International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Using Latent Class Analysis to Identify Political Behavior of Moroccan Citizens on Social Media

Zakia Acharoui, Altaf Alaoui, Mourad Azhari, Abdallah Abarda, Badia Ettaki, Jamal Zerouaoui

***Abstract:** The high use of social media has led to a new form of political involvement and participation. In this paper, we use Latent Class Analysis to identify participants' behavior regarding political participation and engagement based on the nature of their interaction on social media. The LCA findings reveal three statistically distinct and behavioral classes regarding political interaction on social media. The profiles were ranged from 'Activist' that show more engagement in political activity, such as following candidates and political parties, posting and participating in discussions related to economic, social or political issues or, encouraging others to debate their point of view, to 'Agitator' and 'Outsider' profiles that show a low probability of interacting on social media and engaging in political actions. The LCA technique has provided meaningful and distinct information on the participants' political profile than clustering classical techniques.*

***Keywords**: Latent Class Model, Clustering, Social Media, Political Commitment.*

I. INTRODUCTION

Clustering or cluster analysis is a technique of unsupervised classification used to divide data into groups of similar items. Dissimilar to the classification that analyses data by learning from predefined labels, clustering is mainly used when the classes are not known.

In clustering, data on the same cluster have similar characteristics that distinguish it from other clusters, as the data grouping is based on the principle of maximizing the intraclass similarity and minimizing the inter-class similarity.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author

Acharoui Zakia *, Laboratory of Engineering Sciences and Modeling, Faculty of Sciences, Ibn Tofail University Kenitra, Morocco. Email: zakia.acharoui@gmail.com

Altaf Alaoui, Laboratory of Engineering Sciences and Modeling, Faculty of Sciences, Ibn Tofail University Kenitra, Morocco. Email: altaf.alaoui@gmail.com

Mourad Azhari, Laboratory of Engineering Sciences and Modeling, Faculty of Sciences, Ibn Tofail University Kenitra, Morocco. Email: azharimourad@yahoo.fr

Abdallah Abarda, Laboratory of Mathematical Modelling and Economic Calculations, FSJES, Université Hassan 1er, Settat, Morocco. Email : abardabdallah@gmail.com

Badia Ettaki, Laboratory of Research in Computer Science, Data Sciences and Knowledge Engineering, Department of Data, Content and knowledge Engineering School of Information Sciences, Rabat, Morocco. Email: ettakibadia@yahoo.fr

Jamal Zerouaoui, Laboratory of Engineering Sciences and Modeling, Faculty of Sciences, Ibn Tofail University Kenitra, Morocco. Email: j_zerouaoui@yahoo.fr

Clustering has been applied in a wide variety of research problems where there is a need to sort groups from a huge set of data and that by using several algorithms dedicated to this purpose.

The approaches for clustering can be classified into two types: Hierarchical and nonhierarchical Clustering. Hierarchical clustering is a sequential process that starts by calculating the distances of each individual to all the other individuals and forming a distance matrix for all individuals while, Nonhierarchical methods use some criteria other than the distance to cluster individuals into a certain number of clusters, such as partitioning, k-means, and mixed methods.

Since the clustering solution is not unique and it strongly depends upon the analyst's choices, finding an optimal cluster represents a major challenge for the user. Furthermore, clustering does not provide information about the probability of a particular object belongs to a certain class [1]

In contrast, Latent Class Analysis is based on a statistical model that uses observed data to estimate parameter values and gives to each individual a certain probability of membership to each latent class [1] And by considering the fact that items are discrete in reality and the variables intend to measure those items are likely to be categorical, traditional clustering becomes inappropriate [2] Thus, latent class analysis can be a more valuable tool for grouping individuals within a population.

LCA has been extended to various models such as Latent Class Regression (LCR) model that considers the effect of some covariates on the probability of belonging to a certain class [3][4][5][6].

The Confirmatory LCA (CLCA) is an alternative approach to latent class modeling that requires prior hypotheses regarding the number and the nature of latent classes underlying the date [7][8][9]. Otherwise, Latent Class Growth Model (LCGM) is used to identify distinct subgroups of individuals following a similar pattern of change over time on a given variable [10].

Many research studies have used LCA to classify individuals. For example, in the psychology field, LCA has been used to classify children into classes of behaviorally disordered children [11][12]. Similarly, Klonsky E. D. And Olino M. T. [13] have utilized a latent class analysis to identify clinically distinct subgroups of self-injurers. The finding shows four subgroups of self-injurers, that differ on measures of depression, anxiety, borderline personality disorder, and suicidality. In medicine, Lui Z. et al [14] have adopted LCA to identify subgroups of physicians' competency based on eight competency dimensions, and a four-class model best fit the data, which are excellent competency groups,

lack of professionalism competency group, individual competency-driven group, and lack of competency cognitive group. Also, Rossi A. et al.

[15] have used a Latent Class Analysis (LCA) approach to provide a better understanding of the covariance between psychological distress and cancer-related issues.

In educational studies, Denson N. and Ing M. [16] provides a pedagogical application of LCA on higher education to classify entering freshmen based on their pluralistic orientation in order to help college administrators in their program planning and targeted interventions around issues of diversity. The best-fitting model is a four-class model; high pluralistic orientation; high-disposition, low-skill; low-disposition, high-skill; and low pluralistic orientation. Similarly, LCA has been used by Young and Nylund-Gibson [17] to analyze data from the Longitudinal Study of American Youth to examine student attitudes towards mathematics and science.

In social science, Rhead et al [18] have used LCA to study UK public concern toward the environment and its association with pro-environmental behaviors. The study's results show four classes of people, defined by their concern for the environment: Pro-environment, Neutral Majority, Disengaged and Paradoxical. And this finding has been highly predictive of environmental behavior than most previous research with the environmental attitude classes.

