International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Identification of the Productive Potential of Customers in a Commercial Bank Portfolio: Calculation of the Customer Lifetime Value using Statistical Learning Method

Tchokomeni Hervé

Statistician Economist Engineer, Yaoundé, Cameroon
Mail: *tchokoherve[at]yahoo.fr*

Abstract: ***The high temporal variability in the productivity of portfolio calls into question the approach to monitoring customer relationship at Afriland First Bank Cameroon. The Customer Lifetime Value, calculated based on the specific consumption habits of corporate clients in the portfolio over the period from January 1, 2016 to December 31, 2018, makes it possible to determine the individual productive potential of that customer. The calculation of the transition probabilities by Markov chains between productivity levels over time shows that customers from the least productive segments have a very good probability of migrating to high productivity segments. The combination of the productivities of the segments with the probabilities of migration between the productivity levels shows that the highest Customer Lifetime Value segment is 87,019,958 CFA francs per year for the group of high productivity, and 1, 451, 828 CFA francs for the low productivity group.***

Keywords: Customer Lifetime Value; Markov chain; Productivity; Segmentation; Transition probability.

1. Introduction

The Banking Commission of Central Africa (COBAC) recognizes fifteen banks in Cameroon in 2020. This is almost double the eight banks that were in operation in 2008, the years of the global financial crisis. These banks operate in an environment where the diversity of financial products offered is very low. In this context of weak financial innovation, banks are competing fiercely in offering almost similar financial products and services. Each bank, to gain market share, relies on significant marketing investments. The segment of the corporate customer, which is the most source of resources for the banks, is privileged. Afriland First Bank Cameroon, which has taken on systemic importance in the economy (BEAC, 2019), has more than 27,000 companies in its client portfolio. To maintain this position and confidence in this customer segment, she spends enough on marketing. She is also the promoter of the Business Networking Forum *« le mercredi de la PME »*.

In order to optimize marketing expenses, the bank is obliged to acquire modern marketing tools while reducing costs. These tools must be based on data and guide investments towards customers with high productivity potential. The needs of the customer and his behavior appear as the elements to be scrutinized by these modern tools of customer relationship management (CRM). CRM is a managerial effort to manage business interactions with customers by combining business processes and technologies that seek to understand a company's customers (Kim et al., 2003). One of the most important CRM tools is Customer Lifetime Value (CLV) (M. EsmaeiliGookeh& J. Tarokh, 2013). Dwyer, 1997 defines CLV as the present value of expected benefits less customer charges.

Due to the absence of an automatic customer relationship management tool, important customers enter into a bank relationship with the bank and then leave without their potential being noticed, and therefore the bank don't deploys on it appropriate support measures. This has the effect of causing the inactivity of the latter, because they are poorly accompanied. Over the period from July 2014 to July 2015, more than 40% of the bank's customers were inactive, which represented a 604,825,199,709 FCFA deficit in terms of movements entrusted[1]. Conversely, the bank makes marketing expenses towards customers, which do not represent real productivity for it. This lack of marketing targeting implies significant volatility in the dynamics of customer productivity. Indeed, a good number of clients show a drop in productivity of more than 50,000,000 CFA francs between two years. The Business Fund Manager therefore needs to know the profile of the most productive customers in order to prioritize his marketing resources. This is possible through a tool that makes available to him the potential value of each client in his portfolio. The calculation of customer value in the case of a commercial bank is done using a customer evaluation model. Giving value to the customer is important enough in the commercial banking sector, which must know their customers better, or rather better know their values, and address them individually with an adapted marketing package.

This article aims to implement a statistical learning approach to identify the productive potential of customers in the bank's corporate portfolio. It is done by relying on data retracing customer transactions over the period from January 1, 2016 to December 31, 2018. First, we make use of descriptive data analysis techniques to highlight the main variables which determine the productivity of business customers; then, using a combination of first order Markov chains and regression trees (Haenlein et al., 2007), we

[1]According to a study carried out on the subject at the Research and Investments Department in 2017

propose a computational model of the CLV; finally we use the model to calculate customer values.

2. Literature Review

The issue of customer evaluation in marketing has been of great interest to authors in various industries. The telecommunications sector in particular received the attention of Hwang et al., 2004. In their model, they base the calculation of the CLV around three points: the contribution to past profit, the potential profit and the customer's probability of default. They also offer a set of tools to analyze customer value and segment them based on their values. In their conceptual model, the socio-demographic information of the customers as well as the data on the transactions constitute inputs for the construction box of the customer value. The trio, present value, potential value, and customer loyalty constitute its CLV. The segmentation operated based on customer value makes it possible to offer marketing strategies adapted to each segment.

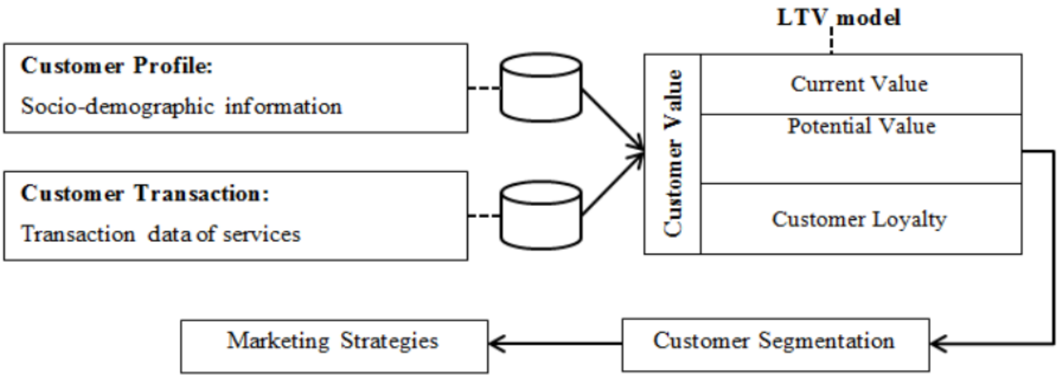


Figure 1: Conceptual framework of the CLV
Source: *Hwang et al. 2004*

Before these authors, Junxiang et al.(1995) presented a model for the calculation of CLV using the duration model. They offer the "customer's monthly margin" and the "customer's survival curve" in the telecommunications sector. For these authors, when competition is viral between companies in a sector, "customer retention" becomes essential. Therefore, it becomes imperative to calculate the "customer survival curve". For the estimation of the "customer survival curve", these authors use the duration model. They inform us that logistic regression and decision trees are important methods for predicting churn and survival rate.

