

## Credit Card Detection System Based on Rudit Approach

Norbaiti Tukiman<sup>#</sup>, Norhaiza Ahmad<sup>\*</sup>, Suhana Mohamed<sup>§</sup>, Zarith Sofiah Othman<sup>#</sup>, CT Munirah Niesha Mohd Shafee<sup>#</sup>, Zairi Ismael Rizman<sup>&</sup>

<sup>#</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 81750 Pasir Gudang, Johor, Malaysia  
E-mail: norbaiti289@johor.uitm.edu.my, zarithsofiah@johor.uitm.edu.my, ctmun518@johor.uitm.edu.my

<sup>\*</sup>Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81200 Skudai, Johor, Malaysia  
E-mail: norhaiza@utm.my

<sup>§</sup>Department of Finance, Faculty of Business Management, Universiti Teknologi MARA, 81750 Pasir Gudang, Johor, Malaysia  
E-mail: suhan291@johor.uitm.edu.my

<sup>&</sup>Faculty of Electrical Engineering, Universiti Teknologi MARA, Dungun, Terengganu, Malaysia  
E-mail: zairi576@tganu.uitm.edu.my

---

**Abstract**— Fraud detection is one of the important agendas in financial and insurance institutions to protect the institutions from fraudsters and loss. The losses to the financial institutions are huge, and the need to detect the fraud at an early stage is critical to the institutions. If the numbers of fraud are not properly managed, the impact may lead to the closure of the institutions. Many predictive analytic systems or models have been proposed to identify and detect the frauds. Hence, this paper examines the effect of different response variables of credit card history as the reference group which used an unsupervised scoring method namely an Identified Distribution (RIDIT) based on a statistically significant test. We illustrate the method using German Credit card dataset retrieved from UCI Machine Learning Data System. The result generates scores and significant value of chi-square test that reflect response variables being classified as reference group or comparison groups, which more or less affected by the response credit card history in fraud detection.

**Keywords**— fraud; RIDIT; score; system; approach

---

### I. INTRODUCTION

White collar crime happened in most of the countries in the world. Among the famous crime is the credit card fraud which causes billions of dollars were robbed from the bank. The process of identifying the credit card fraud in financial data is very crucial, and yet a challenging task since the database in the transactions is large and highly dimensional [1]-[4].

Recently, loss of billions of dollars occurs and arising in credit card fraud for credit card holders and corporate companies. Based on a report by Internet Crime Complaint Centre (IC3) showed an increasing value of complaints with similar dollar loss for five years from 2011 to 2015. In 2015, there are about 288,012 complaints received with dollar loss reports about \$1,070,711,522 million which increase 33.8% or \$270,219,449 million compare to year 2014. We illustrate the complaint and corresponding dollar loss report by IC3 in Fig. 1 and Fig. 2. Thus, from the huge number of loss figure, it is crucial to prevent or fight the fraud.

A fraudster defined as a perverted person whom intent to get some benefit from other source or party without the legal right. Currently, fraudsters may use high and sophisticated methods in electronic commerce and technology tools or equipment in their fraudulent activities to gain more money. The previous literature discusses the issues and challenges in detection fraud which involve the concept of drift, real-time detection, earliness of detection, skewed distribution, big data, misclassification of data, cost sensitivity and so on [5].

Therefore, in dealing with the transaction of data, typically these data are imbalanced, having the small number of fraud compared to non-fraud cases. There are many predictive methods namely supervised method to detect fraud such as support vector machine [6], neural network [7], [8] and decision tree [9], [10]. While, an unsupervised method, a few techniques have been applied such as Self-organizing Map [11], [12], Principal Component Analysis [13], [14] and Peer Group Analysis [15]. However, most of these methods are dependent on distribution which is not appropriate for ordered categorical data. One unsupervised predictive method to counter this issue is using an Identified

Distribution Integral Transformation (RIDIT) [16]-[20]. RIDIT is a transformation of ordered categorical data to a probability scale between groups of observations.

