# Sentimental Analysis on Women Safety Using Criminal Data

S. Sathishkumar[1], R. Reshma[2], Sowmiya kanagaram[3], Soundarya. P[4], Uma Bharathi. R[5]
[1]Professor, [2,3,4,5]UG Students,
[1,2,3,4,5]Department of IT, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India.

***Abstract*** **- The steadily increasing urbanization is causing significant economic and social transformations in urban areas and it will be posing several challenges in city management issues. In particular, given that the larger cities the higher crime rates, crime spiking is becoming one of the most important social problems in large urban areas. To handle with the increase in crimes, new technologies are enabling police departments to access growing volumes of crime-related data that can be analyzed to understand patterns and trends, finalized to an efficient deployment of police officers over the territory and more effective crime prevention. This paper presents an approach based on spatial analysis and auto-regressive models to automatically detect high-risk crime regions in urban areas and reliably forecast crime trends in each region. The final result of the algorithm is a spatio-temporal crime forecasting model, composed of a set of crime dense regions and a set of associated crime predictors, each one representing a predictive model for forecasting the number of crimes that will happen in its specific region. The experimental evaluation, performed on real-world data collected in a big area of Chicago, shows that the proposed approach achieves good accuracy in spatial and temporal crime forecasting over rolling time horizons. Along with this we propose new techniques with CNN and SVM algorithms which classifies the clustering data that has been implements and provides an accurate result and increases the performance evaluation of the entire process. Due to high and better accuracy the crime events has been reduced up to the core. The datasets of entire crime occurring areas and those datasets are analyzed and compared using CNN and SVM algorithms to get an accurate result.**

***Keywords*** **- Spatio-temporal, Support Vector Machine (SVM), Convolutional Neural Network (CNN).**

## I. INTRODUCTION

The 21st Century is frequently referenced as the “Century of the City”, reflecting the unprecedented global migration into urban areas that is happening nowadays. Such a steadily increasing urbanization is bringing huge social, economic and environmental transformations and, at the same time, presenting challenges in city management issues, like resource planning (water, electricity), traffic, air and water quality, public policy and public safety services. Moreover, given that the larger cities the higher crime rates, crime spiking is becoming one of the most important social problems in large urban areas, because it affects public safety, children development, and adult socio-economic status. With the ever-increasing ability of public organizations and police departments to collect and store detailed data tracking crime events, a significant amount of data with spatial and temporal information is daily collected. This offers the opportunity to apply data analytics methodologies to extract useful predictive models related to crime events, which can enable police departments to better utilize their limited resources and develop effective strategies for crime prevention. In particular, extensive criminal justice research studies show that the incidence of criminal events is not equally distributed within a city. In fact, crime rates can change with respect to the geographic location of the area (there are low-risk and high-risk areas) and crime trends can vary (seasonal patterns, picks, dips) with respect to the period of the year. For such a reason, an accurate predictive model must be able to automatically detect both which areas in the city are more affected by crime events and how the crime rate of each specific area varies with respect to the temporal period. This knowledge can enable police departments to more efficiently allocate their resources to specific crime hot spots, allowing for the effective deployment of officers to areas of high risk or removal of officers from areas seeing decreasing levels of crime, thus more efficiently preventing or quickly responding to criminal activity.

## II. RELATED WORK

**M. Cobley; G. Graham:** An overview of resources, facilities and services provided by Canada, as Host Country, to meet the requirements for multiple language versions and access systems. The data set used here is Urban Area. It provides needed facility for people but consumes large amount of time and data. It is used to maintain the people database and can provide efficient information to the government. In the future it can be applied in various fields like automobiles, wildlife etc.

**Venkat Surya Dasari;** Maryam Pouryazdan; Burak Kantarci: It perform a thorough feasibility analysis of two possible data acquisition approaches for crowd-solicited IoT data. The dataset used here is population. It provides detailed information about the citizens but privacy for citizens is a quiet complicated one. It can be applied in various public sector. In the future it can be implemented in numerous fields like corporate etc.

**Mohammad A.Tayebi:** Martin Ester; Uwe Glässer; Patricia L.Brantingham: This phenomenon has drawn attention to spatial crime analysis, focusing on crime areas with higher crime density. In this paper we present crimetracker, a personalized random walk based approach to spatial crime analysis and crime location prediction outside of hotspots. The dataset used here is Crime Analysis. It reduces crime rates in a concerned city but hard to collect the data. The application is designed for police department to reduce the crime rates in the city. In the future it can be enhanced with more efficient features like predicting the locations using tracking options.