In 2005, Fader et al. propose a model for calculating the CLV based on the RFM model. This model creates groups of clients based on three factors, which are Recency, Frequency and Monetary (Gupta et al. 2006). Recency refers to the time taken between the customer's last purchase and the current date, Frequency refers to the total number of purchases the customer has made throughout their lifespan, and Monetary refers to the customer's average spending during their past purchases (Tabei and Fathian, 2011; Jonker et al. 2002). The main assumption of the RFM model is that the future behavior of the customer is based on the pattern of his past and present behavior. They are thus able to propose an evaluation model for calculating customer value by mainly analyzing their consumption habits.

Ahmadi et al.(2011), in their models, consider that the CLV must take into account three elements: the market risk which affects the financial flows of the client, the flexibility of the company to react to changes and to the costs of attraction and customer retention. Considering these elements, this research presents the model for calculating the CLV for four types of business customer relationships (Reinartz and Kumar, 2000; Cannon et al. 2001). (1):: the environment presents a low risk and the company is not flexible; (2): the environment presents a low risk and the company is flexible; (3): the environment presents a high risk and the company is not flexible; (4): The environment presents a high risk and the company is flexible. These authors thus suggest making an upstream diagnosis of the company's macroeconomic environment.

In the commercial banking sector, Michael Haenlein et al.(2007) proposed a model based on Markov chains, classification and regression trees. The first step is to segment clients into sub-groups based on their contributions to productivity. Classification is done by regression trees. The second step is to build the customer transition matrix between the productivity segments. In the third step, the CLV is calculated using the transition matrix and the productivity in each segment. In this industry, these authors argue that a model for determining "Customer Lifetime Value" should satisfy at least three conditions: first, it must be able to handle discrete one-off transactions, which only occur once in a while life or during very long purchasing cycles (eg. mortgages) and continuous income streams (eg. routine account maintenance fees). This is because retail banks generate income in two main ways: by earning a margin on lending and investing activities and by collecting transaction fees for transactions, credit cards, etc. (Garland, 2002). Second, it must focus on the behavior of a homogeneous group of customers. Third, it must be easily understood and parsimonious, to ensure its applicability in different contexts (Michael Haenlein et al., 2007).

3. Data Requirements

The data used as Baseline for this study come from the Data Warehouse of Afriland First Bank Cameroon. We retain the data on corporate clients (GE, ME, PE, and TPE) in bank

Paper ID: SR21123105748 DOI: 10.21275/SR21123105748

relationship with Afriland before January 1, 2016. We observe the transactions on the latter over the period from January 1, 2016 to December 31, 2018. The choice to work in the corporate client segment is guided by the objectives of the Research and Investments Department, which seeks to assess this type of clientele and develop strategies to increase the revenues derived from this clientele. The choice of the chosen period is explained by the fact of having individuals whose product consumption behavior would not be affected by the creation of recent products. We are therefore working over a period when the same products are marketed. More precisely, for this study, we use the basis of the history of accounts, the basis of products, the basis of the productivity of business customers. The account history database records all movements that take place in all customer accounts. This information is recorded with the accounting date, the type of transaction, the direction and the currency. The product database records, for each business customer, their holding status for a bank product (remote banking product, bank card product). The base of the productivity of business customers gives the bank its monthly productivity for each business. The productivity of a business client over a year can be understood as the result produced by this client at the bank. A customer has a high potential for productivity if he belongs to a group of customers who, under the same conditions of marketing support and customer relations, would produce a result equivalent to the average level of productivity of this group and this for the same level of investment.

The raw account history database has over 13 million records. The treatments carried out on this basis made it possible to construct the annual bases of transactions. Over the period from January 1, 2016 to December 31, 2018, we thus obtain three annual databases retracing customer transactions. To alleviate the complexity of our models, the variables Status_Prod1, Satut_Prod2, Status_Prod3, and Status_Prod4 have been grouped together to form the Stat_Bank_dist variable which informs whether the company has subscribed to at least one of the previous remote banking products. The CardType variable of the gross product database has been transformed into Stat_Carte_ban coded 0 if the company has not subscribed to any bank card product and 1 otherwise. The description of the variables of the final databases is given in table 1.

Table 1: Variables of the final database for each year

Variables	Signification	Codification
Age_relation_ban	Provides information on the duration of the company's bank relationship	/
Stat_Banque_dist	Find out if the company owns a remote banking product	0=No; 1=Yes
Stat_carte_ban	Provides information if the company has a bank card product	0=No; 1=Yes
Type_Ent	Information on the type of corporate	1= GE, 2=ME, 3=PE, 4=TPE
Stat_Inactiv	Provides information on the corporate inactivity status on the bank	0=Inactive ; 1 = Active
Prod	Provides information on the company's annual productivity	/

Source: *Our calculations based on data from Afriland First Bank, 2021*

4. Methodology

Descriptive analyzes precede modeling and calculating customer value. One-dimensional analysis refers to the statistical study of the modalities of a single variable, or of several variables considered independently, for describing the sample. In the context of this study, we use it to describe the characteristics of the population covered by our study. Two-dimensional analysis, on the other hand, refers to the statistical study of the relationships that may exist between two variables. The two-dimensional analyzes are used to detect the correlations between the potentially explanatory variables of the productivity of business customers. We will focus on the Kruskal Wallis test (in case of non-normality of the variable of interest), on the chi-square test of independence (in certain cases) in order to shed light on these relationships, and on Cramer' statistics, denoted V, to measure the degree of the possible link demonstrated. This bivariate analysis is necessary because it will serve to better understand the results of analyzes that will follow. The significance level retained for this study is 5%. Thus, any relationship highlighted will be statistically significant if the associated p-value is below this threshold.