Fox B. et al [19] has conducted an LCA to examine the distinct types of individuals who support or oppose police militarization using a national sample of 702 American adults. The findings demonstrate the complexity of the public sentiment toward this controversial topic in contemporary American policing. On the other hand, Bentaleb Y. et al [20] have used latent class analysis to identify unobservable classes of cybercrime victims and propose a measure of the risk of cybercrime based on conditional probabilities resulting from LCA.

In politics, many studies have highlighted the use of LCA as an approach to detect latent variables on a population. For instance, Huang C. [2] has employed LC cluster analysis to determine the number of latent classes in each of the two dimensions (ethnic and national) of politic identity in Taiwan. Hooghe M. and Oser J. [21] have used LCA to assess democratic ideals among European citizens as reported in the 2012 European Social Survey. The findings demonstrate that a majority of Europeans consider political and social rights as equally important, while some citizens predominantly emphasize either political or social rights.

On their study, Alvarez R. M. et al [22] use latent class analysis to identify the four faces of political participation and that include both forms of participation, conventional and unconventional contrary to previous research that generally focus on one form of participation.

This paper aims to examine the relationship between social media use political awareness, engagement and participation by using LCA. In section two, the LC model is introduced. An application of LCA is presented in the third section, with a comparison to a classical clustering algorithm (MC-DBSCAN). Section four and five devoted, respectively, for discussion and conclusion.

II. LATENT CLASS ANALYSIS

A. Latent Class Model Foundation

Let an indicator function I defined as follows:

$$I(y_j = r_j) = \begin{cases} 1 & when\ the\ response\ to\ item\ j\ is\ r_j \\ 0 & Othewise \end{cases}$$

Where y_j corresponds to line my of Vector Y. The following relation shows that the probability of observing a particular vector is a function of the probability of observing each latent class and the conditional probability that an individual chooses the modality r_j of the variable j knowing that it belongs to class C :

$$P(Y = y) = \sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{r_j}^{R_j} \rho_{j,r_j/c}^{I(y_j=r_j)}$$

Therefore, for each individual, the joint probability of choosing an answer y that falls on Class C can be calculated by:

$$P(Y = y, L = c) = \gamma_c \prod_{j=1}^{J} \prod_{r_j}^{R_j} \rho_{j,r_j/c}^{I(y_j=r_j)}$$

Based on the formula $\mathrm{P(AB) = P(B).P(A/B)}$, where A and B are two random events, we can write :

$$P(Y = y, L = c) = P(L = c).P(Y = y/L = c)$$

And we know that:

$$P(Y = y/L = c) = \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j/c}^{I(y_j=r_j)}$$

And that $P(L = c) = \gamma_c$, hence :

$$P(Y = y, L = c) = \gamma_c \prod_{j=1}^{J} \prod_{r_j=1}^{R_j} \rho_{j,r_j/c}^{I(y_j=r_j)}$$

A relationship between marginal probability and joint probability can then be obtained (1):

$$\begin{aligned} P(Y = y) &= \sum_{c=1}^{C} \gamma_c \prod_{j=1}^{J} \prod_{r_j}^{R_j} \rho_{j,r_j/c}^{I(y_j=r_j)} \\ &= \sum_{c=1}^{C} \gamma_c . P(Y = y/L = c) \\ &= \sum_{c=1}^{C} P(L = c).P(Y = y/L = c) \\ &= \sum_{c=1}^{C} P(Y = y, L = c) \end{aligned}$$

Therefore,

$$P(Y = y) = \textstyle\sum_{c=1}^{C} P(Y = y, L = c) \qquad (1)$$

B. Observed Variables and Latent Variables

The quantity that determines the relationship between the observed variables and the latent classes is noted:

$$\rho_{j,r_j/c}$$

It defines the conditional probability that an individual will choose a response modality knowing his class of membership. $\rho_{j,r_j/c}$ varies from one class to another. For a given class C, a value very close to 1 means that the individuals who belong to this class have a high probability of responding to the modality of variable j.

The uncertainty of classification can be assumed to be high when the posterior probabilities are very small across classes. To obtain the expression of a posteriori probability, we will use the Bayes formula:

$$P(L=c/Y=y)=\frac{P(Y=y/L=c).P(L=c)}{P(Y=y)}$$

With $P(L=c/Y=y)$ is the posterior probability that an individual of class C knowing the response structure.

We have already seen that:

$$P(Y=y)=\sum_{c=1}^{C}\gamma_c\prod_{j=1}^{J}\prod_{r_j}^{R_j}\rho_{j,r_j/c}^{I(y_j=r_j)}$$

And $P(L=c)=\gamma_c$, Thus, the posterior probability is equal to (2) :

$$P(L=c/Y=y)=\frac{\left(\prod_{j=1}^{J}\prod_{r_j}^{R_j}\rho_{j,r_j/c}^{I(y_j=r_j)}\right).\gamma_c}{\sum_{c=1}^{C}\gamma_c\prod_{j=1}^{J}\prod_{r_j}^{R_j}\rho_{j,r_j/c}^{I(y_j=r_j)}} \quad (2)$$

C. Parameter Estimation and Model Selection

a. Parameter Estimation

The model parameters can be estimated using the maximum likelihood method, employing the Expectation-Maximization (EM) algorithm, which seeks to converge to a local maximum.

The posterior probability is, written according to two parameters that will be estimated by the maximum likelihood method. An iterative approach to the method will be illustrated by the EM Algorithm. It should be noted that there is another iterative procedure that uses the Newton-Raphons algorithm, developed by Dempster A. P. et al.[23] Whenever an iterative algorithm is used to estimate the parameters, it is necessary to define a stopping criterion. Two stopping criteria will, therefore, be considered: the first corresponds to setting a maximum number of iterations and the second consists in using a convergence index that measures the Maximum Absolute Deviation (MAD) between the estimated parameters of two successive iterations, a very small threshold will be chosen for the MAD (for example MAD<0.000001). The algorithm will converge when the convergence criterion is achieved before reaching the maximum number of iterations set a priori.

