Journal of Engineering Sciences
Vol 11, Issue 1, Jan /2020
ISSN NO:0377-9254

Secure Dynamic Data Mining By Using An Regression Method

[1]RAVI VANI, [2]Mr.A.HARI KUMAR
[1]Stuedent,[2]Assistant Professor
Department of Computer Science and Engineering
Visakha Institute Of Engineering & Technology , Visakhapatnam

ABSTRACT

Data Mining is one of the important techniques for mining the data where we can able to find duplicate copies of repeating data, which has been widely used in cloud storage to reduce the quantity of memory space and save bandwidth. To mine the confidentiality and sensitive information where the repetition of data occurs while supporting, the convergent encryption technique has been projected to encrypt the data before outsourcing. To better protect data security, this process of Data Mining has been implemented. We also introduce several new de duplication constructions supporting authorized duplicate checks in any kind of data by using a logistic regression process. The logistic regression technique involves dependent variable, which can be represented in the binary (0 or 1, true or false, yes or no) values, means that the outcome could only be in either one form of two. For example, it can be utilized when we need to find the probabilit y of successful or fail event.

Keywords:- Data Mining, Knowledge Discovery in Databases, Regression, Regression-ClassMixture

I.INTRODUCTION

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre processing, model and inference considerations ,interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing. Data mining processes are used to build machine learning models that power applications including search engine technology and website recommendation programs.

In simple words, data mining is defined as a process used to extract usable data from a larger set of any raw data. It implies analysing data patterns in large batches of data using one or more software. Data mining has applications in multiple fields, like science and research. As an application of data mining, businesses can learn more about their customers and develop more effective strategies related to various business functions and in turn leverage resources in a more optimal and insightful manner. This helps businesses be closer to their objective and make better decisions. Data mining involves effective data collection and warehousing as well as computer processing. For segmenting the data and evaluating the probability of future events, data mining uses sophisticated mathematical algorithms. Data mining is also known as Knowledge Discovery in Data (KDD).

The term "data mining" is a misnomer, because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (mining) of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence (e.g., machine learning) and business intelligence. The book Data mining: Practical machine learning tools and techniques with Java (which covers mostly machine learning material) was originally to be named just Practical machine learning, and the term data mining was only added for marketing reasons. Often the more general terms (large scale) data analysis and analytics – or, when referring to actual methods, artificial intelligence and machine learning – are more appropriate.

The actual data mining task is the semi-automatic or automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining, sequential pattern mining). This usually involves using database

techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

1.1 PROBLEM DEFINITION

The problem can be defined as," Developer a user-friendly system which is able to detect and mine the data from large sequence of data sets such as (banks, public sectors and hospitals) and captures as the input for the code and execute the path depending upon the set of relations given in the programming ."

1.2 OBJECTIVE OF PROJECT

The objective of the project is to mine the data from the uploaded data set and then implement a code by using an language (Regression method) and need to validate such that the data is validate for the certain given conditions and should trigger a valve when a set of data set is validated.

1.3 LIMITATIONS OF PROJECT

To extract the foreground data set from the background, and should check the eligible criteria and must developed an set of output results first.

The condition must remain constant and should validate for the whole set of data for extracting the data for processing the algorithm .The environment should not be over exposure and conditions must remain constant as much as possible to avoid repetition of same valves.

II.LITERATURE SURVEY

Data Mining is the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the World Wide Web. Regression is a data mining function that predicts a number. Profit, sales, mortgage rates, house values, square footage, temperature, or distance could all be predicted using regression techniques. For example, a regression model could be used to predict the value of a data warehouse based on web-marketing [3], number of data entries, size, and other factors.

A regression task begins with a data set in which the target values are known. For example, a regression model that predicts data warehouse values could be developed based on observed data for many data warehouses over a period of time. In addition to the value, the data might track the age of the data warehouse, size and number of clusters and so on. Data warehouse value would be the target, the other attributes would be the predictors, and the data for each data warehouse would constitute a case. In the model build (training) process, a regression algorithm estimates the value of the target as a function of the predictors for each case in the build data. These relationships between predictors and target are summarized in a model, which can then be applied to a different data set in which the target values are unknown.

Regression m odels are tested by computing various statistics that measure the difference between the predicted values and the expected values. The historical data for a regression project is typically divided into two data sets: one for building the model, the other for testing the model.

2.1EXISTING SYSTEM

In the existing system, the data of the people are displayed randomly with the details of the people who took the loan and who cleared the loan using java.

DISADVANTAGES OF EXISTING SYSTEM

- The complexity is more.
- There is a problem in accurately analyzing the details of the persons.

2.2 PROPOSED SYSTEM

In the proposed system, whenever the data set is entered then the details of the bank is loaded and it checks the data and if the data is validated then it undergoes minning and filters the persons who cleared the loans and it displays the data and by using that data we can check the details and inform to the persons who eligible to apply for loans.

Types of Regression Techniques

Standard multiple regression considers all predictor variables at the same time. For example 1) what is the relationship between income and education (predictors) and choice of neighborhood (predicted); and 2) to what degree

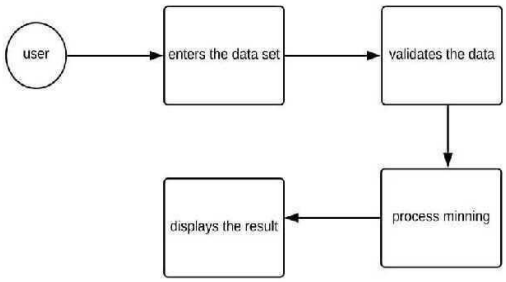
do each of the individual predictors contribute to that relationship?

