DATA ANALYTICS AND APPLIED MATHEMATICS
E-ISSN: 2773-4854
VOLUME 6, ISSUE 2, 2025, 31 – 40
DOI: https://doi.org/10.15282/daam.v6i2.13061

RESEARCH ARTICLE

RIDIT scoring method for ranking finance fraud cases

Norbaiti Tukiman[1]*, Nur Farahin Mohd Fadzil[2], Ahmad Khudzairi Khalid[1], Nur Intan Syafinaz Ahmad[1], Wan Munirah Wan Mohamad[1], Nur Syamilah Ariffin[1], Yuanita FD Sidabutar[3]

[1]Mathematical Department, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Johor Branch, Pasir Gudang Campus, Malaysia.
[2]Mathematical Department, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Seremban, Negeri Sembilan, Malaysia.
[3]Faculty of Engineering, Batam University, Indonesia.

ABSTRACT - The arising number of fraud cases poses a serious threat to financial security. Detecting fraud is challenging due to dataset characteristics, particularly the imbalance between fraudulent and non-fraudulent activities, which may lead to biased results. To address the issue, this study introduces the Relative to an Identified Distribution (RIDIT) scoring method to identify potential financial fraud. RIDIT scores are calculated for different years in financial data, helping rank financiers' preferences based on fraud risk. The study applies this method to the New York Suspicious Activity Report Statistics dataset, which includes 14 types of fraudulent activities such as credit or debit card fraud, wire fraud, Ponzi schemes, and consumer loan fraud. Each category is ranked based on mean RIDIT scores, which range between 0 and 1. A score below 0.5 indicates lower fraud risk, while a score above 0.5 suggests higher potential fraud. Results show that credit or debit card fraud (0.3282) and wire fraud (0.4475) have the lowest potential fraud. In contrast, consumer loan fraud has the highest mean RIDIT score (0.7081), highlighting a greater risk and the need for close monitoring. The novelty of applying RIDIT Scoring in this study contributes to the financial system by providing a useful tool to rank suspicious activities. It helps the financiers and organizations to identify the high-risk of suspicious activity effectively. Thus, allows financiers minimize losses and promote financial transparency.

ARTICLE HISTORY
Received : 4th Sept 2025
Revised : 23rd Sept 2025
Accepted : 26th Sept 2025
Published : 30th Sept 2025

KEYWORDS
RIDIT
Fraudulent
Scores
Fraud risk
Ranking

1. INTRODUCTION

In recent years, fraud has attracted attention and serious concern in many industries such as financial institutions, where fraudulent activity poses a significant risk to organizations. As a result, it can harm market trust and jeopardize the stability of the financial system, leading to severe losses, disrupted resource allocation, and reputational damage, as noted by Brant and Haff [1] and Mao et al. [2]. According to Zhu et al. [3], as cited in Li et al. [4], the detection of fraud becomes more challenging due to the complexity of fraud schemes, which involve several people and transactions, making it difficult to identify abnormal behavior. These challenges include dealing with imbalanced data, overfitting, and a limited ability to precisely simulate fraudulent activities [5].

Other than that, there are several issues that arise when handling financial data, which indirectly make the process of detecting fraud more challenging. Firstly, most institutions are reluctant to expose their real fraud cases, keeping the databases as sensitive personal information. According to Gong et al. [6] and Sanusi et al. [7], financial institutions have become more vulnerable to fraud over the years, even though security precautions have been implemented. This happens because many institutions want to avoid losing profitability, as the disclosure of fraud can severely damage customer trust and loyalty [8]. Ultimately, it poses a significant threat to the stability and integrity of financial institutions. Wang et al. [9], as cited in Ngai et al. [10], defined this action as a planned act that violates policy, rules, or laws with the purpose of obtaining uncertified financial benefit.

Next, in general fraud can be detected by analyzing the Likert Scale survey which required the respondents to state their level of agreement with a series of statements. Data from the Likert Scale usually consists of numerical, ordinal and categorical data types. However, variations in consumer employing the scale might produce several biases that impact the quality of the data. This approach is classified to be suspicious due to the assumption of equal intervals between categories, which is doubtful as cited in Pouplard et al. [11]. Thus, claim data with many categorical attributes might generate multiple large clusters. Hence, Nian et al. [12] stated this scenario makes any single-class outliers detection method ineffective.

