ABSTRACT:

In recent years credit card fraud has become one of the growing problems. It is vital that credit card companies are able to identify fraudulent credit card transactions, so that customers are not charged for the items that they didn't purchase.The reputation of companies will heavily damaged and endangered among the customers due to fraud in financial transactionsThe fraud detection techniques were increasing to improve accuracy to identify the fraudulent transactions.This project intends to build an unsupervised fraud detection method using autoencoder. An Autoencoder with four hidden layers which has been trained and tested with a dataset containing an European cardholder transactions that occurred in two days with 284,807 transactions from September 2013.

Keywords:credit card, fraud detection,Autoencoder network,unsupervised learning

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
AIS	Artificial Immune System
GA	Genetic Algorithm
HMM	Hidden Markov Model
SVM	Support Vector Machine
FNN	Fuzzy Neural Network
ILP	Inductive Logic Programming
CBR	Case-Based Reasoning
SQL	Structured Query Language
BPN	Back Propagation Network
FNNKD	Fuzzy neural network based on knowledge discovery
GNN	Granular Neural Network
SOM	Self Organizing Map
ICLN	Improved Competitive Learning Network
SICLN	Supervised Improved Competitive Learning Network
NSA	Negative Selection Algorithm
KNN	k Nearest Neighbour algorithm
AIRS	Resource-limited Artificial immune System
API	Application Programming Interface
AIN	Artificial Immune Network
AISFD	Artificial Immune System for Fraud Detection
DC	Dendritic Cells
PAMP	Pathogen Associated Molecular Pattern
DCA	Dendritic Cell Algorithm

CHAPTER 1
INTRODUCTION

1.1 Introduction:

A credit card is a thin handy plastic card that contains identification information such as a signature or picture, and authorizes the person named on it to charge purchases or services to his account - charges for which he will be billed periodically. They have a unique card number which is of utmost importance. Its security relies on the physical security of the plastic card as well as the privacy of the credit card number.

There is a rapid growth in the number of credit card transactions which has led to a substantial rise in fraudulent activities. Credit card fraud is a wide-ranging term for theft and fraud committed using a credit card as a fraudulent source of funds in a given transaction. Generally, statistical methods and many data mining algorithms are used to solve this fraud detection problem. Most of the credit card fraud detection systems are based on artificial intelligence, Meta learning and pattern matching.

Fraud detection is a binary classification problem in which the transaction data is analyzed and classified as “legitimate” or “fraudulent”. Credit card fraud detection techniques are classified in two general categories: fraud analysis(misuse detection) and user behavior analysis (anomaly detection).

1.2 Motivation for Work:

At the current state of the world, financial organizations expand the availability of financial facilities by employing innovative services such as credit cards, Automated Teller Machines (ATM), internet and mobile banking services. Besides, along with the rapid advances of e-commerce,the use of credit cards has become a convenient and necessary part of financial life. Credit card is a payment card supplied to customers as a system of payment. There are lots of advantages in using credit cards such as:

- **Ease of purchase** Credit cards can make life easier. They allow customers to purchase on credit in arbitrary time, location and amount, without carrying the cash.Provide a convenient payment method for purchases made on the internet, over the telephone, through ATMs, etc.
- **Keep customer credit history** Having a good credit history is often important in detecting loyal customers. This history is valuable not only for credit cards, but also for other financial services like loans, rental applications, or even some jobs. Lenders and issuers of credit mortgage companies, credit card companies, retail stores, and utility companies can review customer credit score and history to see how punctual and responsible customers are in paying back their debts.
- **Protection of Purchases** Credit cards may also offer customers additional protection if the purchased merchandise becomes lost, damaged, or stolen. Both the buyer‟s credit card statement and the company can confirm that the customer has bought if the original receipt is lost or stolen. In addition, some credit card companies provide insurance for large purchases.

In spite of all mentioned advantages, the problem of fraud is a serious issue in-banking services that threaten credit card transactions especially. Fraud is an intentional deception with the purpose of obtaining financial gain or causing loss by implicit or explicit trick.Fraud is a public law violation in which the fraudster gains an unlawful advantage or causes unlawful damage. The estimation of amount of damage made by fraud activities indicates that fraud costs a very considerable sum of money.Credit card fraud is increasing significantly with the development of modern technology resulting in the loss of billions of dollars worldwide each year.Statistics from the Internet Crime Complaint Center show that there has been a significant rising in reported fraud in last decade. Financial losses caused due to online fraud only in the US, was reported to be $3.4 billion in 2011.

Fraud detection involves identifying scarce fraud activities among numerous legitimate transactions as quickly as possible. Fraud detection methods are developing rapidly in order to adapt with new incoming fraudulent strategies across the world. But, development of new fraud detection techniques becomes more difficult due to the severe limitation of the ideas exchanged in fraud detection. On the other hand, fraud detection is essentially a rare event problem, which has been variously called outlier analysis, anomaly detection, exception mining, mining rare classes, mining imbalanced data etc. The number of fraudulent transactions is usually a very low fraction of the total transactions. Hence the task of detecting fraud transactions in an accurate and efficient manner is fairly difficult and challengeable.Therefore, development of efficient methods which can distinguish rare fraud activities from billions of legitimate transactions seems essential.

