

# Using Unsupervised Machine Learning Techniques for Behavioral-based Credit Card Users Segmentation in Africa

Eric Umuhoza, Dominique Ntirushwamaboko, Jane Awuah, and Beatrice Birir

**Abstract**—Given the fierce competition that has come up because of evolving FinTech and e-payment industries in the global market, the credit card industry has become extremely competitive. To survive, financial institutions need to offer their credit card customers with more innovative financial services that provide a personalized customer experience beyond their banking needs. While we are witnessing this high competition that aims to provide better services to credit card holders, Africa risks remaining behind once again: in 2017, the World Bank reported that only 4.47% of Africans aged 15 and above hold a credit card. In this paper, we define and describe the steps that can be taken to build a behavioral-based segmentation model that differentiates African credit cardholders based on their purchases data. We focus on African customers and African financial institutions as (i) little has been done so far when it comes to understanding the spending behavior of African credit card holders; and (ii) because we believe that this segmentation will allow boosting credit card usage in Africa, thus allowing Africans to fully benefit from credit cards as other parts of the world do. The results of this research can help tailor the market campaign to make them customer-centric and reduce the associated marketing costs. We show the proposed approach at work using anonymized credit card data of one of the leading banks in Egypt, the Commercial International Bank of Egypt.

**Index Terms**—African credit card market, African credit card profiles, Customer segmentation, spending behavior, unsupervised machine learning.

## I. INTRODUCTION

Nowadays, a credit card is one of the widely used tools to pay services and foods, and it is sometimes used to borrow money. Different people use the card to make their daily, weekly, monthly, and even yearly purchases as such, information on what a person buys and when they buy it are excellent opportunities for organizations to explore. Credit cards have some advantages over cash payments and debit cards. They can be used to purchase big ticket items that one may not be able to afford otherwise, perform online transactions, make emergency transactions when an unexpected expense hits and are more convenient compared to carrying cash reducing the circulation of counterfeit money.

Eric Umuhoza was previously with the Department of Electronics, Information and Bioengineering, Polytechnic University of Milan and Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila. He is now with Carnegie Mellon University Africa, Kigali, Rwanda (email: eumuhoza@andrew.cmu.edu).

When it comes to the usage of credit cards, Africa as a continent is way behind other continents and this is mainly because of the rural nature of various regions in the continent. The World Bank report shows that the percentage of people aged 15 and above who have a credit card in Africa, for 2017 was 4.47% [1], which is lower than the world average of 19.21%. The country with highest credit card penetration in Africa was Mauritius with 24.12% and the lowest was Morocco with the penetration of 0.19%; Fig. 1 shows the top 20 countries where data were available [1].

Africa, despite being considered as the big market not served, has all the potential for a quick transition towards the adoption and widespread use of credit cards (see Fig. 2). Companies with a large presence in Africa are Visa with 57% of all credit card purchases and MasterCard which represents 39%. Although these two companies dominate the African market, other companies such as American Express which have 3% of the volume of purchases by credit card [2], are trying their best to penetrate this promising market.

Credit card activity across Africa is the beginning of an increase both in investments and in spending by customers. The McKinsey Global Institute [2] projected rise in consumer spending in Africa from \$0.86 trillion in 2008 to \$1.4 trillion in 2020. Consequently, Africa now boasts the world's fastest-growing middleclass economy with a 100% rise in less than 20

Even though little has been done so far when it comes to understanding the spending behavior of African credit card users, we strongly believe that clustering them based on their spending behaviors will help create more customized marketing strategies for each segment of customers and boost credit card usage in Africa for three reasons:

- 1) Users categorization based on their spending behaviors will allow financial institutions to optimize campaign associated marketing costs through targeted marketing activities instead of mass campaigns and eliminate unnecessary transactional costs of merchant deals and partnerships that are irrelevant to the customers preferences.

Dominique Ntirushwamaboko is a MSIT graduate from Carnegie Mellon University Africa and is currently working as big data scientist at the National Bank of Rwanda.

