Detection of Fraudulent Bank Transactions Using Data Mining Techniques

Parkhi Mohan (S20160010061), Sree Pragna Vinnakoti (S20160010106)

{parkhi.m16, sreepragna.v16}@iiits.in

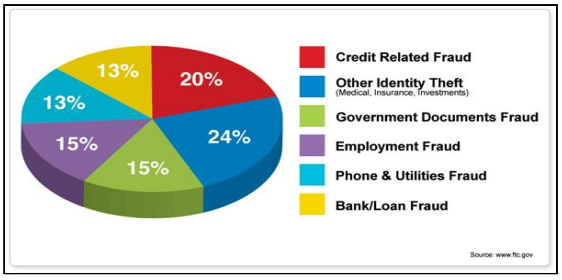
Indian Institute of Information Technology, Sri City

***Abstract*—Fraud is any dishonest act or behavior by which one person gains advantage over the other. Bank fraud is the criminal act of using illegal means to obtain money, assets or other property held by a financial institute. Bank frauds constitute a considerable percentage of white collar offences being probed by the police. Unlike ordinary thefts or robberies, the amount misappropriated in this crimes runs into crores of rupees and is a federal crime. The most important responsibility that a bank has is to protect the integrity of the institution by ensuring that it protects the financial assets it holds. To do so, the bank must be certain to address and resolve fraudulent bank transactions. We attempt to detect the same by implementing various data mining techniques and comparing the results.**

Keywords—fraudulent bank transactions, data mining, decision tree, gaussian mixture model, k-nearest neighbours, logistic regression, naive bayes, neural network, support vector machine.

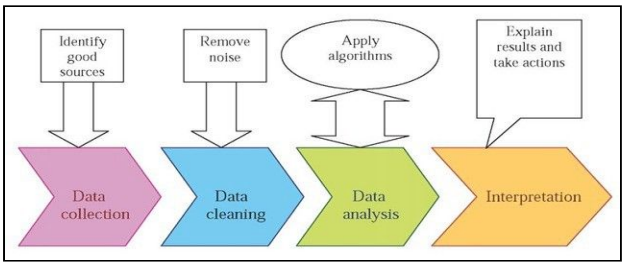
I. Introduction

The banking sector is very important in our present day and age as every human being requires and deals with money. While dealing with banks, many customers have faced the misfortune of being trapped by fraudsters. Some examples of these frauds include accounting frauds, credit card frauds, frauds related to insurance, phishing and e-fraud, etc.



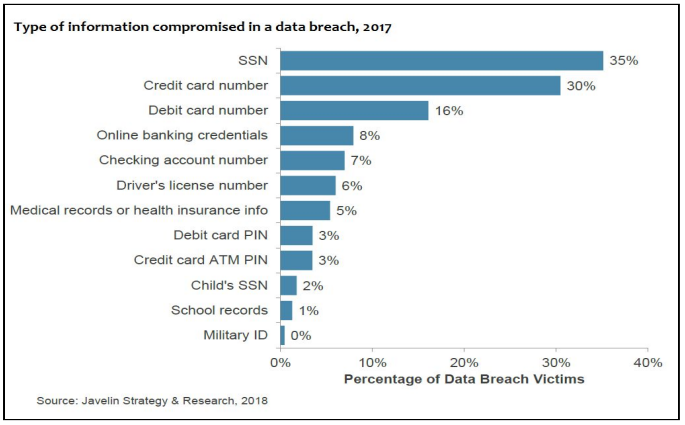
In recent years these banking frauds have been on the rise and present a significant cost to our economy. Detection of such fraudulent activities is of paramount importance so as to maintain integrity. To do this, banks are implementing data mining techniques such that this fraud can be controlled.

Data mining is the process of analysis of data to extract useful information and transform it into a comprehensible structure for further use. The process helps discover patterns in large datasets. It allows users to analyse the data from various dimensions and angles, categorise it and effectively summarise relationships identified. It is the process of finding correlations or associations among a multitude of fields in large relational databases.



Data Mining

For security and fraud detection the banking sector implements data mining techniques where big secondary data like transaction records are monitored and analysed to enhance banking security and distinguish the unusual behavior and patterns indicating fraud, phishing or laundering.



