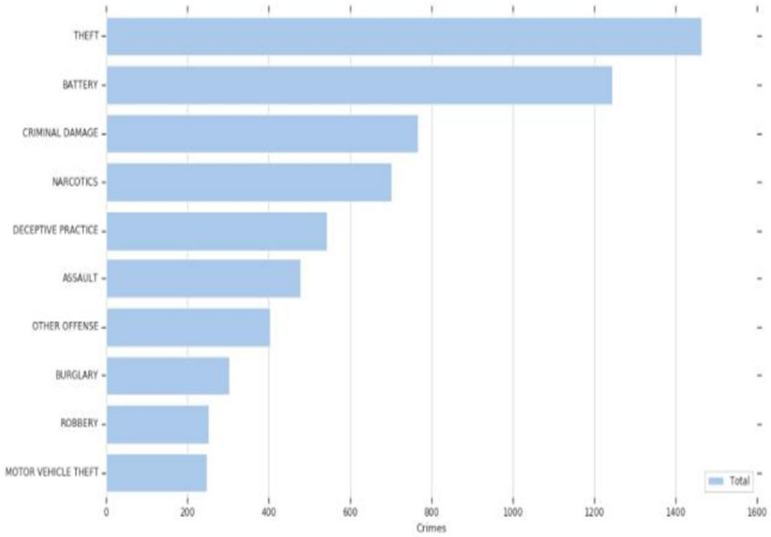
A. Import Function

Import Function for pandas has been used to save dataset and read csv has been created as variables and printing the type crimes to show data frame along with import functions.

B. Crime Count With Crime Types

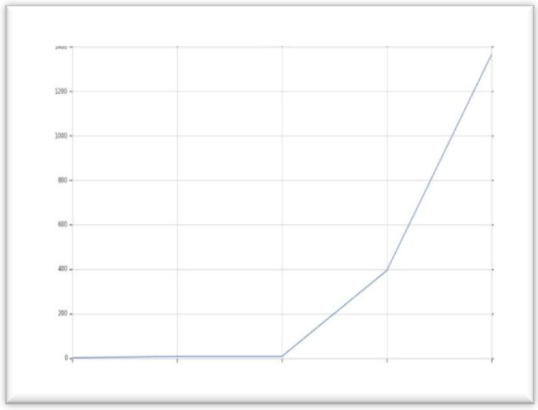
```
>>> crime_count = pd.DataFrame(s.groupby('Primary Type').size().sort_values(ascending=False).rename('counts').reset_index())
>>> crime_count.head()
        Primary Type  counts
0              THEFT    1403
1            BATTERY    1245
2   CRIMINAL DAMAGE     767
3          NARCOTICS     702
4  DECEPTIVE PRACTICE    544
>>> import seaborn as sns
>>> import matplotlib.pyplot as plt
>>> sns.set(style="whitegrid")
>>>
>>> # Initialize the matplotlib figure
... f, ax = plt.subplots(figsize=(6, 15))
>>>
>>>
>>> # Plot the total crashes
... sns.set_color_codes("pastel")
>>> sns.barplot(x="counts", y="Primary Type", data=crime_count.iloc[:10, :],
...             label="Total", color="b")
<matplotlib.axes._subplots.AxesSubplot object at 0x7f66ac817410>
>>>
>>> ax.legend(ncol=2, loc="lower right", frameon=True)
<matplotlib.legend.Legend object at 0x7f66ac25ed60>
>>> ax.set(ylabel="weather",
...        xlabel="Crimes")
[<matplotlib.text.Text object at 0x7f66ac833190>, <matplotlib.text.Text object at 0x7f66ac817d10>]
>>> sns.despine(left=True, bottom=True)
>>>
>>> # Add a legend and informative axis label
... plt.show()
```

Graph



Creating Crime count to get Primary types (crime types) along with crime count and by using group by function we reduce duplicate values from data plotting graph with Crime types and Crime Count.

