An Efficient Privacy-Preserving Credit Score System Based on Non interactive Zero Knowledge Proo

[1] Shaik Abeed Basha, [2] P.Goutham Kumar, [3] Deverakonda Mallikerjuna, [4] Kumbha Ramu,
CSE Department,
[1,2,3,4] Assistant Professor, Dhruva Engineering Collage, Hyderabad.
Shree Engineering Collage, Hyderabad.

Abstract. This study aims to compare the performance of a K-means matrix analysis with that of a larger-sample linear regression approach in identifying instances of personal loan fraud. Tools and Techniques for Scientific Investigation: In order to see how well credit card analysis might detect fraudulent personal loans, we utilized an average accuracy rate of 10-80 samples. We assess how well a linear regression method and an innovative K-means strategy detect personal loan fraud in credit card records. A sample size of 10 is used to find the average accuracy of the two methods, and it is progressively increased to 80. End result: Results show that the K-means algorithm achieved an average accuracy rate of 92.75% with a standard deviation of 3.67849 and a standard error mean of 1.56109. In contrast, the linear regression approach has an average accuracy rate of 86.53%, a standard deviation of 1.56109, and a mean standard error of 0.63603. Evidently, the two sets are statistically different, as a p-value of 0.01 ($P<0.05$) suggests. T-test with two tails applied to distinct samples. The novel K-Means Method outperforms the Linear Regression Process in predicting the detection of personal loan fraud, according to this research.

Fraud involving credit cards, new k-means algorithm, linear regression approach, precision

INTRODUCTION

Credit cards are rapidly gaining popularity as the most effective also stress-free method of acquiring products besides services, by way of they can be utilized both in-store and online. As a result, it has become critical to develop a solution to the credit card data security problem as well as a way for detecting fraud credit card transactions [1]. Numerous Data drawing out approaches utilized in the domain of credit card scam recognition throughout years. Using clustering and anomaly detection, this study aims to build a fraud identification algorithm that can identify credit card fraud as effectively as feasible [2]. The primary aim of this research is to examine the detection correctness of two hybrid techniques [3].

In the past five years there have been about 104 articles published by IEEE Explore, 45 articles by Springer and around 50 Google Scholar for credit card scam discovery. The majority of research on credit card fraud detection has been on pattern matching, the process through which anomalous patterns are distinguished from regular patterns (Kadoka et al. 2021). In recent years, several strategies for detecting credit card fraud have been presented with misrepresentation location [4]. The article [5] highlighted how a vast amount of transaction data might be used to discriminate between fraudulent and authentic models using artificial neural networks with built-in learning capabilities. The Map That Organizes Itself Credit card transactions were classified into four groups using a neural network (an unsupervised artificial intelligence technique used in this study): high-risk, low, high, and risky [6]. When it comes to the results and data that were calculated. Taking into account the most fundamental splitter/differentiator in input components, this method divides the whole population or test into two relatively homogenous subgroups [5].The foremost drawback of the existing research is that only a small amount of loan fraud or credit card fraud detection features are trained with values which leads to less accuracy. A more accurate prediction requires more samples, which results in the model being slower. As an author, the machine learning based

personal loan fraud detection is solution zed and has been carried out in this paper [6]. The major goal of this research is to recognize the customer's personal loan account and to check with the help of an application to detect the bank's fraud account using linear regression and an innovative K means technique.

MATERIALS AND METHODS

Researchers from the Cloud Computing Lab at the Saveetha Institute of Medical and Technological Sciences at the Saveetha School of Engineering in Chennai, India, conducted the study. There were 964 survey takers in all, split evenly between the two groups. Under these conditions—an enrollment ratio of 1, alpha quality of 0.05, beta quality of 0.2, and a confidence range of 95 percent—the G-power pretest yielded an 80% score of 482 [7].

The "Synthetic_Finacial_Prediction_Dataset_For_Fradu_Detection" dataset was sourced from the Google Kaggle Synthetic financial datasets used by Edgar and Kaggle. Learn how to utilize clinicalc.com to identify loan fraud using this dataset [8]. The novel K Means Method uses a split dataset, with half used for training and half for testing.

