Survey on Crime Analysis Using Data Mining Based on Mobile Platforms

Fatima Shaker Hussain [a], Abbas Fadhil Aljuboori [b]**

[a] *Informatics Institute for Postgraduate Studies (IIPS), Iraqi Commission for Computers and Informatics (ICCI), Iraq ,Baghdad, email: Ms201910521@iips.icci.edu.iq*

[b] *College of Engineering , University of Information Technology and Communications, Baghdad, Iraq, email: abbas.aljuboori@uoitc.edu.iq*

A R T I C L E   I N F O

A B S T R A C T

With a large rise in crime globally, there is a necessity to analysis crime data to bring down the rate of crime. This encourages the police and people to occupy the required measures and more effectively restricting the crimes.   Crime analysis and spatial analysis using Geographic Information System(GIS) tools such as hot spot generation, zoning, navigation and crime profiling, detection of mobile locations and various web-based applications are well known and can be used scientifically for citizens advancement, while crime prediction and control can be used effectively. Crime data analysts will enable the law enforcement officials to speed up the crime resolution process. We can analyze previously unknown, helpful information from unstructured data by using the theory of data mining. The purpose of this work is to perform a survey on the crime analysis by using several data mining techniques that has been applied towards criminal activities over the years to be a valuable guide to young researchers that are interested in this field**.**

1. Introduction

Crimes are considered is an unlawful activity of all kinds is punished by law and impacting a society's quality of life and economic development. It is deemed an important laborer that decides whether or not individuals migrate To a new town and where sophisticated geographic information systems and new data mining techniques are still needed to enhance crime analytics and better protect their communities. [1] The wide variety of applications in criminology for approaches to data mining is very valuable this will benefit and assist the police as a result has made it an important area of study. In particular, classification and clustering methods may be used to assist in the field of crime analysis. In the criminal sector,

∗Corresponding author: Fatima Shaker Hussain.

Email addresses: Ms201910521@iips.icci.edu.iq.

data mining systems have played a key role because they assisted humans to make environments that can facilitate this sector.[2]

Crime analysis uses the previous data already available to predict the time and location that the crime will take place. Several good techniques provide quicker evolution of the crime data set, help to forecast the correct place of crime, as well as help to keep track of crime analysis resources.[3] The advancement of information technology expands the horizons for the use of resources by keeping pace with the broad trend in the use of the Internet, mobile devices, positioning techniques, and electronic maps in the development of the concept of security. The Google Maps API provides many merit and solutions on electronic maps for location support and positioning services that play a lusty service in detecting crime and how to tackle offense.[4] Previously, data mining techniques surveys were conducted using certain techniques that covered a Limited amount of data mining. In this survey, we intend to cover all possible techniques used for the analysis of crime. The motivation of this survey provide insights into the crime analysis manner to produces a resource includes various kind of operations for crime analysis that can be used in any police station for crime analysis . The survey is structured in such a way for easy understanding of the concepts. a mobile policing solution is discussed in section 2. crimes type are discussed in the section 3 that include different type of illegal activities .section 4 give insight about data mining in crime analysis and finally, section 5 mention previous studies in this filed.

2. Toward a Mobile Policing Solution

For decades, technology and policing have been connected and this is illustrated by going back to the invention of the telephone, the vehicle, and the two-way radio. As seen by the spread of mobile and wireless technology, ocular and acoustic technology, and high-powered computing, technology today seems to be progressing at an ever-accelerating pace and other technological advancements. These and other innovations are being adopted by several departments to improve execution and improve results, particularly in times of limited capital and increased mass scrutiny and criticism of law fulfillment policies and performance. [5] Many technologies have the best potential for impact on the police, particularly in speeding up processes where reporting speed and accuracy can be enhanced, as well as the ease with which officers can identify suspects, cars and trucks, or places of interest.[6]

Perhaps not surprisingly, the technology credited with having the biggest effect on national police forces are automated record management systems (RMS) and computer assisted dispatch (CAD). CAD and RMS are important for acting out the most basic skilled police operations, responding to service calls, and information administration. RMS/CAD technology is also important for the development of data on which other operations and software applications depend, like Remote sensing, hot-spot policing, and other location-based activities.[5]

