

Data Mining Techniques in Detecting and Predicting Cyber Crimes In Marketplace Sector

Jurisman Waruwu¹, Wira Hadinata², Siska Febriyani³, Rini Wijayanti⁴
Universitas Nias Raya, Institut Teknologi dan Bisnis Bina Sarana Global
Email : ¹rysmanwaruwu@uniraya.ac.id, ²wira@global.ac.id,
³siska@uniraya.ac.id, ⁴rini@uniraya.ac.id

Abstract - Marketplace is one business solution that can be profitable, because it is not bound by time and place. However, marketplace can be misused by irresponsible parties, and can harm others. Then a pattern is needed to predict cybercrime in order to prevent it. To get a pattern, we can use data mining. This paper presents a general idea about the model of Data Mining techniques and diverse cybercrimes in market place applications.

This paper implements data mining techniques like K-Means, Influenced Association Classifier and J48 Prediction tree for investigating the cybercrime data sets. K-means selects the initial centroids so that the classifier can mine the record and also formulate predictions of cybercrimes with J48 algorithm.

The knowledge of K-Means, Influenced Association Classifier and also J48 Prediction tree tends certainly to afford an enhanced, incorporated, and precise result over the cybercrime prediction in the marketplace sectors and prevent the cybercrime.

Keyword : Data Mining, Cyber Crime, Marketplace Sector

I. Introduction

Data is a very important thing. Because of this data can be processed for business needs and considerations the prospect of a company in the future. However, this data can be a general consumption if it is not properly protected. Many hackers take advantage of the vulnerability of a system's security to get data, which can then be misused. However, company data is a diamond that must be secured, and not be consumed by the general public.

The development of Information Technology has generated large amount of databases and huge data in various areas throughout the world. The research in information technology and databases has given rise to an approach to store, process and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from very large data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis [1].

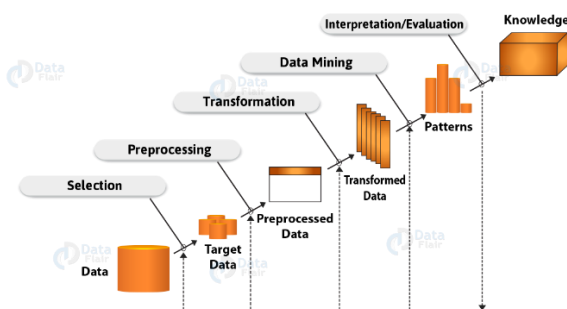


Figure 1. Knowledge Discovery Process

Cybercrime has increased in complexity and financial costs since the company began using computers to run its business. Cybercriminals are becoming more sophisticated in their actions and they target consumers and public and private organizations. Cybercrime analysis has a very momentous responsibility of law enforcement system in any country whole the world. Cybercrime involves the breakdown of privacy, or damage to the computer system properties such as files, folders, financial data and website pages or software [2].

Cybercrime can be reduced from marketplace transactions by applying the latest technology and appointing reliable officers and devices. Cyber attacks in the marketplace sector may be in the form of illegal access, destruction, bribery or amendment of data or any kind of malicious practices for sources of network damage, reboots or slings. The current security system has made cracks very monotonous but not impossible to hack. So some vital security intensities must be recognized before trading on the internet can be carried out sustainably. To protect from cyber crime, intrusion recognition methods must be intended, implemented and regulated. The proposed model was created to access a large number of criminal records so that predictions can be made according to the performance of criminal cyber actors. When notes in the system increase, there is no need to enter information. In the current digital era, the risks that might occur to protect large amounts of data with diverse cyber criminals are major disputes [2].

II. Literatur Survey

Dr. Zakaria Suliman Zubi and Ayman Altaher Mahmud (2013) had proposed a model for crime and criminal data that analyzes using simple K-means algorithm for clustering and Apriori algorithm for data Association rules. It also tends to help the specialist in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. They showed the promising results of their model proposed model from the attributes for crime, criminal and the results of K-means algorithm. They also gave the overall statistical knowledge about the criminal age versus crime type which provided the input to the K-means algorithm [3].

