# Abstract

The banking sector is a very important sector in our present-day generation where almost every human has to deal with the bank either physically or online. In dealing with the banks, the customers and the banks face the chances of been trapped by fraudsters. Examples of fraud include insurance fraud, credit card fraud, accounting fraud, etc. Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data in order to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added.

# TABLE OF CONTENTS

## LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 OVERVIEW

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses.Fraud detection is applied to many industries such as banking or insurance. In banking, fraud may include forging checks or using stolen credit cards. Other forms of fraud may involve exaggerating losses or causing an accident with the sole intent for the payout. With an unlimited and rising number of ways someone can commit fraud, detection can be difficult. Activities such as reorganization, downsizing, moving to new information systems or encountering a cybersecurity breach could weaken an organization's ability to detect fraud. Techniques such as real-time monitoring for fraud is recommended. Organizations should look for fraud in financial transactions, locations, devices used, initiated sessions and authentication systems.

Fraud is typically involves multiple repeated methods, making searching for patterns a general focus for fraud detection. For example, data analysts can prevent insurance fraud by making algorithms to detect patterns and anomalies.Fraud detection can be separated by the use of statistical data analysis techniques or artificial intelligence (AI).

Statistical data analysis techniques include:

- calculating statistical parameters
- regression analysis
- probability distributions and models
- data matching

AI techniques used to detect fraud include:

- Data mining classifies, groups and segments data to search through millions of transactions to find patterns and detect fraud.

- Neural networks learn suspicious-looking patterns and use those patterns to detect them further.
- Machine learning automatically identifies characteristics found in fraud.
- Pattern recognition detects classes, clusters and patterns of suspicious behavior.

According to The American Heritage dictionary, second college edition, fraud is defined as a deception deliberately practiced in order to secure unfair unlawful gain. Fraud detection is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. Examples include insurance fraud, credit card fraud and accounting fraud. Data from the Nigeria Inter-Bank Settlement System (NIBSS) has revealed that fraudulent transactions in the banking sector at its peak. Fraud has evolved from being committed by casual fraudsters to being committed by organized crime and fraud rings that use sophisticated methods to take over control of accounts and commit fraud. Some 6.8 million Americans were victimized by card fraud in 2007, according to Javelin research. The Nilson Report estimates the cost to the industry to be $4.84 billion. Javelin estimates the losses at more than six times that amount – some $30.6 billion in 2007. Of course, fraud is not a domestic product as it's everywhere. For instance, card fraud losses cost UK economy GBP 423 million in 2006. Credit card fraud accounts for the biggest cut of the $600 million that airlines lose each year globally.

The banking sector is a very important sector in our present-day generation where almost every human has to deal with the bank either physically or online. In dealing with the banks, the customers and the banks face the chances of been trapped by fraudsters. Examples of fraud include insurance fraud, credit card fraud, accounting fraud, etc. Detection of fraudulent activity is thus critical to control these costs. This project hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data in order to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added.

# CHAPTER 2

# LITERATURE SURVEY

## 2. LITERATURE SURVEY:

Fraud detection has been usually seen as a data mining problem where the objective is to correctly classify the transactions as legitimate or fraudulent. For classification problems many performance measures are defined most of which are related with correct number of cases classified correctly.

A more appropriate measure is needed due to the inherent structure of credit card transactions. When a card is copied or stolen or lost and captured by fraudsters it is usually used until its available limit is depleted. Thus, rather than the number of correctly classified transactions, a solution which minimizes the total available limit on cards subject to fraud is more prominent.

Since the fraud detection problem has mostly been defined as a classification problem, in addition to some statistical approaches many data mining algorithms have been proposed to solve it. Among these, decision trees and artificial neural networks are the most popular ones. The study of Bolton and Hand provides a good summary of literature on fraud detection problems.

However, when the problem is approached as a classification problem with variable misclassification costs as discussed above, the classical data mining algorithms are not directly applicable; either some modifications should be made on them or new algorithms developed specifically for this purpose are needed. An alternative approach could be trying to make use of general purpose meta heuristic approaches like genetic algorithms.

**Hussein A Abdou, Ahmed A. El-Masry** in the paper on "**Neural nets versus conventional techniques in credit scoring in Egyptian banking**" found

that the number of Non-Performing Loans has increased in recent years, paralleling the current financial crisis, thus increasing the importance of credit scoring models. This study proposes a three stage hybrid Adaptive Neuro Fuzzy Inference System credit scoring model, which is based on statistical techniques and Neuro Fuzzy. The proposed model's performance was compared with conventional and commonly utilized models. The credit scoring models are tested using a 10-fold cross-validation process with the credit card data of an international bank operating in Turkey. Results demonstrate that the proposed model consistently performs better than the Linear Discriminant Analysis, Logistic Regression Analysis, and Artificial Neural Network (ANN) approaches, in terms of average correct classification rate and estimated misclassification cost. As with ANN, the proposed model has learning ability; unlike ANN, the model does not stay in a black box. In the proposed model, the interpretation of independent variables may provide valuable information for bankers and consumers, especially in the explanation of why credit applications are rejected.