In our project, we have used two different datasets for the implementation of various data mining techniques. The first dataset is the australian dataset consists of 690 rows and 15 columns consisting of PCA performed unskewed data. The second dataset is the credit card fraudulent transaction dataset consisting of 31 columns with 28 PCA performed features and 284807 entries in rows of skewed data. On both datasets we have implemented multiple clustering and classification data mining techniques and recorded the results. The techniques applied are:

- Decision Tree
- Gaussian Mixture Model
- K-Nearest Neighbours
- Logistic regression
- Naive Bayes
- Artificial Neural Network
- Support Vector Machine

It must be noted that though big banking data consists of large volumes of unstructured data the access is limited due

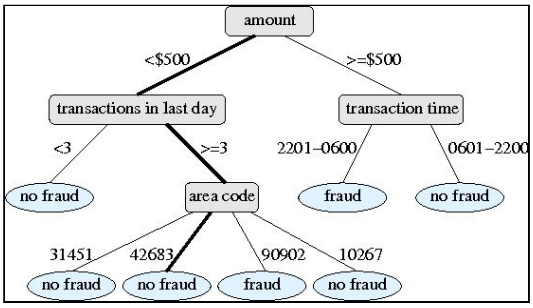
to confidentiality restrictions. After implementing above stated methods, we have compared the results.

II. Data Mining Techniques Used

A. Decision Tree

Decision tree algorithms are a method for approaching discrete-valued target functions that use inductive learning. The tree structure contains a root node, branches and leaf nodes. The upper most node in the tree is the root node. Every internal node indicates a test on attributes and every branch indicates the outcome of the test. Each leaf node indicates the class tag.

The tree organises events by sorting them down the tree from the root to a leaf node that specifies the classification class of that instance. Each node in the tree specifies a test of an attribute of that instance and each branch descending from that node links to the possible values for that attribute [1]. Decision trees are formed by a collection of rules based on variables in the model training dataset. A rule is selected and the node is split into two. This recursive function is performed for all child nodes. The stopping criteria occurs when no further gain can be made or some pre-set defined rules are met. Each branch ends in a terminal node that is defined by a unique set of rules [2].



Decision Tree Example

A decision tree is a tree like graph representing the relationships between the set of variables. To solve decision tree, classification and prediction are used and instances are classified into one of two types: positive or negative. The model is based on a top down approach and involves two phases:

1. Tree building: Starts from the root node that represents a feature of the class that needs to be classified.
2. Tree pruning - Further divided into:
    a. Pre-pruning: Halt the construction in the early stage.
    b. Post-pruning: Remove branches from the fully grown tree.

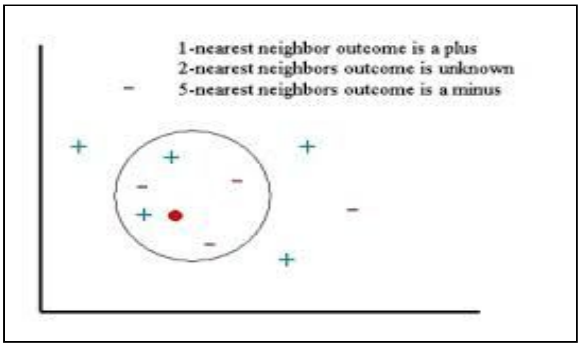
The pruning process is done not only to produce a smaller tree but also for better generalisation. The process involves identifying and removing the branches that contain the largest estimated error rate. The main purpose is to reduce the decision tree complexity and improve accuracy of the model [3].

The advantage of using a decision tree is that it is easy to understand and display. It can also handle non-linear and interactive effects of input variables. However, a disadvantage of this system is that each transaction needs to be checked one by one and the splitting criteria is difficult to choose [4].

Using a decision tree we can even trace the mail and IP address through which the credit card transaction was made based on the location where the card was used previously and currently [5].

B. K-Nearest Neighbours Classification

K-Nearest neighbours algorithm is a non-parametric in which the input consists of the k closest training examples in the feature space. In k-NN classification the output is the class membership. An object id classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours. k-NN is a type of instance-based learning or lazy learning where the function is only approximated locally and all computation is deferred until classification.



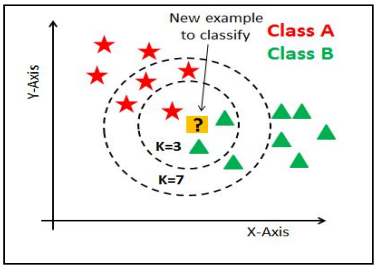
KNN Example

Another useful technique implemented is to assign weight to the contributions of the neighbours, so that the nearer neighbours contribute more to the average more than the distant ones. A peculiarity of the k-NN algorithm is that it is sensitive to the local structure of the data.