In response to the challenges in fraud detection, the Relative to an Identified Distribution (RIDIT) scoring method has been used to rank fraud cases in financial data. The RIDIT score was first introduced by Bross [13] for application in ordinal data. According to Wu [14], ordinal data can be either qualitative scales such as good, better, best or in the form of numerical values. Kumar and Bhattacharyya [15] further highlighted that RIDIT scoring can also be applied to discrete data, which consists of countable data points. In this method, an appropriate score could be assigned to each category.

***CORRESPONDING AUTHOR | Norbaiti Tukiman**| ✉ norbaiti289@uitm.edu.my

RIDIT analysis then become one of the popular statistical method used in analyzing ordered qualitative data which it is suitable for comparing an ordinal data scale such as Likert scale used in questionnaires [16]. RIDIT scoring method was used to rank and determine the order of the items based on the respondents' opinion and perceptions. Fundamentally, according to Pathak and Kumar Panda [17], RIDIT assigns a value to a response group based on their probability of being in the reference population. This method is particularly useful for analyzing the items containing ratings on at least three or more-point scale. By using RIDIT scoring method, the result obtained can be applied to organize the Likert scale items either in ascending or descending order based on their priority as cited in Kumar and Bhattacharyya [15]. This method is a convenient option for researchers due to its ease of implementation and ability to handle discrete and non-normal data [18].

Previous studies addressed the different applications of RIDIT scoring method; for examples, Senapati and Panda [18] used RIDIT scoring method by computing the mean RIDIT to identify distinct patient preferences based on their experience in Indian corporate hospitals. This method is used to rank patients' preference based on the survey questionnaires conducted among 220 patients across eight corporate hospitals. Furthermore, Mandal and Dey [16] employed RIDIT scoring method by evaluating Mean RIDIT to evaluate the socio-economic vulnerability indicators. Meanwhile, Pathak and Kumar Panda [17] use RIDIT scoring method for identifying the best service quality of the management education setting in the public universities of the North Eastern region of India. The data was analysed from the management graduates' perceived service quality in public universities of North Eastern region of India. The highest RIDIT value will be the priority in determining the best service quality.

Other than that, Kalvakolanu [19] also used RIDIT scoring method to assess the levels of data sophistication in HR functions. The data sophistication levels of HR functions with 13 items questionnaire were examined by using this method. The results obtained concluded that the lesser value of the Mean RIDIT contributes to the higher priority placed. In contrast, the higher value of Mean RIDIT shows the lower preference. Tukiman et al. [20] and Kumar and Bhattacharyya [15] employed RIDIT scoring method to examine the effect of credit card dataset in different response variables and to model consumer opinion, respectively. Consequently, most of the previous studies required to evaluate the Likert scale items to determine the best cases based on their customer or client preference. In this case, RIDIT scoring method is applied and used since it performs well for ordinal data. Although there is an alternative method such as conducting a traditional chi-square analysis that can be used to rank the cases, but it is inappropriate way which may lead to omission of important information on the natural ordering of categories as noted by Uwawunkonye [21].

RIDIT scoring provides a simple and accurate method for ranking the cases based on their severity or probability of being fraudulent throughout the given data set. This method is considered as a simpler method which does not require complex calculation as it allows comparison of a set of products to a reference product. Thus, the main objectives of this study are to propose RIDIT scoring method for ranking fraud cases in financial data. This study also aimed to calculate RIDIT score for suspicious activity across different years, which then the results obtained were used to rank the financiers' preferences.

2. RIDIT SCORING FRAMEWORK

There are numerous methods for ranking fraud cases especially in financial data. Nevertheless, this study seeks to compare the different response category levels based on RIDIT scoring method. Figure 1. shows the step-by-step framework of the RIDIT scoring method.

