International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

Intelligent Crime Analysis System Using Pyspark

A. Ponmalar, P. Leela Jancy, V.R. Barath Kumar, B.K. Akshathaah
and P. Pavithra

***Abstract*---** *Crime analysis is one of the most important activities of the majority of the intelligent and law enforcement organizations all over the world. Generally they collect domestic and foreign crime related data (intelligence) to prevent future attacks and utilize a limited number of law enforcement resources in an optimum manner. A major challenge faced by most of the law enforcement and intelligence organizations is efficiently and accurately analyzing the growing volumes of crime related data. The vast geographical diversity and the complexity of crime patterns have made the analyzing and recording of crime data more difficult. Data mining is a powerful tool that can be used effectively for analyzing large databases and deriving important analytical results. This paper presents an intelligent crime analysis system which is designed to overcome the above mentioned problems. The proposed system is here is we find weather analysis along with the crime happened and we proposed Pyspark here to store large amount of data's for crime analysis. The proposed system consists of a rich and simplified environment that can be used effectively for processes of crime analysis.*

***Keywords*---** *Pyspark, Bigdata, Data Mining.*

I. INTRODUCTION

Crime analysis has become one of the most vital activities of the modern world due to the high magnitude of crimes which is a result of technological advancements and the population growth. Law enforcement organizations and the intelligence gathering organizations all around the world usually collect large amounts of domestic and foreign crime data (intelligence) to prevent future attacks. As this involves a large amount of data, manual techniques of analyzing such data with a vast variation have resulted in lower productivity and ineffective utilization of manpower. This is one of the most dominant problems in many law enforcement and intelligence organizations.

There are several significant reasons for crime analysis such as to identify general and specific crime trends, patterns, and series in an ongoing, timely manner, to maximize the usage of limited law enforcement resources, to access crime problems locally, regionally, nationally within and between law enforcement agencies, to be proactive in detecting and preventing crimes and to meet the law enforcement needs of the changing society. There are various crime data mining techniques available such as clustering techniques, association rule mining, sequential pattern mining, and classification and string comparison.

Several web based crime mapping systems are available on the Internet such as narcotics network in Tucson police department, but majority of them have been custom made for legislative authorities in different countries and those systems are not accessible to parties outside that particular law enforcement or legislative authorities.

A. Ponmalar, Assistant Professor, Sri Sai Ram Institute of Technology, Chennai.
P. Leela Jancy, Assistant Professor, Sri Sai Ram Institute of Technology, Chennai.
V.R. Barath Kumar, Sri Sai Ram Institute of Technology.
B.K. Akshathaah, Sri Sai Ram Institute of Technology.
P. Pavithra, Sri Sai Ram Institute of Technology.

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

II. EXISTING SYSTEM

The system is a web-based system which comprises of crime analysis techniques such as hotspot detection, crime comparison and crime pattern visualization.

System cannot be directly validated using records of the police department because police records include both major and minor crime incidents. The system is based on newspaper articles so it includes only a subset of total crime incidents. So individual components of the system are evaluated and results of that evaluation are used to measure the effectiveness.

Crime data mining is the application of data mining techniques for crime analysis. Various researches have been carried out in this domain. Crimes can be divided into subcategories based on different criteria. In eight crime categories are given. They are traffic violations, sex crimes, theft, fraud, arson, drug offenses, cybercrimes and violent crimes.

There were many efforts to analyze different types of crimes using automated techniques but there is no unified framework describing how to apply those techniques to different crime types. In, they have used a framework which includes a relationship between the crime data mining technique and crime type characteristics. There are several existing systems which use crime data mining techniques for crime analysis such as, regional crime analysis program, data mining framework for crime pattern identification and narcotics network in Tucson police department.

An intelligent crime identification system is described in which can be used to predict possible suspects for given crime. They have used five types of agents namely, message space agent, gateway agent, prisoner agent, criminal agent and evidence agent.

III. DISADVANTAGE

1. Crime analysed here is by types of crimes happened.crime has not been analysed with any aspects
2. Storing of dataset is less by compared to proposed system accuracy will not be perfect while prediction analysis.

IV. LITERATURE SURVEY

Research on weather, a new variable will disclose how weather affects crime and influence certain criminal activity. The various variables such as climate, temperature, season and weather condition in particular area are considered for analysis. The studies in 1990's shows that there is a positive relationship between crime and weather. The research on weather will further update in study of criminology and helps in discovering better ways to control crime rates.

