JEAS

JOURNAL OF ENGINEERING AND APPLIED SCIENCES

Journal of Engineering and Applied Sciences, Volume 19, Number 1, December 2021, 569 - 585

Enhanced Predictive Data Mining Algorithm for Fraud Detection and Churn Behaviour Modelling In Telecommunication Systems

E. O. Nonum[1*], A. N Isizoh[2], B. Ugiagbe[3,] and K. C. Okafor[4]

[1&2]Dept. of Electronic & Computer Eng, Nnamdi Azikiwe University, Nigeria

[3]Dept. of Mechatronics Engineering FederalPolytechnic Nekede, Owerri, Nigeria

[4]IEEE MSR-Labs, Dept. of Mechatronics Engineering Federal University of Tech, Owerri, Nigeria

[*]CorrespondingAuthor's E-mail: oyibow@gmail.com

Abstract

This work presents an enhanced predictive model for fraud detection and churn behaviour modelling in a telecommunication network. Computational analytic modelling was employed by using probabilistic models; Naïve Bayesian model, linear discriminant function and neural prediction networks to achieve adaptive control policy for fraud/churn detection. A critical threshold discriminant function (CTDF) Value of 0.00229 was obtained from a multivariate analysis of samples of call detail record (CDR) data sets. From the neural network training and validation plot, the proposed data mining predictive model gave 1.7562 Mean Square Error (MSE) for the CTDF. Also, an evaluation was carried out to determine an optimal algorithm/model with accurate, consistent and reliable results. Hence, three algorithms were analysed for Fraud and churn behavioural mining/detections, viz Decision Tree (DT), Logistic Regression (LR) and Enhanced Neural Discriminant Analysis (Proposed). These gave 14.29%, 30.00%, and 55.71% respectively. It was therefore concluded that the proposed algorithm offers the best and most reliable prediction threshold for churn/fraud attrition.

Keywords: Decision Tree, Machine Learning, Computational Science, Predictive Analytics, Edge Networks

1. Introduction

Today's world, especially in Nigeria, telecommunications network management requires extremely rapid decision-making methods that are data mining driven. The churn and fraud decision-making is based on information extracted from large amounts of data that are continuously collected from networks. The complexity of networks and the amount of monitoring data they provide are rapidly growing making room for churn and fraud behaviours. The mobile telecommunications industry has rapidly developed with new services for end users. The growing communication eco-systems, mobility trends and improved/flexible services are some driving forces behind data growth. All around the world, operators are updating and developing their networks to increase their capacity and to facilitate new kinds of services (Kimmo, 2009). Telecommunications network operation and management is based on the data that network elements provide (Gisela, 2002), (Ugwoke *et al* 2014). The elements create log entries and alarms about events, system states and disturbances and record time series about their performance. Network management systems move data to the operation centres, where it is monitored and analysed to detect any defects or suboptimal states in performance or service quality. An average-sized network can produce several thousands of alarms and tens of gigabytes of log and performance data per day. This data contains information about the performance of all network elements and services that the operator offers.

In a real time network, the volume of collected datasets creates a challenge for analysis methods and tools supporting network management tasks (Mahajan & Som, 2016). For example, how to recognise and identify sudden fraud and churn behaviour problems or issues that could prevent large amounts of customer traffic, and how to find network regions and elements that require optimisation. These issues are found in routine network management processes. Lately, a new paradigm called data mining (DM) and knowledge discovery (KD) has been developed (Timo & Bernhard, 2007), to handle complex analytics in systems. This paradigm combines several research areas like databases, on-line data analysis, artificial intelligence, neural networks, machine learning and statistics, among others. Telecommunications systems that produce large amounts of data were among the first application domains of data mining methods (Prashant et al 2017). Since then, many methodologies have been developed and applied to management and operation of telecommunications systems in terms of fraud and churn behaviour. Such management requires thorough understanding of network infrastructure, communication technologies, customers and their behaviour. Knowledge discovery and data mining technologies have been applied in several related application areas like churn analysis (Hadden et al 2007; Xiu et al 2009) and fraud detection and prevention.

However, most of the presented works have been executed as separated data mining designs, whose results were used in decision making. One of the greatest challenges for the knowledge discovery and data mining technologies seems to be how to support efficient classification with lower error margin. In the telecommunications industry, fraud continues to affect profitability as the problem results mainly in damages in the financial field since fraudsters are currently "leaching" the revenue of the operators who offer these types of services (Berson et al 2015). The main definition of telecommunication fraud corresponds to the abusive usage of an operator infrastructure, this means, a third party is using the resources of a carrier (telecommunications company) without the intention of paying for them. Other aspects of the problem that cause a lower revenue, are fraud methods that use the identity of legitimate clients in order to commit fraud, resulting in those clients being framed for a fraud attack that they never committed. This will result in client's loss of confidence in their carrier, giving them reasons to leave. Besides being victims of fraud, some clients do not want to have any business to do with a carrier that has been victim of fraud attacks. Since the offering of the same services is available by multiple carriers, the client can always switch very easily between them (Berson et al 2015).Observing Table 1, it is possible to verify the dimension of the "financialhole" generated by fraud in telecommunications industry and the impact of fraud in the annual revenue of the operators.

Table 1: Annual Global Telecommunications Industry Lost Revenue

	2005	2008	2011	2013	2015
Estimated Global Revenue (USD)	1.2 Trillion	1.7 Trillion	2.1 Trillion	2.2 Trillion	2.5 Trillion
Lost Revenue to Fraud (USD)	61.3 Billion	60.1 Billion	40.1 Billion	46.3 Billion	50.4 Billion
% Representative	5.11%	3.54%	1.88%	1.81%	1.15%

There are a couple of reasons which make the fraud in this area appealing for some fraudsters. One is the difficulty in the location tracking process of the fraudster; this is a very expensive process and requires a lot of time, which makes it impractical if trying to track large amount of individuals. Another reason is the technological requirements to commit fraud in these systems. A fraudster doesn't require particularly sophisticated equipment in order to practice fraud in these systems (Bhattacharya, 2011). All these evidences generate the need to detect the fraudsters in the most effective way in order to avoid future damage to the telecommunications industry. Data mining techniques are suggested as the most valuable solution, since it allows identifying fraudulent activity with certain degree of confidence. It also works especially well in great amounts of data, a characteristic of the data generated by the telecommunications companies. Acquisition and retention of new clients are one of the most significant concerns of businesses. While recipient companies concentrate on acquiring new customers, mature ones try to focus on retention of the existing ones in order to provide themselves with the opportunity of cross – selling. According to Freeman (2015) one of the most significant ways of increasing customers' value is to keep them for longer period of time.

Faced with this threat companies should be equipped and armed with the most efficient and effective methods of examining their client's behaviour predicting their possible future failure. In accordance with (Lejeune, 2016) churn management consists of developing techniques that enable firms to keep their profitable customers. The mobile telephony market is one of the fastest-growing service segments in telecommunications, and more than 75% of all potential phone calls worldwide can be made through mobile phones and as with any other competitive markets, the mode of competition has shifted from acquisition to retention of customers (Kim & Yoon, 2016). Examining the existing statistics concerning churn magnitude and its costs in this realm would be beneficial for gaining an appropriate mental picture of the importance of this area of research.

• SAS Institute, (2012) reported that the telecommunications sector endures an annual rate of churn, ranging from 25 per cent to 30 percent this churn rate could still continue to increase in correlation with the growth of the market.

• Churn costs for European and US telecommunications companies are estimated to amount to US$4 billion annually (SAS Institute, 2012)

• The ratio (customer acquisition costs/ customer retention or satisfaction costs) would be equal to eight for the wireless companies (SAS Institute, 2012).

While the annual rate of customer churn in telecommunications sector is around 30 percent and it costs US$ 4 billion per year for European and US telecommunications companies, it would seem reasonable to invest more on churn management rather than acquisition management for mature companies especially when we notice that the cost of acquiring a new customer is eight times more than retaining an existing one (SAS Institute, 2012). Most models used to predict churn behaviour such as: C5.0, decision tree, artificial neural network, discriminant algorithm, naive Bayes algorithm, logistic regression are very useful for services provisioning by telecommunication systems but we are limited to the number of telephone calls per unit time and its duration. This calls for the use of Non-Homogenous Poisson Process (NHPP) to estimate the prior probabilities of the customers' churn behaviour. By developing the Bayesian Statistics, i.e., Naive Bayesian algorithm, posterior probability can be ascertained.