C. Yearly Arrest

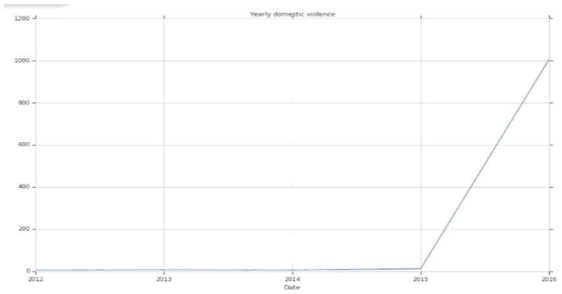


Creating type from to 2012 to 2016 with arrest column and heading the column and plotting with yearly arrest.

D. Domestic Violence

Yearly Domestic Violence

```
>>> domestic_yearly = crimes[crimes['Domestic'] == True]['Domestic']
>>> print domestic_yearly.head()
Date
2016-05-03 23:40:00    True
2016-05-03 21:40:00    True
2016-05-03 22:00:00    True
2016-05-03 22:30:00    True
2016-05-03 23:11:00    True
Name: Domestic, dtype: bool
>>> import seaborn as sns
>>> import matplotlib.pyplot as plt
>>> sns.set(style="whitegrid")
>>> plt.subplot()
<matplotlib.axes._subplots.AxesSubplot object at 0x7fa71e137a10>
>>> # yearly domestic violence
... domestic_yearly.resample('A').sum().plot()
<matplotlib.axes._subplots.AxesSubplot object at 0x7fa71e137a10>
>>> plt.title('Yearly domestic violence')
<matplotlib.text.Text object at 0x7fa71e17559b>
>>> plt.show()
```