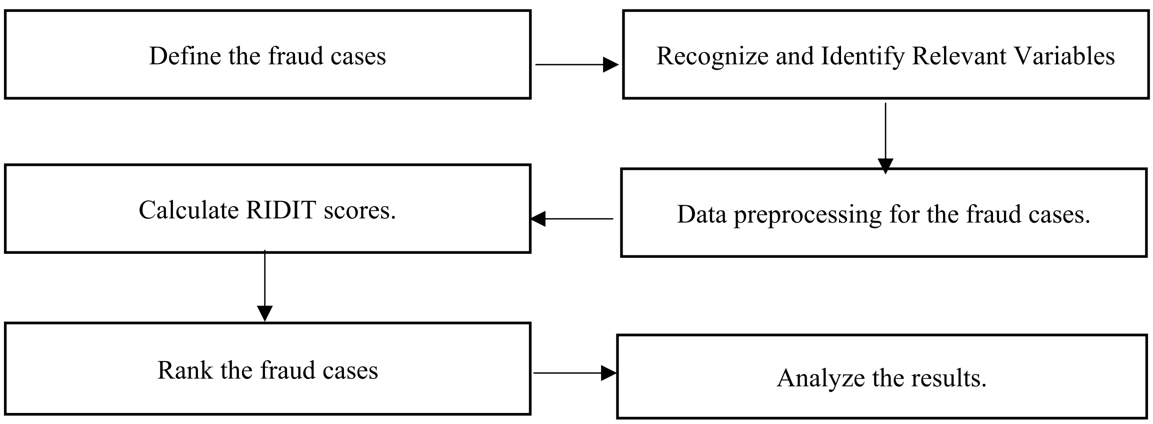

Figure 1. RIDIT scoring framework

This methodology framework outlines the steps involved in identifying and ranking fraud cases using the RIDIT scoring method. Table 1 shows breakdown of each step:

a) Define the Fraud Cases

Select a specific scenario, such as financial fraud, to focus on in the study. This step establishes the scope and context of the analysis.

b) Recognize and Identify Relevant Variables

Identify key variables that contribute to fraudulent activities. These variables could include a few variables such as transaction amounts, transaction methods, frequency, and other financial indicators.

c) Data Preprocessing for Fraud Cases

Clean, filter, and transform the dataset to prepare it for analysis. This may involve handling missing values, normalizing data, and addressing class imbalances in fraud cases.

d) Calculate RIDIT Scores

Apply the RIDIT scoring method to the dataset. This involves computing relative scores for different types of fraud cases based on their likelihood and severity.

e) Rank the Fraud Cases

Using the calculated RIDIT scores, rank the fraud cases based on their potential risk levels. The higher scores indicate a greater likelihood of fraud and vice versa.

f) Analyze the Results

Interpret the ranked fraud cases and derive insights. This analysis helps in understanding potential fraud possibility and making decisions to mitigate risks.

This framework provides a systematic approach to detecting and prioritizing fraudulent activities, enhancing financial security and transparency. The detail calculation of RIDIT score is explained in section 2.1.

2.1 RIDIT Score Calculation

In this study, there are eight steps that are used in order to calculate the RIDIT score as follows:

Step 1: Select a reference data set.
Determine the reference data set. If a specific population cannot be identified, the whole survey responses will be used as reference data set.

Step 2: Compute the sum of frequencies for each response category $(\boldsymbol{f_j})$
Sum up all of the frequencies for each response category.

$$f_j = \Sigma_j \tag{1}$$

where

$j = 1,\ldots,n$; n =Response category

Σ_j : The summation of frequencies for scale item j across all categories.

Step 3: Calculate the midpoint of the sum of frequencies (F_j).
Compute the midpoint of the sum of frequencies by multiplying the sum of frequencies for each response category with $\frac{1}{2}$.

$$F_j = \frac{1}{2} f_j \tag{2}$$

where

$j = 1,\ldots,n$; n =Response category

Step 4: Evaluate the midpoint of the accumulated frequencies (M_j).
Compute the midpoint of the accumulated frequencies by adding the midpoint of the sum of frequencies with the summation of the previous midpoint of the sum of frequencies and the previous value of accumulated frequencies.

$$M_j = F_j + \left(F_{j-1} + M_{j-1}\right) \tag{3}$$

where

$j = 2, \ldots, n$; n = Response category

Step 5: Calculate RIDIT value for the reference data set (r_j).
The computation of RIDIT value for reference data set is the operation on the midpoint of the accumulated frequencies divided by the total frequency of the responses.

$$r_j = \frac{M_j}{N} \tag{4}$$

where

$j = 1, \ldots, n$; $n =$ Response category

$N =$ Total frequency of responses.

Step 6: Calculate RIDIT value for the comparison data set (R_{ij}).
RIDIT value for the comparison data set can be obtained by using the RIDIT value of the reference data set.

$$R_{ij} = \frac{r_j \times f_{ij}}{\Sigma_i} \tag{5}$$

where

$i = 1, \ldots, m$; $j = 1, \ldots, n$;

f_{ij} : Frequency of category j for the i_{th} scale item.

Σ_i : The summation of frequencies for scale item i across all categories.