Jane Awuah, and Beatrice Birir are MSc Students in Information Technology at Carnegie Mellon University Africa, Kigali, Rwanda.

- 2) Tailored marketing activities will increase customer years [3]. This progress will lead to further demand in credit card related systems satisfaction of credit card users.
- 3) The number of credit card holders will continue to grow as satisfied customers will continue to use their credit cards and, most importantly, they will become ambassadors of those products or services.

In this paper, we define and describe the steps that can be taken to build a model to cluster African credit card users based on their purchase data. The key challenge faced in making this clustering and creating the customer groups is that the data are not easily available especially for the African context. In addition, when the data are found, we realized that on the continent, most people do not use the cards often as such it is difficult to find out if the details from the card are accurate or show the customers genuine spending pattern. In this paper, we will show how such sparse data sets, generated from the irregular spending of African customers, can be used to generate customer segmentation.

The proposed approach has been validated by using anonymized data provided by the Commercial International Bank of Egypt CIBEG (see Section II).

This paper is organized as follows: Section II provides the CIBEG 's background, a detailed literature review on different credit card users segmentation models; Section III describes our approach and its implementation; Section V presents segments profiles(clusters) and discusses the proposed marketing strategies; and Section VI gives the conclusion.

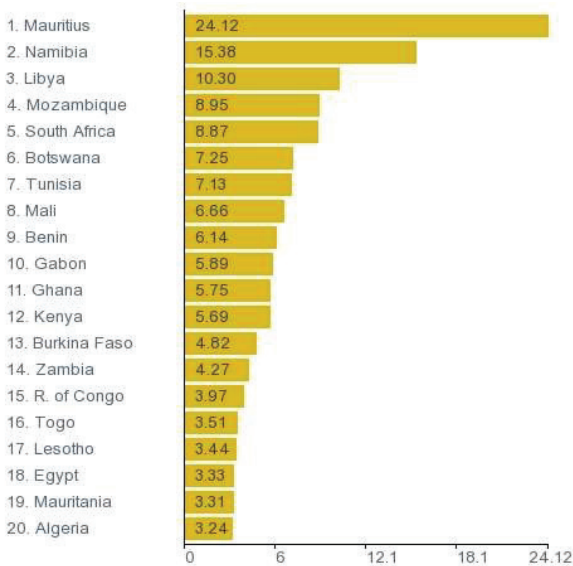


Fig. 1 Percentage of People Aged 15+ with Credit Card in Africa,2017

## II. BACKGROUND

The first part of this section provides the background of the Commercial International Bank, an Egyptian financial institution which provided us the data used to validate our approach. The detailed literature review on different credit card users segmentation models is reported in Subsection II-B.

### A. Commercial International Bank (CIB)

Established in 1975, the Commercial International Bank (CIB) is Egypt's leading private-sector bank. CIB is North

Africa's leading bank in FinTech, data analytics and innovation; It is Egypt's largest sponsor of fin-tech startups and it is one of the most valuable banks in Africa as reported in Fig. 3.

CIB pioneered to build an advanced analytics and big data lab in 2015 with a vision to evolve from a successful yet local traditional bank into a leading data driven, customer-centric organization that fully understands its existing and potential customers needs and tailors financial products and services accordingly. This data lab currently supports managerial decision making across the bank. It is in conjunction with this lab that this research has been carried out.

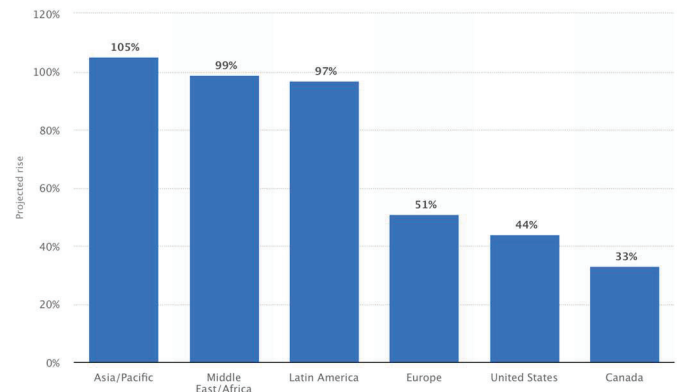


Fig. 2 Percentage growth in payment card transactions worldwide from 2011 to 2016, by region.