2.1 Feasibility Through Available Mobile Technology

Mobile solutions are very important For several different purposes, But this is primarily due to the ubiquity and mobility of these computer platforms. The recent updates in communication, mobile devices and technology facilitate developing a system . In fact, combining a mobile phone with the any system has   many advantages. For example, mobile phones are of daily use and easy to be understood by anyone. Further, they are powerful and quickest means of using technology. Moreover, with the internet enabled mobile phone, it becomes easily benefit of an internet server. [7] it can be used from many aspects, for example, A mobile solution allows law enforcement to report crime and incident location data in real-time using global positioning system(GPS) and cellular data networks to rapidly and reliably respond to crime scene reporting.]8[

2.2 Benefits of The Mobile Policing

Identified long-term potential benefits of mobile policing include [9]:-

- Can acts as a force multiplier
- it advances officer safety
- it advances time efficiency to both police officers and staff
- it improved rates for detection of crime

2.3 Mobile Application Platform

Mobile application creation is openly challenging the production of quality applications with less susceptibility to failure. As the feasibility of mobile apps has risen at an unprecedented rate, reliance on them has also increased. The growing number of mobile devices and potential applications is pushing developers to use sound practices to address various development and implementation barriers. [10] Mobile apps can be divided widely into two groups, namely native and hybrid apps. In the last few years, the field of mobile application growth has grown rapidly. The only type of applications that were common in the past were native mobile applications. Native mobile apps certainly have the best user interface, but when the application is made for multiple platforms, hybrid mobile device applications would be favored and time and expense are the primary factors. There are various pros and cons in using native or hybrid application development technology, we are going to state comparison between them. [11]:-

Table (1) The comparison between native and hybrid

Type of Application/ Features	Native mobile Apps	Hybrid mobile Apps
Programming Language	Java(Android), ObjectiveC(iOS)	HTML5,JavaScript and css
Development Cost	Expensive	Reasonable
Tools and Debugging	provides better tools for testing and debugging	There are not many tools available
Performance	Better performance	Slower performance

3. Crime Type

Crime analysis is a role in law enforcement that involves Organized analysis that describes the pattern of crime and defines it. Crimes can be categorized into various categories .

3-1 Fraud Discovery

A fraud misdirects or unfairly benefits other individuals. Fraud includes any action, exclusion or concealment, inclusive violation of confidence in a lawful or equitable obligation, resulting in damage to others. Check fraud, credit card fraud, internet transactions(means selling falsify items), insurance fraud(means fake insurance reported for car damage, Health Services Care costs, and other expenses) are different forms of fraud. [12]

3-2 Cyber Crime

Since businesses began using computers in the process of doing business, cybercrime has increased in scope and financial costs. Cybercriminals are getting more advanced and targeting customers and public and private

companies. In any country, cybercrime detection has so significant accountabilities for the law enforcement system. Cybercrime comprises breaking down privacy or destroying the resources of the computer system, such as files, website pages, or apps. [13]

3-3 Violent Crime

A violent offences is a felony in which an accused side attempts to use force against a victim. it includes, for example, killing or disgracing, both crime crimes of the rough act called a goal. [12]

3-4 Traffic Violence

Traffic offences occur when laws are damaged by drivers on roads and highways that regulate the movement of vehicles. The rising number of cars in towns is causing high volumes of traffic, This means that traffic violations are being further common, which can cause significant possessions damage and further incidents that can endanger people's lives. Irregularities detection systems are required to fix this issue and obviate such consequences. [12]

4. Data Mining in Crime Analysis

Crime analysis is an analytic tool for the crime detection process. It is used to decrease the crime rate in places where the trends, patterns and sequence of crimes are known to crime enforcers. Analyze crime to meet an evolving society's law enforcement needs and also to understand general criminal patterns.[14]

Data mining techniques can be used to resolving most complicated criminal cases in other words, to facilitate the process of analyzing the crime in general. Data Mining is the process by which large pre-existing datasets are analyzed and checked in order to produce newfangled data that perhaps important to the organization. Data mining techniques are useful for improving the accuracy, execution, and speed of the analysis process.[12] Data Mining has been used in different ways, for example to predict crime-prone areas, to determine the safe path, to photograph the crime scene, etc.

