Contemporary Engineering Sciences, Vol. 8, 2015, no. 32, 1529 - 1534
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2015.58261

Application of Latent Class Analysis to Identify the Youth Population Who Risk Being Cybercrime Victim on Social Networks

Youssef Bentaleb

EECOMAS-LAb, National School of Applied Sciences
Ibn Tofail University, Kenitra, Morocco

Abdallah Abarda

EECOMAS-LAb, National School of Applied Sciences
Ibn Tofail University, Kenitra, Morocco

Hassan Mharzi

EECOMAS-LAb, National School of Applied Sciences
Ibn Tofail University, Kenitra, Morocco

Said El Hajji

LabMIA, Faculty of sciences
Mohamed V University, Rabat, Morocco

Abstract

Cybercrime is one of the latent phenomena where it is difficult to estimate the risk of being a cybercrime victim. In this paper, we propose the application of latent class analysis to solve the problem of identification of youth population who risk being cybercrime victim on social networks. For this reason, we devised a questionnaire which was filled out by a sample of nearly 165 young internet users from six different

regions in Morocco. The data resulted is used to form and test a model that can estimate the proportion of youth population who risk being a cybercrime victim on social networks whether consciously or unconsciously.

Keywords: LCA, data analysis, risk, cybercrime

1 Introduction

Latent Class Analysis (LCA) [1] is a statistical solution for Identifying latent class membership using observed variables. The adaptation of this method to the problem of cybercrime was already mentioned in a scientific article [2].
The purpose of the present paper is to focus on the analysis of cybercrime in social networks. And apply LCA to identify the youth population who risk being cybercrime victim on social networks whether consciously or unconsciously. For this reason, we use the package (poLCA) [3] to build two models: The first concerns the determination of youth population who risk being a cybercrime victim, and the second concerns the determination of young people who are conscious of the danger of cybercrime.

2 Model to estimate the young people who risk being cybercrime victims on social networks

The data used in the study are derived from a survey and a questionnaire which was filled out by a sample of nearly 165 young internet users from six different regions in Morocco. Eleven dichotomous variables were selected by analyzing the scatter plot obtained by multiple correspondence analysis. The choice of variables Was aussi done on the basis of social networks usage behaviors.

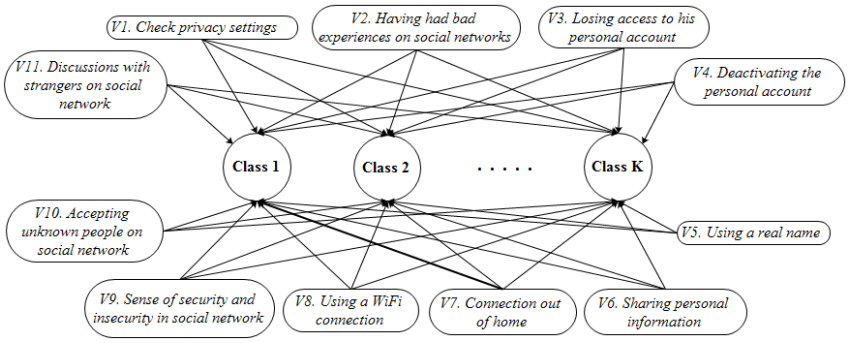

Figure 1: Selected manifest variables

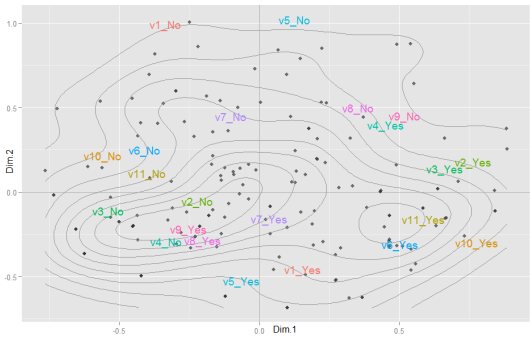

Figure 2: MCA for selected variables

We tested models with different number of classes. The best model specified three latent classes. Indeed, the information criteria AIC3[4], AICc[5], HT[5], CAIu[6], HQ[7] are minimum for this number of classes. The young population can be divided into three different groups or classes.