## II. MATERIAL AND METHOD

The RIDIT analysis is a transformation a set of response variable of ordered categorical data to a probability scale to be more suited applied for further statistical analysis test. The first introduction of RIDIT was to classify borderline cases in a behavioural science study of a categorical dataset [16]. The reference group is a random data taken from large sample observations [23]. The evolution of RIDIT then continued by other researchers. For examples, the calculation of the RIDIT score and applied Principle Component Analysis (PCA) to assign the weighted values based on the importance of the variables in insurance claim data is modified [16]. Furthermore, the RIDIT analysis to cater for ordered continuous variables is extended [22]-[24]. Newly, the categorical and continuous response variables in the detection of fraud in the microinsurance industry are mixed [15].

As mentioned earlier, RIDIT analysis requires choosing a reference group from a set of response variables in the study. After selecting the reference group, the RIDIT score will be calculated for each category. The RIDIT score for each group observation is calculated by taking one-half of the frequency for the response variable plus the cumulative frequency of all response variables before the current response variable and then divide all by the total frequency of the reference group.

### A. Calculation of RIDIT Score

Let  $y_1, y_2, \dots, y_k$  indicate a single response variable with a categorically ordered scale where  $p_j, j = 1, 2, \dots, k$  be a probability function that defined on the set  $\{y_1, y_2, \dots, y_k\}$  which is transform as follows

$$p_j = P(\{y_j\}) \text{ and } \sum_{j=1}^k p_j = 1$$

The interval scale for the probability function is in between 0 to 1. To calculate the RIDIT score from the reference distribution  $\{p_j, j = 1, 2, \dots, k\}$  is quite easy. The score  $r_1, r_2, \dots, r_k$  are to be assigned respectively to the categories  $y_1, y_2, \dots, y_k$  as follows

$$r_1 = 0.5p_1 \quad (1)$$

$$r_j = 0.5p_j + \sum_{k < j} p_k \text{ for } j=2, \dots, k \quad (2)$$

Thus, RIDIT score is related to the cumulative distribution function associated with the probability function  $\{p_j, j = 1, 2, \dots, k\}$ . For single response variable, we do not need a reference group. However, by using more than one response variables, we need a reference group to make a comparison with other groups. RIDIT score of the  $i$ th category of the reference group (A) and the comparison group (B) can be calculated as shown in Table 1.

