iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** I **Month of publication:** January 2018

DOI: http://doi.org/10.22214/ijraset.2018.1008

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

Application of Web Mining in Fraud Detection

D. Santhi Jeslet [1]
[1]*Asst. Professor, Department of Computer Science, M.G.R.College,Hosur,TN, India*

Abstract: Web mining is one of the major applications of data mining techniques which automatically extract information from the web documents and services. This is a technique that can be used to study the behavior of the user. It is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web. Now-a-days the usage of credit cards dramatically increases. As credit card becomes the most popular and stylish mode of payment for both online as well as offline purchases, fraud associated with it are also increasing. Fraud Detection System(FDS) deals with identification of such fraud using various techniques like classification, clustering and aprior. These techniques of web mining can be integrated to represent the sequence of operations in credit card transaction processing and show how it can be used for the detection of frauds.
Keywords: Web mining, FDS

I. INTRODUCTION

With the rapid development in the field of computer networks and multimedia technology, World Wide Web (WWW) has swell up. The explosive growth of information sources available on the WWW has increased the necessity of automated tools in order to find, extract, filter, and evaluate the desired information and resources for the users. For addressing the above said tool, we are entering into a new concept called "*Web mining*".

Web mining is one of the major applications of data mining techniques which automatically extract information from the web documents and services. It is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web. Classical data mining techniques that are used in web mining includes classification, clustering, and association rule. Web mining can be divided into three categories: content mining, structure mining, and usage mining.

A. Web Content Mining

It is the process of discovering useful information from the web which may be in the form of text, images, audio and video. For the discovery, it uses the techniques of Artificial Intelligence (AI), Database and most specifically Data Mining (DM).

B. Web Structure Minin

It helps to derive knowledge of interconnection of documents, hyperlinks and their relationships. It uses graph theory to analyze the node and connection structuring of a web site.

C. Web Usage Mining

It is also called as web logs mining. This helps to judge about the usage of a web page. It uses computer network concepts, artificial intelligence and database. It helps to understand and better serve the needs of web based applications.

D. Major applications of Web Mining

1) To identify fraud in credit card usage
2) Plays vital role in improving the E-business[5]
3) Helps to maintain Customer Relationship Management (CRM)
4) For enhancing Online Distance Education[2]
5) To analyze computing jobs[6]
6) To evaluate and improve E-government Performance[1]
7) For enhancing Technology Management[3][1]
8) To enhance Knowledge Discovery on the Web[4]

©IJRASET (UGC Approved Journal): All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 6 Issue I, January 2018- Available at www.ijraset.com

II. STEPS IN WEB MINING

The major modules in the Web mining are (depicted in the Fig.1): data collection, data preprocessing, Pattern Discovery, Personalization and Recommendations. In these modules first three are offline modules and remaining are online modules.

A. Data collection

Data collection is the first step of web mining. This is an important and most difficult step. The main sources of data for web mining are server side data, client side data and proxy server data.

B. Data preprocessing

The data collected from various sources are sometimes insufficient, inconsistent and may include noise. The data preprocessing has to be performed in order to make the data to be clean so that application of data mining algorithm becomes easy. The data preprocessing work mainly include data cleaning, user identification, session identification and path completion.

1) *Data Cleaning:* The purpose of data cleaning is to eliminate irrelevant items. Since the target of web usage mining is to get the user's travel patterns, following two kinds of records are unnecessary and should be removed:

a) The records of graphics, videos and the format information are irrelevant. The records have filename suffixes of GIF, JPEG, CSS, and so on, which can found in the URL field of every record

b) The records with the failed HTTP status code. By examining the Status field of every record in the web access log, the records with status codes over 299 or fewer than 200 are removed. It should be pointed out that different from most other researches, records having value of POST or HEAD in the Method field are reserved in present study for acquiring more accurate referrer information.

2) *User Identification:* The task of identification of a single user is fundamental to distinguish her/his behavior in web mining. Various methods have been proposed to automatically recognize the user. Some of the methods are using cookies, based on IP address, path analysis based on the network topology etc.

