Finding Insights between Active Aging Variables: Towards a Data Mining Approach

María-Inés Acosta-Urigüen[1][a], Priscila Cedillo[1,2][b], Marcos Orellana[1][c], Alexandra Bueno[1][d], Juan-Fernando Lima[1][e] and Daniela Prado[2][f]

[1]*Laboratorio de Investigación y Desarrollo en Informática – LIDI, Universidad del Azuay, Cuenca, Ecuador*

[2]*Universidad de Cuenca, Cuenca, Ecuador*

Keywords: Active Aging, Data Mining, Cognitive Evaluation.

Abstract: Several proposals on active aging have been addressed within the psychological field, conceptualizing it satisfactorily as a perspective of aging. Those proposals generate indicators that assess the level of physical health, psychological wellbeing, adequate social adaptation. Physical, cognitive, and functional faculties, interpersonal relationships, and productive activities have been evaluated. Although several technological approaches have been proposed to promote active aging, they have not included a deep understanding of the results obtained from solution implementations. Then, this paper presents the first step towards an approach that uses variables proposed by active aging models (e.g., health, cognition, activity, affection, fitness aspects) to generate knowledge through patterns. These patterns are identified using data obtained through several instruments (i.e., psychological evaluations, health studies, and human experts' contributions). Thus, selecting those variables and evaluating them as future models is necessary. Domain experts perform this evaluation. The evaluation of this proposal has been completed with participants belonging to the health area through a case study. This evaluation generates input data for engineers to apply data mining techniques to reveal strategic knowledge. Finally, from the psychologist's point of view, the results showed that the contribution results are appropriate for achieving healthy aging indicators.

1 INTRODUCTION

When the old age concept was analyzed, it was only related to illness, memory problems, senility, dementia, poverty, and depression (Lupien & Wan, 2004). However, the concept of old age can be addressed, considering the quality of life as a relevant factor contributing to it. The World Health Organization (WHO), in 2015, presented a World Report on Aging and Health (World Health Organization, 2015). This report covers aging as the sum of several changes. Biologically, aging is associated with accumulating a wide variety of molecular and cellular damages that gradually reduce physiological reserves, increasing the risk of many diseases and generally decreasing the individual's capacity (World Health Organization, 2015).

The WHO considers three trajectories of healthy aging: a) a period of relatively high and stable capacity, b) a period of diminished capacity, and c) a significant loss of abilities. It also identifies that there are different ways to quantify active aging. Still, they all keep the exact purpose of promoting and maintaining intrinsic capacity. People with reduced functional capacity can continue to carry out activities that are important to them (World Health Organization, 2015). The trajectory does not depend on chronological age and is not uniform among individuals.

[a] https://orcid.org/0000-0003-4865-2983
[b] https://orcid.org/0000-0002-6787-0655
[c] https://orcid.org/0000-0002-3671-9362
[d] https://orcid.org/0000-0001-7188-1210
[e] https://orcid.org/0000-0003-3500-3968
[f] https://orcid.org/0000-0003-1241-1782

Acosta-Urigüen, M., Cedillo, P., Orellana, M., Bueno, A., Lima, J. and Prado, D.
Finding Insights between Active Aging Variables: Towards a Data Mining Approach.
DOI: 10.5220/0011068100003188
In *Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2022)*, pages 268-275
ISBN: 978-989-758-566-1; ISSN: 2184-4984

Worldwide, the concept of active aging has gained interest. For example, the European Union has focused on transitioning from the perception that older adults are only recipients of retirement to a practical orientation. They are active subjects at the family, community, work, educational level, among others (Walker & Maltby, 2012). Other clear examples are observed in China, Japan, and South Korea, where governments have proposed an integrated, composite index for measuring the contribution of older people to society, their communities, and their families (Um et al., 2019).

Here, the generation of datasets and the application of various techniques becomes an attractive option to evaluate and interpret the concept of active aging among a wide range of variables and their combinations (Nayak, Buys, & Lovie-Kitchins, 2006). Also, the prediction, detection of outliers (anomaly detection), clustering, and decision making are some of the most used techniques belonging to the data mining field (Gachet Páez et al., 2018; Moreira & Namen, 2018).