### Complaints Received By IC3 From Year 2011 to 2015

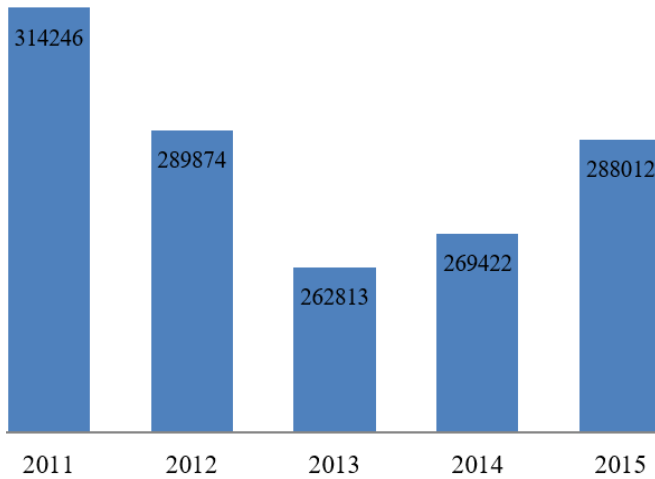


Fig. 1 Complaints received by IC3 from the year 2011 to 2015

### Dollar Loss From Year 2011 to 2015

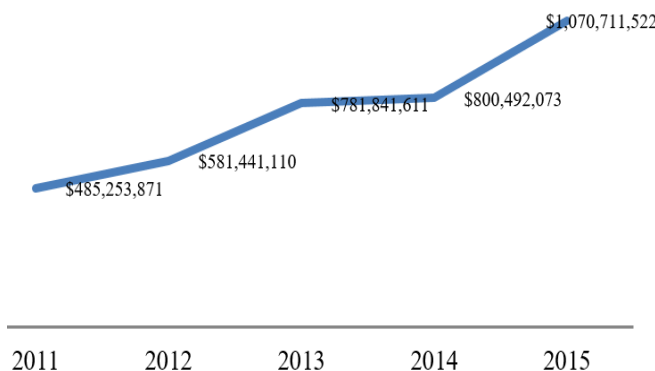


Fig. 2 Dollar loss from the year 2011 to 2015

RIDIT analysis requires selecting a reference group data amongst many groups of observations. However, in the previous literature, choice of reference groups in RIDIT analysis is not chosen objectively [17], [18], [20]. Therefore, the choice of the identified reference group is a critical stage in RIDIT analysis especially for credit card fraud detection as it forms the baseline to subsequent studies. Typically, credit card database consists of several response variables with several categories. In this paper, we introduce a new statistical test to select a reference group for further analysis. We then compare groups of observations and select which is the best reference group from the observations. The outline of this paper is constructed as follows. Firstly, we briefly discuss the methodology of RIDIT Analysis, then we illustrate the methodology by using German credit data and finally the study presented a concluding remark [21].

TABLE I  
THE DATA FORMAT FOR RIDIT ANALYSIS

Ordered Categories ( $C_i$ )	Group A (Reference, $f_{iA}$ )	Group B (Comparison, $f_{iB}$ )	Total
$C_1$	$f_{1A}$	$f_{1B}$	$\sum C_1 = f_{1A} + f_{1B}$
$C_2$	$f_{2A}$	$f_{2B}$	$\sum C_2 = f_{2A} + f_{2B}$
...			** ... represent until
$C_k$	$f_{kA}$	$f_{kB}$	$\sum C_k = f_{kA} + f_{kB}$
Total	$n_A$	$n_B$	$\sum C_n = n_A + n_B$

The RIDIT score for group A is shown as follows

$$r_{iA} = \frac{\left(0.5f_{iA} + \sum_{j=1}^{i-1} f_{jA}\right)}{n_A} \quad (3)$$

Equation (3) can be transformed into Equation (2) which is equivalent to  $0.5p_{iA} + \sum_{j=1}^{i-1} p_{jA}$ , where  $p_{iA}$  and  $p_{jA}$  are the proportion of the total observations in the  $i$ th (row) and  $j$ th (column) categories of the reference group A for  $i = 1, 2, \dots, i-1$  and  $j = 1, 2, \dots, k$ .

For any other comparison groups B, the mean RIDIT,  $\bar{r}_B$  is then calculated as

$$\bar{r}_B = \frac{\left(\sum_{i=1}^k f_{iB}r_{iA}\right)}{n_B} \quad (4)$$

The Equation (4) is equivalent to  $\sum_{i=1}^k p_{iB}r_{iA}$ , where  $p_{iB}$  is the proportion of the group or relative frequency of  $i$ th category in group B for  $i = 1, 2, \dots, k$ . Based on Table 1, group A is selected reference group. Here, we assume that the mean RIDIT score is always 0.5. This implies that if any two subjects are selected at random from this reference group A, one will be expected to be more affected or more serious condition on the criteria of variables or vice versa.

Hence, the interpretation of mean RIDIT for any other group B is explained below.

Given the reference group A and comparison group B, the mean RIDIT  $\bar{r}_B$  for comparison group B is an estimate of  $P(B \geq A)$  that is the probability of randomly selected response variable from group B has more affect compared to the randomly selected response variable from reference group A. If the RIDIT score is more than 0.5, then it will be more affected compared to randomly selected from reference group A. Otherwise, if the mean RIDIT for a given comparison group B is less than 0.5, then it will be less affected than the randomly selected from reference group A. Thus, there is more than one potential reference group might be taken from the response variables.

### B. Statistical Test of Mean RIDIT Score

To test the mean RIDIT of reference group, the null hypothesis is stated as follows

$$H_0 : \bar{r}_B = 0.5$$

$$H_1 : \bar{r}_B \neq 0.5 \quad (5)$$

Assume that 0.5 is the mean RIDIT of reference group, A. If the sample size ( $n_B$ ) is large, then  $\bar{r}_B$  is approximately normally distributed by using the variance equation as follows

$$Var(\bar{r}_B) = \frac{1}{12n_B} \quad (6)$$

Then, the null hypothesis of Equation (5) tested using chi-square test which is suitable for categorical data as follows

$$\chi^2 = \frac{(\bar{r}_B - 0.5)^2}{Var(\bar{r}_B)} = 12n_B(\bar{r}_B - 0.5)^2 \quad (7)$$

For a sufficiently large data set,  $n_B$  also, it approximately value of the chi-square distribution with degree of freedom one which implies  $H_0$  is rejected, and  $H_1$  will be accepted at the half alpha, ( $\alpha/2$ ) level of significance. Otherwise,  $H_0$  is accepted, and  $H_1$  is rejected.

