# Implementation of a Secure Data Mining Technology to Develop Aknowledge Based System in a Business Environment

**ZEYAD HAMID**

**SCHOOL OF ARCHITECTURE, COMPUTING AND ENGINEERING, UNIVERSITY OF EAST LONDON**

**Kuala Lumpur, 71700, Malaysia**

## Abstract

Because the advancement of the internet, several small and large firms have developed their dealings with customers to offer goods through the internet worldwide. Online credit-card deception or no card present deception is gradually rampant in the latest years because that the credit-card is mainly exploited to demand payments by these firms on the internet. So it is necessary to verify safe transactions for credit-card holders when consumption their credit cards to do online payments for merchandise and facilities offered on the internet is a critical issue. Data mining has widely obtained recognition in struggle online credit-card deception thanks to its efficient artificial intelligence (AI) methods and algorithms that can be applied to identify or forecast deception through Information Detection from uncommon behaviour resulted from collected data. In this research, a system's model for online credit card fraud detection is discussed and designed. This system implements and tested using Data mining to detect fraud in a real time transaction on the internet, and thereby classifying the transaction as legitimate, and illegitimate transaction.

*Keywords: Data Mining, Credit Card, Fraud Detection, Transaction, Online Shopping*

## 1. Introduction

The credit card could be a tiny plastic card issued to users as a system of payment. It permits its cardholder to shop for product and services supported the cardholder's promise to pay money for these product and services. Fraud is human behavior of intentional deception for gaining benefits of monetary or different aspects. In later years, credit card fraud has grown up to a significant and threatening money crime. This crime involves counterfeit or hijack of 1 or many legitimate attributes of the user, by means that of fraud, physical stealing of cards, hacking of on-line user credentials etc. and victimization such illegitimate info for own benefits. Data processing is one among the effective ways of dominant credit card fraud. Data mining is technique of extracting unseen and antecedently unutilized info from the offered knowledge. At present, fraud analysis and detection have received major attention thanks to its capabilities of extracting hidden info and victimization it to with efficiency classifying or predicting new info.. Credit card fraud detection is that the method of distinguishing those transactions that square measure dishonest into 2 categories of legitimate (genuine) and dishonest transactions (Maes, S. at all (2008)). Credit card frauds are often generally classified into 3 classes, that is, ancient card connected frauds

## 2. Related work

In a book of information mining, (Witten &amp; Frank, 2005) defines data mining because the process of discovering patterns in knowledge, mechanically of semi-automatically, that is meaty and can cause some blessings ideally economical. Book more adds, data processing is concerning resolution drawback by analyzing already gift knowledge in info. (Palace, 1996) Generalize data mining or data discovery because the process of analyzing knowledge from completely different perspective and summarizing them into helpful data finding correlations or patterns in several fields in giant on-line database.

In articles, (Fayyad et al., 1996) and (Palace, 1996) classifies usages of information mining in principally four tasks, clustering, classification, association rule learning and regression that area unit used severally and/or in combined type in several fields like business and selling, scientific study &amp; technology, security and police work and pattern mining in spatial  knowledge. In their paper, (Phua et al., 2010), they need shortly explained data processing and fraud. In keeping with them, data processing as concerning finding insights that is area unit statistically reliable and unknown earlier, and tools to resolve drawback that can't be done by question and reportage tools. Additionally fraud as business essential drawback and its detection and hindrance is imperative in any competitive atmosphere.

In article, (Palishikar, 2002) author acknowledges fraud as white color crime accountable to bleed many billions from business from worldwide. They additionally determine, data processing as facilitate to mine proof of fraud and dishonorable activities from the massive quantity of information. (Mahdi et al., 2010) Agree in reality that fraud and credit card fraud instances may be similar in content and look however they typically area unit completely different. They are available in numerous form and sizes. Therefore it wants complicated and time overwhelming investigations (Palishikar, 2002). Recognizing the potential, telephone firms were among the primary to use knowledge analysis techniques to forestall fraud. (KUÅžAKSIZOÄž

PhD. thesis, (Paasch, 2008) explains extensively concerning credit card group action and its varied fraud sorts. (Paasch, 2008) and (Bhattacharyya et al., 2011) divides applied math fraud detection methodologies in 2 approaches, supervised and unsupervised. The paper, (Bhattacharyya et al., 2011) more classifies the fraud detection methodologies into 3 data processing technique supplying regression (LR), support vector machines (SVM), random forest (RF). In paper, (Srivastava et al., 2008) authors have instructed that credit card fraud detection have attracted several analysis interest over several techniques. Paper more makes special stress on data processing and use of neural network.