Stepwise multiple regression answers an entirely different question. A stepwise regression algorithm will analyze which predictors are best used to predict the choice of neighborhood — meaning that the stepwise model evaluates the order of importance of the predictor variables and then selects a relevant subset. This type of regression problem uses "steps" to develop the regression equation. Given this type of regression, all predictors may not even appear in the final regression equation.

Hierarchical regression, like stepwise, is a sequential process, but the predictor variables are entered into the model in a pre-specified order defined in advance, i.e. the algorithm does not contain a built-in set of equations for determining the order in which to enter the predictors. This is used most often when individual creating the regression equation has expert knowledge of the field.

Setwise regression is also similar to stepwise but analyzes sets of variables rather than individual variables.

III. CONTENT DIAGRAM OF PROJECT



3.1 Algorithms and Flowchart

Sequence of Steps in Project

- Foreground objects are extracted using background subtraction.
- Static objects are detected by using contour features of foreground objects of consecutive frames.
- Detected static objects are classified into human and non-human objects by using edge based object recognition method.
- Nonhuman static object is analyzed to detect abandoned object.

Flowchart

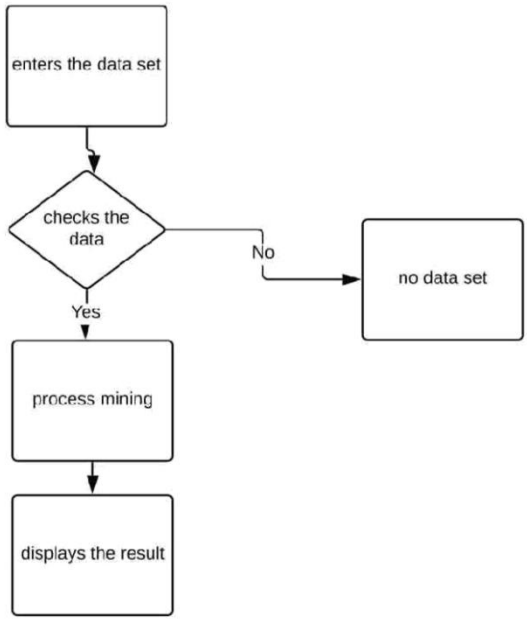


Figure 1 Flow Chart

IV.IMPLEMENTATION & RESULTS

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and changeover methods.

4.1Explanation of Key Functions

Foreground Extraction

In order to extract the region of interest, the foreground should be extracted from the image. This way the background does not have to be considered anymore in further processing. The result of the foreground subtraction will affect the accuracy of the object recognition.

People who cleared the loan

To understand the semantics of the extracted foreground, this process has to distinguish the various kinds of objects that may occur in the scene. These objects include people who cleared the loan. However, before the data set can take place, we have to examine the input first and then extract the features that are useful. Examples of these features are shapes, texture and colour. Using the generated rules from the knowledge based

system, it will be possible to classify the various objects.

VI.CONCLUSION

In this, we propose the approach for the people who qualify to apply for the new loans and who cleared the loans by using bank data set. By using this we can increase the popularity of the banks and includes offers to take the bank loans.

FUTURE ENHANCEMENT

While the concept of the proposed system may be really helpful to improve the level of popularity and security, there are still improvements that are possible in the future. Some of the thinkable improvements are:

- Using different bank sets

  Here we include single Bank data set, in future we can include different bank data sets at a time and here mining takes lots of time to execute and we can reduce the time of execution by using updated methods.

- Automatic message sending

  To use this application we need a person to check the data who clears the previous loan and to apply for new loan. So without a person we need to update the data and automatically it sends the message to person who clears the previous loan and to apply for new loan and the person directly gets the message with offers also.

REFERENCES

[1] Berners-Lee T.J., Cailliau R., Groff J.F., Pollermann B. (1992) *Electronic Networking: Research, Applications and Policy*, 2(1).

[2] *http://dss.princeton.edu/training/Regression101.pdf.*

[3] Anirban Mahanti, Carey Williamson and Derek Eager, *Traffic Analysis of a Web Proxy Caching Hierarchy, University of Saskatchewan.*

[4] *http://www.bren.ucsb.edu/academics/courses/206/readings/ readerch8.pdf.*

[5] Jennrich R.I. (1969) *The Annals of Mathematical Statistics* 40, 633-643.

[6] Blanz B. Scholkopf, ulthoff H.B, Burges C., Vapnik V. and Vetter T. (1996) *ICANN*, 1112, 251-256.

[7] Stromberg A.J. (1993) *J. Am. Stat. Assoc.* 88 (421), 237-244.

[8] *http://www2.tech.purdue.edu/cit/Courses/CIT499d/ODMr%2011g%20Tutorial%20for%20OTN.pdf.*

[9] Heckerman D. (1995) *A Tutorial on Learning Bayeisan Net-works.*

[10] Bekaert Geert, Robert J. Hodrick and David Marshall (2001) *Journal of Monetary Economics*, 48, 41-270.