Table of Contents:-

Abstract

➢ Due to cashless transaction every people use ATM card and credit card for transaction, so fraud can also be increase.

➢ Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms rely on advanced machine learning techniques to assist fraud investigators.

Introduction

➢ The online shopping growing day to day. Credit card are used for purchasing good and services with the help of virtual card and physical card for offline transaction.

➢ In physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. If the cardholder does not realize the loss of card, it can lead to 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.)
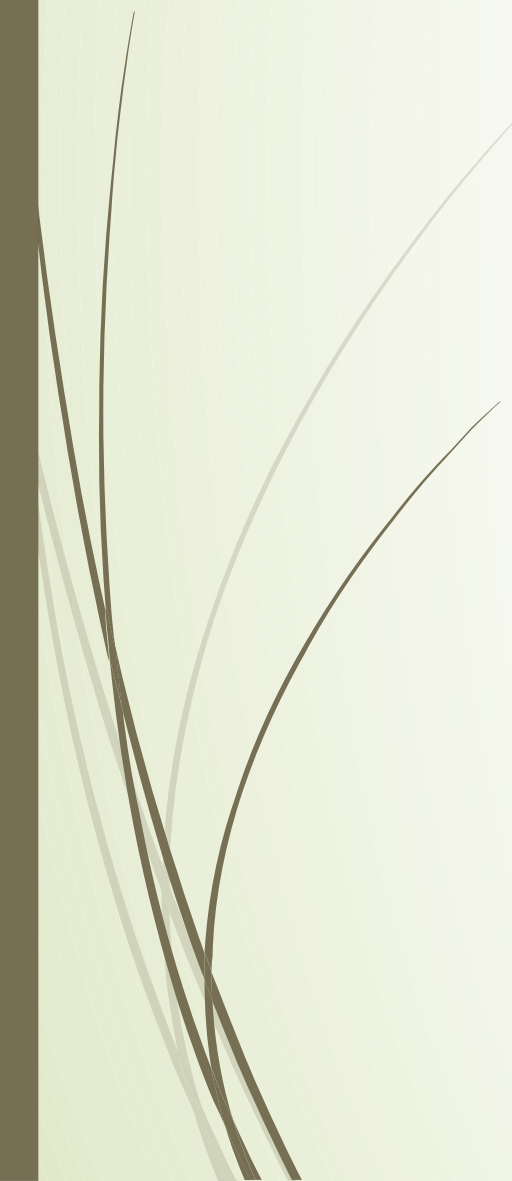
Problem statement

➢ Credit card fraud stands as major problem for world wide financial institutions. Annual lost due to it scales to billions of dollars.

➢ We can observe this from many financial reports. Such as (Bhattacharya 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 2007 and in 2006, fraud in United Kingdom alone was to be £535 million in 2007 and now costing around 13.9 billion a year (Mahdi et al., 2010).

Fraud Detection

- Fraud detection is a topic which is applicable to many industries including banking and financial sectors, insurances, government agencies, and low enforcement and more. Through the use of sophisticated use of data mining tools, millions of transaction can be searched to spot patterns and detect fraudulent transaction.