**Henrik I Christensen:** This course covers the general area of Simultaneous Localization and Mapping (SLAM). Initially the problems of localization, mapping, and SLAM are introduced from a methodological point of view. Different methods for representation of uncertainty will be introduced including their ability to handle single and multi-mode uncertainty representations. The dataset used here is Locating and Tracking Mapping Datasets. It gets a clear status of every information but there is a chance for misspelled data. It can be applied in various fields such as marketing, logistics, crime predictions etc. In the future it can be implemented in attaining accurate and perfect results in crime predictions including current status.

**B. Chandra; Manish Gupta:** M. P. Gupta: Clustering multivariate time series has potential for analysing large volume of crime data at different time points as law enforcement agencies are interested in finding crime trends of various police administration. The dataset used here is Clustering Crime Analysis. Results of different areas can be easily predicted but quiet complicated to acquire data as large amount of data are used. The application is used to predict the criminals in various areas. In the future, the criminal's actors in various fields can be analysed and predicted.

## III. PROBLEM DESCRIPTION

To implement this crime predictor there have been many proposed solutions using spatial analysis approach, where shapes of the detected regions are automatically traced by the algorithm without any pre-fixed division in areas. Then, a specific crime prediction model is discovered from each detected region, analyzing the partitions discovered during the previous step. The final result of the algorithm is a spatiotemporal crime forecasting model, composed of a set of crime dense regions and a set of associated crime predictors, each one representing a predictive model to forecast the number of crimes that will happen in its specific region. As case study, we present here the analysis of crimes within a big area of Chicago involving about two million crime events. Crime data has been gathered a Web framework that provides public access to more than one hundred urban datasets. The results of the experimental evaluation show the effectiveness of the approach, by achieving good accuracy in spatial and temporal crime forecasting over rolling time horizons. However, Due to lack of information gathering the specific action for the occurred crime sequences is a slow process and the accuracy cannot be maintained due to more paper work and manual process, which is an interrupted pattern in the system. Currently, many cities have released crime-related data as part of an open data initiative. Using this as input in our proposed system, we have applied analytics to be able to predict and hopefully prevent crime in the future. In this work, we applied data analytics to the crime dataset, as collected and available through the Open Data initiative. The main focus is to perform an in-depth analysis of the major types of crimes that occurred in the city, observe the trend over the years, and determine how various attributes contribute to specific crimes. Furthermore, we leverage the results of the exploratory data analysis to inform the data pre-processing process, prior to training various machine learning models for crime type prediction. More specifically, the model predicts the type of crime that will occur in each district of the city. We observe that the provided dataset is highly imbalanced, thus metrics used in previous research focus mainly on the majority class, disregarding the performance of the classifiers in minority classes, and propose a methodology to improve this issue. The proposed model finds applications in resource allocation of law enforcement in a Smart City. In addition to the above advantages it reduces the latency up to the core and increases the reliability of the system. This method also enhances the effectiveness and performance of the approach.

## IV. SYSTEM IMPLEMENTATION

Data Acquisition: Initially we collect the crime dataset of a specific city and pre-process the data that has been related with the dataset. The data and information that has been inbuilt in the dataset are analysed and pre-processed with various algorithms. The pre-processed data are extracted according to the features they are related. During the operation the input data is denoted as the dataset of crime data of a specific city, from that dataset the data are pre-processed with salient features of the algorithms.

Pre-processing: The pre-processed data are extracted and compared with the current crime occurrences with the crime occurrence in the dataset. So that the details can be detected in an efficient way. The detected result has been analysed for proper results and in case of occurrences of any errors or any defects the same procedure has been repeated to attain perfect results and evaluation.

Classification: The data are compared with the help of Support Vector Machine Algorithm (SVM) which classifies the given data and compares them for proper predictions. After comparing the data with the help of SVM the perfect result has been gathered and the gathered data are clustered using K-Means Clustering Algorithm to gain the accurate result of the entire system.

Segmentation: The data are compared with the help of Support Vector Machine Algorithm (SVM) which classifies the given data and compares them for proper predictions. The K-Means Clustering provides the accurate value to attain the results. The data can be updated in case of newly occurred crimes and those data are updated to the system which enhances the efficiency of the system in an enormous way. As a result the new crime data are added to the previous dataset which contains the list of crime events of the cities and the entire data can be compared and clustered with the SVM and K- Means Clustering Algorithms to retrieve better results.