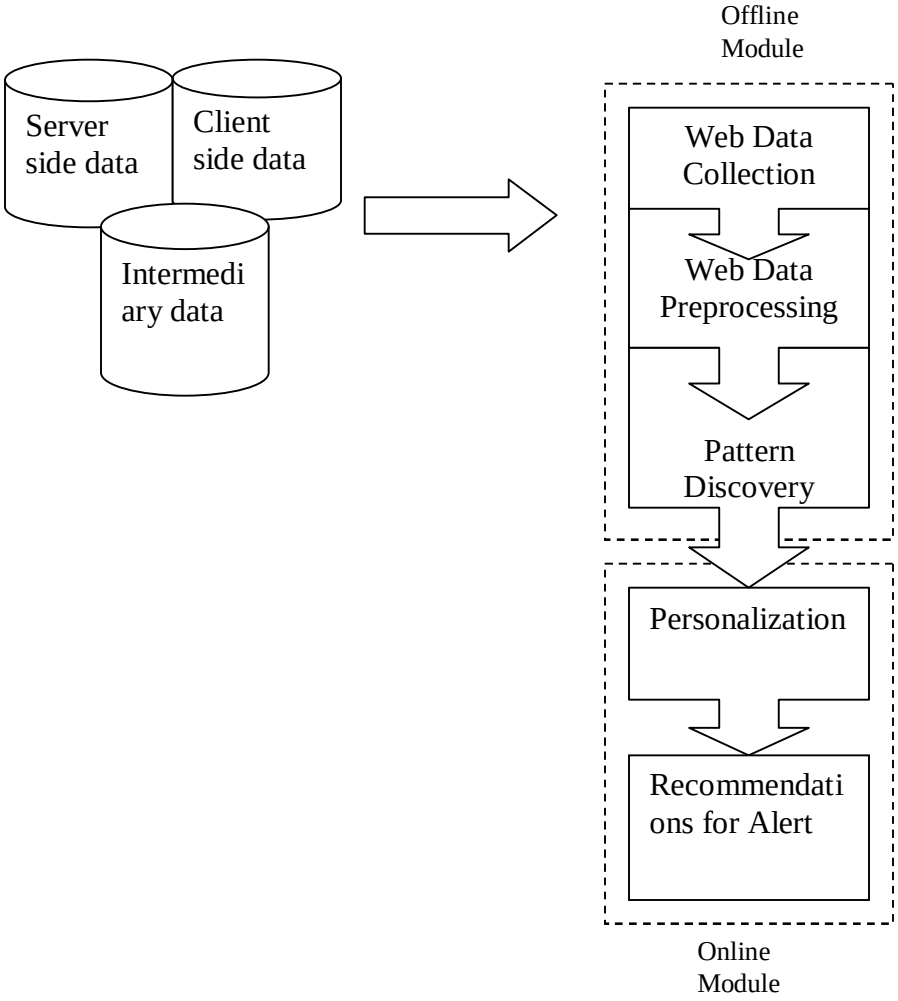
3) *User Session Identificatio:* This is related with identifying the users behavior with a particular web site. The user session is identified by the set of URLs corresponding to the pages visited by a user from the time he/she enters a web site to the time he/she leaves it. Heuristic method based on time and reference can be used for user session identification

4) *Pattern Discovery :* After completing the preprocessing of web data, discover patterns of usage of web site by using statistical, data mining, machine learning and pattern recognition techniques [3]. In particular association rule and clustering techniques of data mining are very frequently used for pattern discovery.

5) *Personalization and Recommendations:* After the identification of patterns, personalization and recommendations can be given to the merchant or on-line payment that the particular usage of card is fraudulent.

©IJRASET (UGC Approved Journal): All Rights are Reserved

Fig. 1 Steps of Web Usage Mining



III. WHAT IS FRAUD?

The usage of credit cards dramatically increases day by day. Credit card becomes the most accepted, popular and stylish mode of payment for both online as well as offline purchases. Because of this popularity, frauds associated with it are also increasing. Fraud is an intentional trick made for personal gain or to damage another user/individual is fraudulent. Fraud is a civil law violation and also a crime. The common purpose of fraud is defrauding people or entities of money.

