MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-1074

A Data Mining Approach to Identify Sexuality Patterns in a Brazilian University Population

Priscyla Waleska Simões[a,b], Samuel Cesconetto[a], Larissa Letieli Toniazzo de Abreu[a], Merisandra Côrtes de Mattos Garcia[a], José Márcio Cassettari Junior[a], Eros Comunello[c], Luciane Bisognin Ceretta[b], Sandra Aparecida Manenti[a]

[a] Research Group in Information and Communications Technology in Health, UNESC, Criciúma, SC, Brazil
[b] Postgraduate Program in Public Health, UNESC, Criciúma, SC, Brazil
[c] Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

Abstract

This paper presents the profile and experience of sexuality generated from a data mining classification task. We used a database about sexuality and gender violence performed on a university population in southern Brazil. The data mining task identified two relationships between the variables, which enabled the distinction of subgroups that better detail the profile and experience of sexuality. The identification of the relationships between the variables define behavioral models and factors of risk that will help define the algorithms being implemented in the data mining classification task.

Keywords:

Medical Informatics; Artificial Intelligence; Data Mining; Sexuality.

Introduction

Data mining consists of applying algorithms to identify and analyze information in order to produce patterns or models of the data [1]. The main topic of this study is sexuality, which is considered inherent to life and health, and is expressed early in humans as one of the indices that measure the level of quality of life [2]. This paper presents the profile and experience of sexuality in a university population in southern Brazil generated from a data mining classification task.

Methods

The database used for this study addressed the sexuality profile and vulnerability to sexually transmitted diseases (STDs) and acquired immune deficiency syndrome (AIDS) in a university population in southern Brazil. First, the data was pre-processed, and then the application for fuzzy clustering under development at UNESC [**3**] was used. The classification of data was carried out using the J48 algorithm in the Waikato Environment for Knowledge Analysis software (WEKA). In addition to the complete base (classification 1), an analysis of the database was examined using only males (classification 2), as well as examining a data set containing both genders but eliminating the variables of the all-male responses (classification 3). Only the rules belonging in trees with at least 70% of the records correctly classified and associations in 5% of the sample were used. We used accuracy for the evaluation of the knowledge generated.

Results

We conducted 204 experiments that resulted in 8728 rules. Classification 1 returned 3727 rules(0.76 accuracy [IC 95%=(0.73;0.80)]); classification 2 returned 1699 rules(0.75 accuracy [IC 95%=(0.71;0.79)]); and classification 3 returned 3302 rules (0.76 accuracy [IC 95%=(0.72;0.80)]). The profile obtained from the analysis was similar to that found in the literature. By observing the age of sexual activity onset, the J48 algorithm not only isolated information (15-19 years of age), but also found that this group was mostly Catholic, and the partner at first intercourse was also 15- 19 years of age.

Conclusion

Our study improved on the profile and experience of sexuality of a university population in southern Brazil by including a distinction between subgroups that allowed for more detail. This distinction may assist in the definition of the actions to be implemented for the profile and experience of sexuality of this population.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) and Universidade do Extremo Sul Catarinense.

References

[1] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008;77(2):81-97.

[2] Heilborn ML. Entre as tramas da sexualidade brasileira. Rev Est Fem 2006: 14(1): 43-59.

[3] Azevedo Simões PW, Martins PJ, Casagrandre RA, Madeira K, de Mattos MC, Manenti SA et al. Using a model of parallel distributed processing associated with data mining in the characterization of sexuality in a university population. Stud Health Technol Inform. 2013;192:1135.

Address for correspondence

Prof PhD Priscyla Waleska Simões - pri@unesc.net