### B. Credit-Card Customer Segmentation Approaches

Segmentation in marketing is a technique aimed at dividing customers or other entities into groups based on attributes such as demographics or behavior. It allows business analysts to identify groups of customers who may respond in the same way to specific marketing techniques.

With the explosive use of machine learning algorithms, the credit card operators are conducting credit card customer segmentation to understand their customers and come up with user-centric products for targeted marketing. Cluster analysis is an effective tool in this effort as one of the purposes of cluster analysis is *understanding*— cluster analysis is used to find meaningful groups in objects sharing a common characteristic that helps analysts and marketers in their daily activities of analyzing, description and utilization of information hidden in groups [4].

There is a large corpus of works that use different machine learning techniques for credit card users segmentation, we present a subset of them in the following paragraphs.

Hui Wu and Chang-Chun Wang [5] have applied self-organizing map (SOM) [6] technique to segment customers as a preliminary step in their credit card default prediction study. SOM is an unsupervised machine learning method that reduces data complexity and dimensionality while keeping its original topology. Their study showed SOM to be superior to other widely used dimension reduction methods, PCA and LDA, when features in data present unclear nonlinear relations.

Z. Bonjak and O. Grljevi [7] used a neural network technique to segment bank customers via credit-scoring. For each of the customers, a binary response variable, *creditability*, that reflects

the client's creditworthiness was assigned. Their study recognized the importance of differentiating customers' characteristics, as it allows to tailor bank policies towards retention and risk minimization.

L. Ying and W. Yuanyuan [8], using credit card customers data, have developed marketing strategies for a commercial bank in Shanghai. Their study uses two algorithms: Analytical Hierarchy Process (APH) [9] for indicator optimization, and K-means for clustering of the credit card users.

C. Sheng-Chai et al. [10], defined data mining as the process of analyzing and discovering useful information i.e., knowledge discovering. They have discussed three main methods of clustering: the principal component analysis (PCA) [11], k-Means and Fuzzy means clustering [12]. The PCA was the preprocessor, meaning that before either K-means or Fuzzy clustering method (FCM) was implemented, PCA would first be implemented. It was found that in as much as k-Means is used in so many different fields, the usage of PCA and FCM has a comparatively lesser computational time as compared to the combination of PCA and K-means.

Various researchers have used the k-Means clustering algorithm in different customer segmentation studies [13], [14]. C. Ezenkwu et al. [13] used k-Means to train an algorithm to segment customers of a retail store, using a z-score normalized of a two-feature dataset containing 100 training patterns. Customer segments were identified with an accuracy of 95%. A. Aziz [14] used the k-Means, using cosine similarity measure, to construct a personal feed for each user in segments with items derived from the user groups. Both researchers agree that k-Means is best used when one wants to cluster data points into x number of segments. The advantage of k-Means is that it gives more weight to bigger clusters and can be computed in batches to improve performances.

### III. APPROACH

This section presents our approach for credit card users segmentation; We will refer to credit card usage data from CIB (Section II-A) but for confidentiality issues, some details will be omitted.

This study seeks to segment credit card users using unsupervised machine learning approaches. We use unsupervised machine learning techniques because credit card usage data are mainly unlabeled. As we want to study the credit card users behavior, we focus on variables that display the purchasing preferences of credit cards holders.

#### A. Data Description and Processing

The data used in this research are of the transactions of credit card customers from the first quarter (Q1) of 2016 to the last quarter(Q4) of 2017. This dataset contains 143,975 observations per quarter and 60 total variables. Basically, these data contain purchase information of customers: where they buy from (local or international), the merchant group they purchased from, frequency and amount of purchase per quarter, etc. In this study, we focus on the behavioral variables that show the purchase preferences of customers. There are 16 variables representing sectors or merchant groups (MG) on which customers spend their money using credit cards. We considered

the customers frequencies and amount spent on each MG. The process of data processing involves checking the data distribution and checking for extreme and missing values. It also involves data transformation and normalization to make data tidy for modeling. After the preprocessing, data modelling was done using k-Means. The model framework below gives the schematic view from data acquisition, transformation, and modeling.