5. Literature Survey

Tahani Almanie et al. [1] Focused on finding spatial and temporal criminal hotspots using a set of real-world datasets of crimes include Los Angeles and Denver cities. Certainly, identifying ties between elements of crime will significantly help to predict possible hazardous hotspots at a clear point in the future. The strategy was therefore aimed at concentrating on three main elements of crime data, which are the kinds of crimes, the timing of incidences, and the place of crimes. Using the Apriori algorithm on datasets to classify all possible patterns of crime often regardless of the type of crime committed, then was used the "Naïve Bayesian classifier and decision tree classifier" to construct two separate classification models, to forecast the possible form of crime in a particular place over a particular period In the future. It achieves an accuracy of 51 percent in Denver crime prediction concerning the Naïve Bayesian classifier, while it hits 54 percent for Los Angeles. On the other hand, with 42 percent for Denver and 43 percent for Los Angeles, the decision tree classifier records less prediction accuracy.

Félix Mata et al. [15] Emphasis was on designing mobile information systems in urban environments for routing and urban planning. It generates a hybrid solution using semantic analysis and classification algorithms to find safe routes with data from social media and official crime reports. The Bayes algorithm uses data submitted by the mobile application (origin and destination points) to return a path that prevents places where crime has happened.

Jakaria Rabbi et al.[16] The linear regression model is used to predict Bangladesh's potential crime patterns. The actual crime dataset is compiled from various sources of the Bangladesh Force Police. The model of linear regression is trained on the real dataset. Crime forecasting for, robbery, murder, persecution of women and children, abduction, burglary, theft, and others for different regions of Bangladesh is carried out after training the

model. This work is beneficial for Bangladesh's police and law enforcement agencies to anticipate, prevent or address potential crime in Bangladesh.

Sivanagaleela and Rajesh.[14] Presented project to examine the crime patterns and classify crime zone and which kinds of crime happen more often in which location. Based on the fuzzy clustering and Decision tree techniques the crime-prone zone is recognized at less time. the technique is training with historical data with different attributes.

Anees Baqir et al.[17] Presented A comparative study of two hierarchical clustering algorithms are Density-based spatial clustering of applications with noise (DBSCAN) and hierarchical agglomerative clustering (HAC) to determine crime hotspots in major cities. The algorithms have been exercised to the New York City(NYC) data, and this data can be used to take necessary steps to efficiently police and use manpower to counter these incidents. In addition, the study showed that in terms of time and memory requirements, HDBSCAN outperformed HAC. Crimes Dataset of NYC in a comma-separated value (CSV) format, It contained 35 features, including incidents of crimes reported over the years.

Bhavna Saini et al.[18] The developed module offers an interactive image to navigate hither and thither the crime scene using Google Maps and can aid the analyst evaluates the protection of an area, which locations can also be the focus for the nearer attack. Presents the methods of visualization and algorithms of classification that can be used to forecast crimes, such as K-Nearest Neighbor and Naïve Bayes, and supports law enforcement agencies. As data is acquired from the official united kingdom(U.K) website. the dataset used for the work is accurate, true, and credible, it includes 11 attributes in total.

Atharva Deshmukh et al.[19] The study provide Application will use high-level machine learning information at entirely different times of day and night to divine crime rates within the zones of the city. With the aid of the latest crime data collection, the application will be able to predict new crime trends in the space. Predict the crime hotspots primarily aimed at helping people to distinguish between safe and dangerous areas when traveling. Django Rest framework ,React Native was used to implement the application.

Shivangi Soni1 et al. [20] A creative way to find the safest route with the lowest risk score is identified. the system that suggests people the path that is safest to travel from source to destination. If only one route is possible between the source and destination point then that route is displayed as the safest route. Else if more than one route is possible, the risk score must be calculated for all the possible routes. The datasets used are the accident dataset and arrest dataset, which are taken from NYC OpenData. Dataset was dealt with using k-nearest and k mean algorithms.