From the posterior probabilities, two classes of subscribers, each having three variates, will be derived. Thus, by applying multivariate data sample analysis methodology and developing a linear discriminant analysis for classifying the posterior probabilities of subscriber churn behaviour, this work established the critical value (bench mark) for subscriber behavioural classification. The overall goal is to finding an efficient, enhanced and accurate predictive data mining algorithm for fraud detection and churn behaviour modelling in pre-paid mobile telephony market segment by utilizing machine learning techniques.

Research Motivation
Telecommunication service providers suffer high revenue losses due to fraud and churn attrition. The negative effect of fraud and churn to business success is high especially in this era of stiff competition in the mobile telecommunication market. The losses range from the cost of convincing a new customer to use the provider's services to the cost of retaining existing customers. Therefore, the need for an enhanced predictive model for fraud detection and churn behaviour cannot be overemphasized. The aim of this work is to develop an enhanced data mining flow model for fraud detection and churn behaviour classification in telecommunication systems. The will be achieved via the following steps:

i. To develop a Non-homogenous Poisson process model (for estimation of prior probabilities) and Bayesian Statistics model (posterior probability analysis)
ii. To establish a critical value (bench mark) for subscriber classification
iii. To propose and designing a telecommunication system in which fraud and churn behaviour are predicted automatically from subscribers' call detail record, based on the probabilistic and statistical models developed in step (i).
iv. To compare the performance of the developed model with existing algorithms, in terms of error precision/accuracy metrics.

This paper is organized in five sections. Section 1 is the introduction and motivation for the research while in section 2, a review of related works was detailed. Section 3 presents the materials and methods used for analysis of the proposed system data and obtaining of results as presented in Section 4 and 5 respectively

2.0 Literature Review

Most works of data mining done in telecommunications fraud detection have the objective of detecting or preventing the methods of superimposed fraud (Hilas, 2009) and subscription fraud (Farvaresh & Sepehri, 2011), because these are the fraud methods that have done more damage to the telecommunications industry. The kind of data that is mostly used in context of data mining are the call details records (or call data records) CDR (Hilas, 2009) (Farvaresh & Sepehri, 2011), and (Augustin *et al.*, 2012). These records have a log like format and they are created every time that the subscriber completes using a service of the operator. The representation of the records may change from one organization to another. However, the most common attributes are: caller and called identification number, date and time, type of service (Voice Call, SMS, etc.), duration, call rate (frequency), network access point identifiers; The authors in (Augustin*et al* 2012) developed a system for fraud detection in new generation networks. The system consists of two modules: one for interpretation and the other for detection. In the model of interpretation, the CDR's raw data is used as an input and the module analyses and converts that data into CDR objects, which will be used to feed the classification module. In the classification model, the CDR objects will be subjected to a series of filter analysis. The filters are actually rule based classification techniques where rules and corresponding thresholds were obtain previously from real fraudulent data. If a CDR object doesn't respect the rules in those filters an alarmed is generated. As a result the author said that the model only classifies as false-positive 4% of the data.

Caigny *et al*, (2018) focused on the highly competitive telecommunication sector using an intelligent rule-based decision-making technique, based on rough set theory (RST), to extract important decision rules related to customer churn and non-churn. Their approach performs classification of churn from non-churn customers, while predicting churn viable customers in the near future. Various evaluations were carried out using four rule-generation mechanisms, viz: Exhaustive Algorithm (EA), Genetic Algorithm (GA), Covering Algorithm (CA) and the Learning from Example Module LEM2 algorithm. The work showed attribute-level analysis for developing a successful customer retention policy in the telecom sector. Other works on churn behaviour are found in (Bell & Mgbemena, 2017) where a classification and regression trees (CART) i.e., a decision-tree method (a churn modelling techniques) was discussed including algorithms such as CART, Chi-squared automatic interaction detector (CHAID), iterative dichotomizer (ID3), and C4.5, which is the successor of ID3 (Bell & Mgbemena, 2017).

Idris *et al* (2012) focused on churn prediction in telecommunication using Random Forest and particle swarm optimization. In (Amin *et al* 2014), the work synthesized customer churn prediction where rough set theory is used as one-class classifier and multi-class classifier to investigate the trade-off in the selection of an effective classification model for customer churn prediction. An experimental study was performed to explore the performance of four different rule generation algorithms (viz: exhaustive, genetic, covering and LEM2). The genetic algorithm classifier gave more suitable performance as compared to the other three rule generation algorithms.

Wong *et al* (2009) estimated the containership arrival rate in harbor operation and management. The work explained that literature discusses arrival processes based on homogeneous Poisson processes, which normally fail to describe the fluctuation status of growth or recession. Their work proposed the Non-Homogeneous Poisson Process in order to analyze the arrival process of containership. In their paper, the Maximum Likelihood Method was used to estimate the parameters and the performance of the models in the case of Taichung Harbor in Taiwan. Goulding *et al* (2016) observed that data streams have events that occur at random arrival times rather than at the regular. Since, event series are continuous, irregular and often highly sparse, differing greatly in nature to the regularly sampled time, the authors derived a new approach to forecasting event series that avoids such assumptions relying upon: the processing of event series datasets in order to produce a first parameterized mixture model of non-homogeneous Poisson processes, and secondly, on application of a technique referred to as parallel forecasting which makes use of the above processes' rate functions to directly generate accurate temporal predictions for new query realizations. Their procedure uses stochastic process to highlight the distribution of future events. The work in (Gugudhvili *et al* 2018), identified Non-Homogeneous Poisson Process (NHPP) and software reliability growth models (SRGM) as the most popular approach to estimate useful metrics for example the number of faults remaining, failure rate, and reliability, etc for a specified period of time. The authors proposed performance-optimized expectation conditional maximization (ECM) algorithms for NHPP SRGM. The work claimed that their ECM algorithm reduces the maximum-likelihood estimation process to multiple simpler conditional maximization (CM)-steps unlike expectation maximization (EM) algorithm.

Shathi *et a*l (2011) presented an effective and efficient method for classifying text documents in order to deliver feasible information retrieval using naïve bayes algorithm. a weight matrix was introduced, during the training of the text documents, which is a combination of term frequency (TF) and inverse class frequency(ICF). The work used the weighted term enabled by a significant number and mapped with the posterior value during the prediction time of Naïve Bayes (NB) algorithm to establish a better and efficient performance of the classification task. While their experimental results show that NB Weight Matrix is efficient, it lacks computational intelligence for accurate classification tasks. Sathyadevan *et al* (2014) discussed Naive Bayes algorithm used to classify documents pushed into knowledge based cloud (information repository) "Kloud". This method offered better accuracy and speed considering a two way sub-classification algorithms namely hierarchical sub-classification and sub-categorization using document similarity method. Zi-Qing *et al* (2017) started by explaining that the Naïve Bayes Classifier (NBC) is a classification method based on the Bayesian theorem which is used to predict the label of data set. It has advantages of high classification efficiency, low cost and low error rate and has been widely used in natural language processing, pattern recognition, machine learning among others. The work used the flower pollination algorithm (FPA) to optimize Naive Bayes classifier. In the work, the NBC algorithm based on improved flower pollination algorithm (NBC-IFPA) was explored.

A blacklist mechanism was introduced to make the FPA transition out of the local optimal solution. Again, a random perturbation term was introduced to increase the diversity of the population and improve the searching ability of FPA. Hence, the improved FPA was used to search for the global optimal attribute weights and use them into the weighted naive Bayesian model for classification. Even though, their simulation results offered (NBC-IFPA algorithm) higher classification accuracy, it still lacks computational intelligence for weightier classifications. Other works studied on Naive Bayes classification algorithm are found in text classification (Bužić & Dobša,, 2018),(Zhang & Gao, 2013), discretisation and structural improvement for classification accuracy (Martinez-Arroyo &Sucar, 2006), fake news detection (data set of Facebook news posts) using naive Bayes classifier (Granik & Mesyura, 2017), Social media Twitter classification (Fatahillah *et al* 2017), Bayesian Classifier with the sparse regression technology (Zhang & Gao, 2013), (Zheng et al 2018) and Solar energy Naïve Bayes application (Bayindir *et al* 2017). From these works, it is clear that the Bayesian classification represents a supervised learning method (Naive Bayes text classification) as well as a statistical method for classification. It assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems.

The Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data. Again, in most works in literature, dimensionality reduction in the domain of pattern recognition and linear discriminant analysis (LDA) remains the most commonly leveraged supervised dimensionality reduction methods. In (Zhao *et al* 2018), Dimensionality reduction and LDA were explored in pattern recognition datasets. In (Rajaguru & Prabhakar, 2017), Bayesian Linear Discriminant Analysis Classifier (BLDA) was used to classify the risk of breast cancer in the patients and the results are shown in terms of classification accuracy, performance index, sensitivity and specificity.
Al-Anzi & AbuZeina (2017) evaluated linear discriminant analysis (LDA) as a dimensionality reduction technique was used in generating effective feature vectors by reducing the dimensions of the original data (e.g. bag-of-words textual representation) into a lower dimensional space. The authors verified LDA to be convenient method for text classification for huge dimensional feature vectors. The work investigated two LDA based methods for Arabic text classification. In (Xie *et al* 2017) a version of classic LDA classification rule was proposed for non-stationary data, using a linear-Gaussian state space model. The Non-stationary LDA (NSLDA) classification rule is based on the Kalman Smoother algorithm used to estimate the evolving population parameters. For unknown system dynamics an Expectation-Maximization (EM) algorithm and the Kalman Smoother were employed to simultaneously estimate population and state space equation parameters. The system performance was assessed in a set of numerical experiments using simulated data. The average error rate obtained by NSLDA was compared to the error produced by a naive application of LDA to the pooled non-stationary data. In (Khandaker *et al* 2017), LDA scheme for feature extraction was used to deal with the discrimination recognition directly. In their work, the issue of absence of data label, enabled the authors to integrate the unsupervised LDA method with Principle Component Analysis (PCA) and K-Means clustering in order simplify the application of LDA for the rotating machinery dataset. Their method was analyzed and verified for the actual rotating machinery dataset collected from rotor shaft under different fault condition. Furthermore, the work adopted K-nearest neighbour to show the effectiveness of the proposed unsupervised LDA method in extracting features from rotor dataset. Similarly, in (Shitole & Jadhav, 2018), the authors worked on improving the performance of recognition system using PCA and LDA. Their system used three different feature extraction methods: chain coding, edge detection using gradient features and direction feature techniques, (for the first raw features extraction) which are reduced by LDA and characters are classified using SVM classifier.

From literature survey, existing methods has the limitation of using resource intensive classification algorithms for identifying fraudulent and /or churning customers. Approximation for automatic computation of threshold using some wired formula for user threshold classification and profiling is another constraint. The proposed Fraud Detection/Churn Predictive Machine Learning Model (FD-CPMLM) offers light-weighted computation that makes prediction based on the previous memory it has on the customer. The proposed scheme leverages its fraud detection architecture to solves complex classification problem using Naïve-Bayesian formula for continuous transactions, and for computing the probability and divergence.

3.0 Materials and Methods

In this section the materials used in this work include Minitab 12, MATLAB and Microsoft excel packages. Below is the descriptions of the methodology used to derive the learning model.

3.1. Methodology

3.1.1. Data Collection

Data were gathered from Globacom Nigeria Limited's database for this study. The examined data of this research is the call records of Globacom Limited, which has been gathered and stored in Oracle data base software. Since Telecom Company's data base is a dynamic one that is being updated and extended every second, the call records' data base is a huge data base with millions of records. Two different files were received from the telecom service provider – one contains the CDR of low income subscribers and the other contains the CDR of high income subscribers

3.1.2. Data Selection and Pre-processing

The data samples (training data) used in developing the predictive churn model for this research was extracted from the Globacom data base in a period of 3 months from 1st November 2016 to 31st January 2017 and it contains the call records of 34523 customers of Globacom Limited and the total number of records is 19500504 for both low income subscribers and high income subscribers. Among various data types that are being saved and gathered in Globacom's database, the following fields were extracted used in building the required and targeted features:

- **i.** **Subsc** stand for subscribers. These subscribers are thirty (30) in each group and are designated by names *1, 2, . . . , 30*. Each of the number represents names such as John Okeke, Henry Okoye, etc.
- **ii.** **n-call** stand for number of calls per subscriber per day. It is represented by number say *2, 8, 5, 7, …,* etc.
- **iii.** **t(min)** stand the time (duration) in minutes spent by each subscriber on the calls.

A total of six clusters / samples (three from the low income subscribers and the other three from the high income subscribers) each containing 30 subscribers were selected from the large CDR. Thus, 180 subscribers' record, 90 for each class of subscriber was used for the training data. The samples were collected randomly, on different days of the week and at different times of the day. Each cluster / sample has unique attributes defined in terms of $fx = At = B$; where A = Average number of calls and B = Total time (duration) taken for those calls. Therefore, the clustering attributes for the three samples of the low income subscribers are $4(t = 3)$,$8(t = 6)$ and $12(t = 10)$ while $3(t = 7), 5(t = 8)$, and $6(t = 9)$ were for the high income subscribers. Table 3 show the extracted and pre-processed data samples for both classes for subscribers.

3.1.3. Data Analysis Technique

In this sub-section, a predictive probability data mining model is established in Table 2. This was done using the Non-Homogenous Poisson Process (NHHP(λt)). NHHP (λt) model was developed to model/determine the prior probability distribution while the Bayesian statistics model was developed to determine the posterior distribution. Linear Discriminant Analysis was use to detect churn behaviour of subscribers so as to classify them accordingly. This classification will enable service providers to determine mitigating measures to discourage their subscribers from leaving their network. At this point, it was generally observed that it is more expensive to attract new subscribers than to retain the existing subscribers. Hence, Poisson distribution provides a realistic model for such random phenomena in Figure 1a. Since the value of Poisson random variables are the non-negative integers, random phenomenon for which a count is of interest is a candidate for modelling by assuming a Poisson distribution.

Such a count includes the number of telephone calls per unit time coming into the switch board of a large business (service providers). If certain assumptions regarding the phenomenon under observation are satisfied, the Poisson model is the correct model (Mood et al 1974). Hence, if $\{X_n, n \geq 0\}$ is a sequence of independent identically distributed exp(λ) random variables, the counting process $\{N(t), n \geq 0\}$ is called a Poisson process with parameter λ and it is denoted by *PP(λ)*.X_n is the probability distribution of the n^{th} customer. Thus, the first event counted by *N(t)* takes place after an exponential amount of time with parameter λ.

Table 2. Extracted and preprocessed data samples for both classes for subscribers (CDR collected from Globacom Ng)

PROCESSED CALL DETAIL RECORD FOR LOW INCOME SUBSCRIBERS								
SUBSCRIBERS X1			SUBSCRIBERS X2			SUBSCRIBERS X3		
Subsc	**n-call /day**	**t(min)**	**Subsc**	**n-call /day**	**t(min)**	**Subsc**	**n-call /day**	**t(min)**
1	2	6	1	9	5	1	6	13
2	3	4	2	7	3	2	7	14
3	5	3	3	3	3	3	11	13
4	5	2	4	5	3	4	12	13
5	9	2	5	10	12	5	12	9
6	2	4	6	9	8	6	13	6
7	4	2	7	5	1	7	9	7
8	4	3	8	8	6	8	13	10
9	9	3	9	8	10	9	8	8
10	2	3	10	6	7	10	8	14
11	4	6	11	5	7	11	16	8
12	4	1	12	11	7	12	14	9
13	3	6	13	10	5	13	14	13
14	1	4	14	8	3	14	12	4
15	4	2	15	12	4	15	9	9
16	3	4	16	11	7	16	9	8
17	4	1	17	8	7	17	12	16
18	2	3	18	9	6	18	12	14
19	3	4	19	13	8	19	15	13
20	4	6	20	10	7	20	11	15
21	3	0	21	6	12	21	15	6
22	2	0	22	7	9	22	13	11
23	3	0	23	8	11	23	15	13
24	3	5	24	10	9	24	16	16
25	3	4	25	9	7	25	10	11
26	4	2	26	7	4	26	14	9
27	8	8	27	6	5	27	16	10
28	6	1	28	8	8	28	9	8
29	3	1	29	5	7	29	14	6
30	4	4	30	9	9	30	15	10