Classification: It performs a spatial clustering of the data set, where each cluster represents a dense region of crimes. The density-based notion is a common approach for clustering, whose inspiring idea is that objects forming a dense region should be grouped together into one cluster. In our implementation, this step is performed by applying, a popular density-based clustering algorithm that finds clusters starting from the estimated density distribution of the considered data. We have chosen the SVM algorithm because it has the ability to discover clusters with arbitrary shape such as linear, concave, oval, etc. and differently from other clustering algorithms proposed and it does not require the predetermination of the number of clusters to be discovered.

Feature Extraction: Given a specific dense region, the crime predictor method discovers a predictive model to forecast the number of crimes that will happen in its specific area. In our implementation, this has been performed by the Seasonal Auto Regressive Integrated Moving Average model, which is defined as a combination of auto regression, moving average and difference modelling. Briefly, having the time series, which is the value of the time series at the timestamp, an ARIMA model is written are the regression coefficient of the moving average part, are lagged values and lagged errors, and it is the white noise and takes into account the forecast error.

Prediction: As described, crime dense regions are detected by applying our ad-hoc modified version of SVM, which exploits a decay factor that gives a higher weight to the recent crime events than the older ones. Moreover, in order to detect high quality crime dense regions, it is necessary to profitable tune the key parameters of the algorithm so as to improve results' performance. In particular, the values of the SVM's parameters and determines the size of the clusters, as they represents the minimum crime density required by an area to be part of a cluster. On the one hand, the larger is the extension of the dense regions detected: this results in the discovery of large regions that actually are no longer dense. On the other hand, the smaller the cluster sizes, resulting in a high number of dense regions detected that could be not significant for the analysis. Conversely, growing the value results in increasing the fragmentation of the produced clustering assignment. The values of a key factor for the accuracy of the dense region detection phase and for the right balance among separability, compactness and significance of clusters.

Evaluation Criteria: To evaluate the performance and the effectiveness of the approach that has been described in the paper, we carried out an experimental evaluation by analysing crimes occurring in a big area of the city. The main goal consists in detecting the most significant crime dense regions and discovering a predictive model for each one to forecast the number of crimes that will happen in the future in each area. In the following sub-sections, we describe the main issues of our analysis: data description and gathering, crime dense region detection, the regressive model training for each region, and the evaluation of the model on the test set. Reports the values of K- Means Clustering for the whole area and the top three largest crime dense regions, by considering one-year-ahead, two-year-ahead and three-year ahead prediction horizons.