Step 7: Compute the Mean RIDIT for the comparison data set (ρ_c).
The Mean RIDIT can be computed by adding all the RIDIT value in comparison data set.

$$\rho_c = \sum R_{ij} \tag{6}$$

Step 8: Rank the fraud cases.
The fraud cases can be ranked based on the computation of RIDIT score. Cases with a higher RIDIT score show a higher relative severity or risk of fraud while the lower RIDIT scores represent a least relative severity or potential being fraud.

2.2 An Illustrative Example of RIDIT Scoring Methodology

In this study, a set of discrete data which contains the consumers' perception levels towards ten activities that lead to fraud is illustrated in Table 1. The perception levels divided into five which are level 1, level 2, level 3, level 4 and level 5. Firstly, level 1 refers to very unaware of activities that result in fraud. Next, level 2 refers to unaware, level 3 is neutral, level 4 is aware and the last is level 5 represents very aware of activities that result in fraud. In this section, we provide an example on the calculation of RIDIT scores to determine the severity of each activity being fraud based on the consumers' perception levels. This method helps to rank and find the best consumers' preference based on the RIDIT scores obtained.

Table 1. Data on the consumers' perception levels towards fraud activities.

No.	Perception Level / Activity	1	2	3	4	5
1.	Check	8	12	6	20	4
2.	Credit/Debit Card	8	12	18	26	8
3.	Fake Invoice	1	7	13	8	5
4.	Financial Statement	9	22	16	30	6
5.	Forex Scams	14	10	7	22	7
6.	Investment	27	6	9	17	10
7.	Money Laundering	4	5	12	4	43
8.	Mortgage	9	13	28	5	8
9.	Payment	1	15	8	16	10
10.	Payroll	4	2	11	27	20

The steps for using the RIDIT scoring methodology are as follows:

Step 1: Select a reference data set.
For this example, consumers' perception levels are used as reference data set to compare with the other sets of data.

Step 2: Compute the sum of frequencies for each response category (f_j).
The sum of frequencies for each response category obtained by using equation (3.1). In this example, we find the sum of frequencies for each consumers' perception levels. For example:

$$f_1 = 8 + 8 + 1 + 9 + 14 + 27 + 4 + 9 + 1 + 4 = 85$$

The calculations are repeated. The sum of frequencies for each consumers' perception levels are shown in Table 2.

Table 2. The midpoint of the sum of frequencies for each consumers' perception levels.

Perception levels	1	2	3	4	5
The sum of frequencies, f_j	85	104	128	175	121

Step 3: Calculate the midpoint of the sum of frequencies (F_j).
Compute the midpoint of the sum of frequencies by using equation (2). Below shows the computation of the midpoint of the sum of frequencies for $j = 1$

$$F_1=\frac{1}{2}(85)=42.5$$

The calculations are repeated to compute the midpoint of the sum of frequencies. The result is shown in Table 3.

Table 3. The midpoint of the sum of frequencies for each consumers' perception levels.

Perception levels	1	2	3	4	5
The sum of frequencies, f_j	85	104	128	175	121
The midpoint of the sum of frequencies, F_j	42.5	52	64	87.5	60.5

Step 4: Evaluate the midpoint of the accumulated frequencies (M_j).
Calculate the accumulated frequencies by using equation (3). Below shows the computation of the midpoint of the sum of frequencies for $j = 2$

$$M_2 = 52 + (42.5 + 42.5) = 137$$

The calculations are repeated to compute the midpoint of the accumulated frequencies. The result is shown in Table 4.

Table 4. The midpoint of the accumulated frequencies.

Perception levels	1	2	3	4	5	
The sum of frequencies, f_j	85	104	128	175	121	613
The midpoint of the sum of frequencies, F_j	42.5	52	64	87.5	60.5	
The midpoint of the accumulated frequencies, M_j	42.5	137	253	404.5	552.5	

Step 5: Calculate RIDIT value for the reference data set (r_j).
RIDIT value for reference data set obtained by using equation (4). For example:

$$r_1 = \frac{42.5}{613} = 0.0693$$

Table 5. RIDIT value for reference dataset.

Perception levels	1	2	3	4	5	
The sum of frequencies, f_j	85	104	128	175	121	613
The midpoint of the sum of frequencies, F_j	42.5	52	64	87.5	60.5	
The midpoint of the accumulated frequencies, M_j	42.5	137	253	404.5	552.5	
RIDIT value for reference data set, r_j	0.0693	0.2235	0.4127	0.6599	0.9013	

Step 6: Calculate RIDIT value for the comparison data set (R_{ij}).
Table 6 indicates the summation of frequencies for each fraud activity across all the categories.