Table 3. Processed call detail record for high income subscribers

SUBSCRIBERS X1			SUBSCRIBERS X2			SUBSCRIBERS X3		
Subsc	**n-call /day**	**t(min)**	**Subsc**	**n-call / day**	**t(min)**	**Subsc**	**n-call / day**	**t(min)**
1	2	9	1	3	9	1	10	4
2	1	7	2	5	4	2	6	7
3	6	4	3	6	9	3	8	10
4	3	11	4	11	8	4	6	8
5	1	7	5	4	6	5	5	5
6	1	12	6	7	9	6	13	11
7	0	7	7	3	10	7	8	8
8	4	5	8	7	8	8	4	5
9	2	4	9	2	6	9	8	6
10	6	7	10	5	6	10	7	8
11	3	11	11	2	1	11	6	10
12	4	10	12	3	4	12	4	6
13	7	8	13	5	6	13	7	10
14	2	11	14	2	10	14	4	10
15	3	11	15	7	7	15	8	12
16	4	3	16	5	6	16	4	10
17	1	4	17	1	7	17	6	13
18	5	8	18	2	11	18	6	8
19	1	12	19	5	8	19	10	5
20	1	4	20	6	8	20	11	9
21	0	5	21	8	8	21	3	14
22	4	8	22	9	13	22	7	12
23	4	6	23	6	8	23	5	11
24	3	7	24	2	7	24	10	11
25	3	5	25	6	10	25	6	7
26	3	4	26	7	8	26	10	15
27	0	3	27	7	10	27	5	11
28	4	6	28	7	5	28	7	4
29	4	5	29	5	4	29	5	12
30	3	9	30	9	12	30	6	12

The rest of the inter event time are independent identically distributed exponential with parameter λ. NHHP(λt) can be thought of as a process that counts events that occur in a non-uniform fashion (Barlow & Proschan, 1975), and (Field & Zidek, 1995), hence, this work will establish critical assumptions for NHHP (λt). Tabel 3 illustrates the processed CRD for high income subscribers.

Now, it is feasible to write *{N(t), t ≥ 0)} ~ NHPP (λ(.))*. This denotes that *{N(t), t ≥ 0)* is a non-homogeneous Poisson process with rate function *λ(.)*. When $\lambda(t) = \lambda$ for all t ≥ 0, then NHPP becomes a HPP. Thus, NHPP is a generalization of HPP. In both HPP and NHPP, events take place one at a time (Field & Zidek, 1995); Hence, the predictive probability models for this work are non-homogenous Poisson process. This is then given in (1), while the Bayesian statistics model is given in (2), (Field & Zidek, 1995)

$$P_n(t) = \exp\{-\lambda t\}\left[\frac{\{\lambda t\}^n}{n\,!}\right] \quad ; n = 0, \ldots 1 \tag{1}$$

Where $P_n(t)$ = the probability of n number of calls at a given time (t), λ is the parameter (intensity) of the model and t is time in minutes. Equ1model will be used to determine the prior distribution, but for posterior distribution, the model is given by Equ 2.

$$P(Y = y/\theta) = \frac{P(y/\theta)P(\theta)}{\sum_{\theta=0}^{n} P(y/\theta)P(\theta)} \tag{2}$$

Where $P(Y = y/\theta)$ = the conditional probability that the random variable Y assumes a specific value y given that its prior probability was θ. Note that $\theta = \lambda$ is now a random variable. $\sum_{\theta=0}^{n} P(y/\theta)P(\theta)$ = the likelihood function of the distribution. The above Bayesian statistics model will be used to determine the posterior probability. Hence, the predictive probability data mining model is the model in Equation (2). The model is called Bayesian statistical model. In context, the Naïve Bayesian Classification (NBC) was used for classifying subscriber records based on numerical or categorical probability. It was used for calculating probability value for each record considering the attributes that best describes the record. The values of each record are numerical. Hence, it implements the formula for continuous values in Naïve-Bayesian method. The discriminant function employ in this work is called Fisher's linear discriminant function given in Equ. 3.

$$W = X^T S^{-1}\left(\bar{X}^{(1)} - \bar{X}^{(2)}\right) \tag{3}$$

Where $X^T(X_1 - X_p)$; where $X^T = \begin{bmatrix} X_{1,1}(f) & \cdots & X_{1,np}(f) \\ \vdots & \ddots & \vdots \\ X_{np,1}(f) & \cdots & X_{np,np+1}(f) \end{bmatrix}$

$S^{-1}$is the inverse of the dispersion (variance – covariance) matrix and $\left(\bar{X}^{(1)} - \bar{X}^{(2)}\right)$ is the difference in the mean vectors between two multivariate samples and W is the linear discriminant function. Consider the subscriber sample events times $X = (X_{1,}X_{2,}X_{3,} \ldots \ldots \ldots \ldots \ldots \ldots \ldots X_{np})$ taken from the time interval $\{t: 0 < t \leq T\}$. This arises from a NHPP with a rate function $\lambda(t)$. Therefore, the log likelihood function is given by Equ.4. (Thompson, 2012).

$$l(\lambda(t)|X) = \sum_{i=1}^{M_x} ln(\lambda(X_i)) - M(t) \tag{4}$$

Where $\mathrm{M(t)} = \int_0^t \lambda(\bar{t})\overline{dt}$.

Now, deriving the functional classification coefficients for NHPP rate function $\lambda(t)$ is a convex optimisation problem in which the objective function and inequality constraint are both twice differentiable.

In terms of classification, let's consider a set of subscriber with $n_{traindsets}$ training observations $T_{train} = \{\{X_1, g_1\}, \{X_2, g_2\}, \ldots \ldots \ldots \{X_{ntraindsets}, g_{ntraindsets}\}\}$ in which $X_l = \{(X_1, X_2, X_3, \ldots \ldots \ldots \ldots \ldots \ldots \ldots X_{np})\}$ is a set of l^{th} observations and $g_1 \in 1, 2, \ldots \ldots \ldots, k$ is the corresponding class label. But the classification task deals with using this training data to make predictive estimation rate function $\lambda_v(t)$ for subscribers with the context of the likelihood in Equ 5.

$$l(\lambda_v(t)|T_{train}) = \sum_{l=1}^{ntrain} 1(G_l = v)\left[-M_v(t) + \sum_{i=1}^{ml} \ln(\lambda_v(X_{l,i}))\right] \tag{5}$$

where the indicator function $1(G_l = v) = 1$, if $G_l = v$ and 0 otherwise. The rate function $\lambda_v(t)$ is given by Equ 5 while the classification of test data for the subscribers, clustering, prior with piecewise constant realisations, posterior distribution of the intensity function and Bayesian asymptotic assumed based on the work on Poisson point processes (Shota *et al*, 2018).

3.1.4. Critical Value Determination

So far, the lightweight models for the proposed scheme has been highlighted above. Now, considering the subscribers telecommunication domain in Figure 1, this work now used Equ 3 to establish the critical value (*Wc*) of the discriminant function and finally classify the samples (subscribers) where they belong based on their posterior probability distributions in Figure 2. Two multivariate sample data with three variates each were derived from the probability predictive model developed for this study. The analyzed two sample multivariate data with three "variates" each were the posterior probabilities of each group (that is, their current probability of churn given that we have prior information about their churn behaviour). Then, this work classifies the samples into the group they belong, either as stable customers (subscribers) or those prone to churn. The classification rule is: classify the subscribers in group I into"A_1; A_2", where A_1 is stable subscribers and A_2 is subscribers prone to churn. Similarly, the same procedure is applied for group II designated by "B_1; B_2".

4.0 Proposed System Architecture and Analysis

For the Fraud/churn behaviour synthesis, Figure 1 was derived as the characteristic telecommunication network having edge devices connected to the BT controller $i_1 \ldots i_n$. Highly intelligent base station controllers connects the BSC to the Fog mobile switching centre which handles computational neural training for all subscribers. The BSS core is the centre of all the Fog controllers which connects to the Linear discriminant service whose processing elements communicates the processed churn or non-churn information by their dynamic state response to the cloud servers. The various layers are discussed below.

i. Mobile Edge Network layer (MENL): This layer offers connectivity and traffic transport between mobile edge users/subscribers and the telecommunication backend cloud data centres; within the Fog neural switches, base station controllers and the base transceiver controllers across multiple sites. It relies on the tree hierarchical structure interfaced into the BTC and the BSCs. Access edge devices via the core BSSS infrastructure allows for latency realization, power conservation and data-offloading in a bi-directional fashion.
ii. Global Telecommunication Compute and Storage (GTCS): This represents the compute and storage infrastructure appropriate for tracking Fraud/churn entities with both an unstructured content and highly structured transaction database.
iii. Fog Mobile Switching Services (FMSS): This hosts the logical application controller streams from the core BSS infrastructure. The intelligent neural layer 3 switch core is the basic unit for traffic distributed coordination. It has potentials for security authentication, subscriber verification, deep traffic inspection (DTI), and load balancing.