Then, the next step follows Equation (8).

$$\chi^2 \geq \chi^2_{\alpha/2, 1} \quad (8)$$

### C. German Credit Data

We illustrate the analysis using German Credit dataset [17]. The German credit data set consisted of a set of 1000 loan applicants, 700 of dataset were creditworthy applicants, and 300 of dataset were classified as bad or not be extended. For each applicant, 20 variables described credit history, account balances, loan purpose, loan amount, employment status and personal information. Each sample contains 13 categorical, three continuous, 4 set/binary features and 1 class feature. The German credit dataset is challenging because of imbalanced with the smaller number of fraud rather than non-fraud. The data also consists of mix type of scale measurement and therefore quite difficult to make analysis in the detection of fraud. Table 1 shows the attribute, data type and measurement of German data set. The scope of this study is to compare the effect of credit history on the status of checking account.

TABLE II  
RESPONSE VARIABLES, DATA TYPE, AND SCALE OF MEASUREMENT OF GERMAN DATA SET

No.	Response Variables	Data Type	Scale Measurement
1	Status of existing checking account	Qualitative	Ordinal
2	Duration in month	Quantitative	Continuous
3	Credit history	Qualitative	Categorical/ nominal
4	Purpose	Qualitative	Categorical/ nominal
5	Credit amount	Quantitative	Continuous
6	Savings account/bonds	Qualitative	Ordinal
7	Present employment since	Qualitative	Ordinal
8	Instalment rate in percentage of disposable income	Quantitative	Continuous
9	Personal status and sex	Qualitative	Categorical/nominal
10	Other debtors/ guarantors	Qualitative	Categorical/ nominal
11	Present residence since	Quantitative	Continuous
12	Property	Qualitative	Categorical/ nominal
13	Age in years	Quantitative	Continuous
14	Other installment plans	Qualitative	Categorical/ nominal
15	Housing	Qualitative	Categorical/ nominal
16	Number of existing credits at this bank	Quantitative	Continuous
17	Job	Qualitative	Categorical/ nominal
18	Number of people being liable to provide maintenance for	Quantitative	Discrete
19	Telephone	Qualitative	Categorical/ nominal
20	Foreign worker	Qualitative	Categorical/ nominal

### III. RESULTS AND DISCUSSION

Therefore, in this study, we only illustrate two response variables from the German Credit dataset which are credit card history and status of existing checking account for analysis because both of the variables have a relationship between them. In this dataset, applicant credit card history

consists of five categories response variables and its ordered categories corresponding four categories response variables of the status of an existing checking account. The description and its code are detailed in Table 3. Table 4 shows the distribution of credit history and status of existing checking account applicants.

TABLE III  
DESCRIPTION OF CREDIT HISTORY AND STATUS OF EXISTING CHECKING ACCOUNT

Status of Existing Checking Account	Credit History
A11 : ... < 0 DM	A30 : no credits taken/all credits paid back duly
A12 : 0 <= ... < 200 DM	A31 : all credits at this bank paid back duly
A13: ... >= 200 DM /salary assignments for at least one year	A32: existing credits paid back duly till now
A14: no checking account	A33: delay in paying off in the past
	A34: critical account/other credits existing (not at this bank)

Note that A11 = less than 0 DM, A12 = 0 and not more than 200 DM, A3 = more than 200 DM and A14 = no checking account. The rest of codes from A30 to A34 are shown details in Table 3.

Based on Table 4, we illustrate the transformation of response variable of no credit card taken in example A30 for calculation of RIDIT score.

$$r_{1,1} = \frac{\left(\frac{1}{2}(13)\right)}{40} = 0.163$$

$$r_{2,1} = \frac{\left(\frac{1}{2}(18)+13\right)}{40} = 0.550$$

$$r_{3,1} = \frac{\left(\frac{1}{2}(2)+18+13\right)}{40} = 0.800$$

$$r_{3,1} = \frac{\left(\frac{1}{2}(7)+2+18+13\right)}{40} = 0.913$$

A summary of the RIDIT score for all file response variables of credit card history and status of checking account is shown in Table 4. These RIDIT scores reflect the overall cumulative distribution function associated with its probability function. For instance, the highest RIDIT score of 0.913 (A30) has a more cumulative frequency, and a lowest RIDIT score of 0.068 (A33) has a less cumulative frequency as shown in Table 5 as referring to original data in Table 4.