Table 6. The summation of frequencies for each fraud activity across all the categories.

No.	Activity	1	2	3	4	5	Total
1.	Check	8	12	6	20	4	50
2.	Credit/Debit Card	8	12	18	26	8	72
3.	Fake Invoice	1	7	13	8	5	34
4.	Financial Statement	9	22	16	30	6	83
5.	Forex Scams	14	10	7	22	7	60
6.	Investment	27	6	9	17	10	69
7.	Money Laundering	4	5	12	4	43	68
8.	Mortgage	9	13	28	5	8	63
9.	Payment	1	15	8	16	10	50
10.	Payroll	4	2	11	27	20	64

RIDIT value for comparison data set obtained by using equation (5). Fraud activities be the comparison data set. For example:

$$R_{1,1} = \frac{0.0693 \times 8}{50} = 0.0111$$

Table 7. RIDIT value for comparison dataset.

No.	Activity	1	2	3	4	5
1.	Check	0.0111	0.0536	0.0495	0.2639	0.0721
2.	Credit/Debit Card	0.0077	0.0372	0.1032	0.2383	0.1001
3.	Fake Invoice	0.0020	0.0460	0.1578	0.1553	0.1325
4.	Financial Statement	0.0075	0.0592	0.0796	0.2385	0.0652
5.	Forex Scams	0.0162	0.0372	0.0482	0.2420	0.1052
6.	Investment	0.0271	0.0194	0.0538	0.1626	0.1306
7.	Money Laundering	0.0041	0.0164	0.0728	0.0388	0.5699
8.	Mortgage	0.0099	0.0461	0.1834	0.0524	0.1145
9.	Payment	0.0014	0.0670	0.0660	0.2112	0.1803
10.	Payroll	0.0043	0.0070	0.0709	0.2784	0.2817

Step 7: Compute the Mean RIDIT for the comparison data set (ρ_c).
Calculate the mean RIDIT by using equation (6). Below shows the computation of the mean RIDIT for the comparison data set:

$$\rho_{check} = 0.0111 + 0.0536 + 0.0495 + 0.2639 + 0.0721 = 0.4503$$

The calculations are repeated to compute the mean RIDIT for comparison data set. The result is shown in Table 8.

Table 8. The Mean RIDIT for comparison data set.

Suspicious Activity	Mean RIDIT, ρ_c
Check	0.4503
Credit/Debit Card	0.4866
Fake Invoice	0.4937
Financial Statement	0.4500
Forex Scams	0.4487
Investment	0.3936
Money Laundering	0.7021
Mortgage	0.4063

Payment	0.5259
Payroll	0.6423

Step 8: Rank the fraud cases
Table 9 below shows the results of calculated mean RIDIT scores and the ranking for each fraud activity.

Table 9. Ranking of the suspicious activities.

Suspicious Activity	Mean RIDIT, ρ_c	Rank
Check	0.4503	5
Credit/Debit Card	0.4866	6
Fake Invoice	0.4937	7
Financial Statement	0.4500	4
Forex Scams	0.4487	3
Investment	0.3936	1
Money Laundering	0.7021	10
Mortgage	0.4063	2
Payment	0.5259	8
Payroll	0.6423	9

The mean RIDIT scores obtained were used to rank all the fraud activities. The scores below 0.5 or closer to 0 represent that the activity has the lowest potential of being associated with fraud. From the table, investment and mortgage activities have the lowest mean RIDIT scores, reaching the values of 0.3936 and 0.4063 respectively. In contrast, the scores above 0.5 and closer to 1 show that the activity has the highest potential of being associated with fraud. According to the results, money laundering activity has the highest mean RIDIT score, achieving 0.7021 which is closer to 1.

Overall, this section discussed the RIDIT scoring methodology which is used to accurately rank fraud cases in financial data. Other than that, it also provides RIDIT scoring framework which helps organize the process of ranking fraud cases in financial data. An illustrative example calculation of RIDIT scores provided to give a clear explanation of this method. As a result, the severity of each activity being fraud can be ranked accurately based on the RIDIT scores obtained.

3. RESULTS AND DISCUSSION

In this study, data was collected from Suspicious Activity Report Statistics (SAR Stats) available online at https://www.fincen.gov/reports/sar-stats, which the data filed by financial institutions to record any suspicious activity. The database includes about 28,087 records of fraud suspicious activity collected over a ten-years period in New York. The variables included in this dataset are the year, suspicious activity, and count of each suspicious activity. Table 10 shows the Suspicious Activity Report Statistic (SAR Stats) from the year 2015 to 2024.