The performance of k-NN is influenced by three main factors:

1. The distance metric used to locate the nearest neighbours.
2. The distance rule used to derive a classification from k-nearest neighbours.
3. The number of neighbours used to classify the new sample.

Among the various methods of supervised statistical pattern recognition, the k-nearest neighbours rule achieves consistently high performance without a priori assumptions about the distribution from which the training examples are drawn. Larger k values can help reduce the effect of a noisy dataset [6].



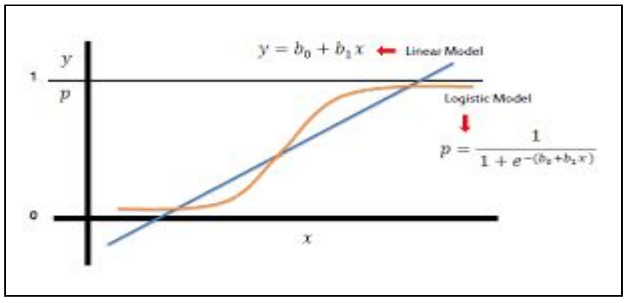
New example belongs to Class B

An advantage of KNN is that it does not require establishing any predictive model before classification. However, disadvantages include that accuracy depends highly on the measure of distance and cannot be implemented in real time [7].

kNN has been used in various anomaly detection techniques [6]. In kNN we classify any incoming transaction by calculating the majority of 3-nearest points to the new incoming transaction. If fraudulent then the transaction is classified as fraudulent otherwise legal.

C. Logistic Regression

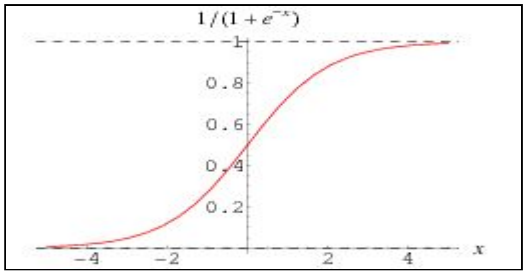
Logistic regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more independent variables. Instead on fitting data in a straight line, logistic regression uses a logistic curve [8].



Logistic regression graph

The interpret curve is using the natural logarithm of the odd of the target variable to predict an outcome that has one of two values: 0 or 1 corresponding to no or yes and false or true. It is used to predict binary valued targets [9].

The logistic function is a sigmoid function that takes any real value between zero and one. The decision boundary helps to differentiate probabilities into the positive or negative class.



Sigmoid function

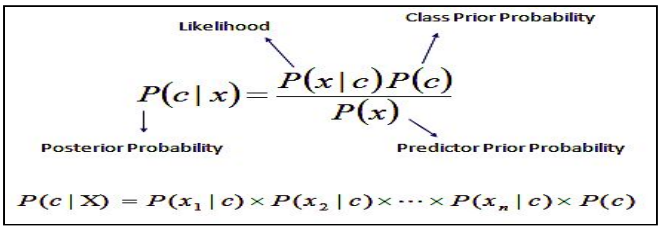
The advantage is that it produces a simple probability formula for classification and works really well with linear data. However, the disadvantages are that it cannot deal with non-linear data and is not capable of real time computations [7].

Logistic regression is widely used in credit scoring and card fraud detection as it is based on an estimation algorithm that requires making less assumptions [10].

D. Naive Bayes

Naive bayes is a simple probabilistic classifier based on applying Bayes' theorem with strong naive independent assumptions. The classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature for the given class variable [2].

The naive bayes classifier considers each feature to contribute independently, irrespective of other features.



Naive Bayes Classifier

The naive bayes classifier makes a conditional independence assumption that the effect of an attribute of a given class is independent of another attributes and is based on bayes theorem [7]. It assumes an underlying probabilistic model and allows us to capture uncertainty about the model in a principles way by determining probabilities of the outcome. It calculates explicit probabilities for hypothesis and is robust to noise in the input data [11].

This model aggregates information using conditional probability with the assumption of independence among features. It is based on finding features describing the probability of belonging to a class given features.