Then, this paper proposes the first step towards an approach that uses variables proposed by active aging models (e.g., health, cognition, activity, affection, fitness aspects) generates knowledge using patterns. These patterns are identified using data obtained through several instruments (i.e., psychological evaluations, health studies, and human experts' contributions). Thus, selecting those variables and evaluating them as future models is necessary. Domain experts perform this evaluation. The review of this proposal has been completed with participants belonging to the health area through a case study. This evaluation generates input data for engineers to apply data mining techniques to reveal strategic knowledge. Finally, from the psychologist's point of view, the results showed that the contribution results are appropriate for achieving healthy aging indicators.

Finally, this study is organized as follows: first, the background and related work on applying data mining in psychology are described; next, the proposed approach which helps to characterize the active aging; finally, conclusions and future work are offered.

2 BACKGROUND

Even though few studies related to data mining techniques applied in active aging, some have shown promissory results. For example, Nayak, Buys, & Lovie-Kitchins (2006) proposed the use of predictive models to identify which variables need to be related, from the groups labeled as work, learning, social, spiritual, emotional, health, home, life events, and demographics, contribute to achieving a positive active aging score. Preprocessing techniques include the value transformation removal of empty fields. Then, the k-means clustering algorithm was used to split the dataset into seven distinct, overlapped clusters. They performed an association analysis and rules to determine the meaning of the correlation of variables.

The recognition of early variables of successful aging can predict long-term survival. The study presented by Swindell et al. (2010) was based on a dataset taken from 4,097 women in the United States of America, with several variables (i.e. demographic, cognitive, familiar, medical variables). Their study proposed a predictive model based on data mining techniques to look for combinations of variables that predict long-term survival. The result was a composition of a 13-variable model.

A system that implements a big data approach was presented to use bio-signal sensors and machine-learning algorithms for recommendations (Gachet Páez et al., 2018). It obtains data from wearable sensors, and prediction, detecting outliers, clustering, and decision making were applied for prediction.

Consequently, although there are several studies, none considers applying data mining techniques to propose models to improve that characterization. Neither do they consider an expert validation performed by psychologists? Therefore, this study aims to provide a method that allows user data to describe insights that help health personnel give the best advice to their patients to reach active aging.

3 PROMOTING THE ACTIVE AGING BASED ON DATA SCIENCE

The Cross-Industry Standard for Data Mining (CRISP-DM) proposes five high-level processes, that contribute to carrying out a data science project. Each step is described briefly as follows: 1) *Business understanding*: it is focused on uncovering essential factors, but in this case, it will be named background understanding, 2) *Data understanding*, which describes the acquisition of data listed in the project resources, 3) *Data preparation*, which seeks to obtain a specific dataset to be used as input during the entire process, 4) *Modeling*, which is focused on selecting the proper technique to be used, and finally, 5)

Evaluation, which allows to accept or decline de the generality of the model (Chapman et al., 2000). Following these steps makes it feasible to reach good results related to promoting active aging in patients.

All the conclusions and found knowledge are reliable since it is based on data that have been collected and can be equivalent to a domain expert opinion; it is because data are a good source of experience.

3.1 Background Understanding

Several theoretical perspectives have emerged that differentiate between the "good" ways of aging in recent years. At the end of the 60s of the 20th centuries, various sociological perspectives appeared that defined these ways of aging. Models of aging based on exact psychological characteristics have now been developed. Although these models have a tremendous everyday basis, each of them represents a different perspective or way of aging successfully.

Figure 1 shows how using models can identify the variables useful for domain experts (i.e., health personnel), which provide insights to improve treatments and promote the well aging of their patients. Those variables can be categorized to obtain knowledge easier to manage depending on relevant topics of preference for each health personnel.

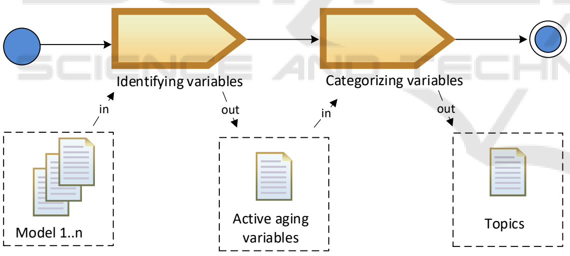

Figure 1: Data comprehension pool.

