Turkish Journal of Computer and Mathematics Education *Vol.10 No.01 (2019), 596-601*
DOI: https://doi.org/10.17762/turcomat.v10i1.13554

Research Article

Implementation of Unsupervised ML Techniques for Behavioural-Based Credit Card Users Segmentation in Africa

Kuljinder Singh Bumrah
Department of Computer Science & Information Technology, Graphic Era Hill University, Dehradun Uttarakhand India 248002

Abstract
A strategy used in marketing is segmentation, which is used to categorise consumers or other entities based on traits like behaviour or demographics. Finding consumer groups that may react similarly to particular marketing strategies, including such email subject lines or display ads, is beneficial. Due to the fact that it allows firms to customise marketing messaging and timing to increase response rates and enhance customer experiences. Customers often exhibit a wide range of behaviours. Businesses frequently employ segments that are based on thresholds. A scientifically based approach to categorise clients is required due to the expanding amount of characteristics and the overall theme of tailored products. The solution is clustering based on behavioural data. The goal of this study is to appropriately classify credit card users so that we can better understand their requirements and behaviour and provide them with relevant marketing offers. We must use the credit card segmentation dataset in our procedure. The input dataset was gathered from a repository of datasets. The pre-processing procedure must then be carried out. After that, feature extraction must be used. The system is then created using the k-means method to come up with the best segmentation plan. We must compute the silhouette score using the k-means technique. The findings of the experiment indicate that the visualisation and customer segmentation based on clustering algorithm.

1. Introduction

An essential component of managing bank credit risk is credit rating. The classic radial basis function network structure is particularly sensitive to the starting centre and class width of the chosen framework since it is highly prone to outliers also not able to efficiently handle the classification data. The incorporation of the radial basis function neural network prototype and the best segmentation algorithm into banks' or other financial institutions' personal loan credit rating models. To be able to raise the centre as well as width of the class also further enhance the centre and width of the RBF network paradigm, the optimum segmentation technique is enhanced and used to the training of RBF neural network characteristics in this research. Furthermore, the dynamic structure adjustment of the radial basis function network model, which is utilised to create the credit rating model, is accomplished by employing differential objective function of the class to select the number of hidden nodes adaptively. The experimental findings demonstrate that the modified model is more accurate while working with data that is not numerical, as well as its resilience has increased. A credit card is one of the instruments that is often used nowadays to pay for goods and services, as well as to borrow money occasionally. As a result of the fact that various individuals utilise the card to make purchases on a daily, weekly, monthly, or even yearly transaction, organisations have a wealth of information about what people buy and when they buy it.

Compared to cash payments and debit cards, credit cards have several advantages. They may be used to make large purchases that one might not otherwise ability to pay for, to conduct online transactions, to make urgent purchases when an unforeseen expenditure arises, and they are handier than carrying cash, which helps to stop the spread of fake currency. Financial institutions will be able to reduce unnecessary transaction costs associated with partnerships and transactions between businesses that are unconnected to the preferences of the customer and that reduce the expenses associated with running a campaign by using targeted marketing efforts rather than bulk campaigns as a result of categorising users based on their spending patterns. Customized marketing initiatives will lengthen client relationships. This development will result in increased demand for credit card-related solutions that will satisfy credit card customers. The number of people who own credit cards will increase as long as happy consumers continue to use them and, more significantly, as long as they act as brand advocates for the goods and services they received. A marketing strategy called segmentation divides consumers or other entities into groups according to characteristics like demographics or behaviour. It enables business analysts to pinpoint client demographics that would react similarly to particular marketing strategies.

In order to better understand their consumers and develop user-centric solutions for targeted marketing, credit card operators are undertaking credit card customer segmentation thanks to the exploding usage of machine

learning algorithms. Cluster analysis is a useful tool in this endeavour because it aids analysts and marketers in their everyday operations of analysing, describing, and utilising information that is concealed in groups by helping them uncover relevant groupings in objects that have a similar attribute.
The application's primary goals are to use clustering methods to categorise clients who use credit cards, to apply principal component analysis and to improve performance analysis.