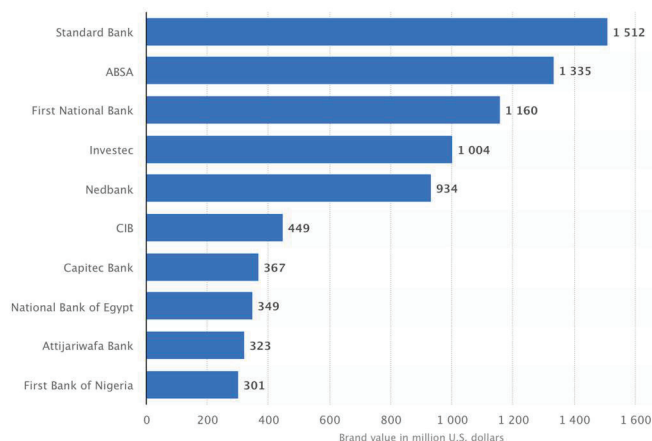


Fig. 3 Most valuable banking brands in Africa as of February 2017 (in million U.S. dollars)

#### B. Data Description and Processing

There are many clustering methods in literature, but the selection of the appropriate method depends on how the objects are structured in space. In our study, we consider a clustering algorithm as an important data mining algorithm, in which the data are divided into different clusters: objects in the same cluster are homogeneous to each other.

In a study with many features as in our case, there was a possibility of using generalized models such as Gaussian Mixture Models (GMM) [15] or a feedforward neural networks also known as multi-layered network of neurons which fit non-linear patterns and are known to fit arbitrarily any complex data. However, we avoided these complex models as they generally tend to fit the noise instead of the true pattern. Specifically, a multi-layer feed forward neural network classifier is best with a reduced number of outputs, especially binary (0,1), which is not our case as we have multiple outputs or clusters [16].

We chose the k-Means algorithm as it is known among all other clustering methods to be fastest and is the most extensively employed in customers segmentation because of its advantages such as being fast, simple and most importantly its capability to handle massive datasets. This algorithm sets the mean value within the cluster objects as the center points, it then determines the place where the non-center points belong by calculating the distance between the objects and the center points and continue in an iterative process until the intra-cluster distance is minimum and the extra clusters distance is maximum [17]. Before clustering with k-means, the researcher needs to set the number k clusters which most of the time is accomplished by methods such as silhouette or elbow to mention a few. The outcome of the k-Means clustering algorithm is to reduce to a minimum the total intra-cluster

variance, which is the squared error function given by the following formula [18]:

$$SME = \sum_{j=1}^k \sum_{i=1}^n ||x_i - c_j ||^2$$

where  $k$  is the number of clusters with  $k$  varies from 1 to the total number of clusters,  $n$  is the number of cases which varies from 1 to  $n$  cases,  $x_i$  is the case and  $c$  is the centroid for cluster  $j$ .  $||x_i - c_j ||^2$  is the distance function.

Before applying the k-Means method, we used the Principal Components Analysis (PCA) for dimensionality reduction. The PCA is usually used to condense information available in many features in a dataset into a smaller set of new synthesis dimensions with a least loss of information. For the PCA, the variables are defined as linear combinations of the original variables  $X_1, \dots, X_k, \dots, X_m$ . We extracted Eigenvectors coefficients based on the following equation [14]:

$$Y_k = C_{k1}X_1 + C_{k2}X_2 + \dots + C_{km}X_m$$

where

- $Y_k$  is the  $k^{th}$  principal component
- $C$ 's are the coefficients

Using PCA, in our study, we found out that nine PCA explains 89% of all the overall variation in the dataset. Therefore, only nine PCAs were used by K-means for segmentation.