TABLE IV  
DISTRIBUTION OF APPLICANT CREDIT HISTORY AND STATUS OF EXISTING CHECKING ACCOUNT IN GERMAN DATA SET

Status of Existing Checking Account	Credit History					
		A30	A31	A32	A33	A34
A11	13	22	160	12	67	274
A12	18	15	146	35	55	269
A13	2	3	37	3	18	63
A14	7	9	187	38	153	394
Total	40	49	530	88	293	1000

TABLE V  
RIDIT SCORE FOR FIVE CATEGORIES OF CREDIT HISTORY

Status of Existing Checking Account	Credit History				
		A30	A31	A32	A33
A11	0.163	0.224	0.151	0.068	0.114
A12	0.550	0.602	0.440	0.335	0.323
A13	0.800	0.786	0.612	0.551	0.447
A14	0.913	0.908	0.824	0.784	0.739

The next step of RIDIT is the selection of the reference group. Here, we have a credit history which has five categorical response variables, A30, A31, A32, A33, and A35. We assume all the response variables have the same chance to be selected as a reference group. In such as a response variable to be considered as a reference group, the mean RIDIT score is calculated for the group. For example, if A30 is a reference group, the mean RIDIT score is 0.5. Then, the other response variable would have a mean RIDIT score of 0.458(A31), 0.578(A32), 0.662 (A33) and 0.666 (A34) based on Equation (4). The summary of mean RIDIT scores for each of the response variables as a reference group

is shown in Table 6. Here, R-A30 refers to A30 as the reference group; R-A31 refers to A31 as the reference group and so on [24].

The bold entries in Table 6 indicate mean RIDIT score less than 0.5. This reflects that they are ineffective to be a reference group. The more bold entries a reference group has indicated that it is more ineffective to be a reference group based on the status of checking account and in contrast, the less bold entries a reference group has indicated that it is less ineffective to be a reference group based on the status of checking account. Here, A30 and A31 appeared better reference group compared to the other groups.

TABLE VI  
SUMMARY MEAN RIDIT SCORE FOR FIVE CATEGORIES OF CREDIT HISTORY

Mean RIDIT	Credit History				
R-A30	0.500	0.458	0.578	0.662	0.666
R-A31	0.542	0.500	0.609	0.689	0.687
R-A32	0.422	0.391	0.500	0.572	0.585
R-A33	0.338	0.311	0.428	0.500	0.522
R-A34	0.334	0.313	0.415	0.478	0.500

We further test the value of mean RIDIT Score by using chi-square test at a level significant  $\alpha = 0.05$ . A summary of the chi-square test statistics is shown in Table 7 and its p-value in Table 8 respectively. Therefore, we can see that the p-value for reference group no credits taken/all credits paid back duly (A30) and all credits at this bank paid back duly (A31) both have three significant value which is no credits taken/all credits paid back duly (A30) shows 0.0107, 0.0314 and 0.0000, and all credits at this bank paid back duly (A31) shows 0.0004, 0.0122 and 0.0000 significant value. For reference group existing credits paid back duly till now (A32), delay in paying off in the past (A33) and critical account/other credits existing (not at this bank) (A34), these reference groups only show one significant value based on the status of checking account.

With that figure, we can select either reference group no credits taken/all credits paid back duly (A30) or all credits at this bank paid back duly (A31) in this analysis. Therefore, we can interpret that both 2 reference groups have more effective to be less fraud rather than other three groups of credit history based on the status of an existing checking account.

#### IV. CONCLUSION

The RIDIT analysis serves as a function to shows the effect of using different reference groups based on the significance test of the mean RIDIT. For the analysis of German credit card data, we only use two response variables which are credit card history and status of checking account with both have five categories and four categories of response group.