- Fraud detection is process of identifying fraudulent transaction

- Credit card fraud detection technic used to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchases.
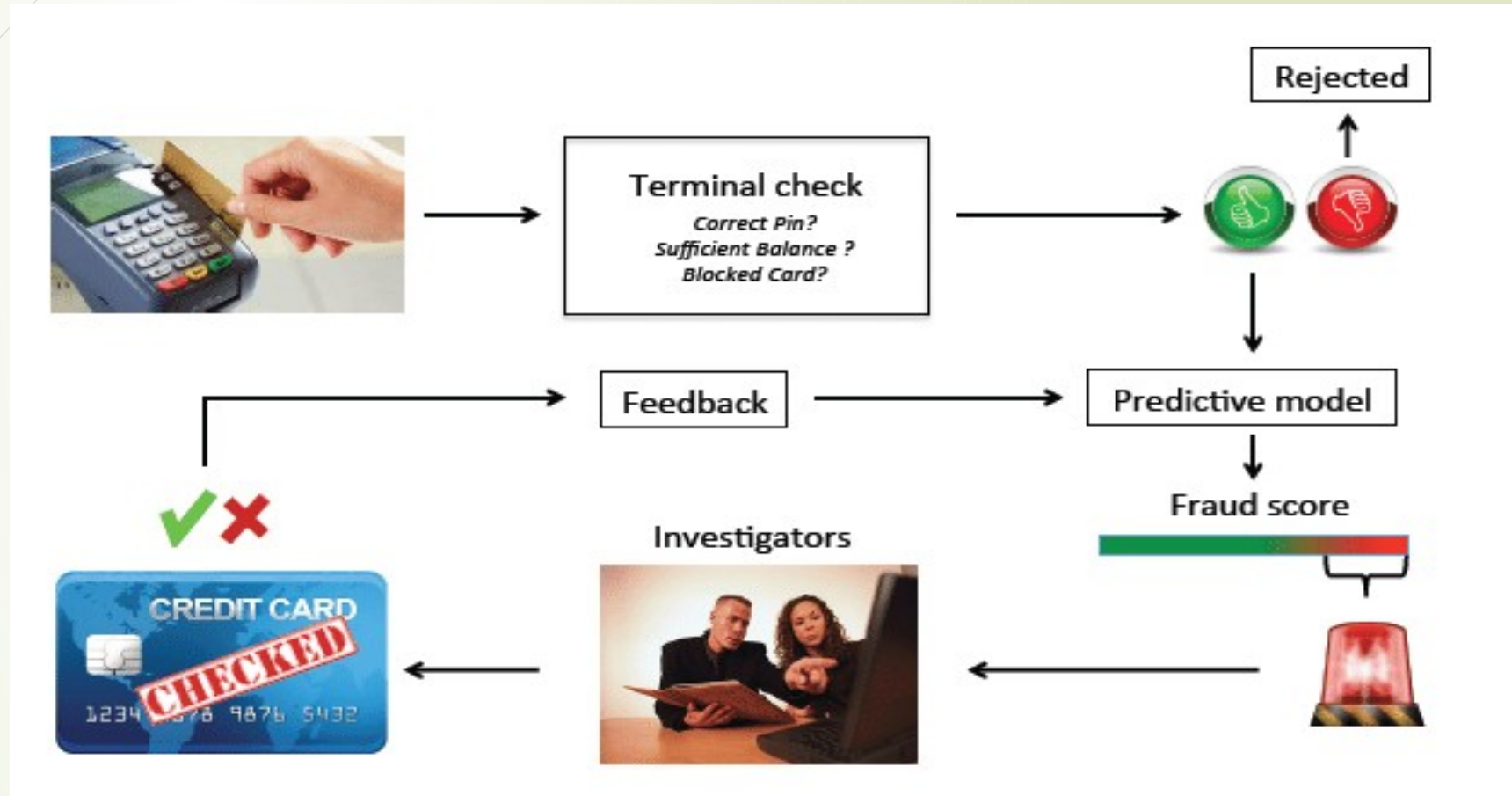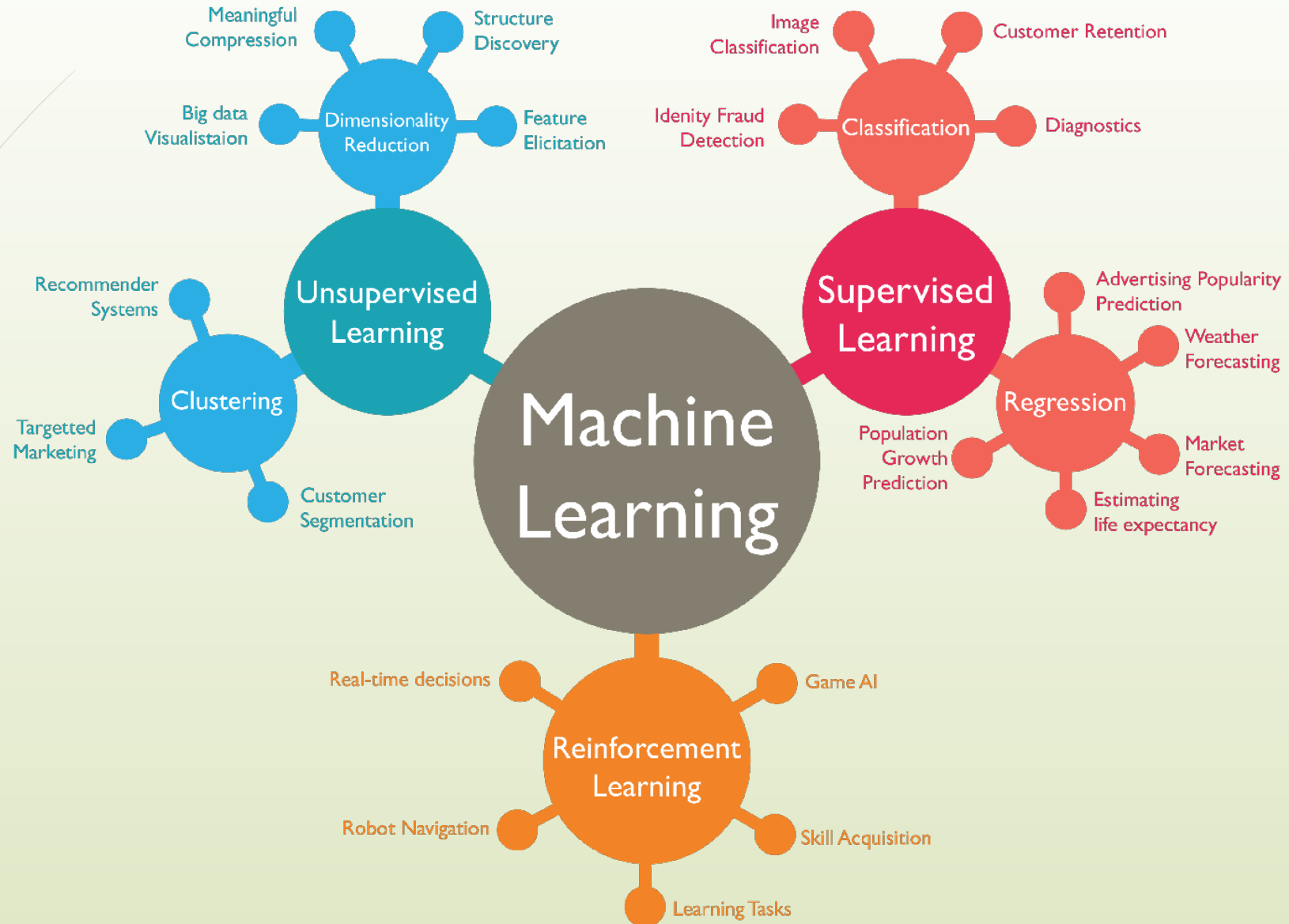
Fraud detection process