Prajakta Yerpude and Vaishnavi Gudur. [21] This study focuses on forecasting using Decision Trees, the Naïve Bayes algorithm, and the Random Forest algorithm. The target is to forecast most of the top features with the precise predictive model that affects the high rate of crime that will ultimately assist constabulary or law fulfillment officials to occupy the appropriate action. The UCI repository crime dataset consisting of crime data in Chicago was used. In order to get an accurate crime data set, removal of missing values was necessary for be clean data. The performance of a model is measured by Accuracy, Precision, Recall, and F1 Score (confusion matrix). confusion matrix calculated after performing 10-fold Cross-Validation awarded the values, which are much more de facto and businesslike. The Random Forest Classifier produces the most equiponderant outcome for the "precision, accuracy, recall and F1 score" of three prediction models. The results will be shown in the following tables.

Table(2) Random Forest Classifier

Random Forest classifier	10-fold Cross-Validation
Accuracy	83.39%
Precision	88.39%
Recall	84.86%
F1 Score	86.54%

Table(3) Decision tree Classifier

Decision Tree classifier	10-fold Cross-Validation
Accuracy	75.9%
Precision	80.62%
Recall	81.53%
F1 Score	81.22%

Table(4) Naïve Bayes Classifier

Naïve Bayes Classifier	10-fold Cross-Validation
Accuracy	77.64%
Precision	92.53%
Recall	96.82%
F1 Score	79.58%

Fantaye Ayele.[22] This study is to create predictive models that could assist in the labor to examine crime trends in order to support crime reduction efforts at the police office of the city of Hossaena. For this research, a six-step hybrid knowledge discovery process model is followed based on the existence of the issue and attributes in the dataset. The algorithms used during predictive model building experimentations are found in Weka 3.8 version. To build the predictive model, J48 and naive Bayes are trained and evaluated. The decision tree of J48 records better performance with 96.34 percent precision.

Biswajit Panja et al.[23] This research paper, identified Crime mapping analysis based on (K – Nearest Neighbor(KNN) and Artificial Neural Network(ANN)algorithms to simplify this process. In tandem with analytic

methods, Crime Analysis uses quantitative and qualitative evidence to address crimes. the data that used is collected from the police department. data is processing to remove Unnecessary data or not required. the authors used two types of classification are supervised and unsupervised to build a model.

Trung et al.[24] This paper explains a system of crime prediction that predicts the types of crimes that will happen based on place and time. The proposed method is a service provided to Portland Police Bureau (PPB). The dataset used is derived from two separate sources: first, the PPB dataset given and the American FactFinder website dataset. Many types of data preprocessing are done include "Data Cleaning, Data Reduction, Data Transformation, Data Discretization, Data Integration". Several machine learning algorithms were used to create models for this crime prediction to perform the job, including Support Vector Machines or SVM, Random Forest, Gradient Boosting Machines, and multilayer neural networks. For the first dataset, SVM outcomes do not seem to be a good model for this task because of the low accuracy compared to other methods used Also training time of the SVM model is higher than the training time for other models. while in the second dataset, the SVM model shows the best accuracy models when compared to other methods

Param et al.[25] A crime predictive model was introduced to forecast crimes related to property burglary in Canada's Vancouver Region. for this model, there are two datasets used are crime and neighborhood. from Vancouver Police Department (VPD) gathered the crime dataset and the neighborhood dataset provides the boundaries for the 22 local areas of the city in the GIS. KNN and decision-tree algorithms were used to train the model. the accuracy obtained of crime prediction is between 39% and 44%. the choropleth mapping is used to map hotspots, this reflects where incidents of crime are more condensed, offering insights into criminal activity.

Shamaila Qayyum et al. [26] Data mining methods, primarily used for crime detection, have been proposed. Based on its strength and weakness, a comparative study of each approach is presented. Each technique is specific to its use. "entity extraction, clustering approaches, association rule mining, sequential pattern mining, deviation detection, classification, string comparator, social network analysis " are current data mining techniques that are employed for crime detection and investigation. The table below illustrates these techniques.