Multivariate analysis is the statistical study of the relationships that may exist between several variables (explanatory methods). It can also lead to structuring the studied variables (descriptive methods). We use the two groups of methods to describe and give a first explanation, in detail, the population we are dealing with and the first factors of productivity. Thus, we will use multiple correspondence analysis (MCA). The MCA is the factorial method best suited to tables in which a set of individuals (in rows) is described by a set of qualitative variables (in columns). From an n-dimensional space, we obtain graphic planes in which it is possible to visualize the proximities between modalities of various qualitative variables. Indeed, the subjects on which we work are divided according to various characteristics. Thus, it is important to represent them in plans that will facilitate their description. These representations can better explain the behavior of customers with regard to productivity. In cases where the independence hypothesis (chi-square test) is not rejected at the significance level of 5%, the characterizing variable, which has been crossed with one of the variables of interest, is automatically put in illustrative mode. In the context of the MCA, in addition to the Kaiser criterion, we rely on the correction of Benzécri (1979) in order to retain a reduced number of axes restoring the maximum of information and thus facilitate the task in our interpretations. This correction is obtained by the following transformation:

$$\mu_B = \left(\frac{p}{p-1}\right)^2 (\mu - \frac{1}{p})^2$$

μ_B = Transformed or corrected eigen value, μ = Eigen value from MCA, p = Number of active variable

In our study, MCA allows us to draw up an initial profile of corporate clients with respect to their contribution to the productivity of the corporate client portfolio at Afriland First Bank Cameroon. This is done by analyzing the modalities of variables strongly correlated with productivity.

This descriptive analysis step is crucial for determining the main explanatory factors for the productivity of business customers.

The approach to calculating customer value in this study draws heavily on the work of Michael Haenlein et al. (2007). After having brought out the major groups of variables determining the productivity of business customers, the modeling is carried out in three stages. First, we construct homogeneous segments of clients in view of their contributions to the productivity of the business portfolio using regression trees, then we use these segments as discrete states used to estimate the transition matrix Markov, finally this transition matrix is used to calculate CLV for each customer segment.

Stage 1

Regression trees (unsupervised learning technique) follow on from previous analyzes. They allow us, thanks to the groups of variables identified above as being the main determinants of the contribution to productivity, to identify homogeneous customer segments. Decision tree learning refers to a method based on the use of a decision tree as a predictive model. Regression tree learning can predict a customer's productivity based on feature predictors that can be both quantitative and qualitative. Breiman et al. (1984) first introduced this technique. Regression trees use separation techniques based on maximizing interclass variance (having subsets whose values of the target variable are as widely dispersed as possible). The prediction of the numeric variable is then the within-class mean. The CART (Classification And Regression Trees) algorithm consists in intelligently choosing a variable, intelligently cutting the data according to this variable, the forecast of productivity in a segment is then the average productivity of the customers of the segment. Then were start on the subtree sobtained.

The goal is to find the divisions, $R_1, \dots, R_J$, which minimize the loss function$\sum_{j=1}^{J}\sum_{i \epsilon R}(y_i - \hat{y}_{R_i})^2$

$\hat{y}_{R_i}$: the mean of the response variable in the region R_j

We seek for each node the division, or more precisely the variable X_j, and the rule of division S, which will contribute to the greatest decrease in the heterogeneity of the child nodes on the left$R_1$and rightR_2.

$$R_1(j,s) = \{X | X_j < s\}\ et\ R_2(j,s) = \{X | X_j \geq s\}$$

The objective is to find the values of j and s which minimize the loss function:

$$\sum_{i:X_j \epsilon R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:X_j \epsilon R_2(j,s)} (y_i - \hat{y}_{R_2})^2$$

As part of our study, $y_i$denotes the annual productivity of company i in the segment R. $\hat{y}_{R_1}$represents the average productivity predicted in the segmentR_1.

This step is crucial and we use post-pruning techniques to control the risk of overfitting our model. This strategy consists in building the tree in two stages: we first produce the tree in an expansion phase, using a first fraction of the data sample (training sample), then we reduce the tree, by relying on another fraction of the data (test sample) in order to optimize the tree's performance. The construction of the sequence of nested trees is based on a penalty for the complexity of the tree.For each value ofα, there is a tree $T \subset T_{max}$ which minimizes$\sum_{m=1}^{|T|}\sum_{i:X_i \epsilon R_m}(y_i - \hat{y}_{R_m})^2 + \alpha|T|$, $\alpha$is chosen by cross-validation by V-sets.

Finally, we improve the performance of the model by bagging (bootstrap aggregating). It consists in building several regression trees by resampling the training set, then by building the trees by a consensus procedure.

Stage 2

Once the segments have emerged well, we take each segment as a state of the nature of our Markov chain. We can therefore build the transition matrix between these different states (transition probability). The homogeneous segments obtained in stage 1 are consideredas states of nature between which clients can migrate according to a first order Markov process. The first-order Markov process implies that migration from one state of nature to another depends only on the properties of the state directly preceding it. This is a memoryless process. We work with this hypothesis. That is to say, the transition of a company from a group to the group directly above or directly below depends only on its characteristics in the group where it was before this migration. Markov chains have long been used in Marketing (Styan and Smith, 1964; Thomson and McNeal., 1967) and more specifically in customer evaluation (Morisson et al. 1982; Pfeifer and Carraway, 2000; Rust et al. , 2004).

