**CREDIT CARD FRAUD DETECTION USING PREDICTIVE MODELLING**

**Index:**

* **Abstract**
* **Introduction**
* **Problem Statement**
* **Existing system**
* **Disadvantages**
* **Proposed System**
* **Advantages**
* **SRS**

**Abstract:-**

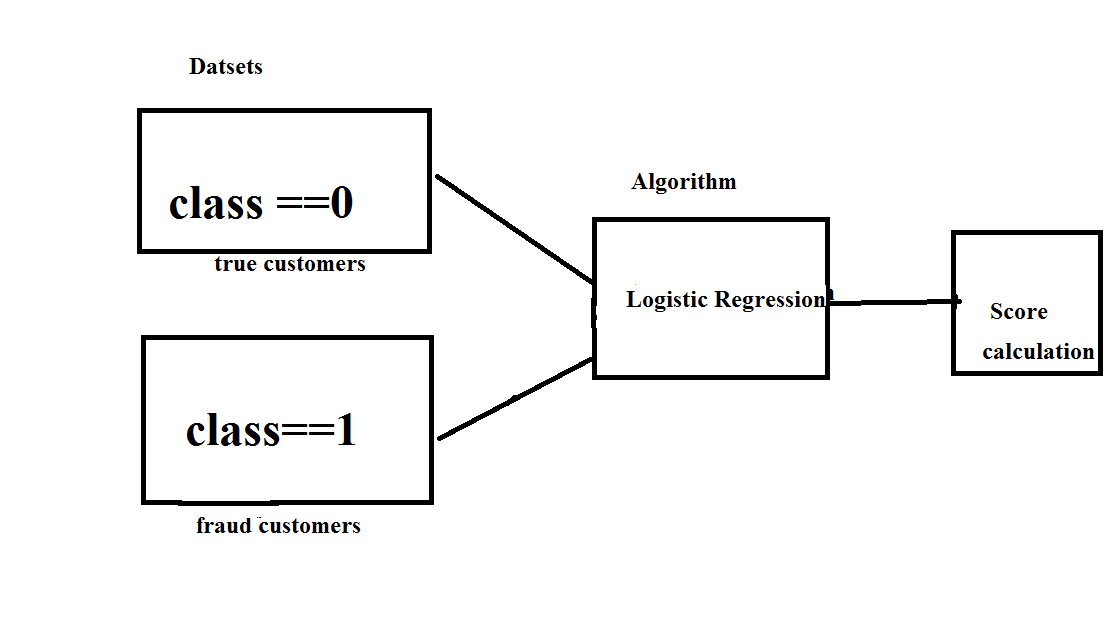
Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators. The design of fraud detection algorithms is however particularly challenging due to the non-stationary distribution of the data, the highly unbalanced classes distributions and the availability of few transactions labeled by fraud investigators. At the same time public data are scarcely available for confidentiality issues, leaving unanswered many questions about what is the best strategy. In this thesis we aim to provide some answers by focusing on crucial issues such as: i) why and how under sampling is useful in the presence of class imbalance (i.e. frauds are a small percentage of the transactions), ii) how to deal with unbalanced and evolving data streams (non-stationary due to fraud evolution and change of spending behavior), iii) how to assess performances in a way which is relevant for detection and iv) how to use feedbacks provided by investigators on the fraud alerts generated. Finally, we design and assess a prototype of a Fraud Detection System able to meet real-world working conditions and that is able to integrate investigators’ feedback to generate accurate alerts.

**Index Terms**: credit card, fraud detection, online shopping, e-commerce, logistic regression.

**Introduction:**

The online shopping growing day to day. Credit cards are used for purchasing goods and services with the help of virtual card and physical card where as virtual card for online transaction and physical card for offline transaction. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.). In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyse the spending patterns on every card and to figure out any inconsistency with respect to the “usual” spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds. Since humans tend to exhibit specific behavioristic profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is a potential threat to the system.

**Design:**



**Problem statement:**

Credit card fraud stands as major problem for word wide financial institutions. Annual lost due to it scales to billions of dollars. We can observe this from many financial reports. Such as (Bhattacharyya et al., 2011) 10th annual online fraud report by Cyber Source shows that estimated loss due to online fraud is $4 billion for 2008 which is 11% increase than $3.6 billion loss in 2007and in 2006, fraud in United Kingdom alone was estimated to be £535 million in 2007 and now costing around 13.9 billion a year (Mahdi et al., 2010). From 2006 to 2008, UK alone has lost £427.0 million to £609.90 million due to credit and debit card fraud (Woolsey &Schulz, 2011). Although, there is some decrease in such losses after implementation of detection and prevention systems by government and bank, card-not-present fraud losses are increasing at higher rate due to online transactions. Worst thing is it is still increasing un-protective and un-detective way.

Over the year, government and banks have implemented some steps to subdue these frauds but along with the evolution of fraud detection and control methods, perpetrators are also evolving their methods and practices to avoid detection. Thus an effective and innovative methods need to be develop which will evolve accordingly to the need.

**Existing system:**

This was on k-means Algorithm implementation, Only the two features with the most variance were used to train the model. The model was set to have 2 clusters, 0 being non-fraud and 1 being fraud. We also experimented with different values for the hyper parameters, but they all produced similar results. Changing the dimensionality of the data (reducing it to more dimensions than 2) also made little difference on the final values.

**Disadvantages:**

The Clustering doesn’t produce the less accuracy when compared to Regression methods in scenarios like credit card fraud detection. Comparatively with other algorithms k-means produce less accurate scores in prediction in this kind of scenarios

**Proposed System:**

Our goal is to implement machine learning model in order to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports.

Logistic regression, as it was a good candidate for binary classification. Python sklearn library was used to implement the project, We used Kaggle datasets for Credit card fraud detection, using pandas to data frame for class ==0 forno fraud and class==1 for fraud, matplotlib for plotting the fraud and non fraud data, train\_test\_split for data extraction (Split arrays or matrices into random train and test subsets) and used Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the algorithm. Finally Confusion matrix was plotted on true and predicted.

**Advantages:**

* The result obtained by the Logistic Regression Algorithm is best compared to any other Algorithms.
* The Accuracy obtained was almost equal to cent percent which proves using of Logistic algorithm gives best results.
* The plots that were plotted according to the proper data that is processed during the implementation

**Hardware Requirements:**

• RAM : 4GB and Higher

• Processor : Intel i3 and above

• Hard Disk : 500GB: Minimum

**Software Requirements:**

• OS : Windows

• Python IDE : python 2.7.x and above

• Pycharm IDE

• Setup tools and pip to be installed for 3.6.x and above