Table 10. The Suspicious Activity Report Statistic (SAR Stats) from 2015 to 2024.

No.	Year / Activity	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
1.	ACH	39	18	21	39	62	377	602	342	715	1016
2.	Advance Fee	0	0	0	3	3	7	5	14	11	11
3.	Business Loan	51	10	1	2	4	9	7	3	2	154
4.	Check	72	79	55	41	43	40	49	72	350	416
5.	Consumer Loan	12	24	21	33	30	35	10	11	10	733
6.	Credit/Debit Card	2052	1810	1506	533	515	522	431	813	925	764
7.	Healthcare Insurance	0	1	1	1	5	0	4	3	5	6
8.	Mail	4	5	8	13	11	7	11	16	21	22
9.	Mass-Marketing	2	6	4	5	8	9	7	15	21	6
10.	Other Fraud	183	279	305	330	282	491	920	1129	1880	1992
11.	Ponzi Scheme	0	0	0	0	10	13	13	59	42	32
12.	Pyramid Scheme	5	3	1	2	2	2	3	47	41	4
13.	Securities	0	0	0	5	19	25	13	10	11	63
14.	Wire	44	104	100	119	104	327	375	185	166	98

There are various types of suspicious activity involve in this dataset such as Automated Clearing House (ACH), advance fee, business loan, check, consumer loan, credit/debit card, healthcare insurance, mail, mass-marketing, Ponzi scheme, pyramid scheme, securities, wire and other fraud. From the Table 10, it can be estimate that credit or debit card activity has the highest reported cases among all the suspicious activities. In 2015, it can see that this activity has reach a value of 2052 cases, where it is higher than "other fraud" activity that only has 1992 cases at maximum value by the year 2024. This indicates that credit or debit activity has the highest potential to be classified as a fraudulent case. In contrast, securities are one of the activities that has the lowest number of reported cases. Thus, this represents that securities activity have the lowest potential to be classified as a fraudulent case.

Generally, the data can be transform into visual clustered bar graph and analyze. The value provides a rough estimation and shows a general understanding on how to rank the suspicious activities based on the occurrence of cases across all the years. However, it is not an accurate method to determine the potential fraud cases simply based on the graph or observation based on highest or lowest value on dataset. Therefore, this study seeks to use RIDIT scoring method, which it helps to accurately rank the suspicious activities based on the computation of RIDIT scores involve in this method. Firstly, all the years were selected as a reference dataset to avoid unbiased comparison. To rank the severity of each activity in financial data, mean RIDIT scores were calculated for each type of suspicious activity based on the cumulative probabilities obtained from the reference dataset using equation (1) to (6) as in section 2.1.

Table 11 below shows the result of calculated mean RIDIT scores and the rank for each type of suspicious activity. All of this was calculated using R Programming Version 4.3.3 to obtain the mean RIDIT scores and ranking for each activity.

Table 11. Rank of the suspicious activity.

Suspicious Activity	Mean RIDIT, ρ_c	Rank
ACH	0.6464	13
Advance Fee	0.6053	8
Business Loan	0.6337	11
Check	0.6203	9
Consumer Loan	0.7801	14
Credit/Debit Card	0.3282	1
Healthcare Insurance	0.5719	5
Mail	0.5246	4
Mass-Marketing	0.5025	3
Other Fraud	0.6038	7
Ponzi Scheme	0.6331	10
Pyramid Scheme	0.5825	6
Securities	0.6417	12
Wire	0.4475	2

The scores represent the relative probability of suspicious activity compared to all years used as reference group, which the values fell between 0 and 1 due to the calculation of cumulative probabilities involved in this method. Scores below 0.5 and closer to 0 represent a lower relative severity to fraud. This implies that the suspicious activity has a lower potential to be fraud. In contrast, scores above 0.5 and closer to 1 indicate a higher severity or likelihood of fraud, illustrating that the suspicious activity has a higher potential to be fraud.

The suspicious activity can be categorized into three categories, which are high, moderate and low potential activity being fraud. According to results from Table 11, credit or debit card and wire activities have the lowest RIDIT scores, reaching the values of 0.3282 and 0.4475 respectively. It shows that both values fall below 0.5 and indicates that the activities have low potential being fraud. Next, mass-marketing, mail, healthcare insurance and pyramid schemes can be considered as moderate risk activities to fraud. Mean RIDIT scores reached the values of 0.5025, 0.5246, 0.5719 and 0.5825 respectively. Moderate risk refers to activity that is neither high potential being fraud nor the lowest potential being fraud. Lastly, consumer loan has the highest mean RIDIT score, achieving 0.7801 which is closer to 1. It indicates that consumer loan is suspicious among all the activities.