Fig. Fraud detection process

Technologies used

Machine Learning

➢ Machine learning is the scientific study of algorithms and static models that computer system use in order to perform a specific task effectively without using the explicit instruction, relaying on patterns and interface instead.

➢ Machine learning algorithms build a mathematical model based on sample data, known as training data in order to make prediction and decisions.

➢ Machine learning algorithms are used in email filtering, face recognition, etc.
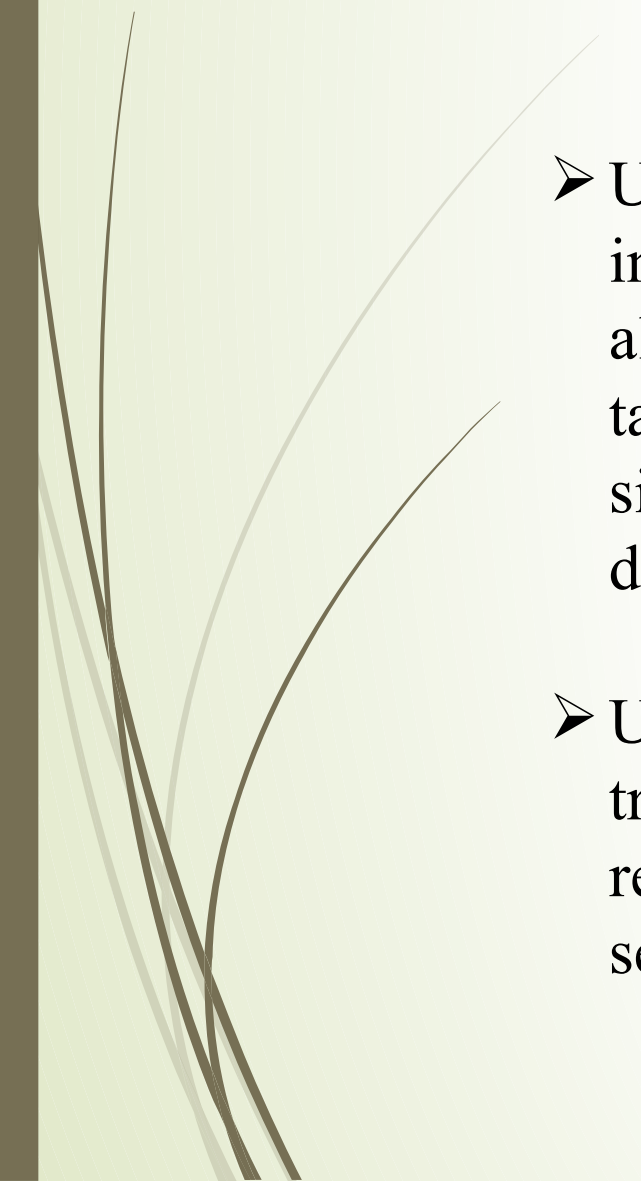
<u>Supervised Learning</u>

➢ Supervised learning as the name indicates the presence of a supervisor as a teacher.

➢ Supervised learning algorithm build a mathematical model of a set of data that contain both input and output.

➢ Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer.

➢ After that, the machine is provided with a new set of examples (data) so that supervised learning algorithm analyses a correct outcome from labeled data.
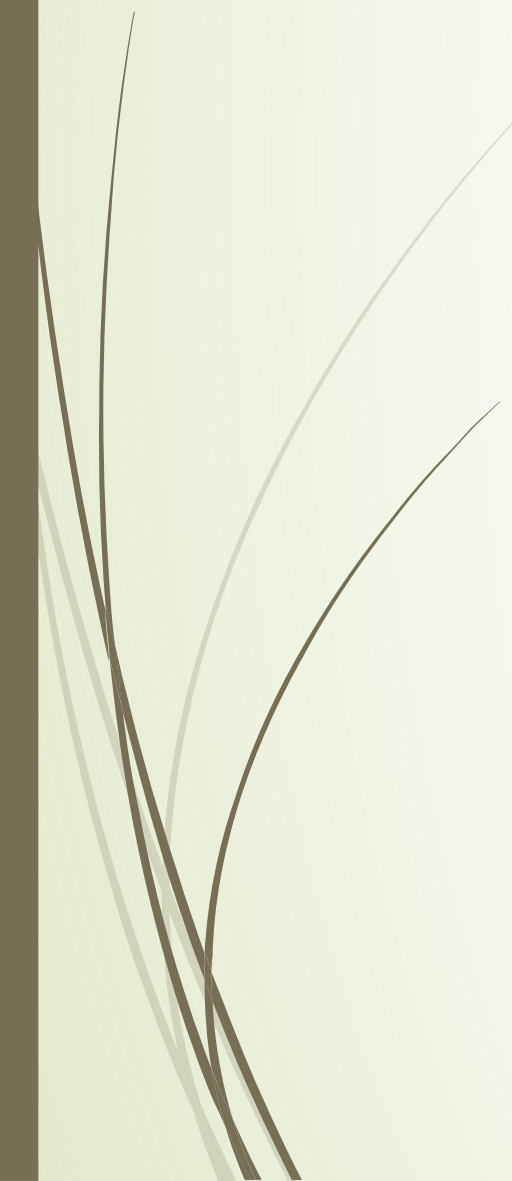
Unsupervised learning :

➢ Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

➢ Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore machine is restricted to find the hidden structure in unlabeled data by our-self.

Reinforcement Learning :

➢ Reinforcement learning is an area of Machine Learning.

➢ It is about taking suitable action to maximize reward in a particular situation.

➢ It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation.