Table(5) strength and weakness of techniques

TECHNIQUE	STRENGTH	WEAKNESS
Entity extraction	Machine learning makes it easier	Large amounts of clean data required
Clustering	Detect outlier without any required label data	Computational cost is high. its effectiveness also depends upon the method used
association rule mining	Support	It is used for the most accurate classification rules
Sequential pattern mining	Wide range of applicable in fraud detection	Large amount of structured data is required
deviation detection	Widely applicable in fraud detection	Sometimes its data dependency becomes a hurdle
Classification	Very less time consumption	Predefined scheme of classification and complete training dataset required
String comparator	Accuracy in terms of numerical value	Large amount of computation required

Social network analysis	Focus on relationships between actors rather than attributes of actor	Won't identify network's true leaders

Huang et al.[27] Focusing on a particular approach to the prediction of illegal activity based on social network interactions based on mining locations . Data can be collected by using these interactions using the spatial interactions of individuals and data sets. They created a working phase in which a number of characteristics from the Foursquare and Gowalla used in the area of San Francisco Bay are classified. Spatial characteristics derived from the map track the patterns of crime and crime incidents and analyze them to classify urban areas with high crime activity. Their work seeks to leverage data from local social networks in order to investigate illegal activity in urban areas.

References

[1]   T. Almanie, R. Mirza, and E. Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 4, pp. 01–19, 2015, doi: 10.5121/ijdkp.2015.5401.

[2]   H. B. Fredrick David and A. Suruliandi, "Survey on Crime Analysis and Prediction Using Data Mining Techniques," *ICTACT J. Soft Comput.*, vol. 7, no. 3, pp. 1459–1466, 2017, doi: 10.21917/ijsc.2017.0202.

[3]   S. R. Bandekar and C. Vijayalakshmi, "Design and analysis of machine learning algorithms for the reduction of crime rates in India," *Procedia Comput. Sci.*, vol. 172, no. 2019, pp. 122–127, 2020, doi: 10.1016/j.procs.2020.05.018.

[4]   A. N. Yousif and A. S. Elameer, "An expert system for the tourism destinations in Iraq based on the google maps API," *Proc. - 2018 1st Annu. Int. Conf. Inf. Sci. AiCIS 2018*, pp. 1–6, 2019, doi: 10.1109/AiCIS.2018.00014.

[5]   K. Strom, "The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource: Document Title: Research on the Impact of Technology on Policing Strategy in the 21st Century, Final Report," 2017, [Online]. Available: https://www.ncjrs.gov/pdffiles1/nij/grants/251140.pdf.

[6]   C. Lum, C. S. Koper, and J. Willis, "Understanding the Limits of Technology's Impact on Police Effectiveness," *Police Q.*, vol. 20, no. 2, pp. 135–163, 2017, doi: 10.1177/1098611116667279.

[7]   I. Shadeed Al-Mejibli and D. Hamed Abd, "Mushroom Diagnosis Assistance System Based on Machine Learning by Using Mobile Devices," *J. Al-Qadisiyah Comput. Sci. Math.*, vol. 9, no. 2, pp. 103–113, 2017, doi: 10.29304/jqcm.2017.9.2.319.

[8]   A. Akpan, B. Barida, and M. Shedrack, "Toward an effective crime mapping solution for Nigeria: Leveraging Emerging Mobile Platforms," vol. 5, no. 9, pp. 108–118, 2018, [Online]. Available: https://www.researchgate.net/publication/329216903_Crime_Mapping_Solution.

[9]   W. Graham, "Benefits of implementation of mobile devices with frontline police officers in Police Scotland," no. January, 2021.

[10]   Ziema Mushtaq, "Mobile Application Development Trends and Challenges," *Int. J. Eng. Technol.*, vol. 3, no. 8, pp. 1096–1099, 2016, doi: 10.1590/S0034-89102005000600008.

[11]   A. Khandeparkar, R. Gupta, and B. S. B.Sindhya, "An Introduction to Hybrid Platform Mobile Application Development," *Int. J. Comput. Appl.*, vol. 118, no. 15, pp. 31–33, 2015, doi: 10.5120/20824-3463.

[12]   S. Prabakaran and S. Mitra, "Survey of Analysis of Crime Detection Techniques Using Data Mining and Machine Learning," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012046.

[13]   K. C. Lekha and S. Prakasam, "Data mining techniques in detecting and predicting cyber crimes in banking sector," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017*, no. August, pp. 1639–1643, 2018, doi: 10.1109/ICECDS.2017.8389725.

