

Various approaches to analyzing Crime and Prediction using data analytics

¹Dhinakaran K, ²R.Nedunchelian, ³Chokkalingam SP, ⁴Manimegalai M, ⁴Sona A, ⁴Yasmeen M

¹Assistant Professor, ⁴UG Scholar, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, India.

²Professor, Department of Electronics and Communications Engineering, Karpaga Vinayaga College of Engineering and Technology, Anna University, Chennai, India.

³Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Chennai.

Abstract: In the epoch, because the technology is increased in all aspects, the speed of crimes and the variety of criminals is additionally increasing proportionately. This causes major threat to the humans. These days, because the crimes keep increasing, therefore is to foresee its prevalence with time potency and accuracy to decrease the amount of crimes within the close to future. We tend to predict the long run crime prevalence victimization historical datasets. When examining the information, we tend to predict the coming crimes with relevancy location, time and different factors. This paper may be a survey of varied methodologies that was enforced in crime information and analysis and therefore the objective of this study may be analyzing and discussing various strategies which are applied on crime prediction and analysis.

Keywords: *Data mining, Machine learning, Big data analytics*

I. INTRODUCTION

Every day the rate of crime is increasing significantly. Since crime is neither systematic nor random, it cannot be expected. The trendy technologies and advanced strategies facilitate criminals in achieving their black act. In step with the Crime Records Bureau, crimes like felony, setting fireplace to a property(arson), etc. are relatively remittent however on the opposite hand, crimes like murder, regulatory offense, etc. are enlarged. Places wherever the rate lies high are thought-about as crime hotspots[2].

Thus, it's necessary to predict the prevalence of future crime hotspots and crime suspects by exploitation crime trends. Crime trend prediction is useful to form a call in crime interference activities [6].The predicted results can not be assured of one thousandth accuracy however the results show that ur application helps in reducing crime rate to an extent by providing security in crime sensitive areas. And so, to develop a strong crime prediction rule we've got to gather the historical crime records and appraise it.

The criminal information or records on the market is proscribed. Therefore, we have a tendency for aggregation of crime information from varied sources like news sites, websites, RSS feeds and lots of additional. This information is used as a criminal offense record info. The main challenge for us is developing an efficient crime pattern detection algorithm to identify crime patterns effectively.The other most challenges we have is the tendency to face are enlarged crime data, issue in analyzing incomplete and inconsistent information, restricted information supply and accuracy depends on however accuracy of coaching set.

To find the crime pattern and trend may be a difficult issue. To spot a pattern, and scanning through the information to seek out whether or not a selected crime fits into a legendary pattern is toughif the pattern does not fit into an existing pattern then the data must be classified as a new pattern. When sleuthing a pattern, it is accustomed predict, expect and stop crime. As a result, this survey targets at providing a short review on the technologies utilized in information analytics and large data to supply acute crime prediction results. This paper is split into many sections wherever, Section two presents an outline of scope and methodologies associated with crime information analytics and hotspot mapping, Section three includes of table that contains previous works-data supply and result and limitation and Section four concludes our survey.

II. SCOPE AND METHODOLOGY

Researchers have planned a range of ways to supply crime knowledge analysis, crime prediction, criminal identification and crime hotspot space identification. A number of the methodologies and papers are mentioned here.

2.1 Data processing

The process of sorting through massive knowledge sets to spot patterns and establish relationships to unravel issues through data analysis is data processing. The long run trends may be expected victimization data processing tools. Gerber Matthew S. [8] has conferred a paper concentration on analysis work the employment of spatio-temporally labeled tweets for crime prediction. He used Twitter-specific linguistic analysis and applied math topic modeling to mechanically determine discussion topics across major towns in the United States. He then incorporated these topics into against the law prediction model and show that, for nineteen of the twenty crime varieties he studied, the addition of Twitter knowledge improves performance of crime prediction versus a regular approach supported kernel density estimation. He haven't analyzed the matter content of tweets on the far side tokenization, part-of-speech tagging, and topic modeling. D.VijayaRohini et al.[23] planned a model that focuses on analyzing and mapping crime through on-line newspaper victimisation text mining techniques to seek out out the matter and solve them using the suitable ways. There are variety of major data processing techniques that are developing and victimization in data mining comes recently as well as association, classification, clustering, prediction, serial patterns and call tree. We are going to concisely examine few data processing techniques within the following sections.

2.2 Clustering

Data objects that are kind of like each other and may sense to 1 another may be found by agglomeration analysis. A lot of doubtless members are found in a very cluster than the members of alternative clusters. To seek out high-quality clusters such the inter-cluster similarity is low and therefore the intra-cluster similarity is high is that the goal of agglomeration analysis. Agglomeration models section the information into teams that weren't outlined antecedently. They are doing not use a target, however classification models section knowledge by distribution it to previously-defined categories that are laid out in a target. Agglomeration is beneficial in knowledge exploration. Agglomeration algorithms may be accustomed realize natural groupings, if there are several cases and no obvious groupings. It additionally is a helpful data-preprocessing step to spot homogenous teams on that before building supervised models. Agglomeration also can be used for anomaly detection. Some cases don't work well into any clusters once the information has been metameric into clusters. These cases are anomalies or outliers. There are numerous approaches to the figure clusters. a number of the agglomeration techniques for grouping objects as totally different clusters are mentioned below:

K-means agglomeration Algorithm:

K-means algorithm is mainly used to partition the clusters based on their means. First, the amount of objects are classified and such that as k-clusters. The distance between the objects is calculated because the average. The relocation repetitive technique is employed to enhance the partitions by movement of objects from one cluster to alternative. Then, the amount of iterations is finished till the convergence happens.