To check the model performance, we had to ensure that an optimum number of clusters was selected. To achieve this, we used various methods including elbow, Silhouette, and Calinski-Harbaz. All the three methods agree on four clusters as the optimum number as depicted in Fig. 4, Fig. 5 and Fig. 6. The number of clusters given by these methods confirms the rule of thumb that suggests that the number of clusters should not be few or many to tailor marketing strategy and optimize resources [19]. It also makes sense in the eyes of the researchers to have optimum clusters of less than five in the banking industry. Other methods of clustering evaluation require to know ground truth which we did not know and compare it with the outcome of the model.

### C. Model Framework

The model framework shows the process from data acquisition, preprocessing, modeling, and production of output. It was inspired by previous works in credit card segmentation and clustering [8][17]. The k-Means first determines the number of optimum clusters, then the algorithm selects the centroids, computing the distance between the centroid and the observation, if the distance that observation is assigned to that cluster is minimum, if not, that observation is moved to the next cluster. After moving the observation to the next cluster, the process starts again by comparing the distance until all observations have been assigned clusters.

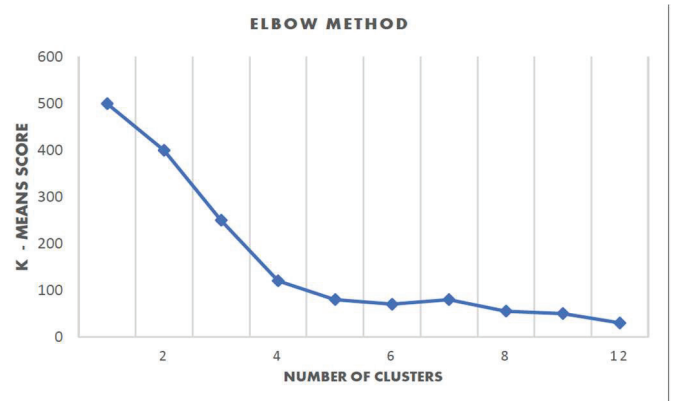


Fig. 4: The elbow method indicates the optimum number of clusters where the bend of the curve in a shape of elbow is observed. In this case, at  $k=4$ , we see a change in the slope, and we consider  $k = 4$  as the optimum number of clusters.

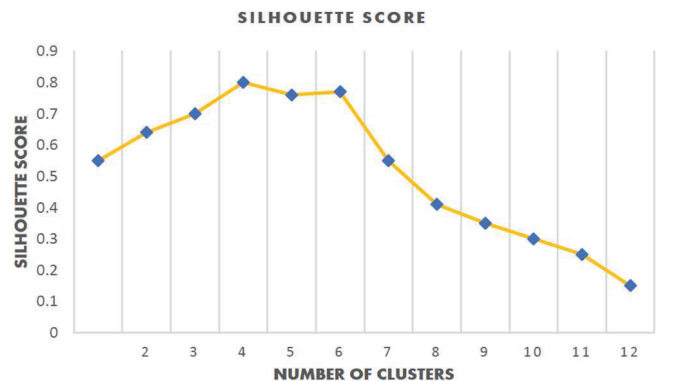


Fig. 5: The silhouette coefficient increases up to the point where  $k=4$  and then decreases. It exhibits a peak at  $k=4$ , which indicates the number for optimum clusters.

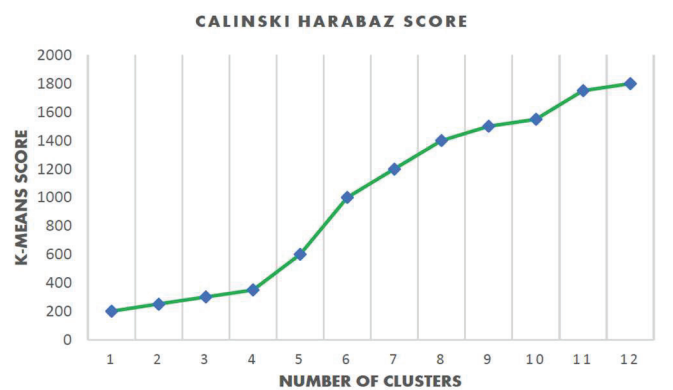


Fig. 6: The optimum number of clusters in Calinski Harabaz method is given by first bend in the curve before a sharp increase. On this graph, the optimum number is at  $k=4$ .