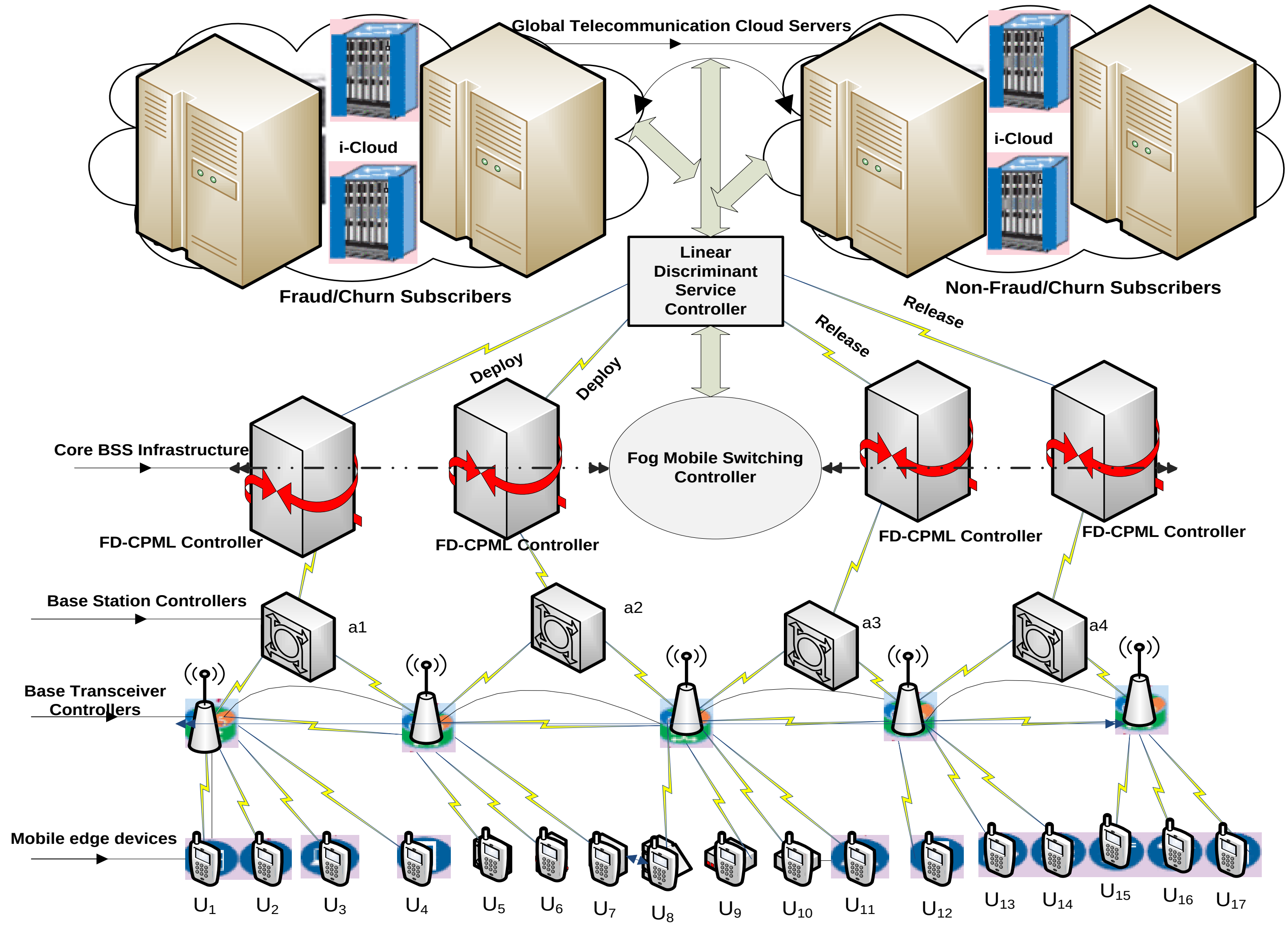

Figure.1: Proposed System Architecture, (Author's Testbed with MS-Visio, 2010)

Again, this could serve as unified management framework that brings together all of the elements of the lower edge devices, and the higher fraud/churn computing infrastructure. It enables efficient and responsive monitoring, management, and planning through its neural intelligence. From Figure 1, each component of the architecture has distinct attributes and specific requirements for intelligent data-mining in the switching algorithm. This meets the system's requirements for performance assessment from the edge domain to the cloud. The neural network dynamic feedback control in discriminant classifiers used to achieve accurate classification for all subscribers. The system flow diagram is given in Figure 2.

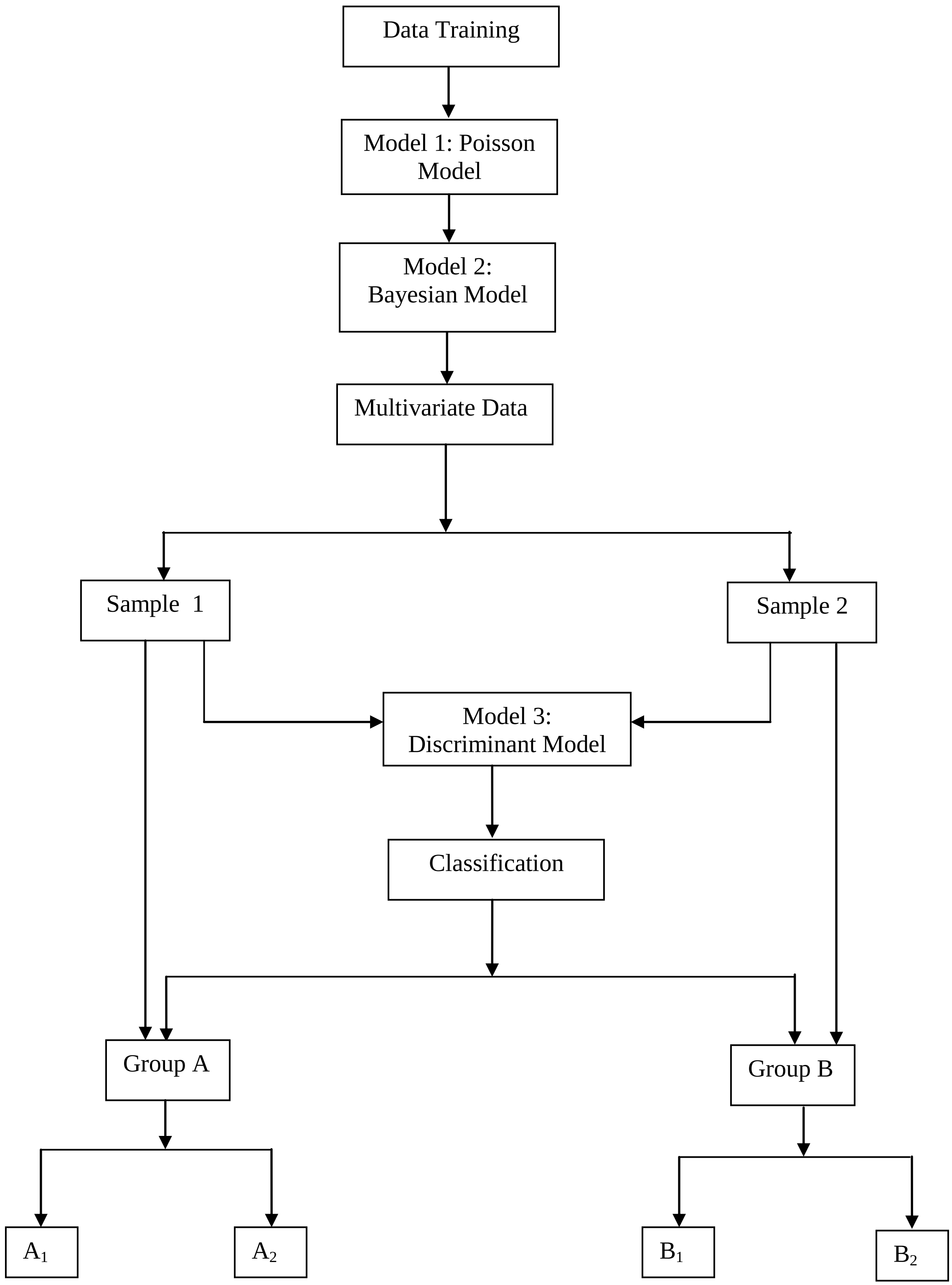

Figure 2: Proposed System flow model for telecommunication fraud /churn

5.0 Results Discussions

In this Section, a neural network classification performance model was achieved with MATLAB 2017 for Figure 1. This was realized using SimEvent and C++ scripts which completed both the training and validation comparisons. The data samples and iteration cycles were synthetically defined also. Figure 3 illustrates data mining predictive controller's (i.e., Fraud Detection/Churn Predictive Machine Learning Model (FD-CPMLM)) mean square error validation. It was observed that out of the 14 Epochs for training, test and best residuals, at the 8^{th} Epoch, the best validation performance was obtained at 1.7562 mean square error. The significance of these results is that the proposed system offers a very reliable/ accurate region of fraud/churn classification. From these results, it is clear that the established Data Mining Predictive controller have 1.7562 mean square error for the 0.00229critical value (threshold) of the discriminant function.