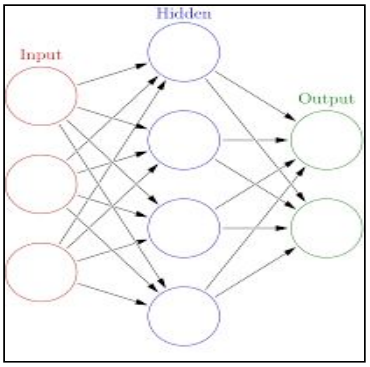
An advantage of naive bayes is that it only requires a small amount of training data to estimate the parameters (mean and variance) necessary for classification. Also, only provides the theoretical justification to the fact but does not use bayes theorem. However, the disadvantage is that in real practice, dependencies exist between the variable [2, 7].

Hence even with less input train data, naive bayes can be used to detect the new transaction as fraudulent or legal.

E. Artificial Neural Network

The artificial neural network is a set of interconnected nodes designed to imitate the functioning of the human brain. Each node has a weighted connection to several other nodes in adjacent layers.

Individual nodes take the inputs received from the connected nodes and use the weights together with a simple function to compute output values [5].

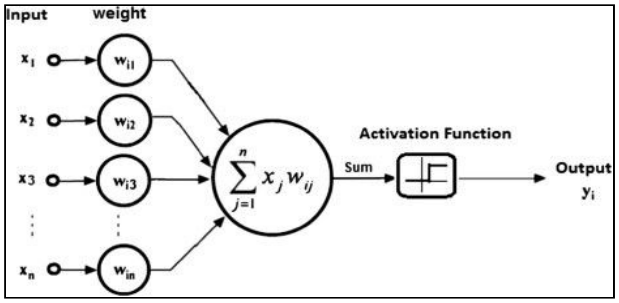


Artificial Neural Network Architecture

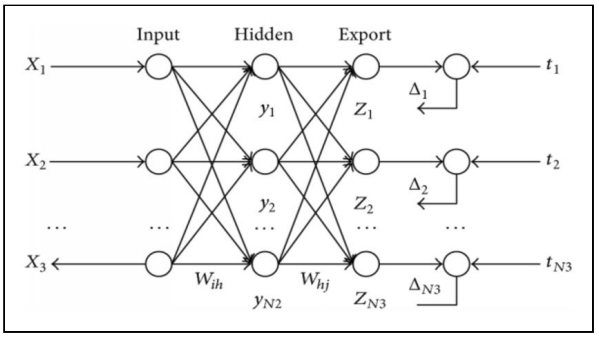
A multilayer perceptron is a feed-forward neural network that consists of at least three layers:

- Input layer: This layer has all input nodes.
- Hidden layers: Form part of the neural network operation
- Output layer: After analysis the output nodes give the output value

Except for input nodes, each node is a neuron that uses a non-linear activation function. MLP utilises a supervised technique called backpropagation for training. It can even distinguish data that is not linearly separable.



Multilayer Perceptron Activation Function



Backpropagation

Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result.

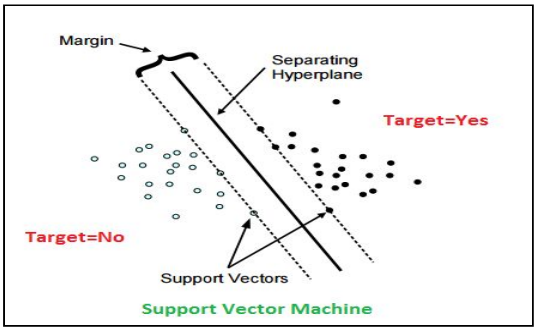
Its advantages include the ability to learn from the past without a need to reprogram. It can extract rules and predict future activities based on current situation and also it can detect fraudulent transaction at the time when transaction is in progress [2]. The model is portable and has high speed and accuracy thus can be used in real time systems. However, the disadvantage is that the model is very expensive, difficult to setup and operate and is sensitive to data format. All non-numerical data must be converted and normalised before it can be used and the number of parameters need to be set before training without any clear rules to set up these parameters.

Using neural network on fraud detection is like the human brain working principle is applied. As each user follows a particular pattern of card usage, by using this data to train the neural network, the model can classify whether the current transaction is fraudulent or legal.