➢ Reinforcement learning differs from the supervised learning in a way there is no answer key but the reinforcement agent decides what to do perform.
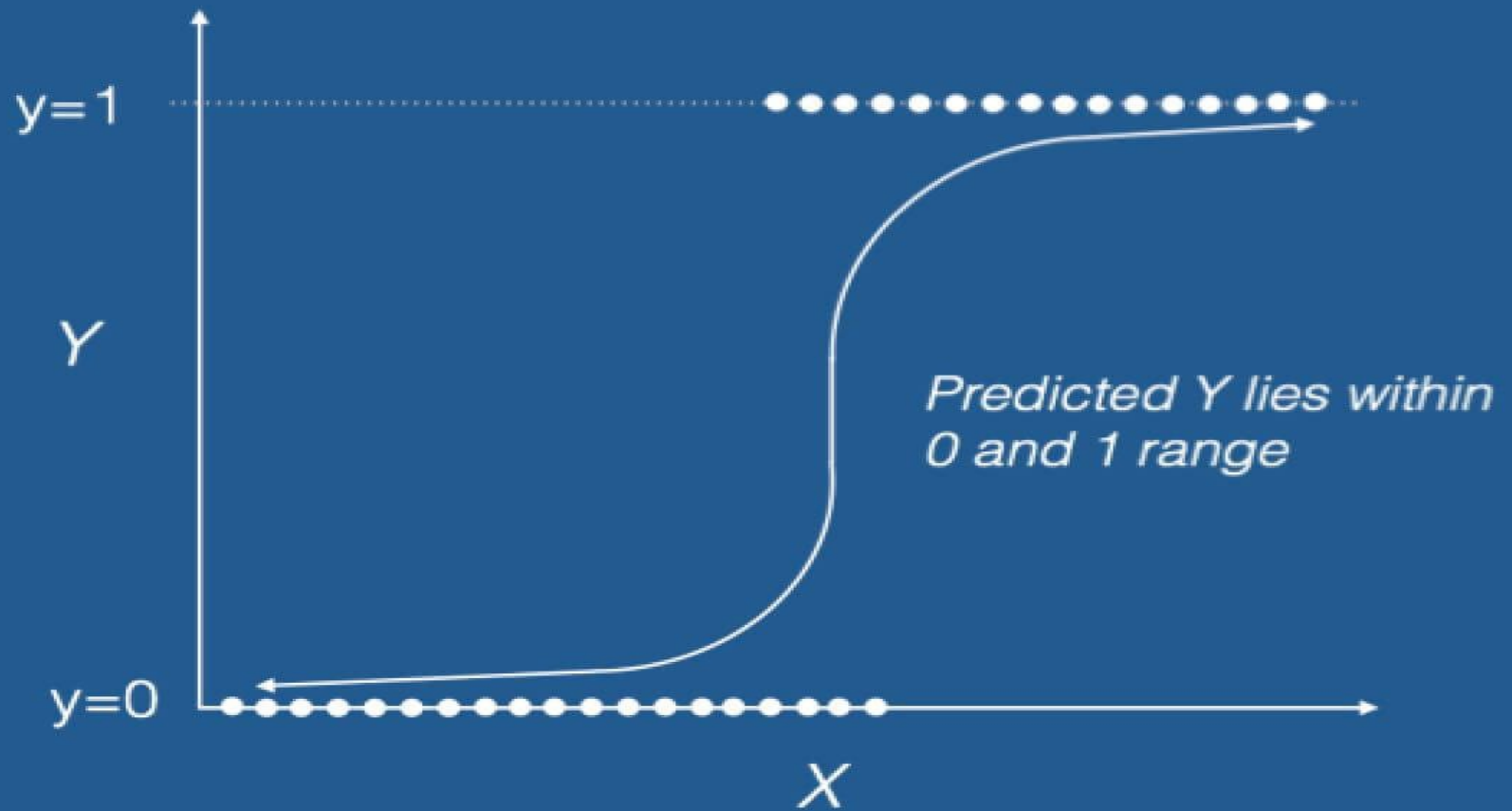
Algorithm used:

Logistic regression: It is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extension exist. In regression analysis, logistic regression[1] (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values.