The publication of article 'British journal of Criminology' by Ellen G analyse the effects of heat and violence.[1]

Cohn stated that the data between 1967-1971 had relationship between temperature and violence. The probability of crime was seen increasing with temperature rising upto 85F.Cohn analysed daily crime in Newyork and New Jersey during his research time of one year.[2]

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

In Journal of Applied social psychology, John Cotton described his article 'Ambient temperature and violent crimes'[3]. He also studied about the research work of Baron and Ransberger in 1978. He collected the data of temperature and divided the crimes into violent and non-violent crimes, correlating them with minimum, average and maximum temperature in two selected Midwestern cities. The overall study showed a positive relationship as aggressive behavior increases as temperature 90° and mean crime was higher during days.

The article of Keith Harries and Stephen Standler titled 'Heat and violence'[4] published in journal of Applied Social psychology in 1998, discussed both a laboratory and field setting research to link temperature and behavior. They studied the city of Dallas, during summer of 1980-1981 by measuring thermal stress as Discomfort index(DI). Their analysis discovered that there is no curvilinear effect even in hot condition. This research is different since the study period is shorter with 3 months of summer and has few data sets to analayse.

James Horrocks and Andrea Menelova conducted a study on crime in Newzealand involving effects of weather[5]. The econometric techniques on daily crimes and static weather from 2000 -2008 were used. The study measured individual cost and benefits. If the expected value is higher than, then crime occurs.

Bell,[6](1992) analyses the theory of correlating violent crime and weather.

Horricks and Menclova analsyed the property of crime and weather giving conclusion[7], that if weather is good, people likely to be out, which increases theft and bulgary. Where, most expensive house on the street is expected to bulgary, by theory of expected benefits of crime. When weather is bad, people will stay in home, rate of crime is also less. This helps to enforcement of Service depending on weather condition.

In British Journal of Criminology, Simon Field discussed his article 'The Effects of Temperature and Crime'[8] focusing areas in England and Wales. He considered rainfall, hours of sunlight, temperature as weather variables and timeframe based on quarter to quarter and month to month. His results confirmed the probability of weather affecting crime is time with different cultural aspects.

Brian Jacob, Enrico Morotti and Lars Lafgren studied the dynamics of criminal behavior[9]. Evidence from weather shocks, that involves examine of dynamic criminal behavior. The data are collected from FBI's uniform crime reports. Their data showed that 10% increase in violent crimes and 2.6% reduction in following week, where weather between 2 weeks did not change. This made negative relationship between weather and crime. Thus overall findings suggest that short term changes as impact of daily or weekly crime rate but long run correlation is not linear (2007)