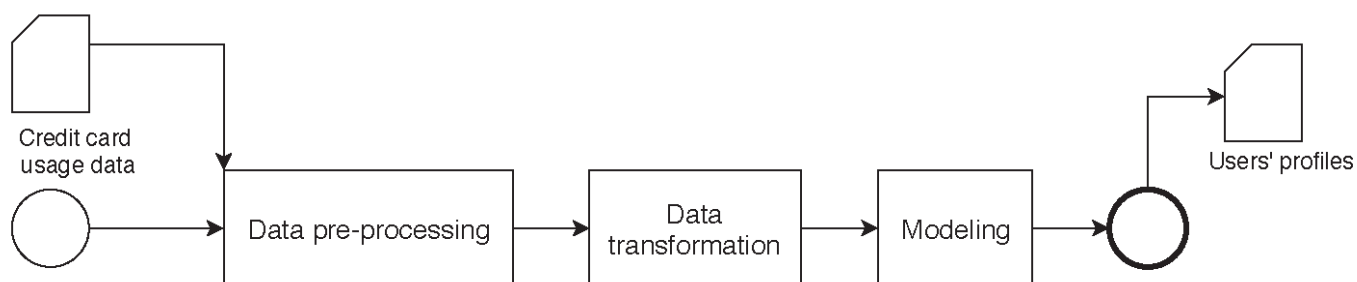


Fig. 7 Credit card users segmentation process: credit card usage raw data are firstly pre-processed to check and remove outliers and missing values. Then data are transformed into a suitable format for modeling phase which produces users' profiles as the final output.

#### IV. IMPLEMENTATION

In this section, we present a step by step implementation of the data segmentation process. As illustrated in Fig. 7, the process starts by inputting the transactional data and ends by generating user profiles as outputs. The implementation is illustrated through the concrete case, using credit card usage data from CIB (Section II-A), some details including the shape of final clusters and their detailed characteristics are omitted for confidentiality reasons.

##### A. Data Pre-processing

This step is comprised of three tasks:

- 1) Missing data management some variables in the dataset had missing observations which had to be deleted for analysis.
- 2) Data distribution checking - after the management of missing values, the data distribution had to be checked to have a general overview of the data skewness and scale to inform data transformation later.
- 3) Outliers detection - this phase detected and removed outliers. In our case, the extreme observations were detected and removed using the interquartile method.

##### B. Data Transformation and Dimension Reduction

In this step we ranked the MG according to their usage to know the most preferred ones. After ranking the MGs, we have selected the top 5 MGs that represented consistency over time. As most credit card users performed fewer transactions during the period under consideration and few had huge transactions, this made the credit card data negatively skewed. The cube root was used to reduce the skewness before modeling. After the transformation of high skewed data, the minimum maximum scaling was used to normalize the data within a range of 0 and 1. This facilitated values to be in the same range and avoid the bias that can be introduced by varying data scale.

To reduce the number of features to be incorporated in the model, a PCA method was used for dimensionality reduction which performed a linear combination of the data to a lower-dimensional space in a way that the variance of the data in the low-dimensional representation is optimum.

##### C. Modeling with K-Means

The clustering with K-Means has been done following the algorithm described in Algorithm 1.