The EM algorithm is an iterative procedure that is used to estimate the two parameters necessary for latent classification. It is based on the maximum likelihood method. The fundamental idea of maximum likelihood estimation is to quantify the adequacy between a probability distribution and a sample: the greater the likelihood of the sample, the better the adequacy. In practice, the EM algorithm makes it possible to converge towards the local maximum likelihood and is carried out in two steps:

- Step E (Expectation).
- Step M (Maximization).

Let us recall that:

$$P(Y=y)=\sum_{c=1}^{C}\gamma_c\prod_{j=1}^{J}\prod_{r_j}^{R_j}\rho_{j,r_j/c}^{I(y_j=r_j)}$$

We look for parameters that maximize the likelihood function, or its logarithm. The Logarithm of the likelihood function is written as follows (3):

$$\iota(\theta)=\sum_{i=1}^{N}log\left(\sum_{c=1}^{C}\gamma_c\prod_{j=1}^{J}\prod_{r_j}^{R_j}\rho_{j,r_j/c}^{I(y_j=r_j)}\right) \quad (3)$$

For the EM algorithm to work properly, it is necessary to start generating random values as initial values of the two parameters such as:

$\sum_{c=1}^{C}\gamma_c=1$ and $\sum_{r_j}^{R_j}\rho_{j,r_j/c}=1$

And, the first step consists of calculating the expectation of the full log-likelihood (i.e., assuming known unobserved classes). The second step consists of finding the local maximum.

The algorithm stops when the difference between the expectations of the two consecutive log-likelihoods is less than a set tolerance threshold. We speak of "local maximum" because the likelihood function can have several maximums. Thus, when all these have been identified, only the largest of them should be selected.

b. Model Selection

The selection of a latent class model is difficult due to the determination of the number of latent classes, and then comes the problem of limiting the number of parameters. According to Agersti A. [24], to find the most appropriate model, the likelihood ratio statistic can be used, noted G^2. By adopting the previous ratings and considering the contingency table of item responses composed of W cells with f_w are the observed frequencies for w, and $\widehat{f_w}$are the expectations of these frequencies, the method proposed by Agersti is defined as follows:

Once the parameters are estimated, the observed frequencies of the different possible vectors - in the contingency table - of the observed variables are compared with their expectations. This is done by comparing the statistics given by (4):

$$G^2=2.\sum_{w=1}^{W}f_w.\log\left(\frac{f_w}{\widehat{f_w}}\right) \quad (4)$$

And a Chi-square at ddl degree of freedom, with $ddl=W-P-1$.

The latent class conditional independence model is acceptable if G^2 it is below the threshold defined by the quantiles of the Chi-square statistic. On the other hand, the comparison between $G^2$and one Chi-square cannot be considered if the quantity $\frac{N}{W}$ is small (less than 5) [25], with N being the sample size. To correct this problem, several solutions have been proposed to obtain a reference distribution for the G^2test statistics. Two different approaches that are conceptually similar are "parametric bootstrap" [26], and a Bayesian procedure called "post predictive controls",

both approaches are based on the generation of several random data sets, and adjustment - for each model - to a random data set and the calculation of the test statistic. When we have two models to compare, we use the difference in the statistics associated with each model: $G_{\Delta}^2 = G_B^2 - G_A^2$ with $G_{\Delta}^2 \approx \chi_{ddl}$ and $ddl = ddl_B - ddl_A = P_B - P_A$. In this case, it is necessary to have a small ddl, otherwise we risk falling into the previous problem [27], a p-value can be obtained if the latent class model is well adjusted.

D. Determination of the degree of freedom

It is assumed in the LCA that P parameters to be estimated are required. The degree of freedom associated with the statistics G^2 is defined by $ddl = W - P - 1$, with W is the number of cells in the contingency table: $W = \prod_{j=1}^{J} R_j$ and P is the sum of the number of prevalence to be estimated and the number of response probabilities.

Since it is necessary to introduce a concept for evaluating model choice, several studies have been carried out to achieve this goal. Nagin D. [28] has proposed a diagnostic criterion for latent classes known as "Odds Of Correct classification, OCC" or "the certain classification ratio", based on the posterior average probabilities. This criterion can also be used to check the fit of a model. Specifically, an OCC of less than five for all classes in a model suggests that there is some instability within that model.

A second criterion is an entropy, proposed by Ramaswamy et al. [29], which assesses the accuracy of classification through the assignment of individuals to classes according to their probabilities. It varies from 0 to 1, so a higher value indicates a better classification; formally entropy is defined as follows (5):

$$E = 1 - \frac{\sum_{i=1}^{n} \sum_{c=1}^{C} -P_{ic} \log P_{ic}}{n \log C} \qquad (5)$$

However, the choice of the entropy criterion for measuring accuracy may be wrong, because E depends on the number of classes.

To choose the models, many criteria can be used such as the Akaike Information Criterion (AIC) [30], The Bayesian Information Criteria (BIC) [31], the Coherent AIC (CAIC) [32] and, the adjust BIC (ABIC) [33].

The Akaiké AIC's information criterion is an adjustment measure based on information theory. For latent classes, AIC is defined by:

$$AIC = G^2 + 2P$$

Schwarz's Bayesian information criterion weights the number of parameters differently. For latent classes, it is equal to:

$$BIC = G^2 + (\log(N)).P$$

Other information criteria have been proposed, namely, adjusted BIC and consistent AIC. It should be noted that, although several model adjustment statistics can be used to evaluate a plausible model, the choice of the final model also depends on certain considerations, the results of previous research, the parsimony of the model and consistent with theory.