1.3 Problem Statement:

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

2.2 Difficulties of Credit Card Fraud Detection

Fraud detection systems are prune to several difficulties and challenges enumerated below. An effective fraud detection technique should have abilities to address these difficulties in order to achieve best performance.

- **Imbalanced data:**The credit card fraud detection data has an imbalanced nature. It means that very small percentages of all credit card transactions are fraudulent. This makes the detection of fraud transactions very difficult and imprecise.
- **Different misclassification importance:** in fraud detection tasks, different misclassification errors have different importance.Misclassification of a normal transaction as fraud is not as harmful as detecting a fraud transaction as normal. Because in the first case the mistake in classification will be identified in further investigations.
- **Overlapping data:** many transactions may be considered fraudulent, while actually they are normal (false positive) and reversely, a fraudulent transaction may also seem to be legitimate (false negative). Hence obtaining a low rate of false positives and false negatives is a key challenge of fraud detection systems[4, 5, and 6].
- **Lack of adaptability:** classification algorithms are usually faced with the problem of detecting new types of normal or fraudulent patterns. The supervised and unsupervised fraud detection systems are inefficient in detecting new patterns of normal and fraud behaviors, respectively.
- **Fraud detection cost:** The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it. For example, no revenue is obtained by stopping a fraudulent transaction of a few dollars [5, 7].
- **Lack of standard metrics:** there is no standard evaluation criterion for assessing and comparing the results of fraud detection systems.

2.3 Credit Card Fraud Detection Techniques

The credit card fraud detection techniques are classified in two general categories: fraud analysis (misuse detection) and user behavior analysis (anomaly detection). The first group of techniques deals with supervised classification tasks at the transaction level. In these methods, transactions are labeled as fraudulent or normal based on previous historical data. This dataset is then used to create classification models which can predict the state (normal or fraud) of new records. There are numerous model creation methods for a typical two class classification task such as rule induction [1], decision trees [2] and neural networks [3].This approach is proven to reliably detect most fraud tricks which have been observed before [4], also known as misuse detection.

The second approach deals with unsupervised methodologies which are based on account behavior. In this method a transaction is detected as fraudulent if it is in contrast with

the user's normal behavior. This is because we don‟t expect fraudsters behave the same as the account owner or be aware of the behavior model of the owner [5].To this aim, we need to extract the legitimate user behavioral model (e.. user profile)for each account and then detect fraudulent activities according to it. Comparing New behaviors with this model, different enough activities are distinguished as frauds. The profiles may contain the activity information of the account; such as merchant types, amount, location and time of transactions, [6].This method is also known as anomaly detection.

It is important to highlight the key differences between user behavior analysis and fraud analysis approaches.The Fraud analysis method can detect known fraud tricks, with a low false positive rate.These systems extract the signature and model of fraud tricks presented in oracle dataset and can then easily determine exactly which frauds, the system is currently experiencing. If the test data does not contain any fraud signatures, no alarm is raised. Thus, the false positive rate can be reduced extremely.However,since learning of a fraud analysis system (i.e. classifier) is based on limited and specific fraud records,It cannot detect novel frauds. As a result, the false negative rate may be extremely high depending on how ingenious the fraudsters.User behavior analysis, on the other hand, greatly addresses the problem of detecting novel frauds. These Methods do not search for specific fraud patterns, but rather compare incoming activities with the constructed model of legitimate user behavior. Any activity that is sufficiently different from the model will be considered as a possible fraud. Though user behavior analysis approaches are powerful in detecting innovative frauds, they really suffer from high rates of false alarm. Moreover, if a fraud occurs during the training phase, this fraudulent behavior will be entered in baseline mode and is assumed to be normal in further analysis[7].In this section we will briefly introduce some current fraud detection techniques which are applied to credit card fraud detection tasks.

2.3.1 Artificial Neural Network

An artificial neural network (ANN) is a set of interconnected nodes designed to imitate the functioning of the human brain [9]. Each node has a weighted connection to several other nodes in adjacent layers. Individual nodes take the input received from connected nodes and use the weights 6 together with a simple function to compute output values. Neural networks come in many shapes and architectures.The Neural network architecture, including the number of hidden layers, the number of nodes within a specific hidden layer and their connectivity, must be specified by the user based on the complexity of the problem. ANNs can be configured by supervised, unsupervised or hybrid learning methods.

2.3.1.1 Supervised techniques

In supervised learning, samples of both fraudulent and non-fraudulent records, associated with their labels are used to create models. These techniques are often used in fraud analysis approach.One of the most popular supervised neural networks is back propagation network (BPN). It minimizes the objective function using a multi-stage dynamic optimization methodthat is a generalization of the delta rule.The back propagation method is