TABLE VII  
CHI-SQUARE TEST STATISTICS SIGNIFICANT FOR THE RIDIT SCORE OF CREDIT HISTORY

Chi-Square	Credit History				
	R-A30	R-A31	R-A32	R-A33	R-A34
R-A30	0.000	0.174	6.511	4.631	16.156
R-A31	0.142	0.000	12.570	6.287	20.457
R-A32	0.491	1.162	0.000	0.911	4.206
R-A33	2.105	3.501	5.487	0.000	0.279
R-A34	2.206	3.421	7.607	0.084	0.000

TABLE VIII  
P-VALUE FOR CHI-SQUARE TEST

P-Value	Credit History				
	R-A30	R-A31	R-A32	R-A33	R-A34
R-A30	0	0.6766	0.0107*	0.0314*	0.0000*
R-A31	0.7063	0	0.0004*	0.0122*	0.0000*
R-A32	0.4835	0.2811	0	0.3398	0.0403*
R-A33	0.1468	0.0613	0.0192*	0	0.5974
R-A34	0.1375	0.0644	0.0058*	0.7719	0

df=1, \* $\alpha = 0.05$

In this analysis, we show how RIDIT analysis works for selecting the reference group. Based on the analysis, two categories of credit card history which are no credits taken/all credits paid back duly (A30) and all credits at this bank paid back duly (A31) showed better reference group compared to the other groups. Both of response groups also have less potential fraud rather than categories of existing credits paid back duly till now (A32), delay in paying off in the past (A33) and critical account/other credits existing or not at this bank (A34) [27].

As a concluding remark, the bank or financial institution may use RIDIT analysis as one of the tools to detect the fraud cases in the institutions. The RIDIT analysis serves as an easy tool to implement and interpret by the user especially the management level.

In future, the result of this study can be extended to which is the best reference group could be chosen if we have more than two groups be the reference groups. Also, as we can see the value of the tests is affected by the size of the observations in the reference groups since it is dependent on mean RIDIT. If the mean RIDIT is more than 0.5 and a number of frequency data is large, it will reflect the value of the tests which show highly significant.

Since fraud is seen as a failure of monitoring among the staffs by the management, the management may opt in the finding the cause by using the optimization problem-solving. The differential evolution as one of the approaches under parameter searching may help any companies to find the reasons for the problems as early as possible [28].

One of the tools that the insurance and financial institutions need to consider in detecting fraud is the Fraud prevention systems (FPSs). Although the system is insufficient to provide adequate security to the electronic commerce systems, the collaboration of Fraud Detection Systems (FDSs) with FPSs maybe the right solution to secure the e-commerce systems. Nevertheless, some issues and challenges delay the performance of FDSs such as concept drift, supports real-time detection, skewed distribution, a large amount of data, etc. [29].

#### ACKNOWLEDGMENT

The authors thankfully acknowledge that this paper research was financially supported by Universiti Teknologi Malaysia (UTM) and Universiti Teknologi MARA (UiTM) as part of author's PhD study.

#### REFERENCES

- [1] T. Xiong, S. Wang, A. Mayers, and E. Monga, "Personal bankruptcy prediction by mining credit card data," *Expert Systems with Applications*, vol. 40, pp. 665-676, Feb. 2013.
- [2] M. Sasirekha, I. S. Thaseen, and J. S Banu, *An Integrated Intrusion Detection System for Credit Card Fraud Detection*, Ser. Advances in Computing and Information Technology. Berlin, Germany: Springer, 2012, vol. 176.
- [3] V. Zaslavsky and A. Strizhak, "Credit card fraud detection using self-organizing maps," *Information and Security*, vol. 18, pp. 48-63, 2006.
- [4] S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert Systems with Applications*, vol. 39, pp. 12650-12657, Nov. 2012.
- [5] A. Abdallah, M.A.Maarof, A. Zainal, "Fraud Detection System : A survey," *Journal of Network and Computer Applicants*, vol. 68, pp. 90-113, 2016
- [6] W. Chen, C. Ma, and L. Ma, "Mining the customer credit using hybrid support vector machine technique," *Expert Systems with Applications*, vol. 36, pp. 7611-7616, May 2009.
- [7] J. Z. Lei and A. Ghorbani, "Improved competitive learning neural networks for network intrusion and fraud detection," *Neurocomputing*, vol. 75, pp. 135-145, Jan. 2012.
- [8] S. Akkoç, "An empirical comparison of conventional techniques, neural networks and the three stage hybrid adaptive neuro fuzzy inference system (ANFIS) model for credit scoring analysis: The case of Turkish credit card data," *European Journal of Operational Research*, vol. 222, pp. 168-178, Oct. 2012.
- [9] M. Mwanza and J. Phiri, "Fraud detection on bulk tax data using business intelligence data mining tool: A case of Zambia revenue authority," *Intelligence*, vol. 5, pp. 793-798, Mar. 2016.
- [10] G. van Capelleveen, M. Poel, R. M. Mueller, D. Thornton, and J. van Hillegerberg, "Outlier detection in healthcare fraud: A case study in the Medicaid dental domain," *International Journal of Accounting Information Systems*, vol. 21, pp. 18-31, Jun. 2016.
- [11] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Knowledge-Based Systems*, vol. 70, pp. 324-334, Nov. 2014.
- [12] M. R. Lepoivre, C. O. Avanzini, G. Bignon, L. Legendre, and A. K. Piwele, "Credit card fraud detection with unsupervised algorithms,"