From the first 90's to late 2010 there are a unit range of different account of researches that area unit dedicated for credit card fraud detection. a number of them area unit strictly sequence alignment algorithmic analysis like, (Kundu et al., 2009) and (Srivastava et al., 2008). Some of area unit strictly subtle system as mentions by (Aleskerov et al., 1997) Clementine, Darwin, Falcon and PRISM. they're supported varied data processing techniques like classification, mental image, segmentation, clustering, profiling, deviation detection, and association rule generation, K-Nearest Neighbor (KNN) technique enforced severally or combinative type in building Neural network, call trees, formal logic and skilled systems etc. more it report of leading to seventieth fraud detection rate.

According to (Bhattacharyya et al., 2011), each supervised and unsupervised fraud detection ways area unit supported prognostic model. Paper additionally accents on active use of prognostic models of credit card fraud detection. It more reports, on comparatively few studies conducted on this subject. Among those few studies, primary focus is upon neural network and its variance of technique of applications (Maes et al., 1993) (O`Dea et al., 2001) (Gupta &amp; Bhargava, n.d.) (Aleskerov et al., 1997) (Zhu et al., 2010) (R. Brause, n.d.).

In analysis paper revealed back in 1994, (Ghosh &amp; Reilly, 1994) researchers had projected use of neural network for police investigation credit card fraud. Researchers had engineered a system to discover by coaching on giant sample of tagged credit card account transactions that contains fraud cases thanks to lost cards, purloined cards, application and counterfeit fraud etc.

Another complete system builds for same purpose victimization meta-learning techniques to find out model of dishonorable credit card transactions (Stolfo et al., 2007). Recent analysis accounts of development of fraud detection system victimization PGNNs (parallel granular neural networks), for increasing data mining and data discovery process (Syeda, 2002).

Artificial neural network, a well-liked term since 90s and mention many time in higher than literatures, is outlined as "A neural network is AN interconnected assembly of straightforward process components, units or nodes, whose practicality is loosely supported the animal neurons" (Gurney, 1997) and neural network once used for &amp; with intelligence then is named Artificial Neural Network (Gupta &amp; Bhargava, n.d.). With definition, authors have explained concerning operating of ANN within the paper. In (Gupta &amp; Bhargava, n.d.), varied sort of

design of ANN is explained like Kohonen's self-organizing maps and straightforward Hopfield network. This paper considerations application of ANN in business applications where principally are predicating anomalies, fraud, bankruptcy etc.

There is a unit range of analysis paper that explains the utilization of neural network for specific credit card fraud detection. As several of researches there is a unit as several variances of the approached. like use of Bayesian network and neural network (Maes et al., 1993), use of combined probabilistic and neuron-adaptive approach (R. Brause, n.d.), feature choice approach before classifying victimization neural network (O`Dea et al., 2001), PGNN for quick learning and detection (Syeda, 2002), improvement victimization Genetic formula for Artificial Neural Network (Paasch, 2008), and study of use of chaos theory and neural network for credit card risk detection (Zhu et al., 2010). From higher than literatures we will observe several researches area unit done and plenty of our current for developing optimized, economical and value effective technique to discover risk and fraud in direct or on-line group action involving credit / debit cards that area unit supported data processing and neural network.

The wide unfold news of the business disasters round the world has indirectly have an effect on the general public confidence in capital market system. The collapse of Enron and Arthur Andersen are a area unit one amongst the key impacts to business atmosphere. Firms like Siebel System, Qwest, WorldCom and Xerox area unit samples of firms that have issues with their accounting. Royal Ahold 2010 that is thought as Dutch retail Trade Company has additionally concerned during this issue. Moreover, UK Baring's Bank and therefore the Japanese Daiwa's Bank have lost legion greenbacks thanks to ineffective finances. All this issue of bankruptcy and money fraud is thanks to inaccurate credit method.