F. Support Vector Machine

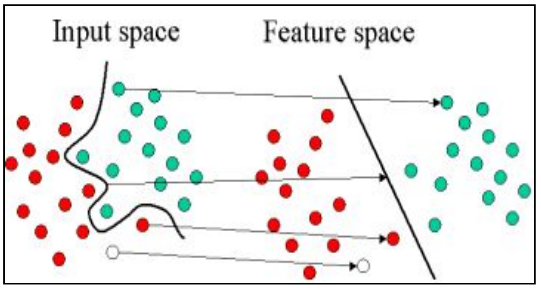
Support vector machine (SVM) is a supervised learning model with associated learning algorithms that can analyze and recognize patterns for classification and regression tasks. SVM is a binary classifier [5].

Support vector machines are based on the conception of decision planes which define decision boundaries. A decision plane separates set of different classes. The algorithm tends to construct a hyperplane as the decision plane which separates the samples into the two classes—positive and negative. This algorithm finds a special kind of linear model, the maximum margin hyperplane, and classifies all training instances correctly by separating them into correct classes through the hyperplane. The maximum margin hyperplane is the one that gives the greatest separation between the classes. The instances that are nearest to the maximum margin hyperplane are called support vectors [12].



Introducing the kernel functions, the idea was extended for linearly inseparable data. A kernel function represents the dot product of projections of two data points in a high dimensional space. It is a transform that disperses data by mapping from the input space to a new space (feature space) in which the instances are more likely to be linearly separable.

In classification tasks, given a set of training instances marked with the label of the associated class, the SVM training algorithm find a hyperplane that can assign new incoming instances into one of two classes. The class prediction of each new data point is based on which side of the hyperplane it falls on feature space [5]. There is always at least one support vector for each class, and often there are more.



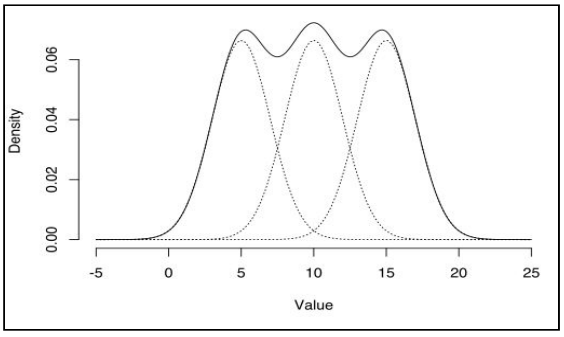
Feature mapping

The advantage of SVM is that it is robust, even when the training sample has some bias and possesses high accuracy. However, it's disadvantages include poor processing of large dataset, low speed of detection [12].

In credit card fraud detection, for each test instance, SVM determines if the test instance falls within the learned region. If yes, it is declared as normal; else anomalous.

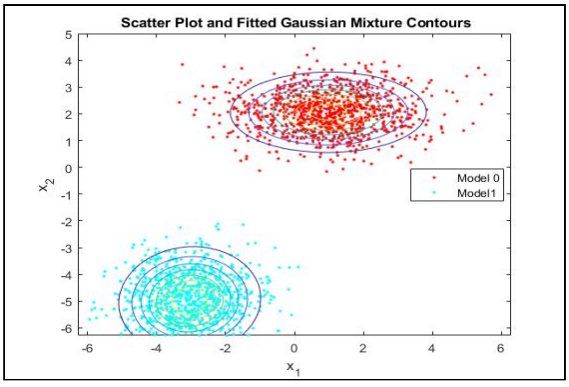
G. Gaussian Mixture Model

The gaussian mixture model is a probabilistic model which states that all generated data points are derived from a mixture of finite gaussian distributions that has no known parameters. The parameters are derived either from maximum a posteriori estimation or an iterative expectation - maximization algorithm from a prior model which is well trained. It is a classification algorithm that can be used to classify a wide variety of N-dimensional signals.



The model consists of covariance matrices, mixture weights and mean vectors from every component density present. Another feature of GMM is the smooth approximations to randomly shaped densities. GMMs are also used for density estimation and are considered most statistically mature for clustering.

The GMM algorithm is a good algorithm to use for the classification of non-temporal pattern recognition. The main limitation of the GMM algorithm is that, for computational reasons, it can fail to work if the dimensionality of the problem is too high.



GMM is used to model the probability density of credit card user's past behavior so that the probability of current behavior can be calculated to detect any abnormalities [13].