A first-order Markov chain, or, more simply, a Markov chain, is a discrete stochastic process whose memory is limited to the last state; that is to say:

$$P(x_{t+1} | x^t{}_{-\infty}) = P(x_{t+1} | x_t)\ \forall\ t \in \mathbb{Z}$$

Let us admit that $\Omega \coloneqq \{\omega_1, \dots, \omega_m\}$represents the m states of the system. The Markov chain is entirely determined by the m × m transition matrix

$$P_{kj} \coloneqq P(X_{t+1} = \omega_k | X_t = \omega_j) = P(\omega_j | \omega_k)$$

Obeying the conditions of consistency

$$P_{kj} \geq 0 \qquad \sum_{k=1}^{m} P_{kj} = 1$$

The construction of the Markov transition matrix is based on data from corporate customer of the portfolio. We use the optimal segmentation model built in the previous step. We take the data over two times: 2016 and 2018 for the same companies. In the first period (year 2016), and using regression tree models, we determine in what level of productivity each corporate customer is. The same process is repeated in the second period (year 2018). It thus becomes possible to observe the migration of corporate customer between levels of productivity. The Markov matrix is thus constructed by taking the frequencies of the firms in each segment at the end of the second period as a proxy for the transition probabilities.

Stage 3

Finally, we determine the CLV for each customer segment as the discounted sum of the dependent contribution margins, weighted with their corresponding transition

Paper ID: SR21123105748 DOI: 10.21275/SR21123105748

probability. So: $CLV_k = \sum_{j=1}^{m} P_{kj} \times \hat{y}_{R_j}(1-r)^{-n}$; Each corporate customer of class k has an estimated CLV_k. We assume the discount rate r to be constant over the modeling period as 3%. Thus, we calculate the annual CLV of each segment in 2 years. This could help to eliminate the effects due to exceptional variations in the productivity of corporate customer in the portfolio.

5. Results

The distribution of individuals according to the type of clientele shows a greater proportion of TPEs in our database. These, which represent more than 70.0% of the population. PEs represent 23.6% of the population and MEs 4.8%. GEs are the least numerous and represent less than 2.0% of the population. The overall rate of remote banking product equipment for our study population is less than 20.0% as of December 31, 2018. The GE subgroup leads the way in remote banking product underwriting. Indeed, in this sub-population, more than 60.0% were equipped with at least one of these products at the end of December 2018. They are followed in the order of ME, PE and TPE with subscriptions within each sub -population of 50.2%, 26.7%, and 13.7% respectively. Regarding bankcard products, the overall equipment rate of our study population was less than 15.00% as of December 31, 2018. Unlike remote banking products, subscription to bankcard products is more done in the sub-group of PE, followed by ME, TPE and finally GE with equipment rates of respectively 19.35%, 14.02%, 13.03%, and 3.11% at the end December 2018. The results show the hyper dominance of GE in terms of productivity. Although the smallest of our study population, this subpopulation displays very high levels of productivity compared to other subpopulations.

The MCA identifies two groups of productivity. Companies that are equipped with remote banking products and / or bankcard products characterize the first group. They seem to be very often active. This group is dominated by PEs. The second group is characterized by low productivity. Companies in this group do not appear to be equipped with remote banking products and bankcard products. These businesses are characterized by a tendency to be inactive. This group is dominated of very small businesses.

Finally, the variables that we retain as determining in the explanation of the productivity of business customers are the status of inactivity, the age of the banking relationship, the status of remote banking and the status of bankcard. These variables are used as a predictor in the segmentation model.

For the segmentation model, we randomly divide the base into training data (70%) and test data (30%). High productivity customers (GM + ME) are treated separately from low productivity customers (PE + TPE).

The optimal segmentation model for business customers in the high productivity group (GE + ME) is obtained for a complexity parameter of 0.0058 (the relative error stabilizes and is minimal) (Figure 2).

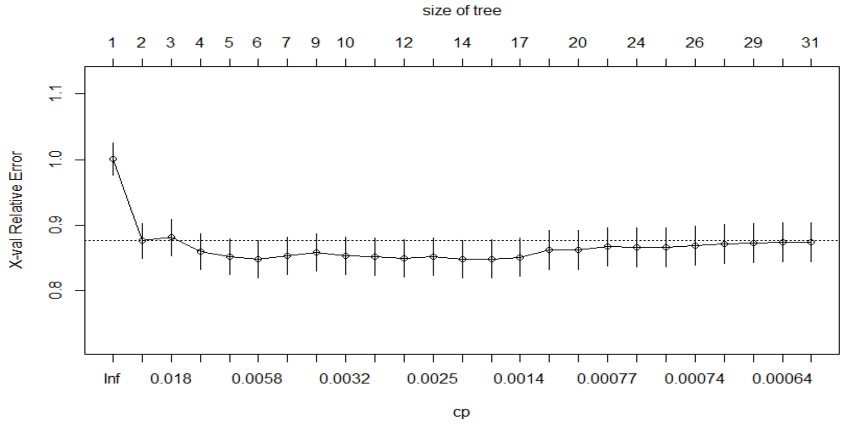


Figure 2: Evolution of the relative error as a function of the complexity parameter for the high productivity model (GE+ME)
Source: *Our calculations based on data from Afriland First Bank, 2021*

For this level of tree pruning, the RMSE no longer drops significantly. The size of the tree is five. The bagging process controls the risk of overfitting the model. We use a bagging model with a replication of 32 trees. (figure 3).