III. APPLICATION

This paper examines the relationship between social media use on the one hand and political awareness and political participation on the other by using LCA to retrieve the behavior regarding political participation.

In this analysis, we consider that, in addition to the factors usually taken into account (age, educational background, socioeconomic status, level of interest in politics, economic developments), informational practices and knowledge of current political and social events affect citizens' political participation.

Numerous studies have been carried out on the relationship between the Internet and politics, with a particular focus on the following topics: online political discussions to fathom the dynamics of political orientation in the public sphere; citizen participation in public debates and decision-making processes; online electoral campaigns, as well as the transformations of political parties with the rise of social media in political campaigns.

With the advent of the Internet, those most interested in politics have been able to increase their consumption of political information.

Likewise, soft news broadcasts had allowed citizens to be informed in an incidental way about political news and to gain knowledge about major social issues.

In the next section, We analyze the effects of different Internet practices on the political commitment and behavior regarding vote action of a sample of the Moroccan population.

A. Study Sample

A survey was conducted online and offline on a group of 500 citizens, able to vote, to regroup them into categories based on measures of their political interest and their interaction on social media. The idea is to use a probabilistic approach to define the political commitment of citizens and their behavior regarding vote action.

The questions were constructed from to meet the objectives of the study and adapted for the diligence of the models to latent classes, they are grouped into six main items as shown in Fig.1.

The first focal point was related to participation in social media (Q.2) by taking into account the nature of social networks used and topics of interest (social, economic, political, cultural or religious).

The second and third axis concerns respectively the sharing behavior of respondents (Q.4 and Q.6) on social media and the possibility to change their opinion (Q.7) by testing the influence of the discussion and post consulted on their change of opinion.

The three last points were mainly dedicated to capture if a participant encourages others to share their opinions and debates their ideas regarding different topics (Q.8 and Q.9), recognize the area of interests to comprehend the information practices of participants (Q.11) and, ensure if the respondent is interested in following candidates, political figures, and political parties (Q.10).

The last five questions of the survey were mainly dedicated to collect information about the participant profile (Q.14, Q.15, Q.16, Q.17, Q.18).

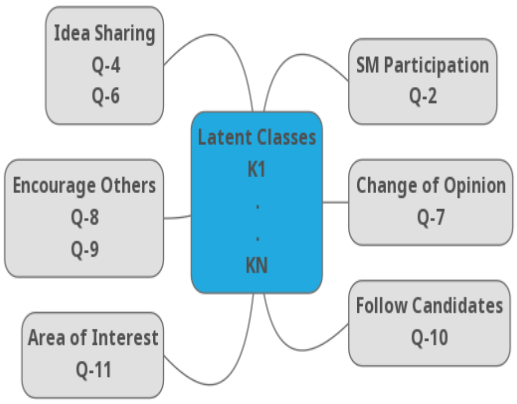


Fig. 1 Selection of manifest variables

Beforehand, we aim to apply an MC-DBSCAN classification to compare the profiles retrieved by both methods (LCA and MC-DBSCAN).

B. Classification with MC-DBSCAN method

The classification of the survey dataset, using the MC-DBSCAN method, resulted in two groups (Fig.2).

To facilitate the extraction of profiles and a readable projection of the classification results, the literature recommends representing the distribution of the variables on two dimensions. This approach offers on the one hand, the possibility of an easy reading of the results, which allows a rigorous analysis for a sharp extraction of hidden profiles.

For this purpose, multiple factorial analysis is the appropriate method to reduce the dimensions without compromising the initial information.

Once the new dimensions have been calculated, the variables are then projected separately onto the results of the classified observations. It's worth noting that the choice of the dimension in which the variables are represented depends on the quality of the representation of that variable within the dimension under consideration.

Fig. 2 Participant Distribution Using MC-DBSCAN

Fig. 3 displays the distribution of the variables Q.2, Q.4, Q.5, Q.6.1 to Q.6.6, the transcription of the corresponding questions generates a central question that provides information on the number of respondents who use social networks to disseminate their opinions on topics of interest to them. Therefore, concerning this question, Group 2 includes profiles that do not use mainly social networks to express their opinions. However, group 1 presents a profile opposite to group 2, i.e. it includes people who are active in discussions on many current topics that affect the political and social life of the country.

Fig. 4 shows the distribution of the variables Q.8, Q9.1, Q9.2, Q.9.3 and Q.10 on the two axes representing 'dim.1' and 'dim.2' respectively. Those variables measure the behavior of participants regarding their change of opinion, attitudes concerning encouraging others and interest of following candidates, political parties and figures.