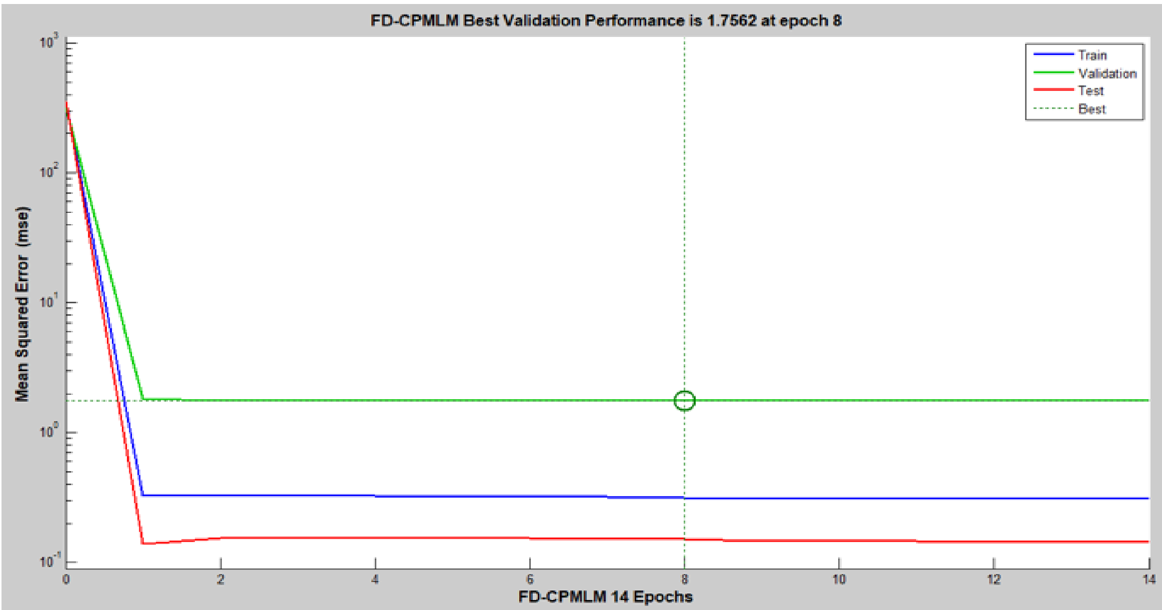

Figure 3: Data Mining Predictive controller (FD-CPMLM) mean square error validation.

The Figure 4 shows the performed comparative evaluation to discover the optimal model with accurate, consistent and reliable results. In this regard, three algorithms namely decision tree (DT), Logistic Regression (LR) and Enhanced Neural Discriminant Analysis (Proposed) (ND) were analysed for Fraud and churn behavioural mining/detections. From the model graph in Figure 4, the total mean square accuracy was obtained as 0.7. The decision tree gave 14.29%, the logistic regression gave 30% while the proposed neural discriminate gave 55.71%. This means that the proposed algorithm offers the best and most reliable prediction threshold. Also, the data mining predictive algorithm having a mean square error of 1.7562 for the 0.00229 critical value (threshold) of the discriminant function is very significant.

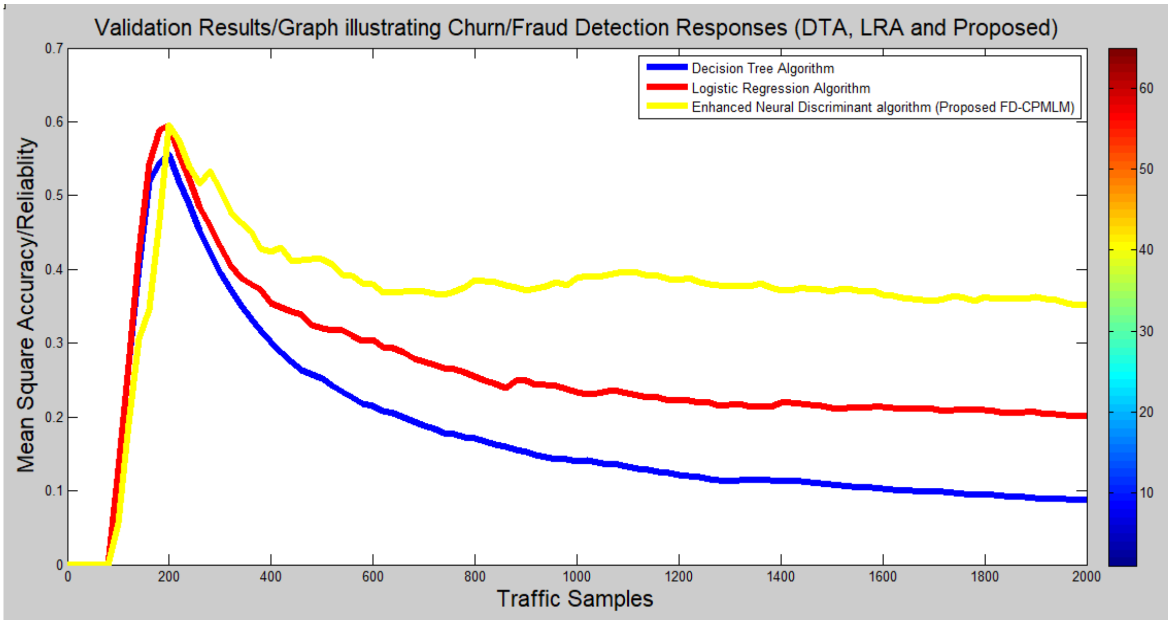

Figure 4: Validation Plots for enhanced data Mining algorithms for fraud / churn prediction

It is obvious that the enhanced computational algorithm offers better prediction of churn/fraud than existing decision tree and regression algorithms. However, by identifying the proper attributes and mapping exact thresholds, this may even offer more precise results. This work has shown that with the established computational models, it is feasible to predict and even analyse critical retention scheme from the global trained datasets.

From the analysis so far, it is obvious that the enhanced computational algorithm offers better prediction of churn/fraud than existing decision tree and regression algorithms. However, by identifying the proper attributes and mapping exact thresholds, this may even offer more precise results. This work has shown that with the established computational models, it is feasible to predict and even analyze critical retention scheme from the global trained datasets. The generated synthetic datasets, anomalous call sources were injected as fraudsters, which results in transaction records. The sequence and interaction data of both normal and anomalous instances were generated according to the neural generative process with varied parameter settings. Figure 5 shows the validation plots for data clustering mean square latencies for the three algorithms. Essentially, data clustering refers to the grouping of CDRs in similarity groups known as clusters. It is a statistical distribution by multi-objective optimization. This work observed the data clustering latency which is physically a consequence of the limited velocity with which data interaction propagates. It was observed that three algorithms namely decision tree, logistic regression and enhanced neural discriminant analysis(Proposed) (ND) gave latency hit-clustering as 40.00%, 33.33% and 26.67% respectively for fraud behavioural mining/detections. The implication is that the building block of designing an optimal fraud prediction system will be very efficient with the enhanced neural discriminant analysis (ENDA).Its basic principle is to detect/predict abnormal behaviour that deviates from normal behaviour. Even when the fraudulent users seem to concentrate on different behavioural pattern, the ENDA offers a more accurate scheme for fraud prediction in telecommunication networks.