**Thi Huyen Thanh Dinha Stefanie Kleimeier** in the paper on **A credit scoring model for Vietnams retail banking market.** As banking markets in developing countries are maturing, banks face competition not only from other domestic banks but also from sophisticated foreign banks. Given the substantial growth of consumer credit and increased regulatory attention to risk management, the development of a well-functioning credit assessment framework is essential. As part of such a framework, we propose a credit scoring model for Vietnamese retail loans. First, we show how to identify those borrower characteristics that should be part of a credit scoring model. Second, we illustrate how such a model can be calibrated to achieve the strategic objectives of the bank. Finally, we assess the use of credit scoring models in the context of transactional versus relationship lending.

**D.J Hand and W.E Henley** in the paper on **Statistical classification methods in consumer credit scoring.** Credit scoring is the term used to describe formal statistical methods used for classifying applicants for credit into 'good' and 'bad' risk classes. Such methods have become increasingly important with the

dramatic growth in consumer credit in recent years. A wide range of statistical methods has been applied, though the literature available to the public is limited for reasons of commercial confidentiality. Particular problems arising in the credit scoring context are examined and the statistical methods which have been applied are reviewed.

**Vijay S.Desai, Jonathan N.Crook, George A.Overstreet Jr** in the paper on **A comparison of neural networks and linear scoring models in the credit union environment** said that the purpose of the present paper is to explore the ability of neural networks such as multilayer perceptrons and modular neural networks, and traditional techniques such as linear discriminant analysis and logistic regression, in building credit scoring models in the credit union environment. Also, since funding and small sample size often preclude the use of customized credit scoring models at small credit unions, we investigate the performance of generic models and compare them with customized models. Our results indicate that customized neural networks offer a very promising avenue if the measure of performance is percentage of bad loans correctly classified. However, if the measure of performance is percentage of good and bad loans correctly classified, logistic regression models are comparable to the neural networks approach. The performance of generic models was not as good as the customized models, particularly when it came to correctly classifying bad loans. Although we found significant differences in the results for the three credit unions, our modular neural network could not accommodate these differences, indicating that more innovative architectures might be necessary for building effective generic models.

**M Vojtek, E Koèenda** in the paper on **Credit Scoring Methods. Czech Journal of Economics and Finance.** The paper reviews the best-developed and most frequently applied methods of credit scoring employed by commercial banks when evaluating loan applications. The authors concentrate on retail loans – applied research in this segment is limited, though there has been a sharp increase in the volume of loans to retail clients in recent years. Logit analysis is identified as the most frequent credit-scoring method used by banks. However, other nonparametric methods are widespread in terms of pattern recognition. The

methods reviewed have potential for application in post-transition countries.

**Lyn C.Thomas** in the paper on **A survey of credit and behavioral scoring: forecasting: financial risk of lending to customers.** Credit scoring and behavioural scoring are the techniques that help organisations decide whether or not to grant credit to consumers who apply to them. This article surveys the techniques used — both statistical and operational research based — to support these decisions. It also discusses the need to incorporate economic conditions into the scoring systems and the way the systems could change from estimating the probability of a consumer defaulting to estimating the profit a consumer will bring to the lending organisation — two of the major developments being attempted in the area. It points out how successful has been this under-researched area of forecasting financial risk. 

# CHAPTER 3
# SYSTEM DESIGN

## 3.1 SCOPE OF THE PROJECT

The main objective of this project is to help banking sector to improve its security from fraudsters. Now a days, fraudsters were finding different ways to do frauds, this project helps to minimize some of those fraud techniques.

## 3.2 OBJECTIVE

The main objective of this project is to help banking sector to improve its security from fraudsters.

## 3.3 PROBLEM DEFINATION

The banking sector is a very important sector in our present-day generation where almost every human has to deal with the bank either physically or online. In dealing with the banks, the customers and the banks face the chances of been trapped by fraudsters.

## 3.4 EXISTING SYSTEM

- In case of bank fraud detection, the existing system is detect the fraud after fraud has been happen. Existing system maintain the large amount of data when customer comes to know about inconsistency in transaction he/she made complaint and then fraud detection system start it working. It first tries to detect that fraud has actually occur after that it transactions that was used to fraud detection mechanism developed by master and visa cards
- A machine learning paradigm classification, with Bank Fraud Detection being the base.

- Intrusion detections to track fraud location and so on. In case of existing system there is no confirmation of recovery of fraud and Customer satisfaction.
- Secure electronic system used to analyze the behaviour of legitimate users.
- Data Mining mechanisms to classify and preprocess the user's data.
- Genetic algorithms.

### 3.4.1 DISADVANTAGES OF EXISTING SYSTEM

- Each payment system has its limits regarding the maximum amount in the account, the number of transactions per day and the amount of output.
- If Internet connection fails, you can not get to your online account.
- If you follow the security rules the threat is minimal. The worse situation when the system of processing company has been broken, because it leads to the leak of personal data on cards and its owners.
- The information about all the transactions, including the amount, time and recipient are stored in the database of the payment system. And it means the intelligence agency has an access to this information. Sometimes this is the path for fraudulent activities.

## 3.5 PROPOSED SYSTEM

In proposed methodology, Detection of fraudulent activity is thus critical to control these costs. This paper hereby addresses bank fraud detection via the use of machine learning techniques; association, clustering, forecasting, and classification to analyze the customer data in order to identify the patterns that can lead to frauds. Upon identification of the patterns, adding a higher level of verification/authentication to banking processes can be added. These kinds of frauds can be credit card fraud, insurance fraud, accounting fraud, etc. which may lead to the financial loss to the bank or the customers. Thus, detection of these