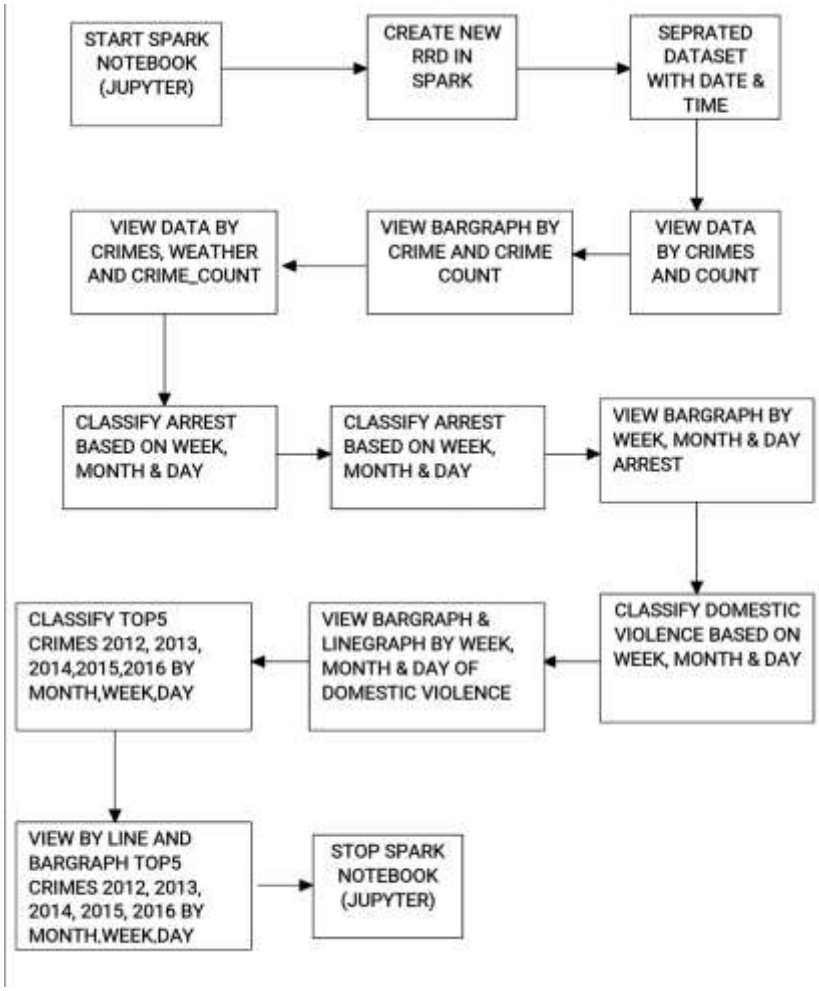
V. PROPOSED SYSTEM

The proposed system consists of a rich and simplified environment that can be used effectively for processes of crime analysis. We proposed pyspark for classification and datasets are stored into sparksql storing and retrival of information will be faster

We proposed weather data analysis along with crime data to analysis and predict which weather takes more crime action and represented crime analysis with graphical representation. We predicted the arrest based on year,

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

month, week and day and top 5 crimes in 2012, 2013, 2014, 2015, 2016 in month, week and day wise and we predicted weather analysis which weather situation had more number of crimes.



Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

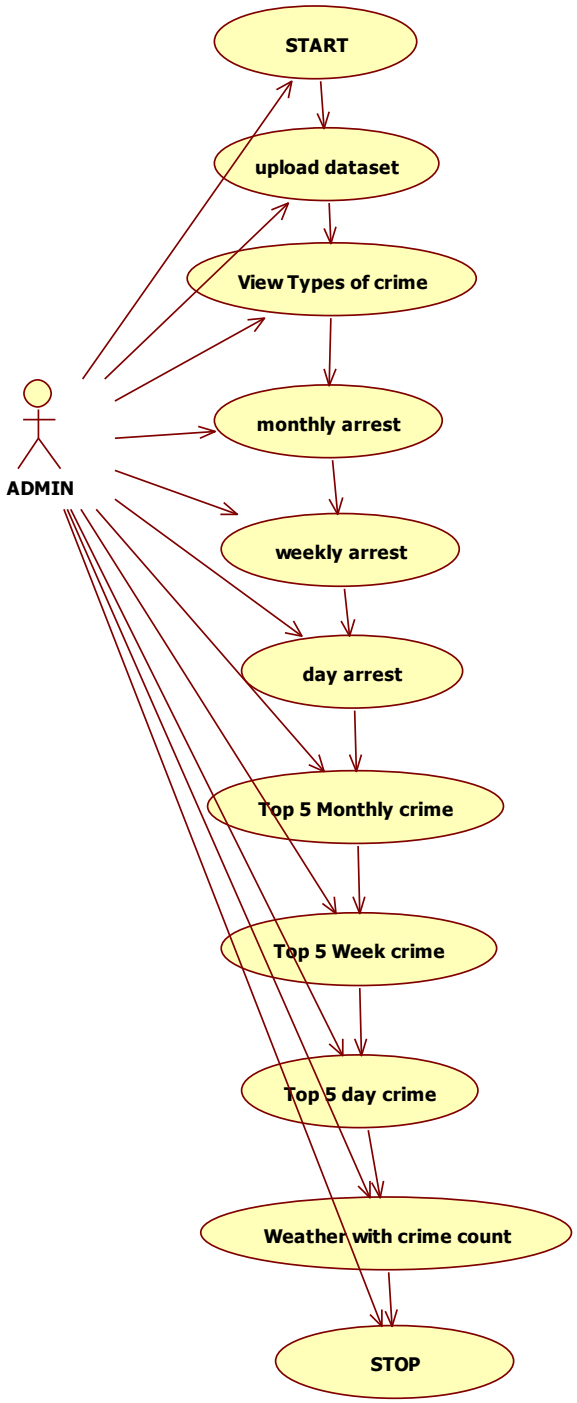
- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

Implementation

1. Tools used here is pyspark, jupyter console, anaconda & SparkSql
2. Spark and jupyter should be connected with bashrc path for notebook to start
3. Command to start notebook is **pyspark --driver-memory 8g --executor-memory 4g --master local**
4. Dataset will load with some spark import. Loaded dataset will be retrieved by omitting first 3 column
5. Retrive primary_type column and used groupby and count to retrive key and values as output
6. Retrieved output will be generated with bar graph which is crime activity and counts
7. Now we will classify Yearly arrest, Monthly arrest and daily arrest by getting year column

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

8. The output of yearly arrest, Monthly arrest and daily arrest will be viewed in line graph
9. Now we will classify Top 5 crimes happened in month,year & day with bar graph in month and other will be line graph
10. Now we will analysis weather data by plotting in bar graph.



DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

Advantage

1. Crime has been analysed with weather aspects to make crime prediction more efficient
2. Pyspark with jupyter console is used to analysis crime prediction

VI. CONCLUSION

The project's objective to analyze the crime data and provide fruitful suggestions to the department of security to protect the precincts and beats where the crime rates are high. Here, we used the different regression, clustering, classification and frequent growth patterns in order to organize and arrange the data in a ordered fashion. Thus I could ascertain that the research if applied on the appropriate data would result in generating patterns that would help to identify the crime rate early.

REFERENCES

[1] Cohn, E. G. (1990).Weather and crime. *British Journal of Criminology*, 30, 51-64.

[2] Cotton, J. L. (1986). Ambient temperature and violent crime. *Journal of Applied Social Psychology,* 16, 786-801.

[3] Field, S., 1992.The Effect of Temperature on Crime. *British Journal of Criminology,* 32, 340-351.

[4] Harries, K. D., &Stadler, S. J. (1988). Heat and violence: New findings from Dallas field data, 1980–1981. *Journal of Applied Social Psychology,* 18, 129-138.

[5] Horrocks, J., & Menclova, A. K. The Effects of Weather on Crime. 1-39.

[6] Jacob, B., & Lefgren, L., & Moretti, E. (2007). The Dynamics of Criminal Behavior: Evidence from Weather Shocks. *Journal of Human Resources,* 42, (3).

[7] G. Jiji-S. Anantharadha, "Automatic Tracking of Criminals using Data Mining Techniques", *Journal of The Institution of Engi-neers* (India): Series B, 2012. 258 *International Journal of Engineering & Technology.*

[8] Devendra Tayal, Arti Jain, Surbhi Arora, Surbhi Agarwal, Tushar Gupta, Nikhil Tyagi, "Crime detection and Criminal identification in India using data mining technique", Ai & Socie-ty, 2014.

[9] Shiju Sathyadevan, Devan S, Surya S, "Crime analysis and pre-diction using data mining", *First International Conference on Networks & Soft Computing (ICNSC2014),* 2014.

[10] Shyam Nath, "Crime Pattern Detection Using Data Mining", *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops,* 2006.

[11] Somayeh Shojaee, Aida Mustafa, Fatimah Sidi, Marzanah Jabar, "A Study on Classification Learning Algorithms to Predict Crime Status", *International Journal of Digital Content Tech-nology and its Applications (JDCTA),* Volume 7, Number 9, 1-3, 2013.

[12] Dawei Wang, Wei Ding, Henry Lo, Tomasz Stepinski, Josue Salazar, Melissa Morabito, "Crime hotspot mapping using the crime related factors—a spatial data mining approach", *Applied Intelligence,* 2012,

[13] Chung-Hsien Yu, Max Ward, Melissa Morabito, Wei Ding, "Crime Forecasting Using Data Mining Techniques", 2011 *IEEE 11th International Conference on Data Mining Workshops,* 2011.

[14] Arunima Kumar, Raju Gopal, "Data mining based crime investi-gation systems: Taxonomy and relevance", 2015 *Global Confer-ence on Communication Technologies (GCCT)* – 2015.

[15] Prajakta Yerpude and Vaishnavi Gudur, "Predictive Modelling of Crime Dataset Using Data Mining", *International Journal of Data Mining & Knowledge Management Process (IJDKP),* Vol.7, No.4, July 2017.

[16] Mohammad Keyvanpour, Mostafa Javideh, Mohammad Ebrahi-mi, "Detecting and investigating crime by means of data mining: a general crime matching framework", *Procedia Computer Sci-ence,* 2011.

[17] Ubon Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns", 2016 *Second Asian Conference on Defence Technology (ACDT),* 2016.

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020

[18] P. Chamikara, D. Yapa, R. Kodituwakku and J. Gunathilake, "Intelligent criminal identification system," *International Journal of Soft Computing and Engineering*, vol. 2, no. 1, pp. 175-180, 2012.

[19] H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin and M. Chua, "Crime data Mining: a general framework and some examples," *IEEE Explore-Computer*, vol. 37, no. 4, pp. 50-56, 2004.

[20] Crime mapping report mobile application using GIS[Online]Available: https://www.crimereports.com/

[21] A Fuzzy Grey Cognitive Maps-based intelligent security system[Online].

[22] R. Krishnamurthy and S. Kumar, "A survey of data mining techniques for analyzing crime patterns," *International Journal of Data Mining Techniques and Applications,* vol. 1, no. 2, pp. 117-120.

DOI: 10.37200/IJPR/V24I5/PR201757
Received: 11 Feb 2020 | Revised: 05 Mar 2020 | Accepted: 18 Mar 2020