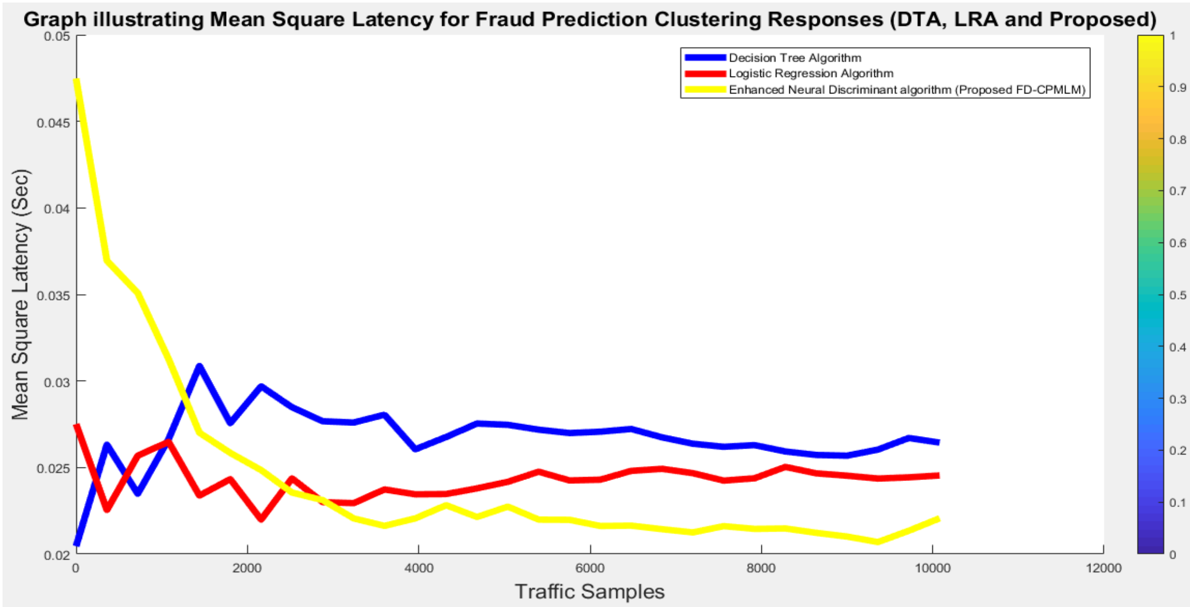

Figure 5: Validation Plots for Data clustering latency.

Figure 6 shows fraud detection linear predictive scalability behaviour of the selected algorithms on the network plane. These algorithms vary significantly in predictive scalability especially as data samples/CDRs are increased. These fraud centriod scalability algorithms namely decision tree, logistic regression and enhanced neural discriminant analysis (Proposed) (ND) offers clustering scalability as 30.12%, 33.73% and 36.15% respectively. This implies that the proposed fraud detection scheme has the capacity to track several CDRs at a given detection cycle accurately.

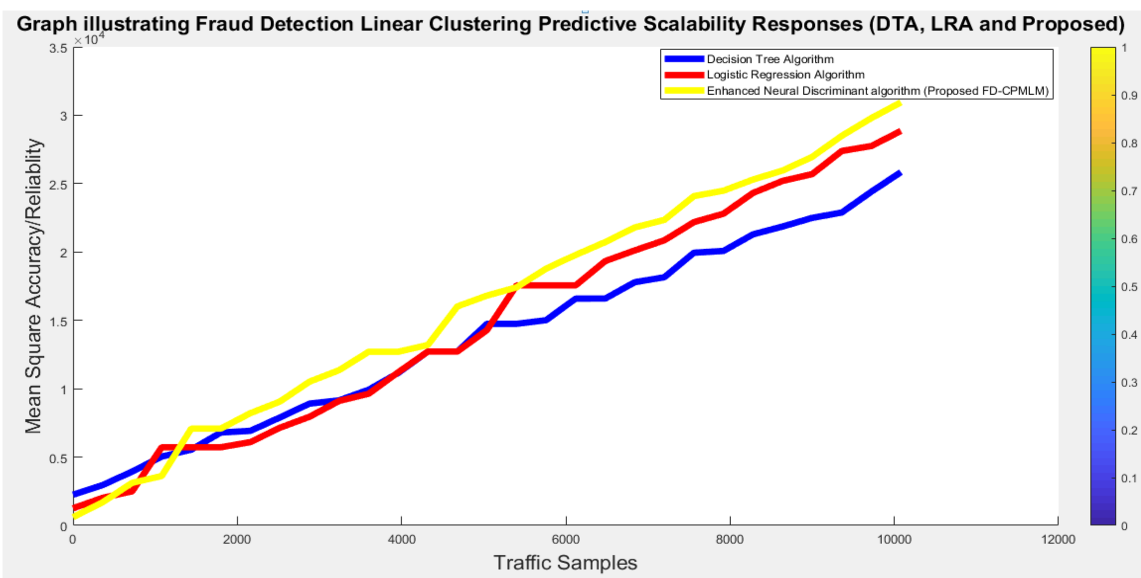

Figure 6: Validation Plots for Data clustering latency.

From the analysis so far, it is obvious that the enhanced computational algorithm offers better prediction of fraud than existing decision tree and regression algorithms. However, by identifying the proper attributes and mapping exact thresholds, this may even offer more precise results. This work has shown that with the established computational models, it feasible to predict and even analyze critical rentention scheme from the global trained datasets. The major significant of this work is that the proposed classification scheme solves most conventional fraud types in Telecommunication industry such as: Subscription Fraud K1, Clip on Fraud K2, Call Forwarding K3, Cloning Fraud K4, Roaming Fraud K5, and Calling Card K6 due to its accuracy prediction level with low computational resource demands.

4.0 Conclusion

This paper has develop a Non-homogenous Poisson process model for estimation fraud in telecommunication systems. We showed that a predictive data mining algorithm offers the best form of mean square accuracy and reliability for subscribers classification. It provides a composite fraud behaviorial prediction. The usefulness of the model was demonstrated through a predictive neural network scenario of experiments in respect of fraud prediction.By using a Non-homogenous Poisson process (for estimation of prior probabilities) and Naïve Bayesian Statistics (posterior probability analysis), the work used Linear discriminant model (multivariate analysis) for classifying the posterior probabilities of subscriber churns behavior. Through the model, a critical value (bench mark) for subscriber classification was derived using multivariate data analysis. The performance of the developed model was compared with existing algorithms, in terms of error precision/accuracy metrics. In the results generated, the predictive model with (No fraud dataset samples) and Machine Predictive model (with fraud samples) were determined. It was observed that out of the 14 epochs for training, test and best residuals, at the 8^{th}epoch, the best validation performance was obtained at 1.7562 mean square error. The comparative evaluation of three algorithms namely decision tree (DT), logistic regression(LR) and enhanced neural discriminant analysis(Proposed) (ND) were analyzed for Fraud and churn behavioural mining/detections.

References

Al-Anzi, F.S. & AbuZeina, D. 2017. Arabic text classification using linear discriminant analysis, "*International Conference on Engineering & MIS (ICEMIS),* 8-10 May 2017, Monastir, Tunisia. DOI: 10.1109/ICEMIS.2017.8272958.

Amin A., Khan C., Ali I., Anwar S. (2014). Customer Churn Prediction in Telecommunication Industry: With and without Counter-Example. In Gelbukh A., Espinoza F.C., Galicia-HaroS.N. (eds) *Nature-Inspired Computation and Machine Learning. MICAI 2014.Lecture Notes in Computer Science*, 8857. Springer, Cham.

Augustin, A., Sajid, A.,Awais, A., Muhammad, N., Khalid, A. &AmirHussain, K. H. 2017. Customer churn prediction in the telecommunication sector using a rough set approach. *Neuro computing*, 237(10), Pp 242-254

Barlow, R.E, & Proschan, F. 1975. Statistical Theory of Reliability and Life Testing probability models.Holt, Rinehart and Winston, Inc. USA.

Berson, W.T., Ngai, L. X. & Chau, D.C.K. 2015. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592 – 2602.

Bužić, D., & Dobša, J. 2018. Lyrics classification using Naive Bayes.41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 21-25 May 2018, Opatija, Croatia, DOI: 10.23919/MIPRO.2018.8400185

Caigny, A, Coussement, K, & Bock, K 2018. A new hybrid algorithm for customer churn. IdeasRePEc , European Journal of Operation Research, 269(2) pp 760-772.

Farvaresh F. &Sepehri M. S. 2011. A data mining framework for detecting subscription fraud in telecommunication.*Engineering Applications of Artificial Intelligence*.24(1), pp.182–194.