Ak- mode agglomeration rule

Ak-mode agglomeration rule may be a 2 step process- attribute advisement part and clustering phase. Within the attribute advisement part, the weights of the attributes are computed victimization info Gain magnitude relation (IGR) price for every attribute. The utmost price of weight is taken as decisive attribute. The space between 2 categorical attributes is calculated because the distinction between two knowledge records that provides the similarity measures. The brink price α is about by the analyst with the assistance of the computation results of similarity measures. Ak-mode agglomeration rule is especially used for categorical attributes.

Expectation-Maximization algorithm:

Expectation- Maximization (EM) rule, is associate extension of K-means rule that is employed to seek out the parameter estimates for every cluster. The information as an entire may be a mixture of constant quantity probabilistic distribution. The burden of attributes is measured within the chance distribution and every object is to be clustered supported the weights rather than distribution the objects to the dedicated clusters in K-means.

B. Chandra et al. [16] have planned a completely unique approach for variable statistic agglomeration supported dynamic time wrapping and constant quantity Minkowski model has been proposed for locating similar crime trends expeditiously and afterwards predict crime trends. The effectiveness of the planned approach over the prevailing agglomeration algorithms has been illustrated victimization Indian crime knowledge. The approach will play a crucial role for wider sort of variable statistic agglomeration issues particularly wherever the scale don't have equal weightages.

Giles C. Oatley et al. [10] have developed a OVER computer code that embodies a model of the method of incidence of burglaries. The power of the model to explain ascertained knowledge and to predict future data is assessed in a very range of the way. These can embody comparison of appropriate functions of the ascertained knowledge with the posterior prognosticative distributions of those quantities. The computer code is additionally capable of refined analysis and predictions and these are once more embedded within the software in such the way that applied math or technology experience isn't needed for interpretation by the top user.

RasoulKiani et al. [18] have conferred a replacement framework for agglomeration and predicting crimes supported real knowledge. During this framework, the Genetic rule was accustomed improve outlier detection within the preprocessing part, and therefore the fitness perform was outlined supported accuracy and classification error parameters. To enhance the agglomeration method, from the weighted options, low-value options were deleted through choosing an appropriate threshold. The strategy planned was enforced, and therefore the results of the optimized and non-optimized parameters were compared to see their effectiveness and quality.

Lianhang Ma et al. [20] have planned associate AK-Modes rule to mechanically realize the similar case subsets while not a given "seed case". Within the attribute-weighting part, we tend to calculate the burden of every attribute associated with associate offender's behavior attribute victimization the construct info Gain magnitude relation (IGR) in classification domain. Later, the results of attribute-weighting part is employed within the agglomeration method to seek out the similar case subsets. The appliance of this model in observe can effectively improve the potency compared with the normal manually reviewing approaches and may assist within the decision-making method.

Jyoti Agarwal et al. [17] have conferred a paper on crime prediction victimization k suggests that agglomeration. it's enforced victimization open supply data processing tool that are analytical tools used for analyzing knowledge. Among the accessible open supply data processing suite like R, Tanagra, WEKA, KNIME, speedy manual laborer. K suggests that agglomeration is finished with the assistance of speedy manual laborer tool that is associate open supply applied math and data processing package written in Java with versatile data mining support choices.

2.3 Classification

Classification is one among the information mining functions that assigns things in an assortment to focus on classes or specific categories. Classification is employed to accurately predict the target category for every case within the knowledge. They are separate and don't imply any order. The classification task begins with an information set wherever the category assignments are identified. Continuous and floating-point values would indicate a numerical, than a categorical, target price. A prognosticative model with a numerical target uses a regression rule and not a classification algorithm. The foremost simplest variety of classification drawback is binary classification. Within the model build, (training) the method, a classification rule finds relationships between the values of the predictors and therefore the values of the target. Totally different classification algorithms use numerous techniques for locating relationships. The relationships are summarized in a very

model, which may then be applied to a distinct knowledge set during which the category assignments are unknown. Classification models are tested by comparing the values predicted to known target values in a set of test data. The historical knowledge for a project victimization classification is usually divided into 2 data sets: one for building the model; the opposite for testing the model. There are several applications victimization classification in client segmentation, business modeling, marketing, credit analysis, and medicine and drug response modeling.