III. Analysis

All seven methods are implemented on two different datasets, i.e. the australian dataset (unskewed data) and the kaggle credit card fraudulent transactions dataset (skewed data). Outputs of the different algorithms are measured based on their three factors: accuracy, F1-score and time taken for the each detection. All results are tabulated.

A. For Dataset - 1 (Australian dataset): Unskewed data

Results	Accuracy (%)	F1 score	Time taken (ms)	Time complexity
Decision Tree	81.739	0.80	0.708	O(mnlog(n))
KNN (k=3)	68.260	0.63	1.942	$O(n^2)$
Logistic Regression	85.217	0.84	0.432	O(n)
Naive Bayes	81.304	0.76	0.571	O(n)
Neural Network	76.521	0.74	3.875	$O(n^5)$
SVM	70.434	0.70	4.042	$O(n^2)$
GMM	55.652	0.20	2.277	$O(nkd^3)$

The following can be inferred:

- It can be observed that KNN has the lowest accuracy and F1-score.
- Maximum time taken is by SVM.
- The method that performs the best is Logistic Regression that has the highest accuracy and F1-score and takes the minimum amount of execution time.

B. For Dataset - 2 (Kaggle credit card fraudulent transactions): Skewed data

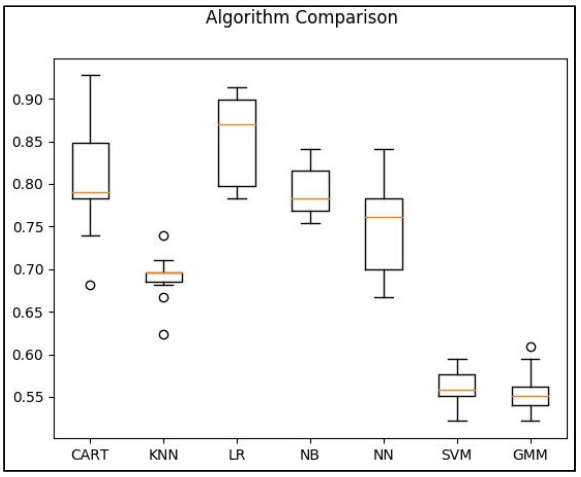
Results	Accuracy (%)	F1 score	Time taken (ms)	Time complexity
Decision Tree	99.846	0.71	5.238	O(mnlog(n))
KNN (k=3)	99.877	0.77	213.96	$O(n^2)$
Logistic Regression	99.871	0.74	1.446	O(n)
Naive Bayes	98.419	0.22	2.473	O(n)
Neural Network	99.891	0.79	200.71	$O(n^5)$
SVM	99.866	0.73	602.52	$O(n^2)$
GMM	99.952	0.81	1.982	$O(nkd^3)$

The following can be inferred:

- It can be observed that Naive Bayes has the lowest accuracy and F1-score.
- Maximum time taken is by SVM.
- The method that performs the best is GMM that has the highest accuracy and F1-score and takes the almost minimum amount of execution time.

IV. Conclusion

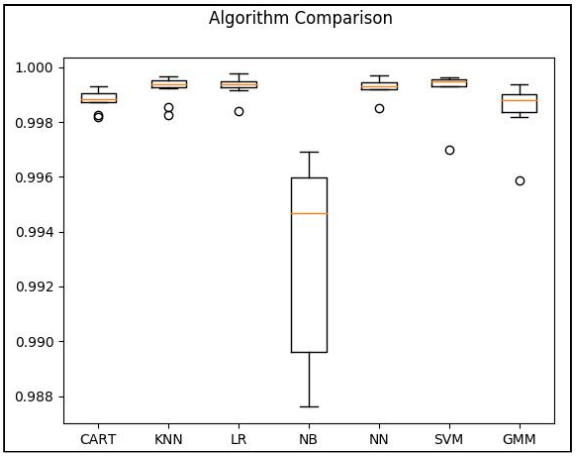
For dataset 1: Australian dataset the Logistic Regression model performs the best in terms of accuracy, F1-score and time taken.



Algorithm comparison graph for dataset 1

Since this is unskewed data, the gaussian mixture model does not perform as well as the other methods.

For dataset 2: Kaggle credit card fraudulent transactions dataset, the Gaussian Mixture model performs the best in terms of accuracy, F1-score and time taken.



Algorithm comparison graph for dataset 1

Since this is skewed data, the gaussian mixture model performs better than other methods.