2. Literature Survey

The credit card sector has experienced intense competition as a result of developing FinTech and e-payment industries on the international market. Financial institutions must supply credit card holders with more cutting-edge financial services that go above and beyond basic banking demands if they are to thrive. The actions that might be performed to develop a segmentation model based on behaviour that separates African credit cards depending on their purchasing data are defined and described in this study. We think that this segmentation will increase credit card usage in Africa, enabling the continent's citizens to fully enjoy the benefits of credit cards just like those in other areas of the world. The findings of this study may be used to adapt marketing campaigns such that they are customer-centric and lower the related marketing expenses [1].
In this research, we use K-means to categorise credit card users into four groups based on actual credit card data from a Chinese commercial bank. Then, based on the background data of the credit card users, we created distinct forecasting models using four data mining techniques, including C5.0, neural network, chi-squared automated interaction detector, and classification and regression tree. Finally, using the best model out of the four, we get some helpful information about decision tree regulation. The information is useful for finding future consumers and implementing focused marketing, as well as helping the bank identify the common traits of various customers [2].
As technology advanced, the financial industry saw the emergence of new business-making techniques. The credit card network is one of them. However, various challenges with the credit card scamming approach have arisen as a result of systemic inadequacies. As a result, both the industry and credit card users are suffering considerably. There is a paucity of guidance on how to look at real credit card statistics in respect to privacy issues. An effort has been made to apply machine learning-based algorithms in the publishing to identify frauds in the credit card industry. Here, two algorithms—Fraud Detection Using Random Forest and Fraud Detection Using Decision Tree for credit card—are employed. Some publicly available data can be used as a sample to determine the model's efficacy. Then, a financial institution's genuine globe credit card details group is analysed. In order to measure the systems' resilience, extra noise is also supplied to the data samples. The study's first strategy is important since it creates a tree against user behaviours that may be used to identify frauds [3].
Large amounts of data are used in business analysis to aid decision-making. The banking sector frequently employs supervised learning techniques, particularly when approving credit cards. The objective is to identify credit fraud using a variety of data mining techniques to forecast whether a new consumer will have good credit or not. The design of a new campaign will be based on this since switching from manual review to automatic review by computer might significantly increase job productivity and save labour expenses. The efficiency of the self-evaluation process for credit card applicants might be improved with the addition of more tools, such as self-organizing maps and algorithms [5].

3. Proposed System

The system's goal is to eliminate every drawback that the current system has. The solution is clustering based on behavioural data. The goal of this study is to appropriately classify credit card users so that we can better understand their requirements and behaviour and provide them with relevant marketing offers. We must use the credit card segmentation dataset in our procedure. The input dataset was gathered from a repository of datasets. The pre-processing procedure must then be carried out. We must deal with the missing values in this phase. The next step is to use feature extraction methods. The Principle Component Analysis (PCA) must be used in this stage to extract meaningful features from the available data. For each sample of various clusters, the silhouette score must be determined using the k-means method. The findings of the trial demonstrate that clustering-based consumer segmentation and visualisation are both effective.

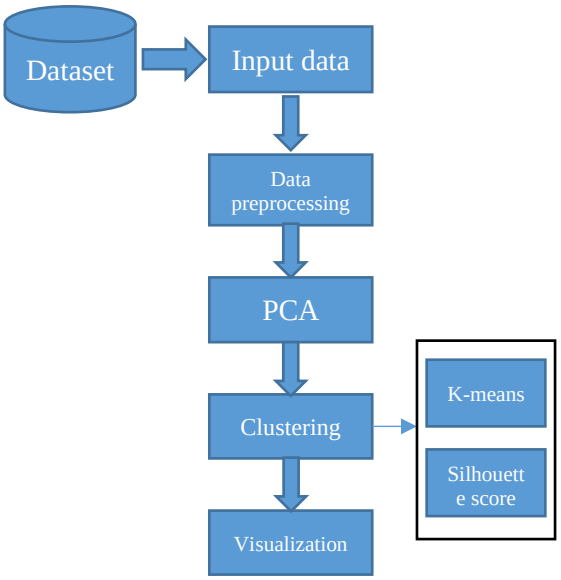


Fig 1: System Architecture

The following are some benefits of the suggested strategy:

- For a lot of data, it is effective.
- The feature extraction is put into practise.
- It figures out the silhouette score.
- The performance is improved.

The next section provides an explanation of the many phases that are involved in putting the suggested technique into practise:

1. Data selection

A dataset repository served as the origin of the input data. The credit card client dataset is utilised in our procedure. The segmentation of credit card clients is done by data selection. The dataset that includes Balance information, One-off purchases, purchasing instalments, cash advances, frequency of one-off purchases, frequency of purchases, credit limit, payments, and minimum payments.