Linear Regression Algorithm

Among the many supervised machine learning methods, one may find linear regression. A regression procedure is carried out by it. A target prediction value is modeled using independent variables in regression. Predictions and the establishment of relationships between variables are its primary applications. The amount of independent variables used by various regression models and the kinds of correlations between them vary. A method for forecasting the value of a dependent variable (y) from the value of an independent variable (x) is linear regression. This regression method creates a straight line between the input (x) and output (y) variables. "Linear Regression" came from this process. What follows is a representation of the linear regression equation, with the stages shown below.

First Step: Put the settings in place.

The second step is to forecast the dependent variable's value using the value of the independent variable.

Step 3: Calculate the prediction error for each data point.

Step 4: Calculate the partial derivative with respect to variables a0 and a1.

Step 5: Calculate and add the costs associated with each number.

Step 6: The values of a0 and a1 should be updated.

Novel K Means Algorithm

One popular and successful method for unsupervised machine learning is K-means clustering. It is used to address several challenging unsupervised machine learning problems. The goal of K-means clustering is to group items that share characteristics into sets. It sorts the items into groups according on how similar they are. The three-stage process that makes up the K-means clustering method is: Choosing the input k-values, initializing the centroid, and calculating the average.

First, get the dataset.

Step2: Transform the information into a data traces strategy

Thirdly, oversample from time to time using the ROSE pack.

The fourth step is to figure out what proportion of the data will go into data preparation and testing.

Step5: Provide 70% of the data for getting ready and the rest for testing..

Step6: Allocate train dataset to the models

Step7: Select the computation between 3 estimations

Step8: For each estimation, make assumptions for the test dataset.

Step9: Compute precision for individual computation

Step10: Apply centroids for separate computation

Step11: Find the optimum value by comparing the computations for all of the variables.

STATISTICAL ANALYSIS

The research is carried out by use of the statistical application IBM SPSS version 26. One example of IBM's data analysis software is SPSS. The Novel K Means Procedure and the Linear Regression Method were compared using two separate sample groups in SPSS version 21. Customer ID and account number are independent variables, while accuracy is the dependent variable. Averages and standard deviations were determined using the data from the independent sample T-test.

RESULTS

In applying various training and testing datasets by adjusting the number of records in a dataset, After comparing a machine learning model that used linear regression with one that used the novel K Means method and 10 samples, the results are shown. The accuracy of both networks was shown to decline as the sample size was increased. Finally, the estimated average accuracy appears to be growing as the number of samples increases, and so the same is reported.

Table 1 shows the accuracy of the innovative K Means procedure and linear regression Method for varying size of sample. The accuracy rates differ for different test sizes from N=10 to N=80 in decimals for both the algorithms and the accuracy values are reported. For the novel K Means algorithm the accuracy values decreased from 88% to 75% with N = 10 to N = 80. Similarly for the linear regression algorithm the accuracy values decreased from 91% to 82% with N = 10 to N = 80. On the other hand, the linear regression technique and the innovative K-means approach both achieve average accuracy levels of 86.53% and 92.75%, respectively. In comparison to the linear regression method, the new K Means technique consistently produces much better results in terms of stated accuracy.

Comparison of linear regression and K-Means algorithms for various sample sizes (TABLE 1)

S.No	No. of Samples	Accuracy of Novel K Means Algorithm	Accuracy of Linear Regression Algorithm
1	10	88	91
2	20	81	82
3	40	76	85
4	60	75	81
5	80	75	82
Avg. Accuracy		92.75%	86.53%

Table 2 represents the group statistics results of the novel K Means procedure and linear regression method. The K Means System has an mean accuracy of 92.75%, with standard deviation of 3.67846 and standard mean error of 1.56109. Whereas for Linear Regression Method the mean accuracy reported is 86.53%, at 0.63603 standard deviation and 2.51059 standard deviation (TABLE 2). With a mean accuracy of 92.75% and a standard deviation of 3.67846, the unique K Means method outperforms the linear regression strategy, which has a mean accuracy of 86.53% and a standard deviation of 2.51559, according to group data.