Modules

1. Frame the problem
2. Collect the raw data
   - Download data from Kaggle
3. Import the libraries
   - Pandas, numpy, sklearn etc.
4. Process the data for analysis.
   - Perform the logistic regression on data
   - Explore the data.
5. Perform in-depth analysis.
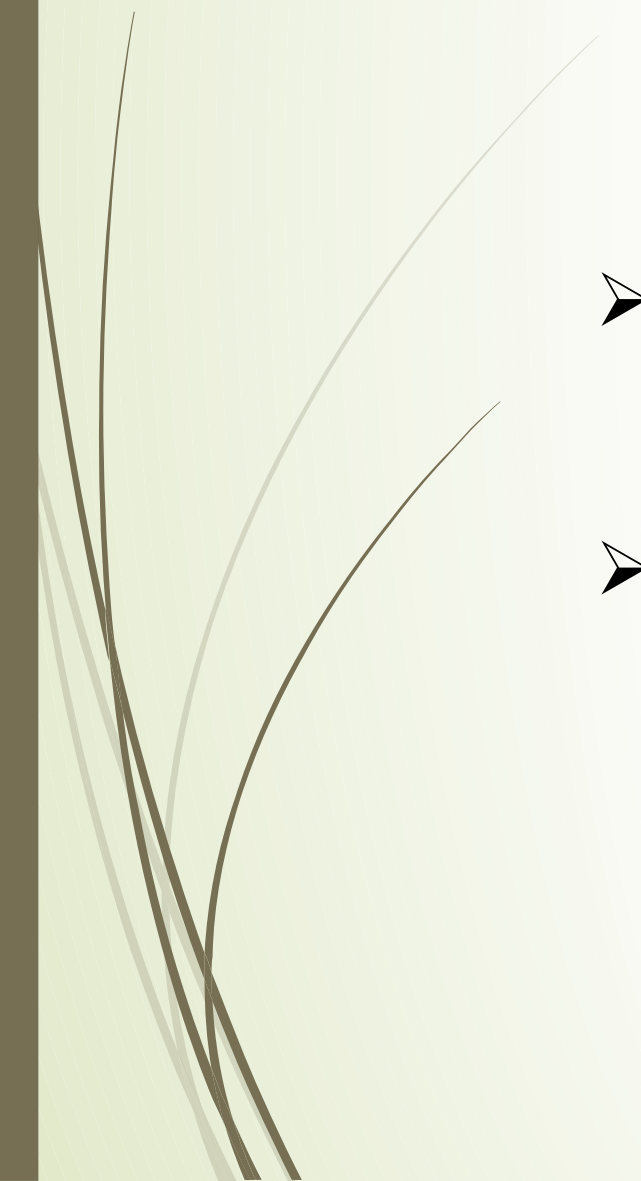6. Communicate results of the analysis.

Dataset for credit card fraud detection

- The Dataset that is used for credit card fraud detection is derived from the Kaggle.

- Kaggle is the online community of data scientist and Machine learners.

- Kaggle allows users to find and publish data sets, explore and build models in a web-based data science project and solve challenges.

Advantages

- ➢ The results 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.

System Requirements

Hardware Requirements:

➢ RAM : 4GB and Higher
➢ Processor : Intel i3 and above
➢ Hard Disk : 500GB Minimum

Software Requirements:

➢ OS: Windows or Linux
➢ Python IDE : python 2.7.x and above
➢ Jupyter Notebook
➢ Setup tools and pip to be installed for 3.6 and above
➢ Language : Python

Conclusion and Future Scope

- Fraud Detection system have become essential for banks and financial institution, to minimize their losses.

- However, there is a lack of published literature on credit card fraud detection techniques, due to the unavailable credit card transaction dataset for researches.

- We designed a system to detect fraud in credit card transaction. This system is capable of providing most of the essential features required to detect fraudulent and legitimate transactions.

- The dataset available on the day to day processing may become outdated, it is necessary to have updated data for effective fraud behavior identification.

References

➢ Aggarwal, Charu c. "Outlier analysis." Data mining. Springer International Publishing, 2015.

➢ Salazar, Addisson, et al. "Automatic credit card fraud detection based on non linear signal processing." Securit Technology(ICCST), 2012 IEE.

➢ Aurelien Geron's, " Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow". Shroff/O'Reilly, 2019.

➢ Gupta, S.C. "Fundamental of Statistics". Himalya Publishing Housing, 2018.

Thank you