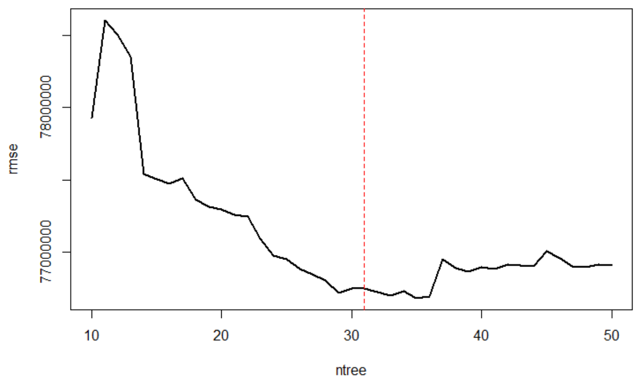
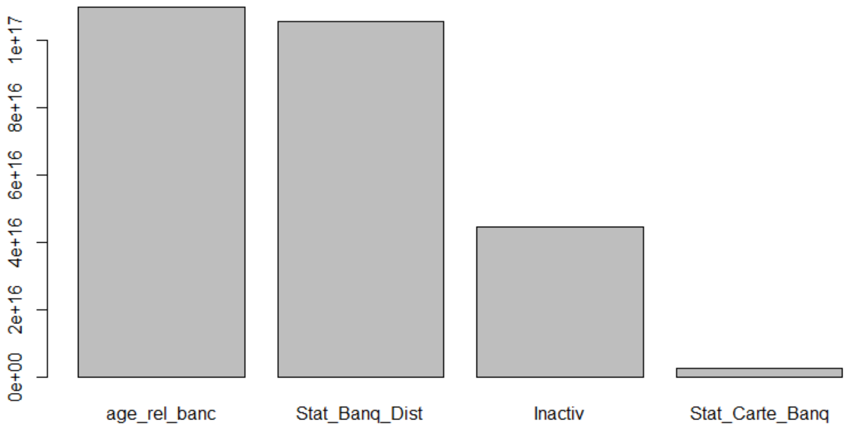


Figure 3: Selection of the number of shafts by Bagging for the high productivity model (GE+ME)

Source: *Our calculations based on data from Afriland First Bank, 2021*

An overview of the importance of the variables of the segmentation model of high productivity (GE+ME) customer reveals the age of the bank relationship and the status of inactivity as having the greatest weight (Graph1). Thus, the age of the bank relationship appears to be very important in predicting the productivity of a portfolio company.



Graph 1: Importance of the variables of the optimal high productivity segmentation tree (GE+ME)
Source: *Our calculations based on data from Afriland First Bank, 2021*

The bank would thus gain by keeping the oldest GE + ME group clients in its portfolio, because they are certainly the most productive. In addition, customers equipped with remote banking products also represent a source of productivity. The optimal segmentation model for business customers of the low productivity group (PE + TPE) is obtained for a complexity parameter of 0.0005 (the relative error stabilizes and is minimal) (figure 4).

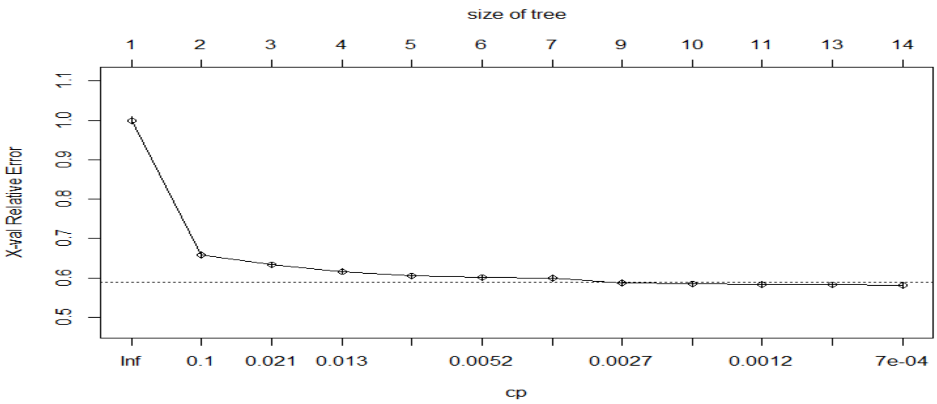


Figure 4: Evolution of the relative error as a function of the complexity parameter for the low productivity model (PE+TPE)
Source: *Our calculations based on data from Afriland First Bank, 2021*

For this level of tree pruning, the RMSE no longer drops significantly. The size of the tree is four. The bagging process controls the risk of overfitting the model. We use a bagging model with a replication of 20 trees (figure 5).

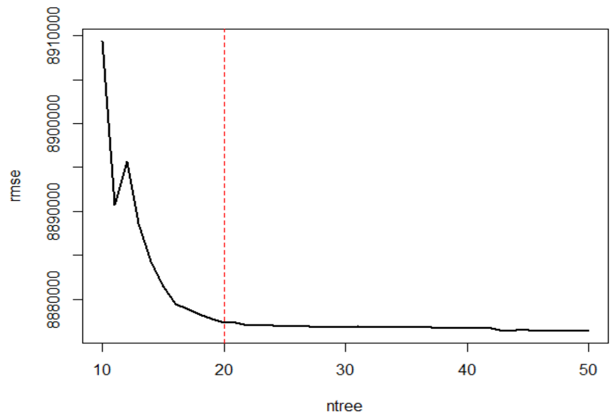
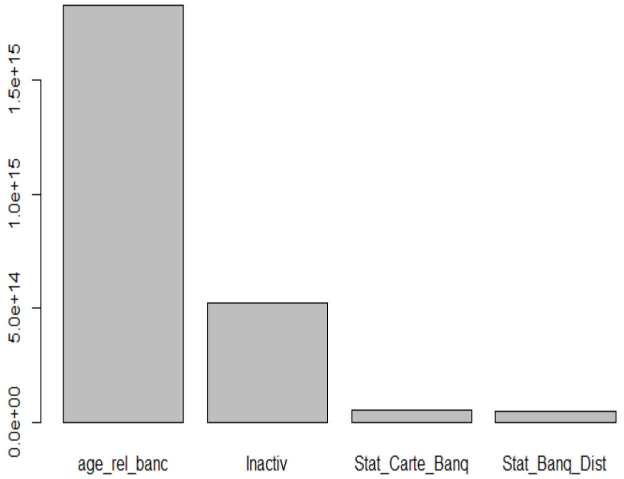


Figure 5: Selection of the number of shafts by Bagging for the low productivity model PE+TPE
Source: *Our calculations based on data from Afriland First Bank, 2021*

An overview of the importance of the variables of the segmentation model of low productivity (PE+TPE) customer reveals the age of the bank relationship having the greatest weight (Graph 2). Thus, the age of the bank relationship appears to be very important in predicting the productivity of a portfolio company.