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	**Novel K Means Algorithm**	10	92.7508	3.67849	1.56109
	Linear Regression Algorithm	10	86.5336	2.51059	0.63603

The data in Table 3 were analyzed using both the more classic linear regression method and the more contemporary K Means System. The significance level for the accuracy rate is 0.001. Different people's samples We will utilize a 95% confidence interval T-test to compare the Novel K Means strategy with the linear regression methodology if the significance level is p=0.001. Other statistics, such as the standard error, differences between the lower and upper ranges, significance (2-tailed), and so on, reveal a significant value, and the unique k-means approach fared better in this independent sample test.

The Novel K Means method and a linear regression approach are compared in this two-tailed sample-independent T-test. With a p-value of just 0.01 and a confidence interval of 91%, the findings are shown in Table 3.

Group		The Levene's Test for Variance Equality		t-test for equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Accuracy	Equal variances assumed	.082	.788	12.255	19	.001	5.96310	.45532	6.39668	8.96601
	Equal variances not assumed			12.255	18.303	.001	5.96310	.45532	6.0446	8.65874

A bar chart displaying the mean accuracy of the novel K Means technique and linear regression algorithm is shown in Figure 1. The bar charts show the comparison of the standard errors and mean accuracy for detecting loan fraud. X-Axis: linear regression algorithm vs. K Means algorithm Axis Y: Accuracy mean. There is a standard deviation of 1 in the error bars. Both the mean accuracy and the standard deviation are improved by the new K Means technique (92.75% vs. 86.53% for the Linear regression algorithm).

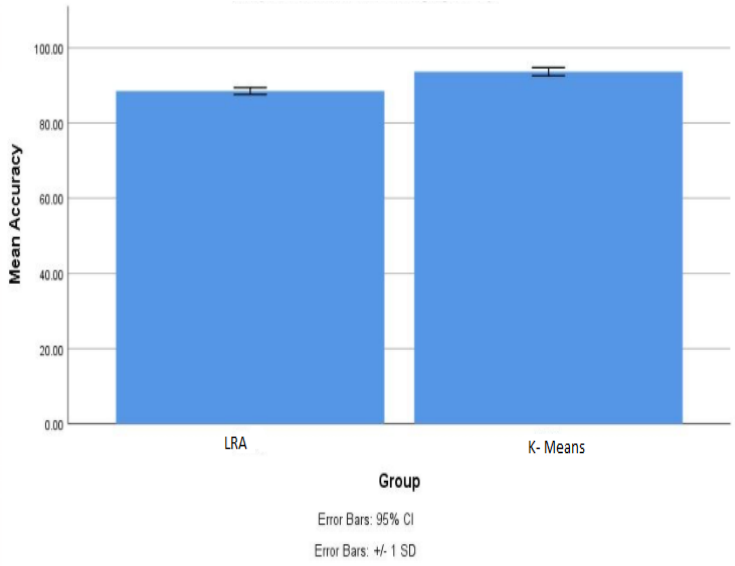

FIGURE 1. Loan fraud detection standard errors and average accuracy are shown in bar charts. By comparing the two approaches' average accuracy and standard deviation, K Means outperforms Linear Regression. On the X-axis, you can see the linear regression and K-Means methods. The accuracy mean is shown on the Y-axis. A one-standard-deviation characterizes the margins of error.

DISCUSSION

The results show that the new K Means technique outperforms the linear regression method by a wide margin, with an accuracy of 92.75%. Based on the methodologies proposed, this study presents a personal loan fraud detection system. A computation is performed that connects the Innovative K-means Procedure with the linear regression Method in order to determine which estimate is best for detection.