Mounjal Sourabh (2016) said when most of the transactions are going on electronically they are also becoming the source of crime. Lots of information shared and money transferred are on the target of cyber criminals. All organizations are increasing their technical base, more and more depending upon the internet for transactions

without analyzing the risk associated with the technology. Cyber criminals manipulate financial data, can move the electronic ownership, interrupted communications with employees or business partners, steal intellectual property, damage an organization's reputation, or bring e-commerce (or an entire business) to shut down [4].

Rasoul Kiani, Silamak Mahdavi and Amin Keshavarzi (2015) had applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales within 1990 to 2011. They employed Genetic algorithm for optimizing of Outlier detection operator parameters using Rapid miner tool [5].

Dr. K. Chitra and B. Subhashini (2013) had analyzed the data mining techniques and its applications in banking sectors like fraud detection and prevention, customer retention, marketing and risk management. They discussed the need of data mining techniques for better targeting and acquiring new customers, most valuable customer retention, automatic credit approval which is used for fraud detection and prevention in real time [6].

Lawrence McClenden and Natarajan Meghanathan (2015) have proven how effective and accurate machine learning algorithms are used to predict patterns of violent crime with data mining analysis. With the help of one of the data mining tools, the WEKA tool, they observed that Linear regression algorithm was very effective and accurate in predicting than Additive regression and Decision stamp algorithms [7].

III. Proposed Model For Cyber Crime Prediction

Discovering and exploring cybercrimes and probing their affiliations with virtual criminals are implicated in evaluating cybercrime progression. The proposed work presents the model over cybercrime prediction with K-Means clustering technique, Influenced Associative classifier and J48 classifier. For the cybercrime prediction in marketplace sectors, the proposed model grants an enhanced prediction outcome. Influenced Associative Classifier affords a well-organized way to utilize the classification method with Association Rule Mining, which enhances the prediction accuracy for classification data. The incorporated implementation of J48 technique with K-Means and Influenced Association Classifier provides the enhanced prediction outcome over the cybercrime hazards in marketplace sectors.

A. Collection of cybercrime dataset

A diversity of cybercrime data has to be collected for the prediction of cybercrime class in marketplace sector by the analysis of crime pattern. So this data has to be collected from various sources news feeds, articles and blogs, police department websites over the internet. The collected cybercrime data will stored in crime database for further handling of data.

B. Pre-processing of cybercrime dataset

The cybercrime dataset stored in Crime database has to be preprocessed before applying data mining techniques on them. Because preprocessing removes noisy data, missing values etc.



Figure 2. Proposed model for cyber crime prediction

C. Data Mining Techniques

For Pre-processed data, Data mining techniques and algorithms are implemented to identify or forecast fraud through Knowledge innovation from abnormal patterns and also it achieves recognition in combating fake transaction fraud Data Mining aids by contributing in solving tribulations in marketplace sector by discovering patterns, relationships and links that are unseen in the business information accumulated in the crime databases.

1. Association Rule mining

Based from frequent occurrences of the crime patterns, Association rule mining produces rules for cyber crime dataset. These generated rules can assist the assessment producer of defense society to take a hindrance action. The procedure comprises the subsequent measures:

- The method of determining commonly occurring some item sets in the cyber crime database.
- The identification of patterns in program implementation and both customer and seller behaviors as association rules known as intrusion recognition.

2. Clustering

Splitting up (divide) of a set of records or items to a number of groups is called clustering. Clustering is implied on discovering interactions linking cyber crime and criminal characteristics having some pattern. For discovering frauds in marketplace sectors, clustering techniques are utilized. Clustering is phrased as unsupervised learning because its classes are not definite and only determined in progress and consortium of data is through exclusive of supervision. K-means partition algorithm is implemented in clustering cyber crime datasets of marketplace sectors because of its minimalism and less computational intricacy. At first, the quantity of data items are assembled and précised as "k" clusters. Between the mean distances of objects, the

mean value is intended. The repositioning iterative method is utilized to recover the partitions by transferring items between clusters. Then until the union occurs, the number of iterations is carried out.

3. Classification

Classification is the most frequently used data mining technique, which executes a set of pre-classified examples to build up a model that can classify the instances of attributes at huge scale. The classification technique creates an association between a dependent variable with an independent variable by mapping the data points. Within the given dataset, Classification is used to bring out in which group each data occurrence is associated [2].