Abba Babakura et al. [19] have conferred a comparison between 2 classification algorithms particularly, Naïve Bayesian and Back Propagation for predicting the 'Crime Category' attribute, having labels, particularly 'Low', 'Medium', and 'High'. For Naïve Bayesian, the Accuracy, preciseness and Recall of cluster one are 90.2207%, 95.8% and 93 and for cluster a pair of are 94.0822%, 98.3% and 94.8%. On the opposite hand, Accuracy, preciseness and Recall values for BP for group1 are 65.9478%, 65.9% and a thousandth and for cluster a pair of are 65.9469%, 65.9% and a thousandth severally. Experimental results for each the algorithms manifest that, Naïve Bayesian performed higher than the BP for the crime dataset victimization wood hen. This experiment was performed victimization 10-fold cross-validation.

Mehmet Sait Vural et al. [15] have planned a sensible model supported the Naive mathematician classifier with novel methodologies applied for the criminal prediction drawback. This model is sensible because of the simplicity caused by the independence assumption of the Naive mathematician. However, the model will moderately work with the independence assumption therein the model achieves reducing the suspect list with eighty the speed. This model may be utilized in sociology with its averagely seventy eight.05 so successful rate so as to assist security forces to seek out the criminal of the incidents. Yet one more important of this model is its ability to require the acquaintances into decision-making method.

2.4 Machine Learning

Arthur prophet, one among the pioneer in machine learning and computing outlined machine learning as a field of study that offers computers the power to find out while not being expressly programmed. Machine learning will be wont to analyze the information and to supply the various subsets of crime analysis [4]. As a result, the machine learning may be a pc system's methodology of learning by manner of examples. There are varied machine learning algorithms offered to users which will be enforced on datasets. But, the 2 major sorts' are supervised learning and unattended learning algorithms. Supervised learning algorithms work by final data or "the right answer" from labelled coaching knowledge. The algorithms are given with a specific attribute or set of attributes to predict. Unsupervised learning algorithms, aim to seek out hidden structures in unlabeled category knowledge. As a result, the algorithms learn additional concerning the information set because it is given more examples to be enforced on.

Machine learning is one among the applications of computing (AI) that has systems the power to find out mechanically and to boost from expertise while not being expressly programmed. Machine learning in the main focuses on the event of pc programs which will access knowledge and use it learn for themselves. The educational method begins with observations or knowledge, like examples, direct expertise, or instruction, so as to appear for patterns in knowledge and build higher choices within the future supported the examples, while not human intervention or help and regulate actions consequently. This additionally permits analysis of huge quantities of information. Whereas it always delivers quicker, additional correct leads to order to spot profitable opportunities or dangerous risks, it's going to additionally need extra time and resources to coach it properly. Combining machine learning with AI and alternative psychological feature technologies makes it even more practical within the process of giant volumes of knowledge.

Lawrence McClendon et al. [14] have proposed WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Un-normalized Dataset. The regression toward the mean, Additive Regression, and call Stump algorithms victimization the identical finite set of options are enforced, on the Communities and Crime Dataset. Overall, the regression toward the mean algorithmic program, a machine learning algorithmic program performed the most effective among the 3 chosen algorithms. This algorithmic program might work with weighted instances. This methodology of regression is simple and provides an adequate and interpretable description of how the input affects the output and could handle randomness in the test samples to a certain extent, also without incurring too much of

prediction error. The scope is to prove however effective and correct the machine learning algorithms utilized in data processing analysis will be at predicting violent crime patterns.

Fateha Khanam Bappee et al. [13] have projected a machine learning model for crime prediction victimization geospatial options for various classes of crime. The reverse geocoding technique is applied to retrieve open street map (OSM) spatial knowledge. This additionally proposes finding hot points extracted from crime hotspots space found by stratified Density-Based spatial clump of Applications with Noise (HDBSCAN). A spatial distance feature is then computed supported the position of varied hot points for various kinds of crime and this worth is employed as a feature for classifiers. Piyush Kakkar et al. [22] have projected a piece to sight and acknowledge faces of criminals in a picture and video stream obtained victimization CCTV. They used Haar feature-based cascade classifiers in OpenCV approach for police work faces. It's one among the machine learning primarily based approach wherever a cascade perform is trained from positive and negative pictures. It additionally won't sight objects in alternative pictures. They need used native Binary Patterns Histograms (LBPH) for recognizing faces.

Mohammad A. Tayebi et al. [25] projected a supervised learning framework for co-offence prediction that may be a link prediction drawback in co-offending networks. The sociology theories concerning social, environmental and homophily roots of offending in a very comprehensive manner, they appointed every combine of offenders to socially-related, geographically-related or expertise connected criminal cooperation opportunities. Victimization these sets, they produce 3 prediction areas to beat the challenge in link prediction (class imbalance).

2.5 Big Data Analytics

Big knowledge is that the term for any cluster of datasets therefore large and composite that it becomes troublesome to follow victimization ancient processing applications. Huge knowledge may be a set of procedures and technologies that entail new varieties of integration to uncover giant unknown values from large datasets that are varied, complex, and of a large scale. Analyzing huge knowledge may be a difficult task because it contains large spread file systems that ought to be fault tolerant, versatile and climbable. There is an immense need of constructions, platforms, tools, techniques and algorithms to handle Big Data. The frameworks utilized by huge knowledge application to handle the large data are Hadoop, Map Reduce, Apache Spark.

Apache Hadoop

Hadoop is an ASCII text file code framework which will run applications on systems with thousands of nodes and terabytes. It is an ecosystem of open source components which primarily changes the way enterprises store, process, and analyze data. In contrast to ancient systems, Hadoop permits multiple kinds of analytic workloads to run on the identical knowledge, at the identical time, at huge scale on industry-standard hardware. Figure 1 shows Hadoop design. The modules in Hadoop are designed with a primary assumption that hardware failures are common and so ought to be mechanically handled in code by the framework. Apache Hadoop's MapReduce and HDFS parts originally derived severally from Google's MapReduce and Google filing system (GFS) papers. parts of Hadoop includes Hbase, Oozie, Sqoop, Avro, Chukwa, Pig, Zookeeper, Hive. Anushka Kumar et al. [24] have projected a model that evaluates and identifies the alternatives to crime incidents, their individual parts and proposes a combinatorial incident-description schema. By victimization Hadoop, they situated the precise town and analyzed the crime patterns, supported that provide preventive measures to folks.

Hadoop Distributed File System

The Hadoop distributed filing system (HDFS), as in Figure 2 is meant for storing terribly massive files with streaming knowledge access patterns, running on clusters of trade goods hardware. It is designed round the concept that the most-efficient processing pattern may be a write-once, scan as repeatedly pattern. It was developed using distributed file system design. The files kept are in a redundant fashion to rescue the system from potential knowledge loses within the case of failure. HDFS additionally makes applications out there to multiprocessing. Hadoop provides a command interface to move with the HDFS. HDFS additionally provides file permissions and authentication.

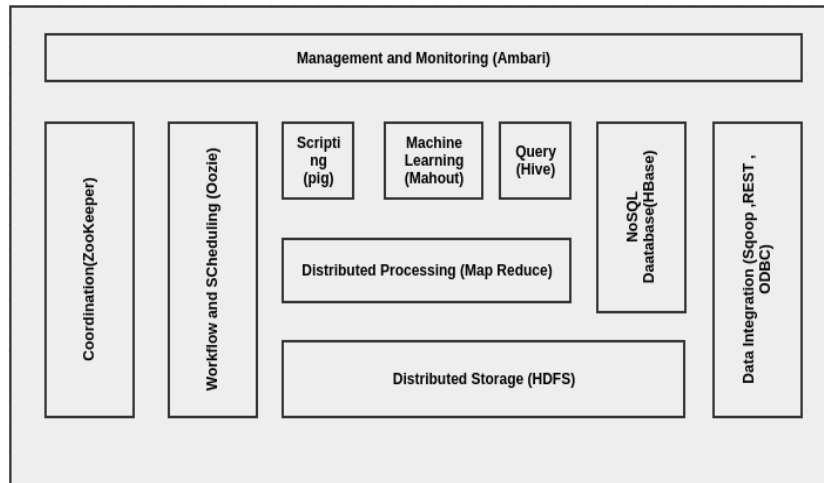


Figure 1 Hadoop Ecosystem

It follows master slave design paradigm having name nodes and knowledge nodes. The name node manages the filing system namespace. It maintains filing system tree and also the data for all the files and directories within the tree. A shopper accesses the filing system on behalf of the user by human action with the name node and also the knowledge nodes. Knowledge nodes are the workhorses of the system. They store and retrieve blocks once they told to and that they report back to the name node sporadically with the lists of blocks that they're storing. The client presents a file system interface like a conveyable software system Interface (POSIX), therefore the user code doesn't have to realize the name node and knowledge nodes to operate.

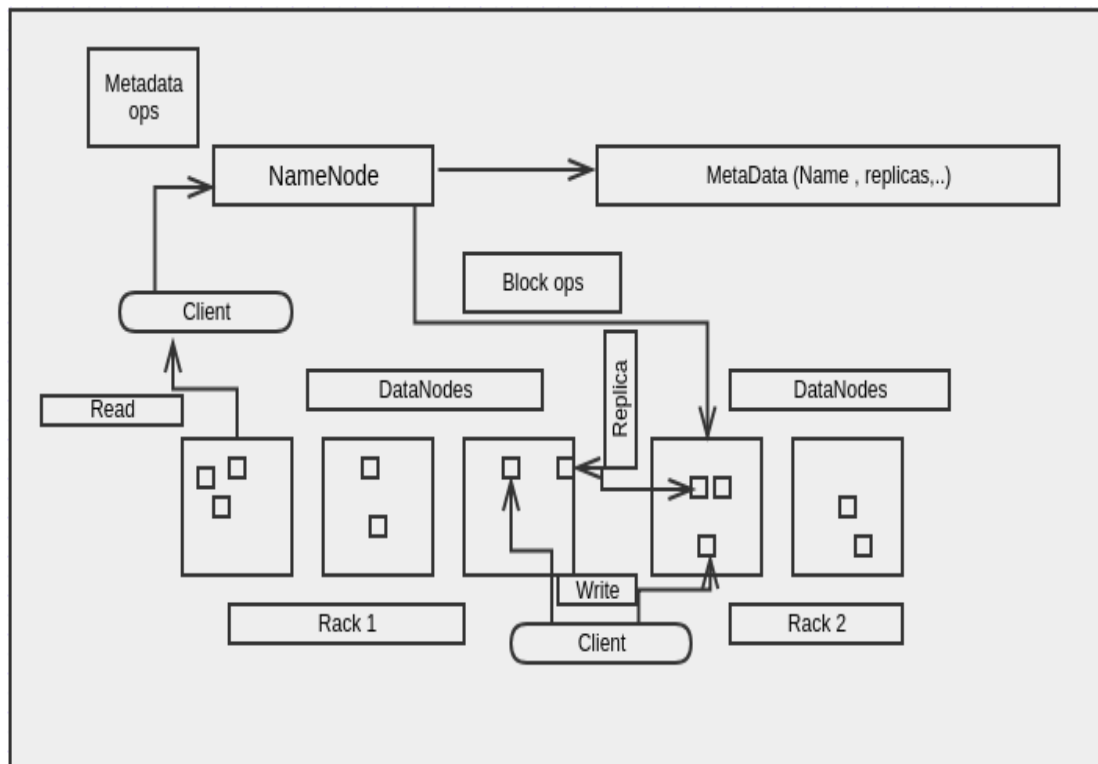


Figure 2 HDFS Architecture

Map Reduce

Map cut back may be a software package framework for distributed process of enormous knowledge sets on laptop clusters. It is initially developed by Google. Map cut back is meant to facilitate and alter the process of huge amounts of information in parallel on massive clusters of artifact hardware during a reliable, fault tolerant manner. Figure 3 shows MapReduce design that may be a programming model and therefore the expectation is multiprocessing in Hadoop. It makes it simple to distribute tasks across nodes and performs type or Merge supported distributed computing. The underlying system, takes care of partitioning the computer file, programing the programs' execution acrossmany machines, handling machine failures and managing inter-machine communication. Input are dividedintomultiple chunks/blocks. Each block of information are processed in several nodes.

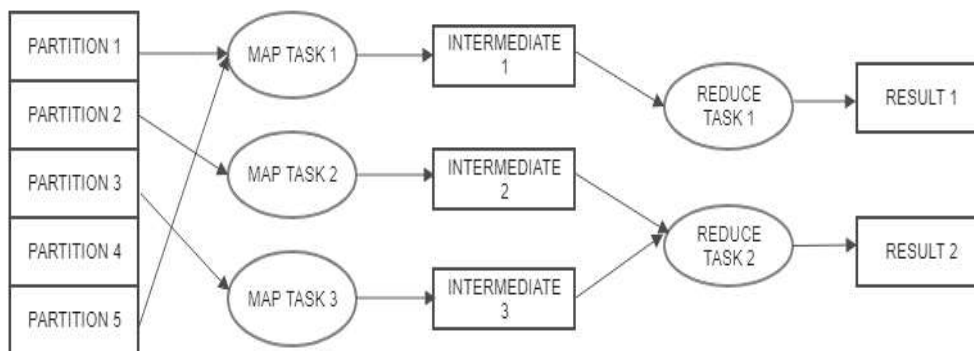


Figure 3 Map reduce back design

Apache Spark

Apache Spark is associate ASCII text file, distributed general purpose, lightning-fast cluster computing technology, designed for quick computation. It supports Hadoop MapReduce and extends the MapReduce model for economical use it for additional forms of computations, which incorporates interactive queries and stream process. The main feature of Spark is its in-memory cluster computing that will increase the process speed of associate application. Spark is meant to hide a good vary of workloads like batch applications, repetitive algorithms, interactive queries and streaming. Apart from supporting all these workload in a respective system, it reduces the management burden of maintaining separate tools. Figure 4 shows three ways of howSpark are often designed with Hadoop elements. Ravi Kumar et al.[21] have projected a model to predict crime sensitive areas, wherever the incidents are categorized victimization SparkR and therefore the association between the hotspots and the crime varieties. Knowledge responsibility and privacy is restricted during this model.

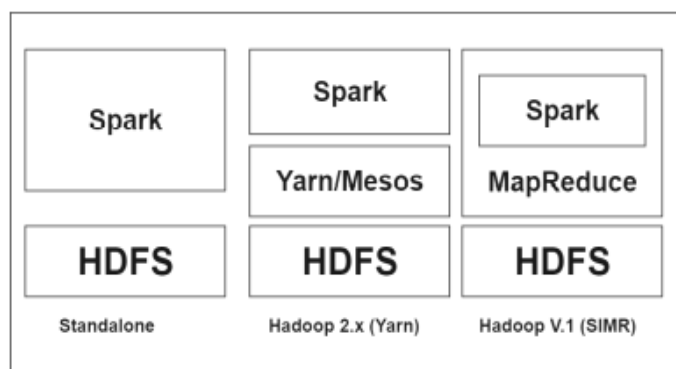


Figure 4 Building Spark on Hadoop

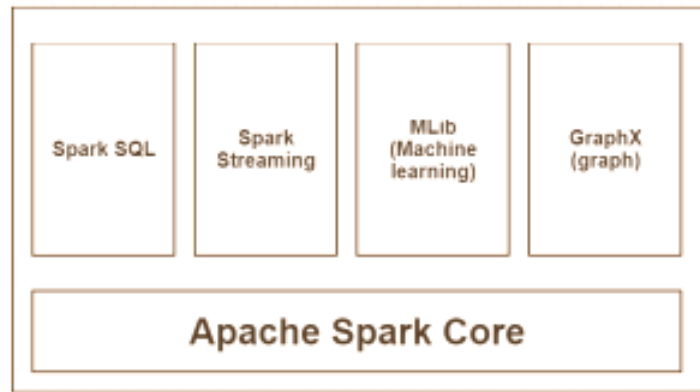


Figure 5. Components of Spark

Apache Spark Core - The Spark Core is that the underlying general execution engine for spark platform wherever all different practicality is made upon. It provides the In-Memory computing and referencing of datasets in memory device systems.

Spark SQL - The Spark SQL may be an element on high of Spark Core that introduces a replacement knowledge abstraction known as Schema RDD, that gives support for structured and semi-structured knowledge.

Spark Streaming - The Spark Streaming leverages Spark Core's quick planning capability to perform streaming analytics. It ingests data in mini-batches and performs RDD (Resilient Distributed Datasets) transformations on those mini-batches of information.

MLlib (Machine Learning Library) - The MLlib is a distributed machine learning framework higher than Spark. It is done by the MLlib developers against the Alternating method of least squares (ALS) implementations. Spark MLlib is ninefold quicker than the Hadoop disk-based version of Apache driver (before driver gained a Spark interface).

GraphX - The GraphX is a distributed graph-processing framework on prime of Spark. It provides an interface for expressing graph computation that may model the user-defined graphs by victimization Pregel abstraction API. There is an optimized runtime for this abstraction.

III. RESULT AND DISCUSSION

Consider the crime knowledge sets of any 2 consecutive years employed in the higher than models. For the any investigations, we have a tendency to compare the placement of the particular hotspots with those found by the results of our prediction models for predictions derived by victimization every specific variety of knowledge one by one and predictions derived by a model trained on options combining all data varieties and sources. Scrutiny of the crime rates of every space between these 2 years, reveals that the deviations are below fifteenth, on average. And so, it's simple to get a extremely correct model that may predict the crime rates of these years, even by trivially predicting the rates to remain the identical. it's troublesome to review monthly combination limits the investigation to a rough level of temporal coarseness, so preventing the possibility for investigation a lot of elaboratetemporal patterns and creating more difficult predictions on this dimension. Instead, the historical crime approach makes a lot of sense once handling knowledge of abundant finer temporal coarseness than ours. Figure 6 shows about the comparisons of the accuracies between the data types.

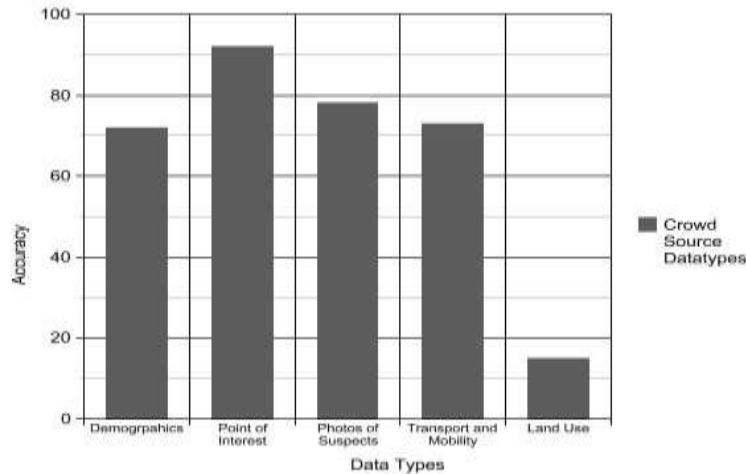


Figure 6 Comparison of accuracies between datatype

TABLE-1 : Performance Analysis Of Various Approaches

| No | Authors | Journal, Year Of Publication | Methodology And Tools | Data Source and Result |
|----|--|--|--|---|
| 1 | Andrey B,BrunoLepri ,Jacopo Staiano,Nuria Oliver ,Fabio Pianesi ,Alex (Sandy) Pentland [1] | ACM 2014 | Random Forest, support vector machines, neural networks, decision trees | Dataset: Smartstep datasets, information from digital telecommunication of individual mobile activity and demographic data or Geo-localized Open information Result: Predicts crime classes by group action to totally different knowledge sets and random forest performs far better Limitations: Limited access to crime events and takes longer time to validate their approach |
| 2 | Lawrence McClendon and Natarajan Meghanathan [14] | Machine Learning and Applications: An International Journal (MLAIJ) 2015 | Machine Learning, Linear Regression, Additive Regression, Decision Stump | Dataset: Communities and Crime Un normalized dataset because they may have plausible connections to one of the possible crime goals. Result: Proves how effective and accurate the machine learning algorithms are used in data mining analysis for the prediction of crime patterns. |
| 3 | Tayebi MA, Ester M, Glässer U, BrantinghamPL [3] | International Technical Conference on Advances in Social Network Analysis and Mining, 2014 | Predictive policing, CrimeTracer model (Personalized Random Walk Model) | Dataset: Aroad network crime datasetCwith an offender who is associated with C Result: A CrimeTracer consisting of a criminal’s daily life, road network and theplaces where criminal carry out a crime, random walk approach is used for crime location forecasting |
| 4 | Cesario, Cesario E, Catlett C, Talia D [5] | 14th International Conference on Dependable, Autonomic and Secure Computing 2014 | Naïve Bayes classifier Neural network | Dataset: Two various data sets from different sources Result: Analyzed and compared Naïve Bayes & Neural network to predict crime level that will occur in particular location and Naïve Bayes classifier performs better with |

| | | | | |
|---|--|---|---|---|
| | | | | 90.2207% of accuracy |
| 5 | Yu CH, Ward MW, Morabito M, Ding W [7] | IEEE Chung- 2011 | Support Vector Machine, Neural Network, Naïve Bayes, 1-NN (One- Nearest Neighbor), J48 and Ensemble learning (Voting) | <p>Dataset: Aggregated counts of crime and crime-related events categorized by the department of police</p> <p>Result: Predictive technique for areas that have a high density of residential burglaries and the probability that will happen in future. Primarily, the temporal information is organized and using the supervised data mining algorithm, residential burglary is predicted efficiently.</p> <p>Limitations: To locate the optimum point at which spatial knowledge is most probable</p> |
| 6 | Giles C. Oatleya, Brian W. Ewart [10] | Elsevier 2013 | Data mining- Bayesian belief network | <p>Dataset: Datasets of repeated victims, Time-To-Crime Expectation Maximization (EM)</p> <p>Result: Describes observed data and predicts future data which can be assessed in a number of ways. It also includes the comparison of suitable functions of the observed data with the posterior predictive distributions of these quantities and predicts the possible outputs.</p> <p>Limitations: Prior specification of the structure is the limitation here</p> |
| 7 | RasoulKiani, SiamakMahdavi, Amin Keshavarzi [18] | (IJARAI) International Journal of Advanced Research in Artificial Intelligence 2015 | Clustering, classification, genetic algorithm, Weighting, Rapid Miner | <p>Dataset: Crime dataset recorded by the police in England and Wales within 1990 to 2011</p> <p>Result: The Genetic Algorithm (GA) was used to improve outlier detection in the preprocessing phase, and the fitness function was defined based on accuracy and classification error parameters. To improve the clustering process, they use weighed features, and the low-value features were deleted through selecting a suitable threshold. The results of the optimized and non-optimized parameters were compared to determine their quality and effectiveness</p> <p>Limitations: Clustering and optimization of the number of clusters is quite difficult.</p> |
| 8 | S.SivaranjaniDr.S.Sivakumari, Aasha.M [11] | [ICETT] International Conference on Emerging Technological Trends | Clustering, Classification, K-Means, KNN, DBSCAN, Agglomerative | <p>Dataset: Crime dataset of crime information of six cities in Tamil Nadu from the National Crime Records Bureau (NCRB) of India.</p> <p>Result: This model is used to predict and forecast crimes and assists law enforcement agencies for an improved and accurate crime analysis using DBSCAN, various clustering and classification techniques.</p> |

| | | | | |
|----|---|---|--|--|
| | | | | Limitations: Privacy issue is the major limitation |
| 9 | Renjie Liao, Xueyao Wang, Lun Li, Zengchag Qinh [9] | Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao 2010 | Gaussian Distribution, Bayesian Learning Theory, Kernel Function, Hausdorff Distance | Dataset: Crime dataset of a serial crime in Gansu, China Result: The area is predicted which will help local police much to arrest the criminals Limitation: The time of next crime is not predicted |
| 10 | Somayyeh Aghababaei, Masoud Makrehchi [12] | IEEE/WIC/ACM International Conference on Web Intelligence 2016 | Sampling, Aggregation, Prediction | Dataset: Crime rates from four cities of the United States, Chicago, Philadelphia, San Francisco, and Houston Result: A crime prediction model based on mining tweets posted from a specific geographic area Limitations: Requires further analysis to examine to incorporate other socio-economic indexes and geographical information |

V. CONCLUSION

Crime prediction and finding relevant data from great amount of crime knowledge is that the dynamic and rising analysis field within the real-world that aims to stop the crime rates. It is important however difficult. A survey is conducted so Crime prognostication will be improved by the employment of economical knowledge assortment and different ways. Combined techniques are needed to make a higher crime prediction by desegregation multiple models to resolve single downside for better performance. From this survey we tend to conjointly learned regarding most supporting factors that may facilitate US to investigate the crime pattern in neighborhood spaces mistreatment demographic factors of that area.