[14]   B. Sivanagaleela and S. Rajesh, "Crime analysis and prediction using fuzzy c-means algorithm," *Proc. Int. Conf. Trends Electron. Informatics, ICOEI 2019*, no. Icoei, pp. 595–599, 2019, doi: 10.1109/ICOEI.2019.8862691.

[15]   F. Mata *et al.*, "A Mobile Information System Based on Crowd-Sensed and Official Crime Data for Finding Safe Routes: A Case Study of Mexico City," *Mob. Inf. Syst.*, vol. 2016, 2016, doi: 10.1155/2016/8068209.

[16]   M. A. Awal, J. Rabbi, S. I. Hossain, and M. M. A. Hashem, "Using linear regression to forecast future trends in crime of Bangladesh," *2016 5th Int. Conf. Informatics, Electron. Vision, ICIEV 2016*, no. June 2020, pp. 333–338, 2016, doi: 10.1109/ICIEV.2016.7760021.

[17]   A. Baqir, S. U. Rehman, S. Malik, F. U. Mustafa, and U. Ahmad, "Evaluating the Performance of Hierarchical Clustering algorithms to Detect Spatio-Temporal Crime Hot-Spots," *2020 3rd Int. Conf. Comput. Math. Eng. Technol. Idea to Innov. Build. Knowl. Econ. iCoMET 2020*, 2020, doi: 10.1109/iCoMET48670.2020.9074125.

[18]   H. K. R. Toppireddy, B. Saini, and G. Mahajan, "Crime Prediction & Monitoring Framework Based on Spatial Analysis," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 696–705, 2018, doi: 10.1016/j.procs.2018.05.075.

[19]   A. Deshmukh, S. Banka, S. B. Dcruz, S. Shaikh, and A. K. Tripathy, "Safety App: Crime Prediction Using GIS," *2020 3rd Int. Conf. Commun. Syst. Comput. IT Appl. CSCITA 2020 - Proc.*, pp. 120–124, 2020, doi: 10.1109/CSCITA47329.2020.9137772.

[20]   S. Soni, V. G. Shankar, and S. Chaurasia, "Route-the safe: A robust model for safest route prediction using crime and accidental data," *Int. J. Adv. Sci. Technol.*, vol. 28, no. 16, pp. 1415–1428, 2019.

[21]   P. Yerpude and V. Gudur, "Predictive Modelling of Crime Dataset Using Data Mining," *Int. J. Data Min. Knowl. Manag. Process*, vol. 7, no. 4, pp. 43–58, 2017, doi: 10.5121/ijdkp.2017.7404.

[22]   F. Ayele, "Appling Data Mining Technique for Crime Prevention: The Case of Hossaena Town Police Office," *Int. J. Adv. Eng. Res. Sci.*, vol. 7, no. 1, pp. 136–140, 2020, doi: 10.22161/ijaers.71.17.

[23]   B. Panja, "Crime Analysis Mapping, Intrusion Detection - Using.pdf," vol. 7, pp. 6–10.

[24]   T. T. Nguyen, A. Hatua, and A. H. Sung, "Building a Learning Machine Classifier with Inadequate Data for Crime Prediction," *J. Adv. Inf. Technol.*, no. January, pp. 141–147, 2017, doi: 10.12720/jait.8.2.141-147.

[25]   S. Kim, P. Joshi, P. S. Kalsi, and P. Taheri, "Crime Analysis Through Machine Learning," *2018 IEEE 9th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2018*, pp. 415–420, 2019, doi: 10.1109/IEMCON.2018.8614828.

[26]   S. Qayyum and H. Shareef, " A Survey of Data Mining Techniques for Crime detection," 2018 university of Sindh journal of information communication technology (USJICT), vol.2, pp.1-6 communication technology (USJICT), vol.2, pp.1-6

[27]   Y.Huang, C.Te Li and S. Jeng, "Mining Location-based Social Networks for Criminal Activity Prediction" , Proceedings of 24th IEEE , Proceedings of 24th IEEE International Conference on Wireless and Optical Communication, pp. 185-190, 2015.