Classification is utilized to create several models of unknown patterns on the basis of the previous decision making. Automatic approve authorization is the nearly major procedure in the marketplace sector and financial organizations. Frauds can be prevented by building a superior assessment for the transaction consents using the classification representation based on decision trees such as J48, CART etc.

4. Influenced Association Classification

To achieve more accuracy, the associative classification is extremely novel and improved method which assimilates the mining of association rule and classifications of the model prediction. This methodology is being implemented for ruling out the link and association over item sets. The associative classification comes under unsupervised learning because it does engage any class characteristic for rule extraction. Two steps employed to extract association rules are:

1. Through cyber crime data set in marketplace sectors, classes are generated based on the association rule.
2. In the class labels, doing analysis on the dataset classification. The Influenced Association classification is entirely novel.

Perception for rule categorization. It also intends weighted confidence and support structure for mining association rules over the cyber crime data set in marketplace sector. Various steps implemented in Influenced Association Classifier has been summarized on below:

First, Pre-process the cyber crime dataset so further mining practices can be achieved on them.

- To replicate the assessment in the replica of prediction, every element is assigned within a range of weight (0-1). Attributes having additional significance are allocated maximum weight (0.9) and having fewer significance are allocated minimum weight (0.1).
- Influenced Association Rule Mining algorithm is implemented on pre-processed cyber crime data set for obtaining fascinating pattern invention. Influenced Association Classification uses weighted support and confidence and the rules spawned by this process are known as Classification Association Rule.

- The extracted Classification Association Rules are stored in the Rule base index.
- At any time if any new cyber crime record is updated, this CAR rule forecast the class label from the Rule base.

5. Cyber crime Prediction using J48

For the classification of issues and problems in the cyber crime prediction analysis in marketplace sector, J48 algorithm is more spiky and precise. Two steps in J48 are:

- Formation of tree.
- Validate the built tree over the cyber crime data set in marketplace sector.

The J48 algorithm uses pruning method for construction of the tree. The pruning technique diminishes the size of the tree by removing appropriate data that guides the terrible concert in prediction. The anticipated J48 algorithm classifies the data till the entire categorization and affords utmost accuracy over the training of cyber crime data in marketplace sector. It also stabilizes the precision and liveness of prediction. The J48 algorithm is the extensive version of decision tree C4.5. The J48 algorithm produces the classifier output in the form of rule sets and decision tree. The rule sets are straightforward to recognize and too easy for employing within the application marketplace system.

IV. Conclusions

The proposed model generates a superior concept over the cyber crime prediction by implementing data mining techniques such as K-Means, Influenced Association Classification with Prediction tree J48. The Influenced Association Classification is an improved model for classification and association with weighted support and confidence measures. From cyber crime datasets in marketplace sector, K-Means algorithm bunches the item sets. The Classification concert and precision can be enhanced with K-Means, Influenced Association Classification with Prediction tree J48. In the marketplace sectors, the clients have to be aided through precise requirements in the application software to discover alert while a stern interruption is recognized. Intrusion tools ought to be established wherever it is practicable and appraised on a standard basis. To prevented beside cyber attacks, customer tutoring must be prepared in association with government and other confidential organizations. Awareness agenda should be put into practice to guarantee that clients recognize data concerns, intensity of privacy and the method to make the marketplace transactions secure.

Bibliography

- [1] M. Ramageri, "Data Mining Techniques And Applications," vol. 1, no. 4, pp. 301–305.
- [2] 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.
- [3] Z. S. Zubi, "Using Data Mining Techniques to Analyze Crime Patterns in the Libyan Using Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data," no. December, 2013.
- [4] S. Munjal and A. Anooja, "Cyber Crimes Threat for the E-Commerce Cyber Crimes – Threat for the e-commerce," no. December 2017, 2016.
- [5] R. Kiani, S. Mahdavi, and A. Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 8, pp. 11–17, 2015.
- [6] K. Chitra and B. Subashini, "Data Mining Techniques and its Applications in Banking Sector," *Int. J. Emerg. Technol. ...*, vol. 3, no. 8, pp. 219–226, 2013.
- [7] N. Meghanathan, "U Sing M Achine L Earning A Lgorithms To," no. March 2015, 2016.