Without having cash in hand selling goods or services is done through the medium called Credit card. A credit card is a simple way of offering credit to a consumer automatically [2]. Credit card has an identifying number that helps in shopping transactions quickly. Credit card fraud detection identifies those transactions that are fraudulent into two classes of legitimate and fraudulent transactions. The credit card fraud detection system developed used four clusters of low, high, risky and high risk. Once the transaction is legitimate, it is processed further. But, if any transaction falls into any of these clusters it is labeled as doubtful/fraudulent. The alert is given off along with the reason. The fraudulent transaction will not be processed but will be committed to the database. Fraud is an intentional deception made for personal gain or to damage another user/individual is fraudulent. Legal definition varies by legal

©IJRASET (UGC Approved Journal): All Rights are Reserved

jurisdiction for fraud [2], [3]. Fraud is a civil law violation and also a crime. Defrauding people or entities of money is a common purpose of fraud.

IV. IMPLEMENTATION OF FRAUD DETECTION SYSTEM(FDS)

Various techniques of web mining like classification, clustering and association [7], [8] will be integrated to represent the sequence of operations in credit card transaction processing and show how it can be used for the detection of frauds.

Initially, the three web mining techniques are trained with the normal behavior of a cardholder. If the credit card transaction is not accepted by the web mining model with high probability, it is considered to be the fraudulent. At the same time, the system gives importance to ensure that genuine transactions will not be rejected. Web mining techniques can be trained on examples of fraud due to lost cards, stolen cards, application fraud and mail-order fraud [6].

Using data from a credit card issuer, FDS is trained on a large sample of labeled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent three month period of time. The FDS detected significantly more fraud accounts with significantly fewer false positives over rule-based fraud detection procedures.

The FDS will receive the card details and the value of purchase to verify, whether the transaction is genuine or not. The FDS is able to detect frauds by considering a cardholder's spending habit without its significance. If the Fraud Detection System confirms the transaction to be of fraud, the transaction will be declined. The number of False Positive transactions is very much reduced which is an added advantage of FDS.