4. CONCLUSIONS AND RECOMMENDATION

This section provides a summary of the significant results from the study on RIDIT scoring method. The first objective of this study has been successfully achieved through proposing the RIDIT scoring method for ranking fraud cases in financial data. This method was chosen due to its advantages in handling high-dimensional datasets and improving fraud detection accuracy. Unlike traditional method like finding the mean, median or standard deviation, which often struggle and perform poorly in large and complicated datasets. Hence, it is clearly showed that the first objective has been achieved since RIDIT scoring method offers a simpler and effective approach for ranking fraud cases in financial data compared to the traditional method. Next, the calculation of RIDIT score was implemented to the dataset that including the suspicious activity reported across multiple years. The result of the computation of RIDIT scores was analysed in section 3. Lastly, the objective has been successfully achieved by ranking the financiers' preferences based on the RIDIT scores obtained. The rank is important for decision-making process, which may avoid the financiers from any losses. Other than that, it also helps the organizations to improve their strategy to fit more closely the financiers' preferences, where it will benefit both parties.

Even though the objectives of this study have been achieved, RIDIT scoring method encountered some limitations. Firstly, the dataset showed an imbalanced number of suspicious activities across different categories. For example, suspicious activity involving a credit or debit card showing prominently higher values compared to other activities. This situation could lead to a significant challenge, which the results may be skewed by the unequal dataset. However, this constraint does not restrict this method to accurately rank the severity each of the suspicious activity. It is because the calculation of RIDIT scores involved dependence to the distribution of the entire dataset. Hence, it helps to balance the categories as the scores are computed from the cumulative counts instead of frequency counts. Furthermore, RIDIT scoring method may produce some conflicting outcomes due to the selection of the years chosen as a reference group. Since the suspicious activities were ranked based on the computation of RIDIT scores involving the calculation of cumulative distributions, this approach allows fair comparisons without considering the raw count for each year. Consequently, RIDIT scoring method was able to handle skewed data without seeking for some adjustments.

In this study, the limitations did not affect the process of ranking by RIDIT scores, but the findings of this study could suggest reaching a more balanced dataset to enhance the analysis in future studies. This demonstrates that balanced datasets provide a clearer comparison between categories of suspicious activity and avoiding from biases. Additionally, researchers could conduct a sensitivity analysis by testing difference reference group. The results obtained can be used to determine on how the selection of reference group influences the score and ranking. Indirectly, this action helps to ensure the robustness of the RIDIT scoring method in future studies.

Other than that, this study also contributes to the financial system by providing a useful tool to rank suspicious activities based on the RIDIT scores. It helps the financiers and organizations to identify the high-risk of suspicious activity effectively. It is crucial for financiers to analyse the severity of each suspicious activity before investing their money, as this helps to avoid any losses. On the other hand, this study also helps the organizations to prevent further damage, take appropriate action and improve their strategy to fit more closely the financiers' preferences, as this will benefit both parties.

ACKNOWLEDGEMENTS

Institution(s)

All the authors would like to express their gratitude to Universiti Teknologi MARA (UiTM) Cawangan Johor Kampus Pasir Gudang for the approval attending conference and financial support.

Fund

The authors would like to thank the Universiti Teknologi MARA (UiTM) Cawangan Johor Kampus Pasir Gudang as financial support under UiTM internal grant (100-UiTMPG(PT.9/6)).

Individual Assistant

NA

AUTHOR CONTRIBUTIONS

Norbaiti Tukiman (Conceptualization; methodology; supervision; writing- original draft preparation), Nur Farahin Mohd Fadzil (Abstract, Literature review; methodology; writing - original draft; Resources), Ahmad Khudzairi Khalid (Literature review; editing), Nur Intan Syafinaz Ahmad (Literature review; data analysis) Wan Munirah Wan Mohamad (data analysis), Nur Syamilah Ariffin (Conclusion), Yuanita FD Sidabutar (Discussion the results). All authors contributed to read and approve final manuscript.

DECLARATION OF ORIGINALITY

The authors declare no conflict of interest to report regarding this study conducted.