Graph 2: Importance of the variables of the optimal low productivity segmentation tree (PE+TPE)
Source: *Our calculations based on data from Afriland First Bank, 2021*

The model of homogeneous segmentation of high productivity customers (GE + ME) into homogeneous productivity segments enables five productivity segments to be identified. Figure 6 gives the optimal regression tree for business segmentation. We can read in the figure, following the decision rules, the prediction of the annual productivity of a company. Table 2 summarizes the information on each segment by giving the average productivity of the segment and the proportion of businesses located there. Figure 6 shows the predominance of the age of the bank relationship in the decision model of the productivity segment of a company. Equipping remote banking products and the client's inactivity status are no exception.

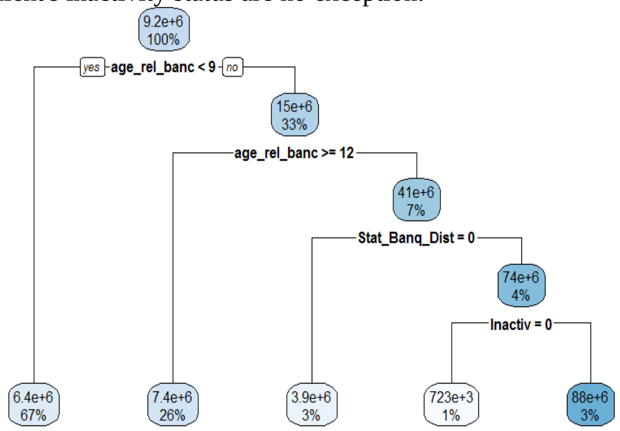


Figure 6: Grouping of high productivity customer (GE+ME) in homogeneous productivity segments
Source: *Our calculations based on data from Afriland First Bank, 2021*

Table 2: Average productivity by segment and proportion of high productivity customer (GE + ME) of the training set

	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Proportion Customer (%)	3.2	26.2	66.6	3.5	0.6
Annual Productivity (CFA F)	87,542,860	7,416,859	6,418,978	3,886,019	722,693

Source: *Our calculations based on data from Afriland First Bank, 2021*

Description of segments

Segment 1: This is the most productive segment; it has an average annual productivity of 87,542,860 CFAfancs. Companies that are equipped with at least one remote banking product characterize it. These companies are very often active. The age of the bank relationship is between 9 and 12 years old. In this segment, in 2018, we find 35.3% of GE, and 64.7% of ME.

Segment 2: This segment has an average annual productivity of 7,416,859 CFA francs. Companies whose bank relationship age is over 12 years characterize it. In this segment, in 2018, we find 30.4% of GE, and 69.6% of ME.

Segment 3: It has an average annual productivity of 6,418,978 CFA francs. Companies whose bank relationship age is less than 9 years characterize it. In this segment, in 2018, we find 19.2% of GE, and 80.8% of ME.

Segment 4: It has an average annual productivity of 3,886,019 CFA francs. Companies that are not equipped with at least one remote banking product characterize it. The age of the bank relationship is between 9 and 12 years old.

Segment 5: It has an average annual productivity of 722,693 CFA franc. Companies that are equipped with at least one remote banking product characterize it. These businesses tend to be inactive. The age of the bank relationship is between 9 and 12 years old. In this segment, in 2018, we find 27.3% of GE, and 72.7% of ME.

Lessons learned from high productivity customer (GE+ME) segmentation

- Age of bank relationship is an important segmentation variable; companies with a long bank relationship display satisfactory levels of dynamism, which would also reflect a dynamism in the development of customer relations. Indeed, the most productive segments are made up of the oldest companies in the portfolio.
- Equipping remote banking products is a significant productivity element. In fact, the segmentation model divides companies into two large groups. The most productive with an average annual productivity estimated at more than 74,000,000 CFA francs and the least productive with an average annual productivity around 3,900,000 CFA francs. The former are equipped with at least one remote banking product. Seconds are not.
- Companies with more than 9 years in the bank relationship have an average productivity of more than 15,000,000 CFA francs against a little less than 7,400,000 CFA francs for those under 9 years old.

The model of homogeneous segmentation of low productivity customers (PE + TPE) into a homogeneous productivity segment makes it possible to highlight four productivity segments. Figure 7 gives the optimal regression tree for business segmentation. We can read in the figure, following the decision rules, the prediction of the annual productivity of a company. Table 3 summarizes the information on each segment by giving the average productivity of the segment and the proportion of businesses located there. Figure 7 shows the predominance of inactivity status and age of the bank relationship in the decision model of the productivity segment of a company.

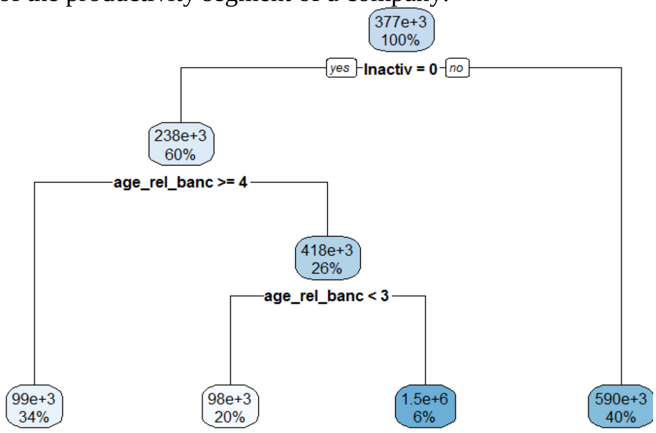


Figure 7: Grouping of low productivity clients (PE + TPE) into homogeneous productivity segments

Source: *Our calculations based on data from Afriland First Bank, 2021*

Table 3: Average productivity per segment and proportion of low productivity companies (PE + TPE) in the training set

	Segment 1	Segment 2	Segment 3	Segment 4
Proportion Customer (%)	6.2	39.5	34.2	20.1
AnnualProductivity (CFA F)	1,460,552	589,778	99,366	97,619

Source: *Our calculations based on data from Afriland First Bank, 2021*

Description of low productivity customer(PE + TPE) segments

Segment 1: This is the most productive segment, with an average annual productivity of 1,460,552 CFA francs. Companies whose bank relationship age is around 4 years characterize it. In this segment, in 2018, we find 15.0% of PE, and 85.0% of TPE.