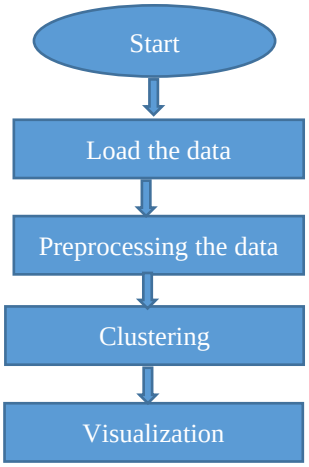


Fig 2: Flow Diagram

2. Data pre-processing

Pre-processing data involves deleting unnecessary information from the dataset. Missing data removal: In this procedure, null values like missing values and Nan values are changed to a value of 0. Data was cleared of any errors and missing values as well as duplicates. Coding Categorization-Possible Data. The great majority of machine learning algorithms prefer numerical input and output variables over categorical data, which are variables with a finite set of label values. Then, using the drop () function, we must additionally delete unprocessed columns from our datasets.

3. Feature extraction

Getting meaningful data from an existing system is referred to as feature extraction. This phase requires the usage of PCA (principle component analysis). Visualization, noise reduction, and feature extraction are all done using PCA. Using PCA, which reduces the number of dimensions, we may find significant links in our data and change the current data in accordance with these relationships. then quantifies the value of these connections so we may maintain the most crucial connections and discard the less crucial ones.

4. Clustering

A population or set of data points must be separated into several groups before clustering can occur, with each group being composed of data points that are more similar to one another than to those in the other groups and that are distinct from those in the other groups. We must utilise the k-means clustering method in this phase. K-means clustering is a vector quantization strategy that has its roots in signal processing. This seeks to group k clusters of each observation, each of which is a part of the n observations, into and the cluster's closest mean acts as a prototype. Finding the silhouette score requires the use of the k-means clustering technique. The effectiveness of clusters made using the clustering method is assessed using it. This score was determined for every sample from various clusters.

5. Visualization

Data visualisation aims to uncover patterns, trends, and relationships that might not otherwise be obvious by placing data in a visual context and attempting to understand it. Numerous excellent graphing libraries with a wide range of functionality are available in Python. A chart for choosing the best data display method in a specific circumstance. Groups of colours in a scatter plot. For the third variable, nation size, a scatter plot with colour groups and size encoding was created. To visualise data in Python, we must utilise the seaborn and matplotlib programmes.

4. Results

The dataset for credit card segmentation must be used in our procedure. Dataset repository was used to get the input dataset. The pre-processing process is then required to put into practise. Applying feature extraction is the next step. In order to generate the best segmentation approach, the system is then created using the k-means algorithm. To determine the silhouette score, we must employ the k-means method. The findings of the trial demonstrate that clustering-based visualisation and client segmentation are effective.

The application's primary goals include segmenting credit card users, segmenting users using clustering techniques, using PCA, and improving performance analysis. According to the findings of segmenting credit card users, users may be divided into four different categories. The clustering technique k-Means was used to achieve this. The worth of each consumer to the business must be maximised, hence it's critical to segment consumers in order to determine how to interact with them. The silhouette score must be determined using the k-means technique. The findings of the trial demonstrate that clustering-based consumer segmentation and visualisation are both effective.

----------Data Selection----------

	CUST_ID	BALANCE	...	PRC_FULL_PAYMENT	TENURE
0	C10001	40.900749	...	0.000000	12
1	C10002	3202.467416	...	0.222222	12
2	C10003	2495.148862	...	0.000000	12
3	C10004	1666.670542	...	0.000000	12
4	C10005	817.714335	...	0.000000	12

Fig 3: Data Selection

---------After Handling Missing Values------

BALANCE	0
BALANCE_FREQUENCY	0
PURCHASES	0
ONEOFF_PURCHASES	0
INSTALLMENTS_PURCHASES	0
CASH_ADVANCE	0
PURCHASES_FREQUENCY	0
ONEOFF_PURCHASES_FREQUENCY	0
PURCHASES_INSTALLMENTS_FREQUENCY	0
CASH_ADVANCE_FREQUENCY	0
CASH_ADVANCE_TRX	0
PURCHASES_TRX	0
CREDIT_LIMIT	0
PAYMENTS	0
MINIMUM_PAYMENTS	0
PRC_FULL_PAYMENT	0
TENURE	0