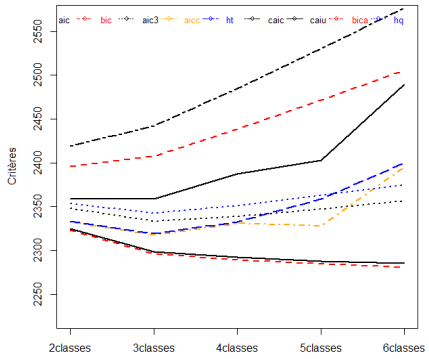

Figure 3: Criteria for selection and validation of class numbers

The conditional probabilities presented in the graph below are used to give an interpretation to the different classes. All manifest variables have two modalities (Yes or No).

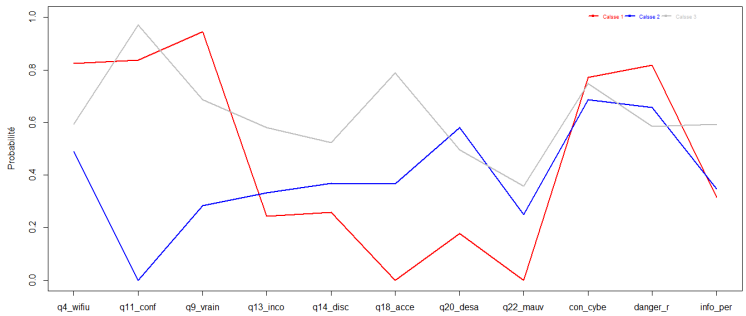

Figure 4: The conditional probabilities to answer "yes" to the behaviors

The classes detailed as follows

- Class 1 (30%): Attentive and not exposed to risk
- Class 2 (25%): At high risk of falling victim to cybercrime
- Class 3 (45%): Attentive and low exposure to risks

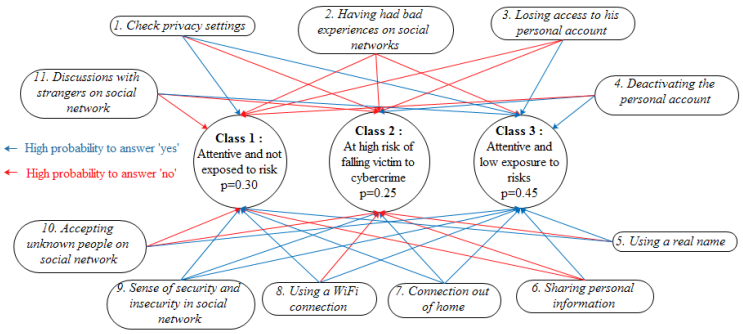

Figure 5: Interpretation of classes

2.1 Model to estimate the class of conscious youths of the danger of cybercrime in social networks

The aim of building this model is to estimate the youth aware of the dangers of cybercrime, to discover their characteristics, and to demonstrate the relationship between "to be a victim", and "to be conscious". We followed the same previous process and found that this model is adjusted to five latent classes presented as follows

- Class 1 (6.7%): Security measures Ignorant
- Class 2 (32%): Not conscious of the danger of cybercrime
- Class 3 (8.5%): Less conscious of the danger of cybercrime
- Class 4 (37.4%): Conscious of the danger of cybercrime
- Class 5 (15.5%): Unable to evaluate the danger of cybercrime

3 Results and Discussion

LCA has identified three latent classes. One among them represents a class of young people at risk of becoming a victim of cybercrime "risky". This class represents 25%. The other classes are regrouped as "Not risky ".
We have concluded that

- One in four young risk to be a victim of cybercrime on social networks.
- About one in three young is conscious of the danger of cybercrime.
- The classes built by LCA showed that the highest risk are those young people who ignore security measures and those unable to assess the risk. (54.5% et 46.2% respectively).
- The risk decreases to 13% for youth conscious of the danger of cybercrime.