Fact is that a lot of parties, like shareholders, investors, creditor, tax authorities, and managers have an interest within the exactness of organizations' money performance. Creditor are the area unit in a very position to watch and management operations in organizations ( Eija Koskivaara (2011)).

The globally calculable average loss per organization from economic fraud is calculable to be $2,199,930 over a biennial amount. in keeping with Association of Certified Fraud Examiners (ACFE), roughly six % of firm's revenues or $660 billion lost per annum because the results of fraud. The continual increase in fraud has cause the increase from many anti-fraud laws. However, several structure anti-fraud efforts don't seem to be current and still have loopholes. Organizations area unit still finding ways that to fight this fraud since red flags approach don't seem to be effective additionally. Though red flag area unit related to fraud it's still not good enough. Moreover, since it focuses attention on specific cues, it inhibits internal and external creditor from distinctive different reasons of fraud prevalence. Another excuse of organization finding other ways to fight back fraud is as a result of entities use impractical way of fraud detection. It's additionally judicious for forestall fraud instead of take plenty of trouble to fix

through it. Most of the businesses and creditor wear down fraud on independent basis instead of implementing it in long run arrange. Yankee Accounting Association (YAA) encourages researches to collaborate with creditor to beat this issue by arising with new technology or approach.

A research was conducted so as to spot what percentage organizations use tools to combat fraud. Studies show that solely five.37 % of organization uses data processing to fight fraud. (James, Richard and Carl (2012)).

## 3. Research Methodology

Research methodology is a technique (s) that an investigator applies to carry out his investigation activities. It contain four types; qualitative, quantitative, Pragmatic approach to research (mixed methods) and Advocacy/participatory approach to research (emancipatory). (Types of research 2014)

### 3.1 Qualitative Approach

This kind provides further details around human manner, emotion and private individualities that quantitative technique might not process (Madrigal and McClain 2012). It perform the process of recording, investigating and sincerity to discover the great meaning and conduct of the human. There are diverse approaches exploited to perform this kind of method, these are: observation, database records and interview (Refer to figure 9).

### 3.2 Quantitative Approach

It is a "statistical investigation that permits investigators to mine significant information from investigation data that comprises favorite directions, divergences amongst sets and population" (Madrigal and McClain 2012). Some of these statistical investigation encompass; average, typical divergence and illative statistics. Furthermore, this method contains techniques such as questionnaire and survey (Refer to figure 9).

### 3.3 Pragmatic Research (Mixed Methods) Approach

It is "apply the better technique in execute a investigate problem without been constraint in the philosophical discussions on the better method to be apply" (Alzheimer 2009). This method permits investigators to apply either qualitative or quantitative technique. This method give researcher chance to use diverse practices. in any case what practices or approaches apply in this methodology, data gathered applying this method must be investigated in a appropriate approach.

### 3.4 Advocacy/Participatory Approach (Emancipatory)

In this method "investigators do their investigation intentions in discovering positive change in their aim investigation topic because they expect that other methods cannot achieve their requirements" (Alzheimer 2009).

### 3.5 Research Strategy

In this study, the writer apply quantitative methodology to execute his investigation. They encompass Data gathering, period and data collection that exploited the observation method. The period for this study to be accomplish is restricted.

### 3.6 Information Collecting

The objective of this investigation is to apply data mining technique in credit card fraud detection to discover illegal transactions. The suggested system will try to overcome on the weaknesses of the traditional credit card fraud detection system,. The necessary information has been gathered from the following sources:

> College (Library & Ebrary).
> E-books, Journals, articles and white papers.
> Google Scholar.
> Seminars Papers.
> Dissertations.
> Expert's forums.