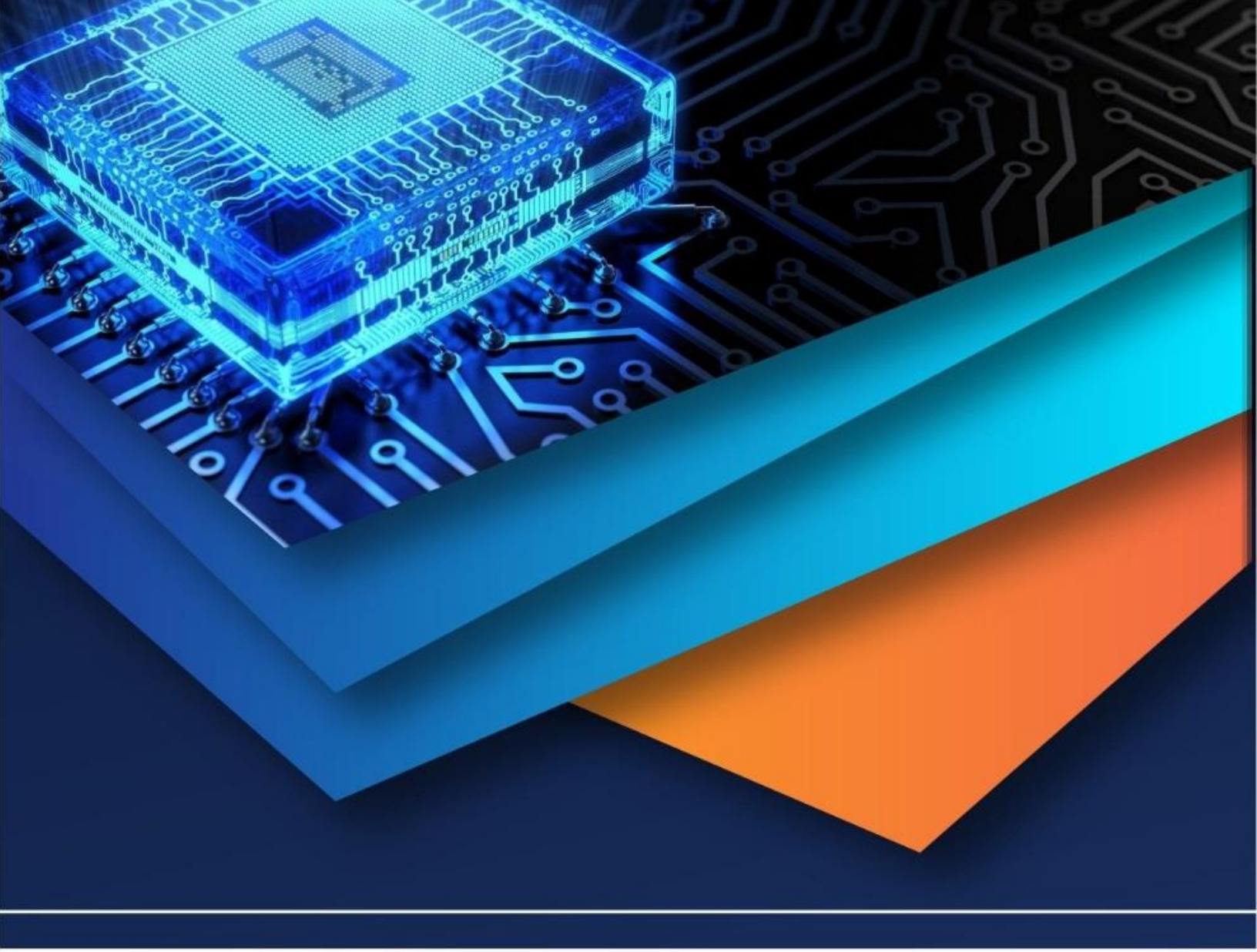


VI. CONCLUSION

The objective is to analyses the crime data and provide fruitful suggestions to the department of security to protect the precincts and beats where the crime rates are high. Different regression, clustering, classification and frequent growth patterns are found and plotted in order to organize and arrange the data in an ordered fashion. Thus the research if applied on the appropriate data would result in generating patterns that would result in generating patterns that would help to identify the crime rate early.

REFERENCE

[1] Nurul Hazwani Mohd Shamsuddin, Nor Azizah Ali, RazanaAlwee "An Overview on Crime Prediction Methods" IEEE study paper 2017 978-1-5090-6255-3/17/$31.00 ©2017 IEEE.
[2] VikasFrover, Richard Adderley, Max Mramer1 "Review of Current Crime Prediction Techniques".
[3] Shiju Sathyadevan, Devan M.S, Surya Gangadhaab. S "Crime Analysis and Prediction using Data Mining" Conference: First International Conference on Networks & Soft Computing (ICNSC), 2014.
[4] Tahani ALmanie, Rsha Mirza and Elizabeth Lor "Crime prediction based on crime types and using spatial and temporal criminal hotspots" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.4, July 2015
[5] Ms. Vrushali Pednekar, Ms. Trupti Mahale, Ms. Pratiksha Gadhave, Prof. Arti Gore "Crime Rate Prediction using KNN" International Journal on Recent and Innovation Trends in computing and communication ISSN: 2321-8169 Volume: 6, Issue: 1
[6] Malathi. A, Dr. S. Santhosh Baboo "An Enhanced Algorithm to Predict a future Crime using Data Mining" International Journal of Computer Applications (0975 – 8887) Volume 21- No.1, May 2011.
[7] Aarathi Srinivas Nadathur, Gayathri Narayanan,IndrajaRavinchandran,SrividhyaS, Kayalvizhi J, "Crime analysis and prediction using Bigdata" International Journal of Pure and Applied Mathematics Volume: 119, No.12 2018, 107-211
[8] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav "Crime Prediction Detection,Analysis and Prediction" International Conference on Electronics, communication and Aerospace Technology ICECA 2017
[9] Hitesh Kumar Reddy, Toppi Reddya, Bhavna Sainia, Ginika Mahajana, "Crime Prediction & Monitoring Framework based on Spatial Analysis International conference on Computational Intelligence and Data Science (ICCIDS 2018) International Conference on Computational Intelligence and Data Science (ICCIDS 2018)
[10] Noora Abdulrahman and Wala Abedalkhader "KNN classifier for crime prediction in San Francisco contect" International Journal of Database Management Systems (IJDMS) Volume 9, no. 4, August 2017
[11] Anisha Agarwal, Dhanasree Chogule, Arpita Agarwal, DivyaChimote, "Application for analysis and prediction of crime data using Data Mining", International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume: 4, Issue: 5, May 2016
[12] Lawrence McClendon, Natarajan Meghanathan, "Using machine learning algorithm to analyse crime data", Machine Learning and Applications: An International Journal (MLAIJ) Volume: 2, No: 1, March 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY