"Robotized Wrongdoing Examples Examination System for Prescient Policing utilizing Information Mining Procedures

[1] C.Maddilety, [2]C.Ishaq Shareef, [3]M.Shaheda Begum
[1,2,3]Assistant Professor
[1,2,3] Department of Computer Science & Engineering,
[1,2,3]Ashoka Women's Engineering College

Abstract: These studies are aimed at studying and developing an automated technology framework that may be utilized to detect contributing characteristics, patterns and trends utilizing data mining methods from reported instances. Because of its efficacy in revealing patterns and trends that are connected, related, and close to one another, a data mining strategy based on categorization and association criteria has been recommended for this investigation.

1. INTRODUCTION

Through the digitization of reported cases by several law enforcement and public oversight agencies, the need for faster and reliable methods of sifting through massive data and cases to identify attributes and patterns that could lead to a future occurrence arises. Currently when policing and conducting security operations in response to reported crimes most agencies either use the previous reports and will have to sift through a lot of records to find patterns or conduct blanket policing which both are inefficient and laborious[3]. This research details a general framework developed from the use of a combination of several data mining techniques to map occurrences to their dominant attributes and combinations.

In this age of vast data generation, it is imperative that we should find new and novel ways of effectively and quickly analyzing data and give appropriate feedback for decision making in any sphere. As processing power rises, so does the potential for new information to be discovered via data mining and machine learning, both of which may be used to mine databases for previously undiscovered and possibly beneficial information. The goal of intelligent data analysis is to extract valuable information, a process that requires a combination of numerous things, such as extraction, analysis, conversion, categorization, organization, and reasoning.This is precisely what this

research has managed to do by creating an intelligent analysis framework that can be universally applied to any data by combining two data mining techniques to leverage on both to create a reliable model applicable to law enforcement through intelligence based policing.

BACKGROUND CONCEPT

Data mining is a relatively recent data analysis approach that is capable of discovering patterns recorded in historical data and is currently regarded a catalyst for improving corporate operations by avoiding failure patterns and utilizing success patterns. A number of data mining methods have been created in the recent decade. There are four main types of data mining techniques: classification, clustering, quantitative prediction, and association rule extraction. Algorithms and methods employed, as well as the way results are conveyed, are the key differences between the various techniques.[11]

***Instance Based Learning*:** With every query, the approximation procedure is repeated until it approximates the target function exactly as it did in the training instances. An instance's connection to previously stored examples is checked to assign a target function value[5] each time it is discovered. The K-Nearest Neighbor and the Radial Basis Functions[12] are two algorithms to keep an eye out for in the field of data mining.

***Classification*:** An existing database may be analyzed to discover new patterns or category divisions using puzzles, which are basically prediction models. Problems with classification include figuring out whether properties of each record in a database point to a certain group or class[9]. In the absence of a predetermined class or group, however, the clustering approach is used to collect things that seem to belong together.Several algorithms are inherently designed and suited for this purpose which include the KNN, ANN, Radial Basis Functions[6].

The data mining technique used in this research will be a combination of instance based association and classification machine learning algorithms. Through the discovery of associations, it is the purpose of association learning to find any noteworthy patterns in the data. They can forecast the value of more than one attribute at a time, unlike classification rules, which can only predict a single attribute's value. The

following is an example of a common association rule:

Cause_1, Cause_2 => Result (or consequence)

A rule of thumb is that for n percent of the instances with x percent confidence, there is a correlation between Cause 1 and Cause 2.

It is common for each rule to be accompanied with a degree of confidence and a reference. For example, one may calculate a rule's statistical probability by multiplying it by the number of cases/projects in which it has been observed. The term "pattern" refers to a set of rules that are similar enough to indicate a trend. When generating rules or patterns, and also when correcting models that rely too heavily on specific records in order to produce rules and patterns, most data mining approaches involve statistical tests[8]. Because the purpose of this research was to identify any potentially beneficial trends in the target sector based on reported incidents and registered complaints, association learning was the data mining approach used in this study.

[10]Conducted a research on data mining methords for thetransport industry user behavior using Smart Card Data withdesirable results. The limitations to their research based on theproposedresearchwerethattheirsfocused oncommutersbehavioral patterns for economic and financed based planningoftheirtransportsystem.This study's participants may be quickly divided into four distinct behavioral categories, regardless of the sort of ticket they use.

[9]Examined the most recent methods for predicting criminal activity and criminals. They were able to focus their investigation on three key types of prospective approaches.:

- Statistical Methods, they generally focus on the path to crime, the age of offending, and the conduct of offenders.
- Crime hotspots, repeat victims, and crime magnets may be identified utilizing Geographic Information Systems.
- Automated systems that detect criminal behavior patterns and research on reoffending are part of the

Crime Generators category.Althoughtheirresearchprovid edagreatinsightonmethords of crime patterns analysis. Their approach wasnot focused on any particular industry and thus followedno formal government policy.

METHODOLOGY

To mimic the framework system, this model will be built in layers and pieces, each of which will be designed and developed independently to reflect the autonomous structure sub processes Depending on the instance of computation, each subcomponent will use an assortment of diverse tactfully regarded and really esteemed data sources and provide actually esteemed results.

- Bayesian Belief Networks (BBNs) from the Bayesian Based Learning (BBL) class are used as the basis for the core framework, with the basic notion of mixing numerous hypotheses, weighted by their probability, to classify new examples being taken from this approach. The MAP Hypothesis and the Maximum Likelihood (ML) Hypothesis are two of the primary principles adopted from this methord, which aims to discover the most likely hypothesis h from a collection of candidate hypotheses H given the observed data D.
- To select the most likely hypothesis, given the data D and any previous information about the prior probability of the different hypotheses in H.
- ***Priorprobabilityofh,P(h):***If we know the probability that h is right (before we have seen the data, irrespective of D), then this value represents our prior knowledge of that probability. It is possible to give the same probability to each hypothesis if we do not know P(h)
- ***Probability of D, P(D):***In the absence of any prior information about the validity of hypothesis h, it represents the likelihood that training data D will be seen.
- ***Conditional Probability of observing D, P(D|h):*** For

example, the likelihood of seeing data D is denoted as D/H.

As well as Concept Learning, which involves inferring a function's input and output based on training instances.

It is possible to see concept learning as a problem of searching across a pre-defined search space for the hypothesis that best matches the training instances. The following is an example of how the hypothesis may be expressed:

- Hypothesishisaconjunction ofconstraintson attributes
- Eachconstraintcan be:
  - Aspecificvalue:e.g.*Gender=Male*
  - RangeValuee.g.*Age=19-25*
  - Adon't carevalue:MaritalStatus=?
  - Novalueallowed(nullhypothesis):e.g.*Height*

Example:hypothesis *h*

theBayesianBasedLearningwhichinprincipleis as below.

FeaturesofBayesian learningmethods:

A hypothesis' estimated likelihood of correctness may be slowly decreased or increased based on each observed training example, unlike algorithms, which dismiss a hypothesis if it contradicts any one example. There will be a two-stage approach to this investigation.

i. Theoretical modelframeworkdesignand
ii. Systemprototypedevelopmentandmodeltesting

TheoreticalModelFramework Design

Thiswillbedonein threesub-stagesasstructured below.

Frameworkresearchand design

This is the first stage of the overall research as it will involvethe research of existing systems. Identification of successfuldesigns in related areas of application. Identification of therelevant frameworks approach development of discreet overallstepsandstagesoftheproposedframework[7].

The outcome of this stage will be a high level description ofthesystemframeworkthatwillbeusedtosolvethesaidproblem. It will be a diagrammatic representation of the

wholesystemframeworkasaflowchartwithg eneralizeddescriptions of the individual components that make up thefinalsystem.

Inputdatastructure andformatdesign

Thisisthesecondsub-stageasitwillinvolvetheidentification of all the possible outcomes (which in this casewill be cases or reported incidences). The identification of allattributes that affect the outcomes. And also the structure oftheattributesandtheirsignificanceinhowa ndbywhatmagnitude they affect the effective outcome[5]. From there theidentified attributes will have to be converted into discreet orsemi-discreet value or variable based data inputs so that theproposed model will be able to mathematically compute theoutcomesbasedontheassociatedattribute s.

The outcome of this stagewillbe the creationof a discreetand continuous value based attributes input data structure thatwilleventuallydictatethefinalsystems databasedesign.

This stage will overlap with the third sub-stage which is theMathematical model design as the model will dictate the dataand typeofdata neededas modelinputs.

Mathematicalmodeldesign

Thiswillbethefinalstageofthetheoreticalsys temframework research process. This stage will involve the actualmathematical computation model formulation. It will involve“mapping”theattributestotheiroccu rrencesandidentificationoftheirassociation stotheoverallresultoutcomes.Eachattribute willbeanalyzedseparatelyandtogether with all others to identify the magnitude effect on theoveralloutcometheyhave[4].Oncetheatt ributes,theirrelationshipsandeffectstothefi naloutcomehavebeendiscretely identified, they will be combined with respect totheir magnitude and effect to the overall outcome to providethe actual mathematical formula that will be mapped on to theframework to provide the complete theoretical blueprint of thesystemthatcanbeimplementedbyavariet yofdifferentprogramming technologies and approaches that observe thealgorithmicprocess.Thisprocesswillinv olveseveralmathematicalprocesses.Thepro cesseswillfocusontwomain things.

1. Attributeincidentassociationcalculation .

Eachassociatedattributeimpactstheoverallo utcomedifferently as some have a greater impact on the outcome thanothers.Thiswillbeachievedbycomputin gthemagnitude

effect of the attribute individually and by combination withotherattributesthat resulttothesameoutcome.

Thegeneralmathematicalruletobeobser vedwillbeasbelow:

$$w_{c_a} = \sum_{1}^{n} a_{c_t}$$

Equation 1: Proposed Attribute incident equationWhere:

- Istheoverallweightoftheattri buteintermsofhowit affects theoutcome.
- Isthenumberofpossiblecombinati onsofthe attributes.
- Isthecombinationinstanceof theattributecombination.

This initial model is subject to alteration and improvementduring the research process. The result of this model will bethemathematicaldeterminationofthei mpactofspecificattributesor conditions toan outcome.

2. AttributeEffectCalculation

This model will also be used to identify the most dominantattribute that leads to an occurrence or incident. The reverseassociation rule will beas below:

$$a_{Dominant} = a_{i_{MAX}}$$

Equation2:AttributeEffectMagnitudeEquat ion

This first main stage will involve the most research effort andtime of implementation as the whole concept and purpose oftheresearchis basedon this stage.

ModelTestingandEvaluation

It will involve the actual development of a "proof of concept"system prototype that will not only be independently testablebut will demonstrate the practicability of thewhole researchas an artificially intelligent computer system. This stage willinvolveseveralstructuredindepende ntanddependentsub-stagesthatwillbefollowedtocomeupwit

hanactualworking model of the system. The stages involved will be asfollows.

PrototypeInterfacedesign

This stage will involve the actual system GUI design. It bedoneintwostages

1. Inputcontroldesign

Thissub-stagewillinvolvetheactualdesignofthemain system user interface and the required controls needed to fortheeasyinteractionandmanipulationbyth euserincomputingtheoutput.Themainfocus ofthisstageistodesign the controls and selection rules needed to compute thefinal outcome based on the selection criteria. This is where theuserwillinteractwiththesystemin performingtheanalysis.

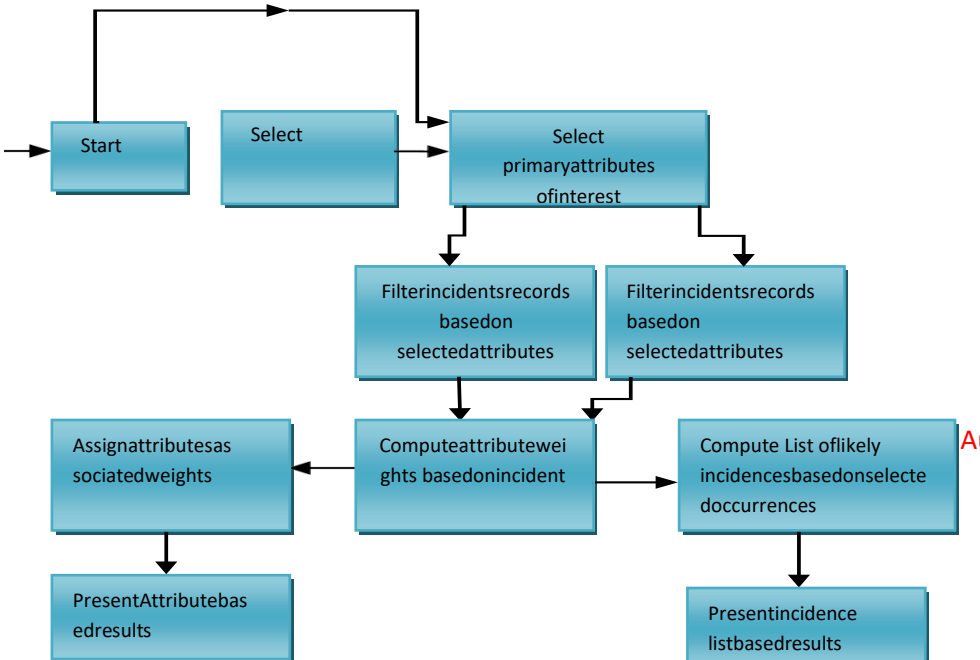
algorithms.

2. Outputresultsdesign

Thissub-stagewillinvolvethedesigningoftheresultsp resentation interface. Factors to be considered in this stageare the ease of interpretation of the results in such a way alayman can deduce the results of the analysis. The output willbepresented as acollectionofcharts and histograms

Systeminternalfunctiondesignandimplem entation

This is the third sub-stage and it is the most vital process ofthisresearchstagethiswillbetheactualizati onofthetheoreticalframeworkasworkingex ecutablecode.Itwillinvolve defining and implementing the relevant functions thatwillautomaticallycomputetheresultsusi ngtheartificialintelligence(MachineLearni ng)



DatageneratiоnPrototypetesting
Thisstage willinvolve:

1. Theactualgenerationofthesimulatio ndataforthemodelaccuracytestingp urposes.
2. Thesystemprototypetestingandcom parisontoexternallypredefinedcase s and conditions.
3. Observationoftheresultsandcompar isontothetestcases foraccuratereflectionofthemodelsv alidity.

ResultsDocumentation

This is the final stage of the research project. This will involvethedetaileddocumentationofthe projectssimulationandtesting observations results and conclusions. This will be thefinalconclusionoftheprojectasitisan dthesummarizationoftheresultsandobs ervationswillbedocumentedforexternal analysis and possible implementation of the proposedresearch product.

KEYOUTPUTS

If the model can be borrowed and implemented effectively onadata set,it shouldimprove:

- Intelligenceanalysisofcrimedatasets
- Strategicpolicyformulationusingre sultsofanalysisabove
- Intelligenceled policing ofmasses

4. KEYACHIEVEMENTS

The research has resulted to some noted achievements whichinclude:

- Mathematical models that may be utilized to create and execute a predictive model are being developed.
- An Automatable framework that can be applied to anydataset tobringout certainpatterns andtrends
- Extendingthebodyofknowledgeby introducinganovelhybriddata miningconcept

5. KEYCHALLENGES

This study, like many others, has had its share of difficulties, both in terms of the technology used and the resources available.

- Inability to calculate real-world occurrences due to a lack of genuine real-world data.
- Due to the scope of this study, there is a limited amount of time available for research, which necessitates both interpretation and analysis.
- Framework testing was limited due to the inadequacy oftestingdatarequired;occasionedb y thesensitivity ofthesector in which the intelligence analysis

frameworkisdesignedtofunctionin.

6. ASSUMPTIONS AND LIMITATIONS OF SCOPE

It is expected that this study and the development of this framework would face a number of problems, which are detailed below.

- The framework will work the same if applied to any data set and prototype
- In the absence of real-world data, there can be no assurance of the framework model's correctness.
- When this framework is adopted and used all policies within the law will allow its usage

7. CONCLUSION AND RECOMMENDATIONS

As a result of this framework's creation, it is now possible to do quick and automated analysis on large data sets. Using the elements that are important to events and persons, stakeholders in the relevant industry will be able to justify their policies based on this information. Data analysis speed, convenience, and reliability would all be much improved if this were put into practice.This will reduce the needless effort to develop half-baked rules and policies while ensuring that every major contributing attributes and entity sets are catered for.

REFERENCES

1. Murdoch, J. 1978. Queuing Theory Worked Examples and Problems. The Macmillan Press. Ltd
2. Pawlish, M., Varde, Aparna S. &Robila, Stefan A. 2012. Analyzing Utilization Rates in Data Centers for Optimizing Energy Management. IEEE International Green Computing Conference
3. Rittinghouse, John W. & Ransome, James F. 2010. Cloud Computing Implementation, Management, and Security. CRC Press, Taylor & Francis Group
4. Shahin, Ashraf A. 2017. Enhancing Elasticity of SaaS Application Using Queuing Theory, International Journal of Advanced Computer Science and Application Vol. 8
5. Velde, V. & Rama, B. 2017. Simulation of Optimized Load Balancing and User Job Scheduling Using CloudSim. 2nd International Conference On Recent Trends In Electronics Information & Communication Technology
6. Atmaca, T., Begin, T., Brandwajn, A., & Castel-Taleb,
7. H. 2016. Performance Evaluation of Cloud Computing Centers with General Arrival and Service. IEEE
8. El Kafhali, Said. & Salah, Khaled. 2017. Stochastic Modeling and Analysis of Cloud Computing Data Center. IEEE Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.
9. Guo, L., Yan, T., Zhao, S., & Jiang, C. 2014. Dynamic Performance Optimization for Cloud Computing Using M/M/m Queuing System. Journal of Applied Mathematics Vol. 2014, Hindawi Publishing Corporation
10. Hurwitz, J., Bloor, R., Kaufman, M. & Halper, F. 2010. Cloud Computing for Dummies. Wiley Publishing Inc.
11. Hwang, K., Fox, Geoffrey C., &Dongarra, Jack C. 2012. Distributed and Cloud Computing. Elsevier (Singapore) Inc.

12. Khazaei, H., Misic, J., &Misic, Vojislav B. 2012. Performance Analysis of Cloud Computing Centers Using M/G/m/m+r Queuing Systems. IEEE Transactions On Parallel and Distributed Systems Vol. 23