dtype: int64

Fig 4: After Handling Missing Values

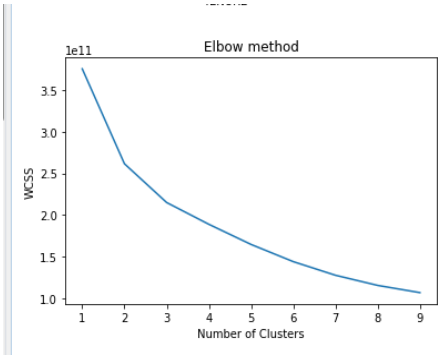
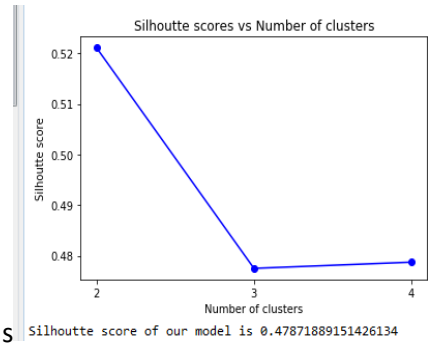


Fig 5: Elbow Method



Silhoutte score of our model is 0.47871889151426134

Fig 6: Performance Analysis

5. Conclusion

Customers are divided into four different categories, according to the findings of the segmentation of credit card users. The k-Means clustering technique was used to achieve this. It is crucial to choose how to communicate with each group of customers in order to maximise each one's worth to the business. We must compute the silhouette

score using the k-means technique. The findings of the experiment indicate that the visualisation and customer segmentation utilizing the clustering algorithm.

Reference

[1] H. Wu and C.-C. Wang, "Customer segmentation of credit card default by self-organizing map," American Journal of Computational Mathematics, vol. 8, no. 03, p. 197, 2018.
[2] S. Bansal, P. Bruno, O. Denecker, M. Goparaju and M. Niederkprn, Global payments 2018: A dynamic industry continues to break new ground, Global Banking McKinsey, 2018.
[3] Z. Boˇsnjak and O. Grljevic, "Credit users segmentation for improved customer relationship management in banking," in 2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI). IEEE, 2011, pp. 379–384.
[4] L. Ying and W. Yuanyuan, "Application of clustering on credit card customer segmentation based on ahp," in 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), vol. 3. IEEE, 2010, pp. 1869–1873.
[5] L. Ying and W. Yuanyuan, "Application of clustering on credit card customer segmentation based on ahp", 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM), pp. 1869-1873, 2010.
[6] S.-C. Chi, R.-J. Kuo, and P.-W. Teng, "A fuzzy self-organizing map neural network for market segmentation of credit card," in Smc 2010 conference proceedings, vol. 5. IEEE, 2010, pp. 3617–3622.
[7] H. Abdi and L. J. Williams, "Principal component analysis," Wiley interdisciplinary reviews: computational statistics, vol. 2, no. 4, pp. 433– 459, 2010.
[8] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," Computers & Geosciences, vol. 10, no. 2-3, pp. 191–203, 2004.
[9] C. Ezenkwu, S. Ozuomba, and C. Kalu, "Application of kmeans algorithm for efficient customer segmentation: A strategy for targeted customer services," International Journal of Advanced Research in Artificial Intelligence (IJARAI), vol. 4, 10 2015.
[10] Aziz, "Customer segmentation based on behavioral data in emarketplace," 2017.
[11] D. A. Reynolds, "Gaussian mixture models." Encyclopedia of biometrics, vol. 741, 2009.
[12] Malinowski, T. Cholewo, and J. Zurada, "Capabilities and limitations of feedforward neural networks with multilevel neurons," 01 1995, pp. 131 – 134 vol.1.
[13] J. M. Pena, J. A. Lozano, and P. Larranaga, "An empirical comparison of four initialization methods for the k-means algorithm," Pattern recognition letters, vol. 20, no. 10, pp. 1027–1040, 1999.
[14] Y. Thakare and S. Bagal, "Performance evaluation of k-means clustering algorithm with various distance metrics," International Journal of Computer Applications, vol. 110, no. 11, pp. 12–16, 2015.
[15] S. Tripathi, A. Bhardwaj, and P. E, "Approaches to clustering in customer segmentation," International Journal of Engineering & Technology, vol. 7, p. 802, 07 2018.