- Journal of Advances in Information Technology*, vol. 7, pp. 34-38, Feb. 2016.
- [13] B. Lebichot, F. Braun, O. Caelen, and M. Saerens, "A graph-based, semi-supervised, credit card fraud detection system," *Complex Networks and Their Applications*, vol. 693, pp. 721-733, Nov. 2016.
- [14] U. Bandyopadhyay and A. Biswas, "Fixed-width confidence interval for two-stage response-adaptive designs in Ridit analysis," *Journal of Statistics and Probability Letters*, vol. 108, pp. 45-51, Jan. 2016.
- [15] B. P. J. F. Agostinho and C. J. Cherry, "The significance of claims fraud in microinsurance and a statistical method to channel limited fraud identification resources," in *Proc. ASSA'14*, 2014, p. 7.
- [16] P. L. Brockett, R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, "Fraud classification using principal component analysis of Ridits," *Journal of Risk and Insurance*, vol. 69, pp. 341-371, Sep. 2002.
- [17] F. Buonora, R. Senesi, and O. Thonnard, "An application of unsupervised fraud detection to passenger name records," in *Proc. IEEE DSNW'16*, 2016, p. 54.
- [18] E. G. Uwawunkonye and O. I. C. Anaene, "A comparative study between Ridit and modified Ridit analysis," *American Journal of Theoretical and Applied Statistics*, vol. 2, pp. 248-254, 2013.
- [19] S. Mahapatra and S. Kumar, "A Ridit approach to evaluate factors influencing online shopping behaviour," *International Journal of Internet Marketing and Advertising*, vol. 7, pp. 83-99, Jan. 2012.
- [20] I. D. J. Bross, "How to use Ridit analysis," *International Biometric Society*, vol. 14, pp. 18-38, Mar. 1958.
- [21] M. Lichman, *UCI Machine Learning Repository*, California, USA: University of California, School of Information and Computer Science, 2013.
- [22] S. R. Alenazi and K. Ahmad, "Record duplication detection in database: A review," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, 2016.
- [23] S. V. Gaikwad, S. Chougule, and S. Charhate, "Detection and prevention of sensitive data from data leak using shingling and Rabin filter," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 663-667, 2016.
- [24] J. Ai, P. Brockett, L. Golden, and M. Guillén, "A robust unsupervised method for fraud rate estimation," *Journal of Risk and Insurance*, vol. 80, pp. 121-143, Mar. 2013.
- [25] I. M. Yassin, A. Zabidi, M. S. A. M. Ali, N. M. Tahir, H. A. Hassan, H. Z. Abidin, and Z. I. Rizman, "Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 630-637, Oct. 2016.
- [26] M. N. M. Nor, R. Jailani, N. M. Tahir, I. M. Yassin, Z. I. Rizman, and R. Hidayat, "EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 793-798, Oct. 2016.
- [27] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90-113, Jun. 2016.
- [28] J. H. Ong and J. Teo, "A time-based investigation of parameter tuning in differential evolution for non-linear global optimization," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, no. 4, pp. 426-436, Jul. 2016.
- [29] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey", *Journal of Network and Computer Applications*, vol. 68, pp. 90-113, Jun. 2016.