Supervised learning computation techniques often use a previously defined goal variable in order to solve classification issues. Any combination of known and unknown variables, as well as yield factors, may be used using this strategy. By taking into account the simplest splitter/differentiator in input components, this approach divides the complete population or test into two similar groups. The beneficial point to highlight in the present work is that it detects fraudulent activities with higher accuracy. Moreover, with increasing samples the accuracy drops, however, the accuracy reported with different samples can be averaged for a total prediction. In other words, with higher samples, it is challenging to perform better performance.

The limitations of the present study lies in data acquisition, where the public dataset or the real world dataset may consist of skewed dataset. Hence, it will challenge the accuracy as there is high imbalance in the data gathered for personal fraud detection which is a crucial aspect to be addressed. Thus, the future scope of the present work can be extended in applying machine learning techniques for addressing the data imbalance with some pre-processing techniques and making the data clean and balanced.

CONCLUSION

In this paper, implementation of credit card fraud detection is done with several statistical factors relying on K Means algorithm (92.75%) and linear regression algorithm (86.53%). The accuracy of loan fraud detection reported with K Means algorithm is observed to be higher than the linear regression algorithm.

REFERENCES

1. H. Wen and F. Huang, “Personal Loan Fraud Detection Based on Hybrid Supervised and Unsupervised Learning,” 2020 5th IEEE International Conference on Big Data Analytics (ICBDA). 2020. doi: 10.1109/icbda49040.2020.9101277.
2. T. Pousada, J. Garabal-Barbeira, C. Martínez, B. Groba, L. Nieto-Riveiro, and J. Pereira, “How Loan Bank of Assistive Technology Impacts on Life of Persons with Amyotrophic Lateral Sclerosis and Neuromuscular Diseases: A Collaborative Initiative,” Int. J. Environ. Res. Public Health, vol. 18, no. 2, Jan. 2021, doi: 10.3390/ijerph18020763.
3. L. M. Fuhrer, M.-A. Ramelet, and J. Tenhofen, “Firms’ participation in the Swiss COVID-19 loan programme,” Swiss J Econ Stat, vol. 157, no. 1, p. 2, May 2021.
4. T. Baer and S. Schnall, “Quantifying the cost of decision fatigue: suboptimal risk decisions in finance,” R Soc Open Sci, vol. 8, no. 5, p. 201059, May 2021.
5. F. Ahmadi, H. Farrokh-Eslamlou, H. Yusefzadeh, and C. Alinia, “Incidence of household catastrophic and impoverishing health expenditures among patients with Breast Cancer in Iran,” BMC Health Serv. Res., vol. 21, no. 1, p. 327, Apr. 2021.
6. S. Malovaná, M. Hodula, and J. Frait, “What Does Really Drive Consumer Confidence?,” Soc. Indic. Res., pp. 1–29, Feb. 2021.
7. S. Mayordomo, A. Moreno, S. Ongena, and M. Rodriguez-Moreno, “‘Keeping It Personal’ or ‘Getting Real’? On the Drivers and Effectiveness of Personal versus Real Loan Guarantees,” SSRN Electronic Journal. doi: 10.2139/ssrn.2731833.
8. X. Dong et al., “An untargeted metabolomics approach to identify markers to distinguish duck eggs that come from different poultry breeding systems by ultra high performance liquid chromatography-high resolution mass spectrometry,” J. Chromatogr. B Analyt. Technol. Biomed. Life Sci., vol. 1179, p. 122820, Jul. 2021.
9. G. Anitha, P. Nirmala, S. Ramesh, M. Tamilselvi and G. Ramkumar, "A Novel Data Communication with Security Enhancement using Threat Management Scheme over Wireless Mobile Networks," 2022 International

Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752584

10. Prabu, R.T., Benisha, M., Bai, V.T. (2020). Characteristics of Alpha/Numeric Shape Microstrip Patch Antenna for Multiband Applications. Intelligent Systems Design and Applications. ISDA 2018, Advances in Intelligent Systems and Computing, vol 941. Springer, Cham. https://doi.org/10.1007/978-3-030-16660-1_86