REFERENCES

- [1]. Andrey Bogomolov A, Lepri B, Staiano J, Oliver N, Pianesi F, Pentland A, "Once upon a crime: towards crime prediction from demographics and mobile data", 2014 Nov 12 (pp. 427-434). ACM.
- [2]. Zhang Q, Yuan P, Zhou Q, Yang Z, "Mixed spatial-temporal characteristics based Crime Hot Spots Prediction", In Computer Supported Cooperative Work in Design (CSCWD), 2016 May 4 (pp. 97-101). IEEE.
- [3]. Tayebi MA, Ester M, Glässer U, Brantingham PL, "Crimetracer: Activity space based crime location prediction", In Advances in Social Network Analysis and Mining, (ASONAM), IEEE/ACM 2014 Aug 17 (pp. 472-480). IEEE.
- [4]. Mr. Sufiyan Tamboli, Mr. Ashish Wankhede, Mr. Prathamesh Puranik, Mr. Vinayak Natarajan, Prof G.B. Deshmukh: "A Survey on Crime Analysis Using Machine Learning Algorithms", IJAERD-2018.
- [5]. Cesario, Cesario E, Catlett C, Talia D, "Forecasting Crimes Using Autoregressive Models", In Dependable, Autonomic and Secure Computing, 2016 IEEE 14th Intl C 2016 Aug 8 (pp. 795-802). IEEE.
- [6]. Retnowardhani A, Triana YS, "Classify interval range of crime forecasting for crime prevention decision making", In Knowledge, Information and Creativity Support Systems (KICSS), 2016 Nov 10 (pp. 1-6). IEEE.
- [7]. Yu CH, Ward MW, Morabito M, Ding W, "Crime forecasting using data mining techniques", In Data Mining Workshops (ICDMW), 2011 Dec 11 (pp. 779-786). IEEE.
- [8]. Gerber Matthew S., "Predicting Crime Using Twitter and Kernel Density Estimation", Decision Support Systems, Volume 61, pp. 115 – 125, ELSEVIER, 2014.
- [9]. Renjie Liao, Xueyao Wang, Lun Li, Zengchag Qinh, "A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, 2010 IEEE.
- [10]. Giles C. Oatley, Brian W. Ewart, "Crimes analysis software: 'pins in maps', clustering and Bayes net prediction", Elsevier, 2003.
- [11]. S. Sivarajani, Dr. S. Sivakumari, Aasha M., "Crime Prediction and Forecasting in Tamilnadu using Clustering Approaches", 2016 International Conference on Emerging Technological Trends [ICETT].
- [12]. Somayyeh Aghababaei, Masoud Makrehchi, "Mining Social Media Content for Crime Prediction", 2016 IEEE/WIC/ACM International Conference on Web Intelligence.
- [13]. Fateha Khanam Bappee, Amilcar Soares Junior and Stan Matwin, "Predicting Crime Using Spatial Features", Bagheri E., Cheung J. (eds) Advances in Artificial Intelligence. Canadian AI 2018. Lecture Notes in Computer Science, vol 10832. Springer, Cham.

- [14]. Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning Algorithm To Analyze Crime Data", An International Journal (MLAIJ) 2015.
- [15]. Mehmet Sait Vural , Mustafa Gok , "Criminal prediction using Naive Bayes theory", 2016 Springer on Neutral Computer and Applications, 2017.
- [16]. B. Chandra, Manish Gupta, M.P. Gupta, "A Multivariate Time Series Clustering Approach for Crime Trends Prediction", IEEE International Conference on Systems, Man and Cybernetics, 2008.
- [17]. Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal, "Crime Analysis using K-Means Clustering", International Journal of Computer Applications, 2013.
- [18]. Rasoul Kiani, Siamak Mahdavi, Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.8, 2015.
- [19]. Abba Babakura, Md Nasir Sulaiman and Mahmud A. Yusuf, "Improved Method of Classification Algorithm for Crime Prediction", International Symposium on Biometrics and Security Technologies (ISBAST), 2014.
- [20]. Lianhang Ma, Yefang Chen, Hao Huang, "AK-Modes: A Weighted Clustering Algorithm for Finding Similar Case Subsets", IEEE, 2010.
- [21]. Ravi Kumar, Bharti Nagpal, "Analysis and prediction of crime patterns using big data", Springer, 2018.
- [22]. Piyush Kakkar, Mr. Vibhor Sharma, "Criminal Identification System Using Face Detection and Recognition", International Journal of Advanced Research in Computer and Communication Engineering, 2018.
- [23]. D. Vijaya Rohini, Dr. P. Isakki, "Crime Analysis and Mapping through Online Newspapers: A Survey", International Conference on Computing Technologies and Intelligent Data Engineering, 2016.
- [24]. Anushka Kumar, Vishnudas S, R. Kayalvizhi, "Using Mapreduce Techniques to Predict and Examine Crime Pattern", International Journal of Engineering & Technology, 2018.
- [25]. Tayebi MA, Ester M, Glässer U, Brantingham PL, "Spatially Embedded Co-offence Prediction Using Supervised Learning", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014.