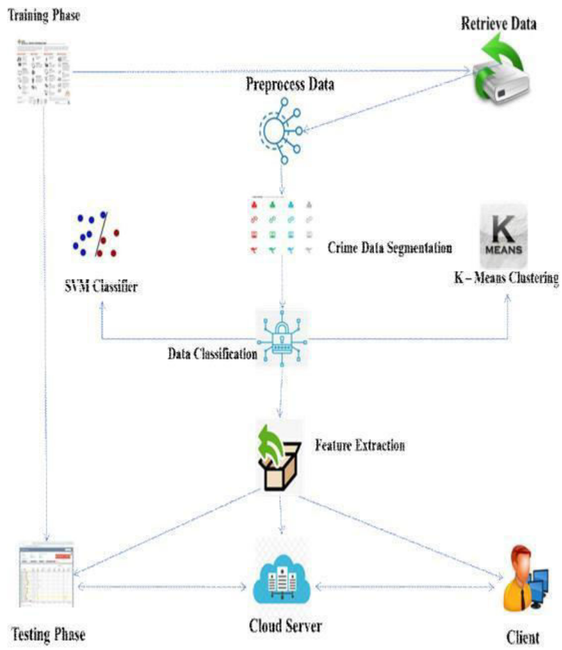


Fig. 1. Architecture Diagram

## V. RESULT AND EXPECTED OUTCOMES


Fig. 2. Login

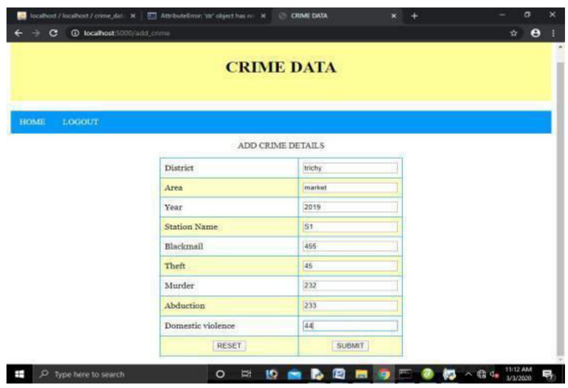


Fig. 3. Crime data

DATABASE CRIME DATA

TABLE I. TABLE STRUCTURE FOR TABLE ADD_DATA

| Field | Type | Null | Default |
|---|---|---|---|
| id | int(11) | Yes | NULL |
| district | varchar(50) | Yes | NULL |
| area | varchar(50) | Yes | NULL |
| year | int(11) | Yes | NULL |
| crime1 | int(11) | Yes | NULL |
| crime2 | int(11) | Yes | NULL |
| crime3 | int(11) | Yes | NULL |
| crime4 | int(11) | Yes | NULL |
| crime5 | int(11) | Yes | NULL |
| rdate | varchar(50) | Yes | NULL |

TABLE II. TABLE STRUCTURE FOR TABLE ADMIN

| Field | Type | Null | Default |
|---|---|---|---|
| id | int(11) | Yes | NULL |
| username | varchar(20) | Yes | NULL |
| password | varchar(20) | Yes | NULL |

TABLE III. TABLE STRUCTURE FOR TABLE REGISTER

| Field | Type | Null | Default |
|---|---|---|---|
| id | int(11) | Yes | NULL |
| name | varchar(50) | Yes | NULL |
| mobile | bigint(20) | Yes | NULL |
| email_address | varchar(50) | Yes | NULL |
| address | varchar(100) | Yes | NULL |
| username | varchar(20) | Yes | NULL |
| password | varchar(20) | Yes | NULL |
| rdate | varchar(20) | Yes | NULL |

## VII. CONCLUSION AND FUTURE WORK

This paper presented a general algorithm for Spatio-Temporal Crime Prediction in urban areas, that takes advantages from the partitioning of the whole analyzed area by detecting crime dense regions (of arbitrary shapes). Such regions are then analyzed and a different forecasting autoregressive model is tailored specifically for each detected region. Experimental evaluation, performed on crime data of a wide area of a city, showed that the proposed methodology can forecast the number of crimes with an high accuracy. Furthermore, the approach gives fine-grained information about where crime events are expected to occur.

In future work, other research issues may be investigated. First, we may further explore the application of other spatial analysis approaches for the detection of crime dense regions and for modelling and forecasting crime trends on such regions. Second, we may perform an extended experimental evaluation on an wider urban territory, to assess the results obtained in the case study reported here. Third, we may apply such an approach for spatials-temporal prediction of other kind of events, different than crimes.

## REFERENCES

[1] R. Billions, "Authorship of clinical trial documents," Medical Writing, vol. 25, pp. 33 – 35, 2016.

[2] “Cities: The century of the city,” Nature 467, 900-901 (2010), vol. 467, pp. 900–901, 2010.

[3] F. Cicirelli, A. Guerrieri, G. Spezzano, and A. Vinci, “An edge-based platform for dynamic smart city applications,” Future Generation Comp. Syst., vol. 76, 2017.

[4] M. Tayebi, M. Ester, U. Glasser, and P. Brantingham, “Crimetracer: Activity space based crime location prediction,” in Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, 2014, pp. 472–480.

[5] H. Wang, D. Kifer, C. Graif, and Z. Li, “Crime rate inference with big data,” in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’16. ACM, 2016, pp. 635–644.

[6] C. Catlett, T. Malik, B. Goldstein, J. Giuffrida, Y. Shao, A. Panella, D. Eder, E. van Zanten, R. Mitchum, S. Thaler, and I. T. Foster, “Plenario: An open data discovery and exploration platform for urban science,” IEEE Data Eng. Bull., vol. 37, no. 4, 2014.

[7] T. Wang, C. Rudin, D. Wagner, and R. Sevieri, “Learning to detect patterns of crime,” in Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2013, 2013

[8] D. E. Brown and S. Hagen, “Data association methods with applications to law enforcement,” Decision Support Systems, vol. 34, no. 4, pp. 369–378, 2003.