### 3.7 Time Horizon

"There are two time horizons for research works: Either cross-sectional studies or longitudinal studies. Cross-sectional studies are applicable if projects are time-constrained. The study of particular phenomena is conducted over a short period. Similarly, longitudinal study employs a method that investigates, for example, change and development. To do that, researchers require a longer timeframe" (Saunders 2009). In this investigation the writer has selected time horizon cross-sectional way since to the constraint of investigation period.

### 3.8 Research Data Collection

Because credit card fraud data consider very sensible to the banks and they won't to purplish this type of data. The writer select databases examination, annual fraud reported by the banks as an approach to collect research data. As well as there are some other approaches such as, observation and interviews. So, because research time is short, the writer will remain only using databases examination approach.

### 3.9 Software Development Approach

(Dennis et al. 2012) define software development life cycle (SDLC) as "a process of determining how information system (IS) can support business needs, designing the system, building it and delivering it to users". There are many techniques practice in SDLC. In this research, the writer apply RAD (Rapid Application Development). RAD technique develop system quicker during short period. It apply to counter the shortcoming of the traditional software development life cycle such as, waterfall, by applying particular procedures and computer tools to raise process speed and deliver outcomes on time (Dennis et al. 2012).

RAD practice can be categorized into three kinds, iterative development, system prototyping and throwaway prototyping. The system prototyping has selected with the intention of develop this system. It carry out with stages, planning then (analysis, design and implementation) simultaneously to complete a rapid version that send to the customers for their view to improve and then produce a new version. This method (repetition) will carry on till achieve customer requests and pleasures (Dennis et al 2012).

### 3.10 System Prototyping Methodology

As stated before this approach carry out by development stages, analysis, design and implementation (Ref 1). These stages will be detailed in next sections.

### 3.10.1 Planning stage

(Dennis et al. 2012) describe this stage as "The planning phase is the vital process of understanding why an information system should be constructed and determining how the project team will go about constructing it". The writer will explain the project scope which is (applying data mining technique to discover credit card fraud), assessed period for the suggested system to be finished, tasks and attributes anticipated to be provided form the suggested system and the appropriate tools and methods that applied to building prototype. Then he will register all those requirements in an appropriate manner. As well, the writer will perform the project feasibility to monitoring the project practices. Project feasibility can be explained as the following:

- **Technical feasibility**

    It shows how can the suggested system analyzed, designed, developed and set up effectively. In outline, it could be as a response for this question "Can we construct this system?"

- **Financial feasibility**

    (Dennis et al 2012) states, "This type of feasibility can also be referred to as cost-benefit analysis. It identifies the cost and benefits associated with the system, assigning possible values to it, calculating future cash flows and measuring the financial worthiness of the project". Furthermore, it offers a reply for the question "Should we construct the system?"

    Organizational feasibility

    It illustrates how the suggested system will be accepted. Nevertheless, it provides a reply to this question "will they approve it?"

### 3.10.2 Analysis stage

In this stage the gathered data will be investigated. The writer is will perform this process in the following chapter. There are several matters require engrossed realization earlier beginning design phase. These matters involves system requirements, user requirements, business tasks and attributes, as well as, the current systems require full examination to recognize their strong point and flaws. The purpose of this investigation is to achieve credit card fraud detection system using data mining technique. The anticipated system does not involve any extra configuration from the customer side.

### 3.10.3 Design stage

In this stage the writer will build the prototype exploited UML (Unified Modeling Language) and CASE (Computer Aided Software) tools. The design will contain, system architectural design, database design and user interface design. The author will use Rational Rose and Microsoft Visio in order to draw the models.

3.4.1.4 Implementation stage

In this stage all intended objects are converted into codding that realized by the computer. As well as prototypes programing and checking of the diverse tasks will be done in this stage. Some of functions are:

User registration and login.
Make transactions.
Other relevant functions.

Each dependent prototype will be verified as a component, this component will be integrated with the previous ones to be verified once more in a practice that named integration test. Full explanations will be offered in chapter 5 (Implementation).