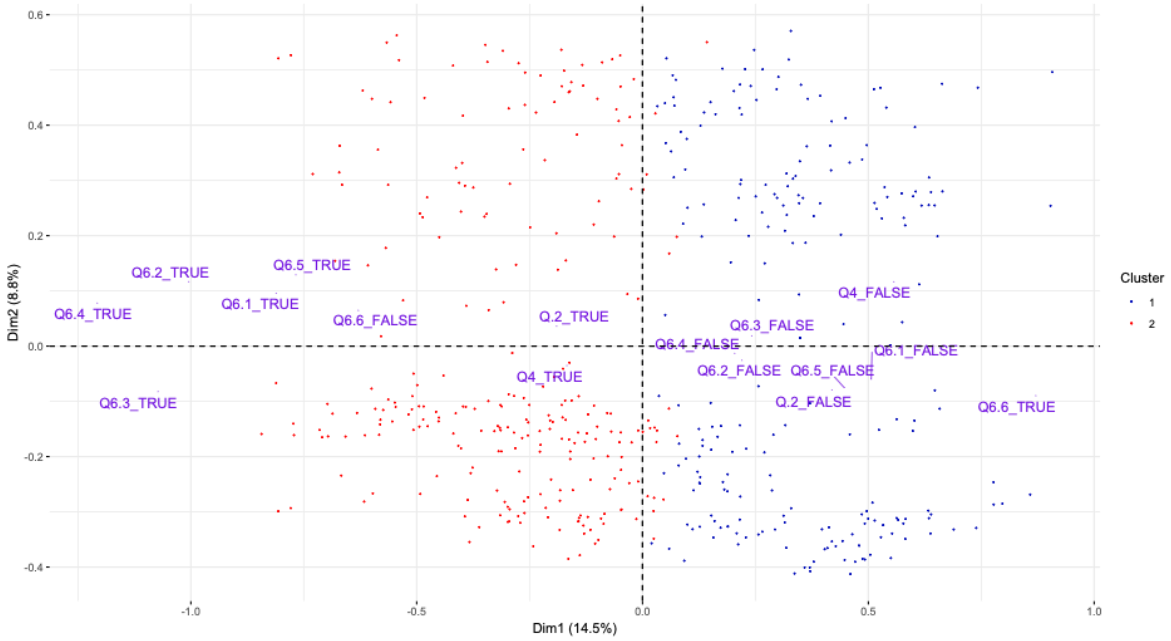


Fig. 3 Participant Distribution according to Q-2, Q-4, Q-5, and Q-6

The projection shows that group 2 includes the observations characterized by the value 'False' for the variables 'Q.9.1', 'Q.9.2', 'Q.10' and 'Q.8' and the value 'True' for the variable 'Q.9.3'.

However, the observations in group 1 are characterized by the values 'True' for the variables 'Q.9.1', 'Q.9.2', 'Q.10' and 'Q.8' and 'False' for the variable 'Q.9.3'.

In essence, this set of questions relates to the involvement of third parties in the citizen contribution and the communication channels they use for this purpose. Therefore, the transcription of the results gives the information that group 2 is composed of people involved in citizen mobilization through both channels, with a predominance of the oral route over social networks. However, the first group is dominated by people showing no involvement in encouraging citizen contribution.

Fig. 4 Participant Distribution according to Q-8, Q-9, and Q-10

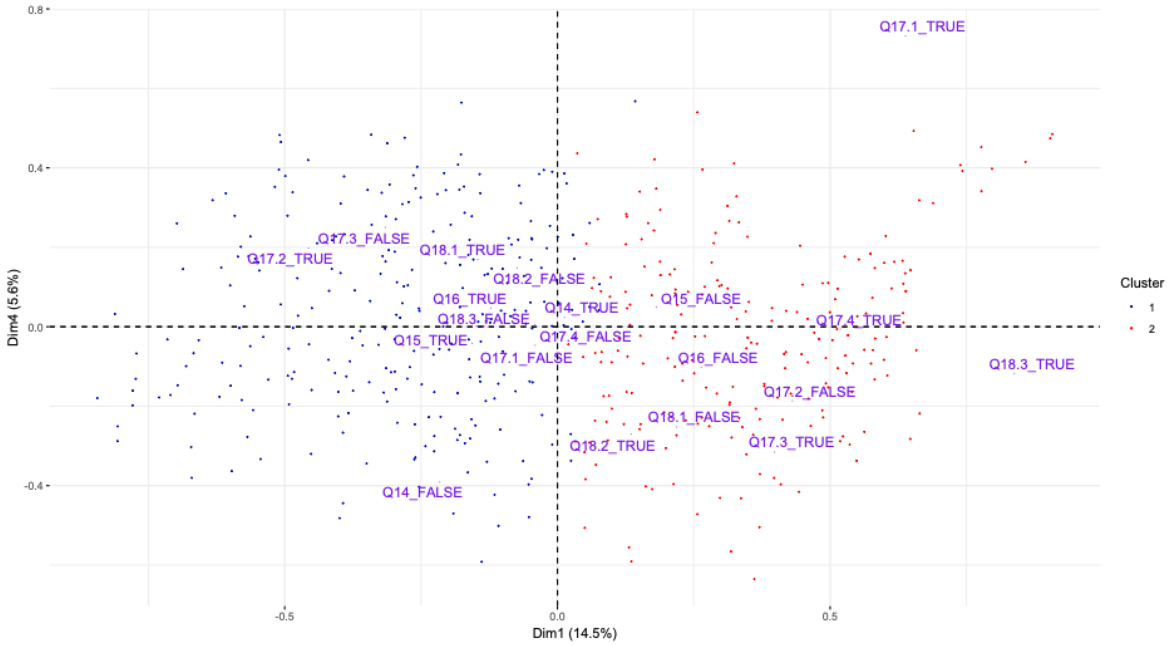


Fig. 5 Participant Distribution according to Q-14, Q-15, Q-16, Q-17, and Q-18

Fig. 5 shows the distribution of the variables Q.14, Q.15, Q.16, Q.17 (1-4) and Q.18 (1-3) relate to the basic profile of the respondents. In concrete terms, they aim to find out the age, gender, education level and employment status of the respondents.

Group 1 shows the profile of people, most of whom are female, with an average age of less than 30 years, who have at least a university education or are still students. In addition, they naturally identify with a profession. On the other hand, group 2 is composed mostly of men over 30 years of age whose level of education is divided between secondary and higher education. And some of them have not had any education. On the other hand, people in this group consider themselves to be either employed or unemployed.

C. Latent profiles: LCA Analysis

To determine the optimal number of classes, we used the Akaike Information Criterion (AIC) which has been recommended in situations such as this [34].

A sequence of models was fitted to the data with different numbers of classes and a three-class model was found to provide the best fit to our data. The AIC, BIC, and likelihood ratio statistics were markedly lower for the three-class model compared to other models.