Fatahillah, N.R., Suryati, P. &Haryawan, C. 2017. .Implementation of Naive Bayes classifier algorithm on social media (Twitter) to the teaching of Indonesian hate speech. *In Proc. International Conference on Sustainable Information Engineering and Technology (SIET)*,24-25 2017, Malang, Indonesia.DOI: 10.1109/SIET.2017.8304122.

Zidek, J.V. .& Field, C 1995. Modelling and Analysis of Stochastic Systems.Chapman and Hall, London, UK.

Freeman, M. 2015. The two customer lifecycles. *Intelligent Enterprise , 2* (16), 9.

Goulding, J. Preston, S. & Gavin Smith, G. 2016. "Event Series Prediction via Non-Homogeneous Poisson Process Modelling"2016 IEEE 16th International Conference on Data Mining (ICDM), 12-15 Dec. 2016, Barcelona, Spain, DOI: 10.1109/ICDM.2016.0027

Granik, M., & VolodymyrMesyura, V. 2017. Fake news detection using naive Bayes classifier.*IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 29 May-2 June 2017, Kiev, Ukraine, DOI: 10.1109/UKRCON.2017.8100379.

Gugudhvili. S. M. 2018. A Knowledge Management Framework for imbalanced data using Frequent Pattern Mining based on Bloom Filter, 2016 11th International Conference on Computer Engineering & Systems (ICCES), 20-21 Dec. 2016, Cairo, Egypt, DOI: 10.1109/ICCES. " 2018.7822004.

Hadden, J., Tiwari, A., Rajkumar, R., &Dymitr R. 2007. Computer assisted customer churn management: state of the art and future trends. *Journal of Computers and Operations Research,* 34(10), 2902 – 2917.

Hilas, K. 2009a. Data mining for telecommunications network log analysis.PhD Thesis, Series of Publications A, Report A-2009-1, Department of Computer Science, University of Helsinki, Finland.

Idris, B. Y. Yanuar S. R. HartantoKusuma H. W. 2012. "Location prediction model using Naïve Bayes algorithm in a half-open building", *In Proc. 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE),* Pp.18-19, 2012, Semarang, Indonesia, DOI: 10.1109/ICITACEE.2012.8257667

K. C. Okafor, I. E. Achumba, G.A Chukwudebe, G.C. Ononiwu, 2017. Leveraging Fog Computing For Scalable IoT Datacenter Using Spine-Leaf Network Topology", Journal of Electrical and Computer Engineering, Vol.2017, Article ID 2363240, Pp.1-11, 2017, Egypt.

Khandaker, N., Zhang, R., Dong, S., Xu, B. Zhang, Z. & Wen, G. 2017. Application of unsupervised linear discriminant analysis for condition monitoring of rotating machinery.*In Proc. 9th International Conference on Modelling, Identification & Control (ICMIC),*10-12 July 2017, Kunming, China, DOI: 10.1109/ICMIC.2017.8321687.

Kim, Y.; Yoon, R.H.&Dodier, R. 2016. Predicting customer behaviour in telecommunications", IEEE Intelligent Systems, 19(2), Pp. 50 –58.

Kimmo,S. H 2009. Designing an expert system for fraud detection in Private Telecommunication's networks.*Expert Systems with Applications.* 36(9), pp.11559– 11569.

Lejeune., J. 2016. *User Profiling and Classification for Fraud Detection in Mobile Communications Networks*.PhD thesis, Helsinki University of Technology, Laboratory of Computer and Information Science. Helsinki, Finland.

Mahajan, R., & Som, S. 2016. Customer behavior patterns analysis in Indian mobile telecommunications industry", *In Proc. 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 16-18 March 2016, New Delhi, India.

Martinez-Arroyo, M. & Sucar, L.E. 2006. Learning an Optimal Naive Bayes Classifier*18th International Conference on Pattern Recognition (ICPR'06),*20-24 Aug. 2006, Hong Kong, China, DOI: 10.1109/ICPR.2006.749.

Prashant, Y., Niranjan, H., Kenchannavar, H.&Ramesh M. 2017. Data mining techniques in Telecom Sector *In Proc. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS),* Chennai, India, DOI: 10.1109/ICECDS.2017.8389950.

Rajaguru, R. & Kumar S. P. 2017. Bayesian linear discriminant analysis for breast cancer classification. *In Proc. 2nd International Conference on Communication and Electronics Systems (ICCES),* 19-20 Oct. 2017, Coimbatore, India, DOI: 10.1109/CESYS.2017.8321279.

Sathish, M., Santhosh, K, Naveen, K.J. &Jeevanantham, V. A. 2011. Study on Consumer Switching Behaviour in Cellular Service Provider. A Study with reference to Chennai.*Far East Journal of Psychology and Business*, 22, pp. 71-81.

Sathyadevan, S. Sarath, P.R., Athira U. &Anjana, V. 2014. Improved document classification through enhanced Naive Bayes algorithm.*In Proc. International Conference on Data Science & Engineering (ICDSE),*26-28 Aug. 2014, Kochi, India, DOI: 10.1109/ICDSE.2014.6974619.

Shitole, S. & Jadhav, S. 2018. Recognition of handwritten devanagari characters using linear discriminant analysis", *In Proc. 2nd International Conference on Inventive Systems and Control (ICISC)*, 19-20 Jan. 2018, Coimbatore, India, DOI: 10.1109/ICISC.2018.8398991.

Timo, H. &Bernhard, S. 2007. Collaborative Knowledge Discovery & Data Mining: From Knowledge to Experience. In Proc. 2007 IEEE Symposium on Computational Intelligence and Data Mining, 1 March-5 April 2007, Honolulu, HI, USA, DOI: 10.1109/CIDM.2007.368905.

Ugwoke Ç., Erdem Z., Gökhan, A., Ozgur Y. Sahingoz, K. &Buldu, A. 2014. Network anomaly detection with deep learning", *In Proc. 26th Signal Processing and Communications Applications Conference (SIU),* 2-5 May 2014, Izmir, Turkey, DOI: 10.1109/SIU.2014.8404817.

Shota Gugushvili, Frank Van Der Meulen, Moritz Schauer, and Peter Spreij, 2018. Fast And Scalable Non-Parametric Bayesian Inference For Poisson Point Processes, arXiv:1804.03616v2 [stat.ME] 19 Jul 2018.

Wong, Z, Haifeng Zhao, H. . &Nie, F. 2009. A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction.*In IEEE Transactions on Knowledge and Data Engineering*, 30 May 2018, Pp.1-1, DOI: 10.1109/TKDE.2018.2842023.

W. Thompson, 2012. Point process models with applications to safety and reliability. Springer Science & Business Media.

Xie S., Imani, M. & Dougherty, E.R 2017. Ulisses M. Braga-Neto, "Nonstationary linear discriminant analysis" *In Proc. 51st Asilomar Conference on Signals, Systems, and Computers*, 29 Oct.-1 Nov. 2017, Pacific Grove, CA, USA, DOI: 10.1109/ACSSC.2017.8335158.

Xiu, P, Vidhyashree, N., Fiondella, L., Zeephongsekul, P., Jayasinghe, C.L. & Thierry Wandji, T 2009. Performance Optimized Expectation Conditional Maximization Algorithms for Non-homogeneous Poisson Process Software Reliability Models. *IEEE Transactions on Reliability*, 66(3), Pp. 722 – 734.

Zhang, Y., Liang, R., Li, Y. Zheng, Y. & Berry, M. 2011. Behavior-Based Telecommunication Churn Prediction with Neural Network Approach.*In Proc. International Symposium on Computer Science and Society*, 16-17 July 2011, Kota Kinabalu, Malaysia, DOI: 10.1109/ISCCS.2011.89.

Zheng, Z, Cai, Y., Yang Y. L. 2018. Sparse Weighted Naive Bayes Classifier for Efficient Classification of Categorical Data.*In IEEE Proc. Third International Conference on Data Science in Cyberspace (DSC),*18-21 June 2018, Guangzhou, China, DOI: 10.1109/DSC.2018.00110.

Zi-Qing, W., Ming-Xin, Z., Jin-Long, Z. & Guo-Hai, Z. 2017.Optimization of Bayesian Classifier Based on Flower Pollination Algorithm. *In Proc. 14th Web Information Systems and Applications Conference (WISA)*, 11-12 Nov. 2017, Liuzhou, China, DOI: 10.1109/WISA.2017.28