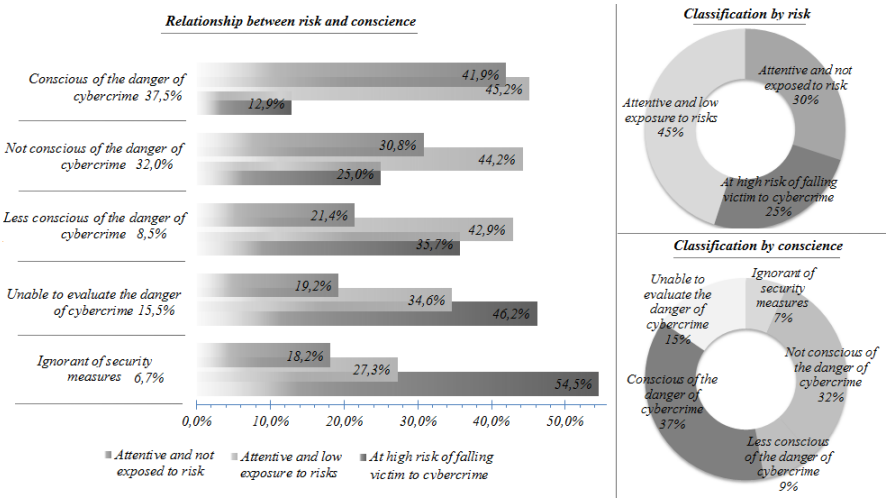

Figure 6: Relationship between risk and conscience

4 Conclusion

In this paper, we applied LCA to estimate the proportion of youth who risk being cybercrime victims on social networks and the proportion of youth conscious of the danger of this phenomena. Using LCA, we tested several models, the best model identified three latent classes, one of these classes represents the class of young Internet users who risk being cybercrime victims on social

networks. Similarly, we built a model to estimate the youth class conscious of the danger of cybercrime.
Although this approach is used to classify individuals according to the risk of cybercrime in social networks, we can generalize this approach for the risk assessment from a set of behaviors related to the use of the Internet . The difficulty encountered in the implementation of LCA is how to select the manifest variables[8].

Acknowledgements. The authors would like to thank the CMRPI (Moroccan Centre polytechnic research and innovation) for its support.

References

[1] P. Lazarsfeld, N. Henry, *Latent Structure Analysis*, Houghton-Mifflin, New York, 1968.

[2] Y. Bentaleb, A. Abarda, H. Mharzi, Said El Hajji, Probabilistic approach to estimate the risk of being a cybercrime victim, *Applied Mathematical Sciences*, **9** (2015), 6233-6240.
http://dx.doi.org/10.12988/ams.2015.58559

[3] Drew A. Linzer, Jeffrey B. Lewis, poLCA: An R Package for Polytomous Variable Latent Class Analysis, *Journal of Statistical Software*, **42** (2011), 1-29. http://dx.doi.org/10.18637/jss.v042.i10

[4] H. Bozdogan, Model Selection and Akaike Information Criteria (AIC): The general theory and its analytic extensions, *Psychometrika*, **52** (1987), 345-370. http://dx.doi.org/10.1007/bf02294361

[5] C.M. Hurvich, C. Tsai, Regression and time series model selection in small samples, *Biometrika,* **76** (1989), 297-307.
http://dx.doi.org/10.1093/biomet/76.2.297

[6] A. McQuarrie, R. Shumway, C. L. Tsai, The model selection criterion AICu, *Statistics and Probability Letters,* **34** (1997), 285-292.
http://dx.doi.org/10.1016/s0167-7152(96)00192-7

[7] E.J. Hannan, B.G. Quinn, The determination of the order of an autoregression, *Journal of the Royal Statistical Society,* **41** (1979), 190-195.

[8] Nema Dean, Adrian E. Raftery, Latent class analysis variable selection, *Annals of the Institute of Statistical Mathematics,* **62** (2010), 11-35.
http://dx.doi.org/10.1007/s10463-009-0258-9

Received: September 15, 2015; Published: November 28, 2015