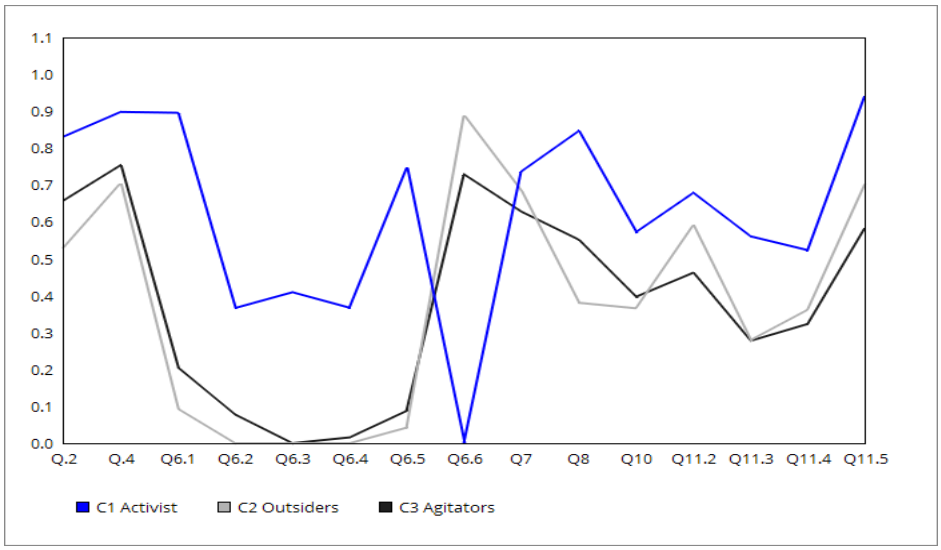


Fig. 6 Qualitative Differences Averages by Latent Class

A three-class model was more conceptually meaningful. People were assigned to each of the three latent classes based on their response profile and the highest calculated probability of membership.

To facilitate interpretation, Fig.6 shows the probabilities associated with the three-class model. As shown, the three latent classes characterized by different qualitative differences.

The first class (column 1) on Tab.1 has the outcome of class one membership, Activist who shows a high probability to participate in social media and shows high engagement on political and social actions. Results indicate that those members participate in social media (0.8311) and post mainly about cultural and social issues.

Tab. 1 Characteristics of Participants Assigned to Three Classes

			0.475	0.2635	0.2615
	Question main purpose	Modality	C1: Activist	C2 : Outsiders	C3 : Agitators
Q2	SM Participation	Yes	0.8311	0.5296	0.6578
Q4	Idea Sharing	Sharing on SM	0.8981	0.7052	0.7538
Q6.1		Cultural issues	0.8958	0.0935	0.2057
Q6.2		Economic issues	0.3665	0.0000	0.0777
Q6.3		Political issues	0.4093	0.0000	0.0000
Q6.4		Religious issues	0.3672	0.0000	0.0158
Q6.5		Social issues	0.7466	0.0425	0.0881
Q6.6		Not sharing	0.0000	0.8886	0.7285
Q7	Change of Opinion	Yes	0.7373	0.6836	0.6276
Q8	Encourage Others	Yes	0.8468	0.3807	0.5518
Q9.1		Via Face to face	0.3522	0.1628	0.3647
Q9.2		Via Social Media	0.6529	0.2635	0.2177
Q9.3		None	0.1198	0.5892	0.4331
Q10	Follow Candidates	Yes	0.5726	0.3657	0.3971
Q11.2	Area of Interest	Economic issues	0.6790	0.5912	0.4624
Q11.3		Political issues	0.5608	0.2797	0.2777
Q11.4		Religious issues	0.5232	0.3623	0.3235
Q11.5		Social issues	0.9427	0.7026	0.5843
Q14	Resident	Morocco	0.9326	0.9849	0.9859
Q15	Age	Woman	0.5319	0.5387	0.4485
Q16	Gender	Under 30	0.7527	0.9699	0.2200
Q17.1	Education	Secondary	0.0286	0.0000	0.1150
Q17.2		University	0.6120	0.3219	0.3849
Q17.3		Higher Education	0.3427	0.6781	0.4394
Q17.4		Without	0.0167	0.0151	0.0607
Q18.1	Profession	Student	0.6984	1.0000	0.0000
Q18.2		Employee	0.2786	0.0000	0.8597
Q18.3		Without	0.0230	0.0000	0.1403

Those class members consent to change their point of view (0.7373) after a discussion on social media and they undertook the quest to encourage others to debate and discuss their beliefs and position regarding social issues principally by means of social media. Besides, they are interested in following candidates and political parties in Social media and get informed on a different area (economic, political, religious and, social issues). Student females who are under thirteen with a university background were more likely to belong to this class.

The second class has the outcome of outsiders (column 2), which shows a negative behavior regarding political participation and had less interaction on social media. The results point out that the class members participate and post online, but not about issues related to the society. They admit changing their point of view after a discussion on social media, however, they do not encourage others to debate and discuss their point of view. The respondents don't follow political candidates' actions and political parties on social media (0.3657l), but they show interest in social and economic issues. A female student under thirteen with higher education backgrounds are more likely to belong to this class.

The third class has the outcome of agitators (column 3), that are more likely to be interested in issues related to society and encourage others to be active but, they do not post and share their own opinions.

The result shows that the members of this class participate and post on social media, but not on the named issues on the survey. They accept that they have changed their opinion after a discussion and they encourage others to debate their opinion. The class members don't follow political candidates' actions and political parties on social media, but they show interest in social issues. Male employees above thirteen with higher education backgrounds were more likely to belong to this class.

IV. DISCUSSION

In the light of the synthesis of the elements describing the three profiles resulting from the analysis carried out, a certain consistency emerges.

Profile 1 is young, under thirteen, with a good grasp of social reasons and their uses, and is at the heart of current social and political events. Now it is through debates on social networks that young people seem to draw the society of their aspiration. As a result, they want to be an active player in the life and the future of the country. Moreover, the fact that statistics at the national level show a strong presence of women in Moroccan universities, explains the predominance of women in this profile.