---

#### Algorithm 1 K-Means Clustering

---

- 1: Select the optimum number of K based on the elbow, silhouette and Calinski methods.
  - 2: Partition the objects into k sub-sets.
  - 3: Identify the centroids (mean point) of the current partition.
  - 4: **Repeat**
  - 5:   **for** each data point  $x_i$  **do**
  - 6:     Find the nearest centroid ( $C_1, C_2, \dots, C_k$ )
  - 7:     Assign the point to that cluster
  - 8:   **end for**
  - 9:   **for each** cluster  $j = 1 \dots k$  **do**
  - 10:     New centroid = mean of all points assigned to that cluster
  - 11:   **end for**
  - 12: **Until** convergence or until the end of a fixed number of iterations
- 

#### V. FINDINGS AND DISCUSSION

The purpose of this study was to come up with homogeneous segments that represent customer purchasing behaviors. Armed with this segmentation, financial institutions can not only craft better value propositions for their customers, but they can also identify groups that are not well served by current offers.

The analysis revealed that the customers are grouped into four distinct segments, their profiles, and a marketing strategy for each segment are reported in the next paragraphs.

##### A. Segment 1: The Ordinary Joe / Jane

Customers of this segment seem to use their cards on a general day to day basis for general shopping; healthcare, professional, telecommunication and commercial services; and electronics.

*Marketing Strategy:* Acquiring and keeping these customers calls for a balancing act as they tend to spend in diverse yet common areas. They are your everyday average Joes and they tend to represent most people's buying patterns. Keeping this group and reaching more people in the market requires royalty rewards, personalization of services and an easy account management system where they can manage their cash. The bank could also foster a sense of partnership by making these customers feel they are getting the best deal.

### B. Segment 2: Fashion Lovers (Fashionistas)

Customers in this segment tend to spend their money on all things fashion while occasionally making purchases in grocery stores and supermarkets, on telecommunication services and for electronics.

*Marketing Strategy:* Acquiring and keeping these customers would be through plans such as partnering with huge popular retail clothing shops and having special cards that enter them into drawings to win amazing gifts, giving them shopping discounts when they shop above a certain amount and having seasonal promotional sales especially towards big public holidays or celebrations a time when people tend to go all out in shopping.

### C. Segment 3: The Prosperous (Executives)

Customers in this category, tend to be the big spenders, using their cards in several different segments from education to jewelry purchases. They are based off customers who tend live large and tend to travel; thus, they end up spending money on jewelry, food stores, restaurants, beverages, sports, leisure, transport, airline, and accommodation. They also perform government, legal and social services transactions together with education, either paying fees for someone else or investing in their education.

*Marketing Strategy:* Getting a card into this segment's wallet involves offering rewards for the amount they spend. To differentiate their card, we should not merely see them as a spending instrument but as a tool that facilitates financial success through ease of use. For instance, if a bank could offer them priority services such as direct international call lines and special banking lounges more users would be attracted to this and would not mind spending extra.

### D. Segment 4: Limited Spenders

Individuals here seem to be those who use their cards to perform home constructions and purchase furniture. Beyond that, they use it for transport, sports, and leisure. They are not big spenders and use their card in a limited way.

*Marketing Strategy:* To build and encourage this segment, the bank could respond to them through simplicity and transparency in fees, rates, and terms. This will bolster their confidence and will make a credit card be more viable. One such way is by giving them low interest rates. In addition, seasonal marketing advertisements to spur their spending will boost their spending behaviors.

## VI. CONCLUSIONS AND FUTURE WORKS

The results of credit card customers segmentation revealed that customers are grouped into four distinct segments. Most customers belong to the ordinary joe, followed by fashion lovers, limited spenders and last are the prosperous. This was achieved from using the k-Means clustering algorithm.

The purpose of this study was to transform the existing credit card business model and campaigns strategy at CIB from relying mainly on traditional value-based campaigns to targeted campaigns that are based on the customer's needs, lifestyle, and usage preferences. We were able to extract customer purchasing behaviors and suggest tailored marketing strategies.

Using the results of this study will improve customer satisfaction, boost customer activity, optimize the campaign associated marketing costs through targeted marketing activities and eliminate unnecessary transactional costs of merchant deals and partnerships that are irrelevant to the customers preferences.

We recommend the CIB to use the results of the project in their marketing plans and update this model on quarterly basis to get new behaviors from customers.