### 3.10.4 System testing stage

In this stage the completed system need to be verified depending on functional and non-functional requirements. Functional means what the system should do (Features) while non-fictional defines the system quality (Performance, availability, integrity, etc.). Also, there is consumer acceptance test for together usual customers and managers.
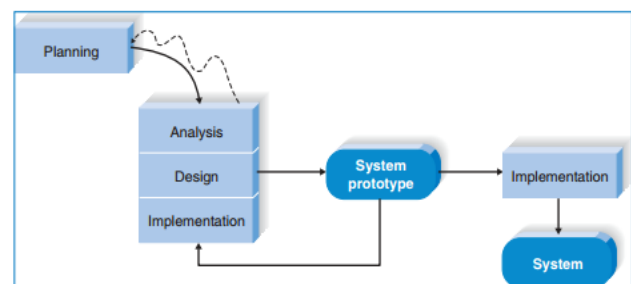


Figure1: System Prototyping Methodology (Dennis et al 2012)

## 4. Finding and System analysis

This chapter will focus on the Findings of the database examination of the majidi mall in Iraq under the study performed by the researcher, Requirement Analysis as well as functional and nonfunctional for developing system were determined in this chapter.

### 4.1 Research Findings

The writer select Baghdad, Basra and Mosul branches of Majid mall in Iraq to carry out his research. These branches use data mining technique in both visa and PayPal payment to discover fraud but they are not use this technique in credit card payment to detect fraud they are still use traditional ways. The outcomes of the investigation will be analysis and will be stated in a statistical data design.

### 4.2. Presentation and Analysis of the Result

Researcher choice three branch Baghdad, Basra and Mosul from majidi mall in Iraq to carry out his research. Through examine their databases and annual reports about their client's losses from 2011 to 2013, author has fond following results.

## 4.2.1 Losses of Baghdad branch's customers

The author start his research by examine and investigate database and annual reports of Baghdad branch he has found the total losses of money for period 2011 to 2013 in in credit card payment due to fraud is larger than another two payments as shown in table (1)

*Table 1: losses of Baghdad branch's customers*

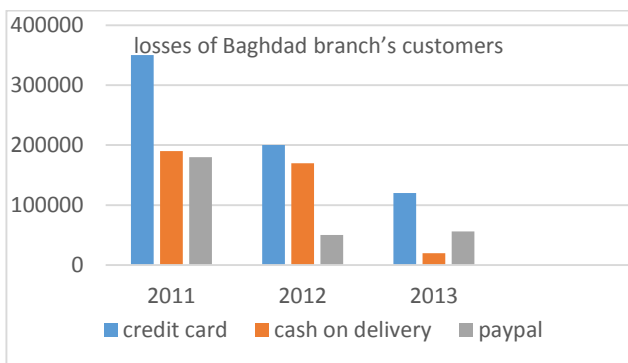| Type of payment / years | Credit card | Visa card | PayPal |
|---|---|---|---|
| 2011 | 350000 $ | 190000$ | 180000$ |
| 2012 | 200000$ | 170000$ | 50000$ |
| 2013 | 120000$ | 20000$ | 56000$ |
| total | 670000$ | 380000$ | 286000$ |



Figure 2: losses of Baghdad branch's customers

## 4.2.2. Losses of Basra branch's customers

The researcher after finish examine database and records in the Baghdad branch he has started his investigation in Basra branch and he found following facts:

*Table 2: losses of Basra branch's customers*

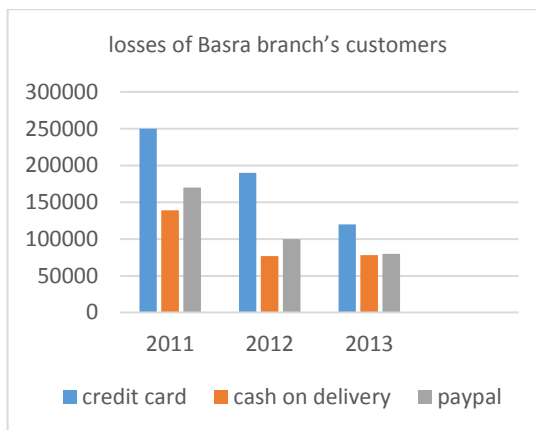| Type of payment / years | Credit card | Visa card | PayPal |
|---|---|---|---|
| 2011 | 250000 $ | 139000$ | 170000$ |
| 2012 | 190000$ | 77000$ | 100000$ |
| 2013 | 120000$ | 78000$ | 80000$ |
| total | 560000$ | 294000$ | 350000$ |



Figure 3: losses of Basra branch's customers

As shown in table (2) and figure (3) credit card losses compare to others two type of fraud is larger.