However, profile 2 is dominated by people on the fringes of social networks, which suggests that they would also be disconnected from the political news of the country. This predisposition would be either a lack of a minimum level of interest in or understanding of the political substance or an assumed posture.

The third profile, shows a little positive attitude regarding being interest in current issues, especially social and takes the time to encourage others to participate and debates their point of view on social media.

Our findings revealed several important and distinct patterns of political behavior in a sample of Moroccan citizens. The LCA identified three statistically distinct and behavioral classes regarding political interaction on social media. The profiles were ranged from participants actively engage in political action, such as following candidates and political parties' actions on social media, posting and participating on discussions related to economic, social or political issues or, encouraging others to debate their point of view, to "agitator" and "outsider" profiles that shows a low probability of interacting on social media and engaging on political actions.

The LCA analysis has revealed a distinct class (class 3) from MC-DBSCAN, that is moderately active on social media, but shows a discrete interest to the political context and behaves contradictory to what they act.

On the other hand, the Moroccan political terrain reveals a certain lack of interest on the part of citizens that stems from the influence of political culture on the nature of their political behavior. This can be explained by the difficulty of grasping the real impact of political action, understanding the role of political actors or losing the sense of community.

The Moroccan political terrain is characterized by a high rate of abstention due to voter disinterest and the use of the electoral act for individual gain.

In substance, concerning this study, even the profiles that are interested in political activities and engage actively in social media, they still reluctant regarding vote action.

Consequently, the consistency shown in this study indicates the relevance of the methodology adopted. The data sources used and the algorithms deployed made it possible to highlight the potential of Big Data in prospective analyses in the Moroccan socio-political context.

V. CONCLUSION

This paper focused on regrouping individuals into political profiles, based on their political interest and their interaction on social media, by using latent class analysis. The survey analyses a situation at a given point in time and does not allow an accurate evaluation of the causality between social media use and political participation. The LCA technique has provided meaningful and distinct information on the participants' political profile than clustering classical techniques (MC-DBSCAN).

The LCA results reveal three behavioral classes regarding political interaction on social media. The first profile retrieved, 'Activist', show more engagement in political activity, such as following candidates and political parties, posting and participating in discussions related to economic, social or political issues or, encouraging others to debate their point of view; the second and third profile, respectively 'Agitator' and 'Outsider', show a low probability of interacting on social media and engaging in political actions.

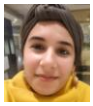
REFERENCES

1. M. Porcu, & F. Giambona, F. "Introduction to Latent Class Analysis With Applications". J. of Early Adolescence, Vol. 37(1), 2016, pp.129–158. doi: 10.1177/0272431616648452
2. C. Huang, "Dimensions of Taiwanese/Chinese Identity and National Identity in Taiwan". J. of Asian and African Studies, vol. 40(1-2), 2015, pp: 51–70. doi: 10.1177/0021909605052941
3. K. Bandeen-Roche, D. L. Miglioretti, S. L. Zeger & P. J. Rathouz. "Latent Variable Regression for Multiple Discrete Outcomes". J. of the American Statistical Association, Vol. 92(440), 1997, pp. 1375-1386 . doi: 10.2307/2965407.
4. F. Leisch, "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R". J. of Statistical Software, 2004, Vol. 11(8). doi: 10.18637/jss.v011.i08
5. M. Wedel, & W.S. Desarbo. "A mixture likelihood approach for generalized linear models". J. of Classification, Vol. 12(1), 1995, pp. 21–55. doi: 10.1007/bf01202266.