### A. Future Directions

The key limitation faced in this study is the customers' segments movement throughout quarters. We think that by using monthly, bi-annual, and annual data can solve this problem; We suggest further investigations be done using those data for this end. Furthermore, we plan to conduct similar studies in other African countries, especially in sub-Saharan Africa, to understand the similarities/dissimilarities of card customers' behaviors across Africa.

## ACKNOWLEDGMENT

This research came out of the practicum project submitted by the CIB to CMU Africa in fall 2019. We acknowledge the big role played by the CIB and its staff, especially Ms. Nelly Youssef and Mr. Andrew Raafat of the analytics and big data lab for their tremendous assistance during the project.

## REFERENCES

- [1] A. Demircuc-Kunt, L. Klapper, D. Singer, S. Ansar, and J. Hess, The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution. The World Bank, 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] M. Ncube, C. L. Lufumpa, and S. Kayizzi-Mugerwa, "The middle of the pyramid: dynamics of the middle class in africa," Market Brief April, vol. 20, p. 2011, 2011.
- [4] P.N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining. Pearson Education India, 2016.
- [5] 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.
- [6] T. Kohonen, "The self-organizing map," Proceedings of the IEEE, vol. 78, no. 9, pp. 1464–1480, 1990.
- [7] Z. Bo'snjak 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.
- [8] 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.
- [9] T. L. Saaty, "What is the analytic hierarchy process?" in Mathematical models for decision support. Springer, 1988, pp. 109–121.
- [10] 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 2000 conference proceedings, vol. 5. IEEE, 2000, pp. 3617–3622.
- [11] H. Abdi and L. J. Williams, "Principal component analysis," Wiley interdisciplinary reviews: computational statistics, vol. 2, no. 4, pp. 433–459, 2010.
- [12] 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, 1984.
- [13] 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.

- [14] A. Aziz, "Customer segmentation based on behavioral data in e-marketplace," 2017.
- [15] D. A. Reynolds, "Gaussian mixture models." Encyclopedia of biometrics, vol. 741, 2009.
- [16] A. Malinowski, T. Cholewo, and J. Zurada, "Capabilities and limitations of feedforward neural networks with multilevel neurons," 01 1995, pp. 131 – 134 vol.1.
- [17] 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.
- [18] 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.
- [19] S. Tripathi, A. Bhardwaj, and P. E, "Approaches to clustering in customer segmentation," International Journal of Engineering & Technology, vol. 7, p. 802, 07 2018.



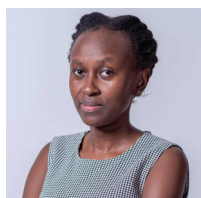
**Eric Umuhoza** was born in Rwanda in 1985. He received the B.Sc. and M.Sc. degrees in computer engineering from the Polytechnic University of Milan, Milan Italy, in 2013 and the Ph.D. degree in information technology from the same university, in 2017. He is Instructor at Carnegie Mellon University Africa. His research interests include model-driven software engineering, user interaction design, smart cities, and big data. Prior joining CMU Africa, Dr. Eric Umuhoza held different research positions in Europe including Senior Postdoctoral Researcher with the Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila. He has been a Postdoctoral Researcher with the Department of Electronics, Information and Bioengineering, Polytechnic University of Milan, Italy; and a Visiting Scholar at 'Ecole des Mines de Nantes, France.



**Dominique Ntirushwamaboko** is a MSIT graduate from CMU and is currently working as big data scientist at the National Bank of Rwanda. He is interested in application of machine learning and artificial intelligence techniques to solve social and economic challenges.



**Jane Awuah** is a data analyst who is currently pursuing a Master of Science degree in Information Technology at Carnegie Mellon University Africa, specializing in Applied machine learning and IT Entrepreneurship. She is interested in the application analytical techniques to improve everyday activities, businesses, and user interactions.



**Beatrice Birir** is a Carnegie Mellon University Africa Student currently pursuing her Master of Science in Information Technology. Her research interests are how data science and machine learning can be used to provide innovative solutions and improve people's lives.