### 4.2.3. Losses of Mosul branch's customers

The facts has found by the researcher is registered in table 3.

*Table 3: Mosul branch losses*

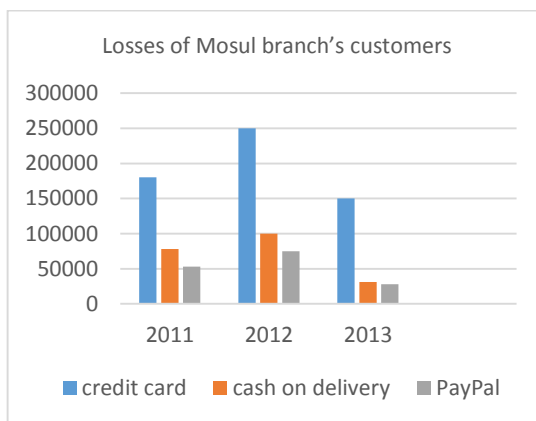| type of payment / years | Credit card | Visa card | PayPal |
|---|---|---|---|
| 2011 | 180000$ | 78000$ | 53000$ |
| 2012 | 250000$ | 100000$ | 75000$ |
| 2013 | 150000$ | 31000$ | 28000 |
| total | 580000 | 209000 | 156000 |



Figure 4: Losses of Mosul branch's customers

As shown in table (3) and figure (4) the researcher can extract the same result in the others two branches, the credit card fraud losses is bigger than others two sectors.

### 4.2.4. Branches total losses from 2011 to 2013

**IJCSI**
www.IJCSI.org

Table (4) shows the total amount of money loses by three brunches in the three sector due to the fraud

*Table 4: total losses of three branches*

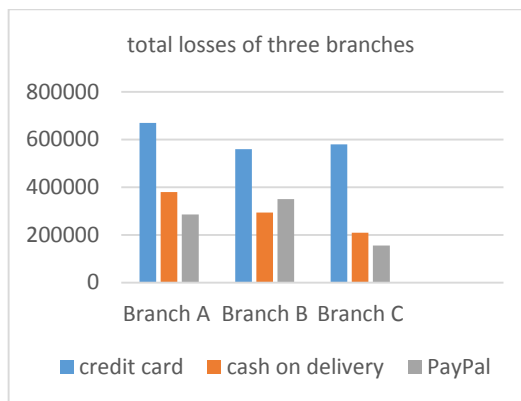| type of payment / branches | Credit card | Visa card | PayPal |
|---|---|---|---|
| Baghdad Branch | 670000 | 380000 | 286000 |
| Basra Branch | 560000 | 294000 | 350000 |
| Mosul Branch | 580000 | 209000 | 156000 |



Figure 5: total losses of three branches

As shown in table (4) and figure (5) we can realize the losses in credit card from 2011 to 2012 is very big compare to others sectors and the reason for that is because the in the visa card and PayPal payment the mall use fraud detection system depending on the data mining technique which measure the behavior of the customers, so the majidi mall need to develop system to detect fraud in credit card payments using data mining technique.

REFERENCES

1. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm. [Accessed 29 March 2013].

2. A. Srivastava, A. Kundu, S. Sural and A. K. Majumdar, "Credit Card Fraud Detection Using Hidden Markov Model," IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, vol. 5, no. 1, pp. 37-48, 2008.

3. E. Aleskerov, B. Fieisleben and R. Bharat, "CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection," Department of Electrical Engineering and Computer Science, University of Siegen, pp. 220-226, 1997.

4. Stolfo, S.J.; Fan D.W.; Lee, W.; Prodromidis, A.; and Chan, P.K. (2007). Credit card fraud detection using meta-learning: Issues and initial results. Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Risk Management, 83-90.