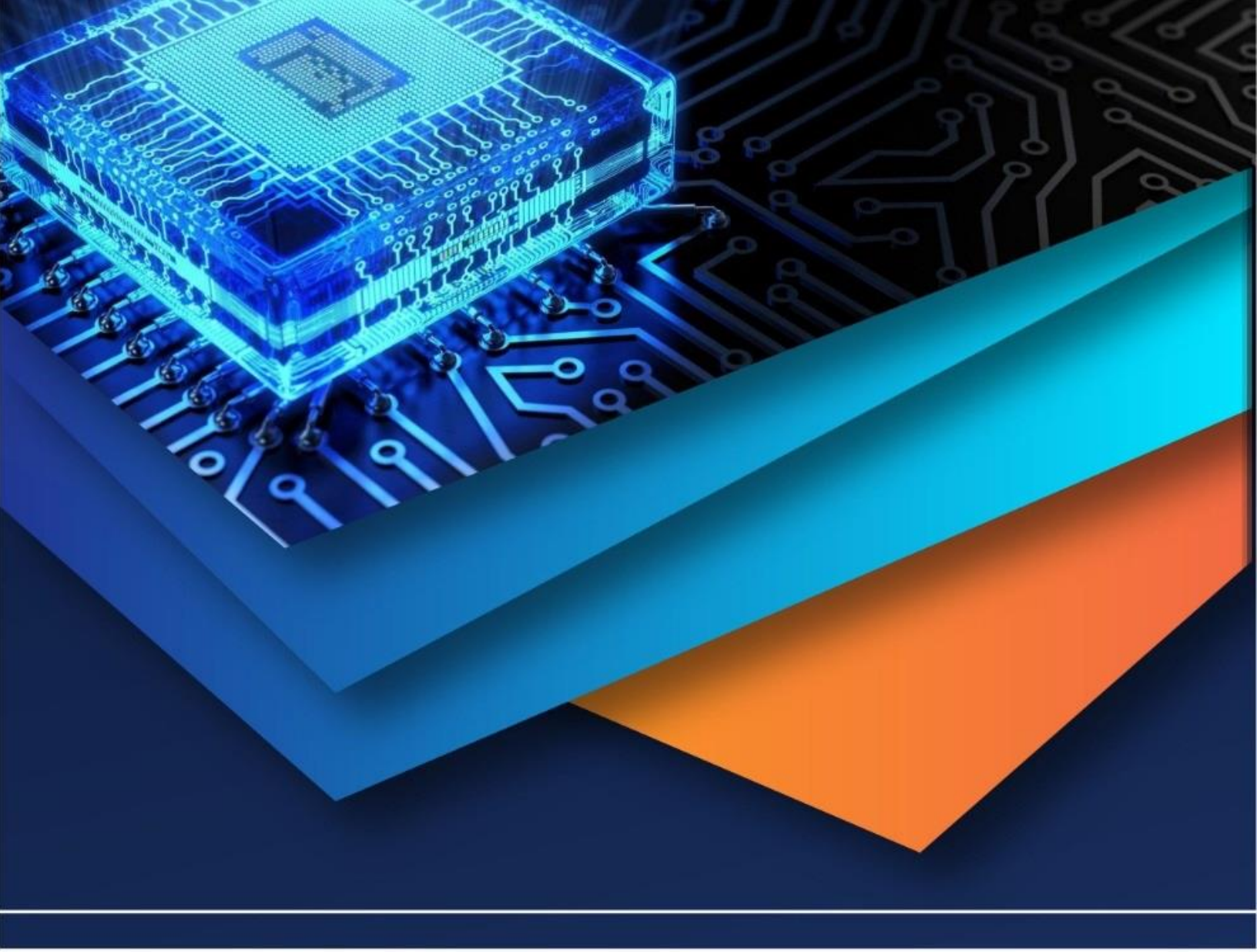
V. CONCLUSION

This paper dealt with an introduction to Web mining, its major steps and major application areas. Since usage of credit cards increases every day the fraud associated with it also increases. So, this paper also dealt with one of the application called Fraud Detection System (FDS). The three major techniques classification, clustering and association are integrated to represent the sequence of operations in credit card transaction processing and used for the detection of frauds. The FDS is initially trained properly with set that consisted of all account activity over a subsequent three month period of time and then used for identifying the fraud related to credit card usage.

REFERENCES

[1] Zhou Xiaobin **"**A Practical Approach to E-government Performance Evaluation Based on Web Usage Mining" Third International conference on Knowledge discovery and data mining,2010,WKDD'10, Digital Object Identifier:10.1109/WKDD.2010.85 , Page(s): 154 – 158
[2] "Statistics for General and On-Line Card Fraud," http://www.epaynews.com/statistics/fraud.html, Mar. 2007.
[3] D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams, "Data Mining for Fun and Profit," Statistical Science, vol. 15, no. 2, pp. 111-131,2000.
[4] Gu Chengjian & Huang Lucheng "Web Mining in Technology Management" International Seminar on Business and Information Management,2008, Digital Object Identifier:10.1109/ISBIM.2008.127 , Volume 2,Page(s): 88 - 91
[5] S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results fromthe JAM Project," Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130-144, 2000.
[6] Sunil Kr.Pandey1 & R.B.Mishra2 "Intelligent Web Mining Model to Enhance Knowledge Discovery on the Web" Seventh International conference on Parallel and Distributed Computing Applications and Technologies, 2006, Digital Object Identifier:10.1109/PDCAT.2006.74, Page(s): 339 – 343.
[7] Pieter Adriaans and Dolf Zantinge. "Data Mining"
[8] Jiawei Han & Micheline kamber (2005), "Data mining concepts and techniques

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)