Segment 2:This segment has an average annual productivity of 589,778 CFA francs. Companies that are very often active characterize it. In this segment, in 2018, we find 42.00% of PE, and 58.0% of TPE.

Segment 3:It has an average annual productivity of 99,366 CFA francs. Companies whose bank relationship age is over 4 years characterize it. In this segment, in 2018, we find 7.0% of PE, and 93.0% of TPE.

Segment 4: It has an average annual productivity of 97,619 CFA francs. Companies whose age of bank relationship is over 3 years characterize it. In this segment, in 2018, we find 27.0% of PE, and 73.0% of TPE.

Lessons learned from low productivity customer (PE + TPE)segmentation

- The age of the bank relationship is an important segmentation variable, company with a long bank relationship showing satisfactory levels of dynamism, which would also reflect a dynamic development of customer relations. Indeed, the most productive segments are made up of the oldest companies in the portfolio.
- The inactivity status is a striking element of productivity. In fact, the segmentation model divides companies into two large groups. The most productive with an average annual productivity estimated at more than 590,000CFA francs and the least productive with an average annual productivity around 238,000CFA francs. The former are very often active. Which is not the case with seconds.

Table 4 gives the transition matrix between the different productivity levels of the high productivity customer segments (GE + ME). This matrix presents the probabilities for a company of migrating from one customer segment to another after 2 years. By considering a company in segment 1, which is in the segment of companies with the highest productivity profile, it has a 20% chance of remaining in this segment in period two. It has more than 75% risk of ending up in segment 2, a segment with lower productivity than that in which it was in period 1. In addition to this observation, there appears a rather alarming risk of 4% that it ends up in segment 5. For such customer, if left unchecked, it is most likely headed for the "churn". The bank's CRM must come into action to improve customer relations with these companies.

Table 4: First-order Markov matrix for transitioning between productivity levels of high productivity customers (GE + ME)

	Segment1	Segment2	Segment3	Segment4	Segment5
Segment1	20.0	75.4	0.0	0.0	4.6
Segment2	0.0	100.0	0.0	0.0	0.0
Segment3	4.7	0.0	91.0	3.6	0.8
Segment4	12.9	59.7	0.0	24.2	3.2
Segment5	20.0	20.0	0.0	0.0	60.0

Source: *Our calculations based on data from Afriland First Bank, 2021.*

Table 5 gives the transition matrix between the different productivity levels of the low productivity customer segments (PE+TPE). This matrix presents the probabilities for a company of migrating from one customer segment to another after 2 years. Considering a company in segment 1, that is to say in the segment of companies with the highest productivity profile, it is difficult for it to remain in this segment in period two. It has more than 11% of risk of ending up in segment 2, a segment with lower productivity than that in which it was in period 1. In addition to this observation, there appears a rather alarming risk of 6% that it ends up in segment 4. For a such and such a company, if nothing is done, it is very likely heading towards the "churn". The bank's CRM must come into action to improve customer relations with these companies.

Table 5: First-order Markov matrix for transition between productivity levels of low productivity customers (PE + TPE)

	Segment1	Segment2	Segment3	Segment4
Segment1	0.0	11.0	89.0	0.0
Segment2	7.3	64.5	22.1	6.1
Segment3	0.0	6.9	93.1	0.0
Segment4	12.4	39.3	14.8	33.6

Source: *Our calculations based on data from Afriland First Bank, 2021*

Paper ID: SR21123105748 DOI: 10.21275/SR21123105748

In general, the direction of corporate customer migration can be summarized as follows: businesses from high productivity segments will migrate mainly to low productivity segments while those from low productivity segments migrate to higher productivity segments. This with a few exceptions.

For high productivity customers (GE + ME), segment 1 displays the highest CLV. The CLV of this segment is estimated at 87,019,958 CFA francs (Table 6). A company in this segment thus has a current value in two years estimated at 87,019,958 CFA francs. This segment represents a good productivity niche for the bank. In addition, companies in this segment, in the current state of the bank's marketing policy, have a propensity to remain in high productivity segments. Companies that are equipped with at least one branchless banking product characterize this segment. These companies are very often active. The age of the bank relationship is between 9 and 12 years old. Thus, by strengthening support for segment companies (promotional offers, overdraft and loan facilities), the bank would further increase its productivity and gain more from this segment. Looking at the performance achieved in this segment in 2018, the bank is on a good dynamic in this segment. She should continue there. Segment 5 is that of companies with the lowest customer values in the high productivity customer portfolio (GE + ME). Companies that are equipped with at least one remote banking product characterize it. These businesses tend to be inactive. Although these companies are equipped with remote banking products, the fact that they are inclined to be inactive greatly reduces their values. However, these companies have good tendencies to migrate to high productivity segments. The bank should activate the mechanisms to combat the inactivity of companies in this segment.

Table 6: Customer values by homogeneous productivity segment for high productivity customers (GE + ME)

	Segment1	Segment2	Segment3	Segment4	Segment5
CLV (CFA F)	87,019,958	7,372,558	6,380,638	3,862,808	718,377

Source:*Our calculations based on data from Afriland First Bank, 2021*

For low productivity customers (PE + TPE), segment 1 displays the highest CLV. The CLV of this segment is estimated at 1,451,828 CFA francs (Table 7). A company in this segment thus has a current value in two years estimated at 1,451,828 CFA francs. Unfortunately, companies in this segment have a high probability of migrating to low productivity levels. Companies whose bank relationship age is around 4 years characterize this segment. The young age of companies in this segment could explain their instabilities. The fact remains that these companies represent a niche of future productivity. Segment 4 is that of companies with the lowest customer values in the low productivity customer portfolio (PE + TPE). Companies whose age of bank relationship is over 3 years characterize it. These companies have high probabilities of migrating to high productivity segments. These companies need to be supervised, in order to increase trust in the customer relationship.