6. J.A. Hagenaars & A. McCutcheon. “Applied latent class analysis. Cambridge”, UK: Cambridge University Press, 2002.
7. M. Croon. “Latent class analysis with ordered latent classe”. British Journal of Mathematical and Statistical Psychology, Vol. 43(2), 1990, pp.171–192. doi: 10.1111/j.2044-8317.1990.tb00934.x
8. A. L. Mccutcheon. “Basic Concepts and Procedures in Single- and Multiple-Group Latent Class Analysis. Applied Latent Class Analysis”, Cambridge University Press, 2009, pp. 56–86. doi: 10.1017/cbo9780511499531.003.
9. W. H. Finch & K. C. Bronk. “Conducting Confirmatory Latent Class Analysis Using Mplus”. Structural Equation Modeling: A Multidisciplinary Journal, 2011, Vol. 18(1), pp. 132–151. doi: 10.1080/10705511.2011.532732.
10. H. Andruff, N. Carraro, A. Thompson, P. Gaudreau & B. Louvet (2009). “Latent Class Growth Modelling: A Tutorial”. Tutorials in Quantitative Methods for Psychology, Vol. 5(1), 2009, pp. 11–24. doi: 10.20982/tqmp.05.1.p011.
11. E. Lacourse, R. Baillargeon, V. Dupéré, F. Vitaro, E. Romano & R. Tremblay. “Two-year predictive validity of conduct disorder subtypes in early adolescence: a latent class analysis of a Canadian longitudinal sample”. J. of Child Psychology and Psychiatry, Vol. 51(12), 2010, pp. 1386–1394. doi: 10.1111/j.1469-7610.2010.02291.x.
12. J.D. Burke. “An affective dimension within oppositional defiant disorder symptoms among boys: personality and psychopathology outcomes into early adulthood”. J. of Child Psychology and Psychiatry, Vol. 53(11), 2012, pp. 1176–1183. doi: 10.1111/j.1469-7610.2012.02598.x
13. E.D. Klonsky & T. M. Olino. (2008). “Identifying clinically distinct subgroups of self-injurers among young adults: A latent class analysis”. J. of Consulting and Clinical Psychology, Vol. 76(1),2008, pp. 22–27. doi: 10.1037/0022-006x.76.1.22.
14. Z. Liu, Y. Zhang, L. Tian, B. Sun, Q. Chang &Y. Zhao (2017). “Application of latent class analysis in assessing the competency of physicians in China”. BMC Medical Education, Vol 17(1), 2017.doi: 10.1186/s12909-017-1039-4
15. A. Rossi, M. Marconi, S. Mannarini, L.Minelli,C. Rossini, R. Morena, … C.Verusio. “Profiling cancer-related needs and the role of physical, practical and psychological difficulties: A latent class analysis approach”. J. of Clinical Oncology, Vol. 36(15_suppl), 2018. doi: 10.1200/jco.2018.36.15_suppl.e22161.
16. N. Denson & M. Ing. “Latent Class Analysis in Higher Education: An Illustrative Example of Pluralistic orientation”. Research in Higher Education, Vol. 55(5), 2014, pp. 508–526. doi: 10.1007/s11162-013-9324-5.
17. M. Ing & K. Nylund-Gibson. “Linking early science and mathematics attitudes to long-term science, technology, engineering, and mathematics career attainment: latent class analysis with proximal and distal outcomes”. Educational Research and Evaluation, Vol. 19(6), 2013, pp. 510–524. doi: 10.1080/13803611.2013.806218.
18. R. Rhead, M. Elliot & P. Upham, “Using latent class analysis to produce a typology of environmental concern in the UK”, Social Science Research, Vol. 74, 2018, pp. 210-222, https://doi.org/10.1016/j.ssresearch.2018.06.001.
19. Fox, B., Moule, R. K., & Parry, M. M. (2018). Categorically complex: A latent class analysis of public perceptions of police militarization. Journal of Criminal Justice, 58, 33–46. doi: 10.1016/j.jcrimjus.2018.07.002
20. Y. Bentaleb, A. Abarda, H. Mharzi & S. El Hajji. “Application of latent class analysis to identify the youth population who risk being cybercrime victim on social networks”. Contemporary Engineering Sciences. Vol. 32, 2015, pp. 1529-1534.
21. J. Oser & M. Hooghe. “ Democratic ideals and levels of political participation: The role of political and social conceptualisations of democracy”. The British Journal of Politics and International Relations, Vol. 20(3), 2018, pp. 711–730. doi: 10.1177/1369148118768140
22. R. M. Alvarez, I. Levin & L. Núñez. “The Four Faces of Political Participation in Argentina: Using Latent Class Analysis to Study Political Behavior”. The Journal of Politics, Vol. 79(4), 2017, pp. 1386–1402. doi: 10.1086/692786.
23. A.P. Dempster, N. M. Laird & D. B. Rubin. “Maximum Likelihood from Incomplete Data Via the EM Algorithm”. J. of the Royal Statistical Society: Series B (Methodological), Vol. 39(1), 1977, pp. 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x.
24. A. Agresti. “Categorical data analysis”. New York: Wiley, 1990.
25. K.J. Koehler. “Goodness-of-fit tests for loglinear models in sparse contingency tables”. J. of the American Statistical Association,Vol. 81, 1986, pp. 483–493.
26. G. McLachlan & D. Peel. “Finite mixture models”. New York: Wiley, 2000.
27. N.A.C . Cressie & T.R.C. Read. “Multinomial goodness-of-fit tests”. J. of the Royal Statistical Society, B, Vol. 46, 1984, pp. 440–464.
28. D. Nagin, D. “Group-based modeling of development”. Cambridge: Harvard University Press, 2005.
29. V. Ramaswamy, W.S Desarbo, D. J. Reibstein & W.T. Robinson. “An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data”. Marketing Science, Vol. 12(1), 1993, pp. 103–124. doi: 10.1287/mksc.12.1.103
30. H. Akaike. “Information theory and an extension of the maximum likelihood principle”. In Selected Papers of Hirotugu Akaike. Springer, 1998, pp. 199-213.
31. H. Bozdogan . “ Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions”. Psychometrika, Vol 52, 1987, pp.345-370.
32. G. Schwarz (1978). “ Estimating the dimension of a model. Ann. Statist Vol. 6, 1987, pp. 461-464.
33. S.L Sclove (1987). “Application of model-selection criteria to some problems in multivariate analysis”. Psychometrika , Vol. 52, 1987, pp. 333-343.
34. A. Abarda A. et al. “Solving the problem of latent class selection”. In Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications, 2018, pp. 1-6.

AUTHORS PROFILE

Zakia Acharoui currently a Ph.D. at the Faculty of Sciences, Kenitra, Ibn Tofail University, Morocco. Her main research interests cover the study of data science, Machine Learning, Data Mining.



Altaf ALAOUI is a Ph.D. at the Faculty of Sciences, Kenitra, Ibn Tofail University, Morocco. Her research interests cover the study of data scientist, Machine Learning methods , clustering, estimation and prediction of the time series data.

Morad Azhari is a Ph.D. at the Faculty of Sciences, Kenitra, Ibn Tofail University, Morocco. He obtained an engineering degree in operational research at the National Institute of Statistics and Applied Economics of Rabat in 2013, Morocco. His research interests cover the study of Machine Learning methods in the context of Big Data using the datasets in the physic area.

Abdallah Abarda has an experience of over ten years in the field of data science, applied statistics, survey methodology, and artificial intelligence. Data scientist professor with advanced statistical skills and experience programming with Python and R software.

Badia Ettaki is a professor at the Engineering School of Information Sciences in Rabat, Morocco. She obtained a doctoral degree in the Statistic Physic. His research interests cover the study of Content and Knowledge Engineering, Machine Learning algorithms and Big Data analytics.

Jamal Zerouaoui is a professor at the Faculty of Sciences, Kenitra, Ibn Tofail University, Morocco. He received a doctoral degree in the physic science. Currently, his research interests cover modeling using Machine Learning methods in the frame of the physic area.