References

[1] Patel , Snehal, et al. "Credit Card Fraud Detection Using Decision Tree Induction Algorithm." *International Journal of Computer Science and Mobile Computing*, Rinku Badgujar, Apr. 2015, www.ijcsmc.com/docs/papers/April2015/V4I4201511.pdf.Vol. 4, Issue. 4, pg.92 – 95,ISSN 2320–088X*(references)*

[2] Sharma, Himanshu. *Detection of Financial Statement Fraud Using Decision Tree Classifiers*. Dr.V.Ravi, 2013, www.idrbt.ac.in/assets/alumni/PT-2013/Himanshu Sharma_Detection of financial statement fraud using decision tree classifiers_2013.pdf.

[3] C, Rajabhushanam, and Yasvand S cumaar . "AN STUDY OF DATA MINING APPLICATIONS IN BANKING." *International Journal of Pure and Applied Mathematics*, C.Rajabhushanam, 2017, www.ijpam.eu/. Volume 116 No. 15 2017, 265-271 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (online version)

[4] Delamaire, Linda, et al. "Credit Card Fraud and Detection Techniques: a Review ." *Research Gate*, Hussein Abdou, 2014, www.researchgate.net/publication/40227011_Credit_card_fraud_and_detection_techniques_A_review. Banks and Bank Systems, Volume 4, Issue 2

[5] Sorournejad, Samaneh, et al. "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective ." *Research Gate*, Reza Ebrahimi Atani, Nov. 2016, www.researchgate.net/publication/310610856_A_Survey_of_Credit_Card_Fraud_Detection_Techniques_Data_and_Technique_Oriented_Perspective. 9 citations, 23 references

[6] C, Sudha, and Nirmal Raj T. "CREDIT CARD FRAUD DETECTION IN INTERNET USING K-NEAREST NEIGHBOR ALGORITHM." *IPASJ International Journal of Computer Science (IIJCS)*, Nirmal Raj .T, Nov. 2017, www.ipasj.org/IIJCS/IIJCS.htm. Volume 5, Issue 11,ISSN 2321-5992

[7] Patel, Twinkle, and Ompriya Kale. "A Secured Approach to Credit Card Fraud Detection Using Hidden Markov Model." *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Ompriya Kale, May 2014, pdfs.semanticscholar.org/8306/40276430c90d1e7c18f4da4e38d4a1091321.pdf. Volume 3 Issue 5

[8] K, Chitra, and Subhashini B. "Data Mining Techniques and Its Applications in Banking Sector." *International Journal of Emerging Technology and Advanced Engineering*, Subhashini.B, Aug. 2013, ijetae.com/. ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 8

[9] Perantalu, Varre, and Bhargav Kiran K. "Credit Card Fraud Detection Using Predictive Modeling: a Review." *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY*, Bhargav Kiran.K, Feb. 2017, ijirt.org/master/publishedpaper/IJIRT144240_PAPER.pdf.Volume 3 , Issue 9 , ISSN: 2349-6002

[10] Kibekbaev, Azamat, and Ekrem Duman. "Profit-Based Logistic Regression: A Case Study in Credit Card Fraud Detection." *DATA ANALYTICS 2015 : The Fourth International Conference on Data Analytics*, Ekrem Duman, 2015. ISBN: 978-1-61208-423-7

[11] R, Mallika. "Fraud Detection Using Supervised Learning Algorithms." *International Journal of Advanced Research in Computer and Communication Engineering*, Mallika R, June 2017, ijarcce.com/upload/2017/june-17/IJARCCE 2.pdf. Vol. 6, Issue 6

[12] Bhatia, Sunil, et al. "Analysis of Credit Card Fraud Detection Techniques." *International Journal of Science and Research (IJSR)*, Sunil Bhatia, Mar. 2016, pdfs.semanticscholar.org/89fb/942d03f8edd0f55424e8737332b2ad3f37fe.pdf. Volume 5 ,Issue 3, ISSN (Online): 2319-7064

[13] Agarwal, Surbhi, and Santosh Upadhyay. "A Fast Fraud Detection Approach Using Clustering Based Method." *Journal of Basic and Applied Engineering Research*, Oct. 2014, www.krishisanskriti.org/vol_image/03Jul201510071210.pdf. Print ISSN: 2350-0077; Online ISSN: 2350 Online ISSN: 2350-0255; Volume 1, Number 10