Table 7: Customer values by homogeneous productivity segment for low productivity customers (PE + TPE)

	segment1	Segment2	segment3	segment4
CLV (CFA F)	1,451,828	586,255	98,773	97,036

Source:*Our calculations based on data from Afriland First Bank, 2021*

6. Conclusion

The issue of customer value is decisive in the management of customer relations and the conduct of marketing policy in a commercial company. In Cameroon, telecommunication companies like MTN Cameroon and Orange Cameroon, have already integrated this issue into their daily Marketing. In the banking sector, on the other hand, it remains even less popular. However, it is obvious that the banking ecosystem in Cameroon is becoming increasingly harsh. There is fierce competition from banks for both customer acquisition and retention. The survival of banks in this ecosystem will depend on their ability to "know" the customers. Knowing customers therefore implies that banks should have a high-performance instrument that can allow them to assess them. This instrument must drink in data tracing customer transactions, and therefore be able to anticipate their needs.

By taking advantage of uni and multivariate analysis tools, variables such as inactivity status, status of equipment in bankcard and remote bank product and age of the bank relationship were revealed. as determining factors in the prediction of the productivity potential of the business portfolio. Equipping as a remote banking product discriminates productivity in the group of high productivity customers (GE + ME) into two broad segments. The most productive with an average annual productivity estimated at more than 74,000,000 CFA francs and the least productive with an average annual productivity around 3,900,000 CFA francs. On the other hand, in the group of low productivity customers (PE + TPE), it is the inactivity status that appears to be the most important discriminating factor. He divides these clients into two large groups. The most productive with an average annual productivity estimated at more than 590,000 CFA francs and the least productive with an average annual productivity around 238,000 CFA francs. The transition matrices built, the calculation of the customer values of each customer group of similar productivity behavior and for each segment results in the highest CLV in the group of high productivity customers (GE + ME) at 87,019,958 CFA francs. That of the group of low productivity customers (PE + TPE) is estimated at 1,451,828 CFA francs.

References

[1] Ahmadi, K., Taherdoost, H., Fakhravar, S., Jalaliyoon, N., 2011, “A New Model for Evaluating Customer Lifetime Value in High Risk Markets”, International Conference on Social Science and Humanity, IPEDR vol.5, IACSIT Press, Singapore.

[2] Benzécri, Jean Paul, 1979,« Pratique de l'analyse des données », Abdi, H. et al. Linguistique&lexicologie. Dunod, 1979.

[3] Cannon, J.P., Christian, H., 2001 "Buyer–Seller Relationships and Customer Firm Costs," Journal of Marketing, 65 (January).
[4] Dwyer, F. R., 1997, "Customer lifetime valuation to support marketing decision making",Journal of Interactive Marketing, 11(4), 6–13.
[5] Fader, P. S., B. G. S. Hardie, and K. L. Lee, 2005, "RFM and CLV: Using Iso-value Curves for Customer Base Analysis." Journal of MarketingResearch, XLII, 415-30.
[6] Firmin Bossali et al., 2015 Le protocole de recherche : étape indispensable du processus de recherche garantissant la validité des résultats, Hegel Vol. 5 N° 1
[7] Garland, Ron, 2002,"Estimating customer defection in personal retail banking", International Journal of Bank Marketing.
[8] Gupta, S. and Zeithaml, V., 2006, "Customer Metrics and Their Impact on Financial Performance", Marketing Science, Vol. 25 No. 6, 718-739.
[9] Hwang, H., Jung, T., & Suh, E., 2004, "An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry", Expert Systems with Applications, 26(2), 181–188., 2004
[10] Jonker, Piersma, Poel, 2002 "Joint Optimization of Customer Segmentation and Marketing Policy to Maximize Long-Term Pro Tability", Econometric Institute Report.
[11] Junxiang, Lu, Overland, Park and Kansas, 1995, "Modeling Customer Lifetime Value Using Survival Analysis − An Application in the Telecommunications Industry", Data Mining Techniques.
[12] Kincaid, J.W.,2003, "Customer Relationship Management: Getting It Right!", Prentice Hall Professional, 480 pages.
[13] Kim, J., Suh, E., & Hwang, H.,2003, "A model for evaluating the effectiveness of CRM using the balanced scorecard". Journal of Interactive Marketing, 17(2), 5–19.
[14] Kotler, P., 1997**,** "Marketing management: Analysis, planning, implementation and control (9th ed.)". UpperSaddle River, NJ, USA:Prentice-Hall.
[15] MahsaEsmaeiliGookeh and Mohammad JafarTarokh, 2013,"Customer Lifetime Value Models : A literature Survey", International Journal of Industrial Engineering & Production Research, Volume 24, Number 4, pp. 317-336.
[16] MahsaTavakolijou, 2009, "A Model to Determine Customer Lifetime Value in Iranian Banking Industry", Luleå University of Technology – Department of Business Administration, Technology and Social Sciences - Master's Thesis.
[17] Michael Haenlein, Andreas M. Kaplan, Anemone J. Beeser, 2007, "A Model to Determine Customer Lifetime Value in a Retail Banking Context", European Management Journal Vol. 25, No. 3, pp. 221–234.
[18] Paul D. Berger, Nada I. Nasr**,**1998, "Customer Lifetime Value : Marketing Models and Applications", Journal of Interactive Marketing, Volume 12/Number 1/Winter.
[19] Reinartz, Thomas., Kumar, V., 2005 "Balancing Acquisition and Retention Resources to Maximize Customer Profitability", Journal of Marketing, 69, 63–79.
[20] Tabaei, Z., Fathian, M.,2011, "Product Recommendation Based on Customer Lifetime Value —An Electronic Retailing Case Study", International Conference on Networking and Information Technology, IPCSIT Vol.17, IACSIT Press, Singapore.
[21] Y. Ekinci, F. Ülengin, N. Uray, and B. Ülengen, 2014,"Analysis of customer lifetime value and marketing expenditure decisions through a Markovian-based model".European Journal of Operational Research, 237 :278-278.