GENERATIVE AI DECLARATIONS

The authors claim that artificially intelligent-assisted technologies in the form of generative AI were not used to generate content, ideas, or theories. We have just utilised AI to enhance readability and refine the language. This was used with extreme human control and oversight. The authors take full responsibility for reviewing and approving the content.

REFERENCES

[1] Brant SB, Haff IH. The Fraud Loss for Selecting the Model Complexity in Fraud Detection. Journal of Applied Statistics. 2023;50(10):2209–27.

[2] Mao X, Liu M, Wang Y. Using GNN to detect financial fraud based on the related party transactions network. Procedia Comput Sci. 2022;214(C):351-8.

[3] Zhu X, Ao X, Qin Z, Chang Y, Liu Y, He Q, Li J. Intelligent Financial Fraud Detection Practices in Post-Pandemic Era. The Innovation. 2021;2(4):100176.

[4] Li J, Li N, Xia T, Guo J. Textual Analysis and Detection of Financial Fraud: Evidence From Chinese Manufacturing Firms. Economic Modelling. 2023;126:106428.

[5] Zhao C, Sun X, Wu M, Kang L. Advancing financial fraud detection: Self-Attention Generative Adversarial Networks for Precise and Effective Identification. Finance Research Letters. 2024; 60:104843.

[6] Gong Y, Li J, Xu Z, Li G. Detecting Financial Fraud using Two Types of Benford Factors: Evidence from China. Procedia Comput Sci. 2022; 214:656-63.

[7] Sanusi ZM, Rameli MNF, Isa YM. Fraud Schemes in The Banking Institutions: Prevention Measures to Avoid Severe Financial Loss. Procedia Economics and Finance. 2015;28:107-13.

[8] Van Deventer M, Sharp K-L, Żelazny R, Kot S. Predicting South African Consumers' Intention to Continue Using Their Preferred Retail Bank's Services: A Model Validation. Entrepreneurial Business and Economics Review. 2024;12(4):199-214.

[9] Wang Y, Ashton JK, Jaafar A. Financial Statement Fraud, Recidivism and Punishment. Emerging Markets Review. 2023;56:101033.

[10] Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X. The Application Of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and An Academic Review of Literature. Decision Support Systems. 2011;50(3):559-69.

[11] Pouplard N, Qannari EM, Simon S. Use of Ridits to Analyse Categorical Data in Preference Studies. Food Quality and Preference. 1997;8(5):419-22.

[12] Nian K, Zhang H, Tayal A, Coleman T, Li Y. Auto Insurance Fraud Detection Using Unsupervised Spectral Ranking for Anomaly. Journal of Finance and Data Science. 2016;2(1):58-75.

[13] Bross IDJ. How to Use Ridit Analysis. Biometrics. 1958;14:18-38.

[14] Wu CH. On the Application of Grey Relational Analysis and Ridit Analysis to Likert Scale Surveys. In International Mathematical Forum. 2007;2(14).

[15] Kumar RV, Bhattacharyya S. Modeling Consumer Opinion Using Ridit and Grey Relational Analysis. In: Handbook of Research on Intelligent Techniques and Modeling Applications in Marketing Analytics. IGI Global; 2016. p. 185-201.

[16] Mandal K, Dey P. Coastal. Vulnerability Analysis and Ridit Scoring of Socio-Economic Vulnerability Indicators – A Case of Jagatsinghpur, Odisha. International Journal of Disaster Risk Reductio. 2022;79:103143.

[17] Pathak VK, Kumar Panda R. Application of Ridit Analysis in Prioritizing Perceived Service Quality Dimensions of Management Graduates in Indian Universities. 2022. Available from: https://www.researchgate.net/publication/361756781

[18] Senapati S, Panda RK. Prioritizing the Precursors of Patients' Experience in Indian Corporate Hospitals: Application of Hybrid Ridit-GRA Approach. Asia Pacific Journal of Health Management. 2023;18(3).

[19] Kalvakolanu S. Assessment of Data Sophistication in HR Functions by Applying Ridit Analysis. Journal of Mechanics of Continua and Mathematical Sciences. 2019;14(5).

[20] Tukiman N, Ahmad N, Mohamed S, Othman ZS, Shafee CMNM, Rizman ZI. Credit Card Detection System Based on Ridit Approach. International Journal on Advanced Science, Engineering and Information Technology. 2017;7(6):2071.

[21] Uwawunkonye EG. A Comparative Study Between Ridit and Modified Ridit Analysis. American Journal of Theoretical and Applied Statistics. 2013;2(6):248.