3.1.1 Successful Aging

Successful aging refers to "a high level of physical health, psychological wellbeing and adequate adaptation" (Nadler et al., 1997). Successful aging allows the person to reach an advanced age in full use of their physical, cognitive, and functional faculties. The Rowe and Kahn model was born and proposed to include three main components for the definition of successful aging: a) low probability of disease or disability; b) high functional capacity at the cognitive and physical level; and c) active commitment to life (Stowe & Cooney, 2015).

3.1.2 Optimal Aging

One of the first proposals, presented by (Baltes & Baltes 2010), indicates a development model throughout the entire life cycle with the appropriate interaction between three main elements: selection, optimization, and compensation, adjusting their goals and objectives to their vulnerability and the reduced reserve capacity, typical of their age.

Brummel-Smith (2007b) points out that this form of optimal aging should be considered a biopsychosocial model. Happiness, an active social life, or financial solvency contribute to healthy aging in the psychosocial area. Finally, social support facilitates coping and adaptation to changes and protects individuals from stress-induced pathologies.

3.2 Data Comprehension

To accomplish this phase, the understanding of the below phase is essential. The fields from the proposed models are extracted and synthesized into one collection of topics (see Figure 1).

To construct the topics and related variables, the proposal of Lak et al. (2020) was considered. It presents an iterative systematic review in the field of active aging. The authors identified several studies and summarized the most relevant variables into different categories; these variables were included in other models of active aging explained in section 2. The considered topics are personal information, personal behaviour, land use, social process, policymaking, mental and physical health, and social health.

Although there is no single model to measure effective aging, the literature indicates that the tools used by the authors are constituted in its measurement based on the Likert scale (Echauri et al., 2013). The Likert scale is an additive scale with an ordinal level compound of a series of items to measure the subject's reaction (Namakforoosh, 2002). Data mining can play an important role by having quantitative data, understanding the data, and finding clues by mining information from surveys and tests (Koufakou et al., 2016).

3.3 Data Preparation

The best way to avoid untrusted results data considered in the study must be cleaned before using data in the *Modeling stage*. The main activities of data preparation are shown in Figure 2, and each activity is part of the flow until to get a proper input dataset.

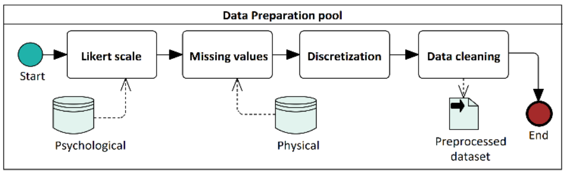

Figure 2: Preprocessing the input dataset.

3.3.1 Likert Scale

This paper expects multiple fields based on the Likert scale due to the input data. This point reflects the available ways to deal with this kind of data. Likert-Scale is commonly used as comfortable feedback, and it is categorized as an ordinal data type. Also, machine learning models evaluate and try to generate the best goodness in their results (Kandasamy et al., 2020); conversely, it is possible to deal with these data only as ordinal data types and convert them to their relative ordinal numbers.

3.3.2 Missing Values

No matter the type of data, it could be numerical or not, discrete or continuous. If there are missing fields, the Neuropsychologists experts prefer to backup up and drop the records and all the linked records to avoid the noise in the results (Nayak, Buys, & Lovie-Kitchins. Literature such as Swindell et al. (2010) shows that its variables are continuous and categorical, proposing to fill in missing data with the "Imputation approach" method. The missing data are imputed based on the average value between the k = 20. In this stage, the choice about missing data depends on the amount of data collected and the opinion of experts in the domain (Enders, 2010).

3.3.3 Data Discretization

The variables are transformed into discrete data to standardize continuous data inputs, establishing ranges that cover data to be changed. As mentioned in Nayak et al. (2006), one of the standard data to be transformed is peoples' ages, who are part of the group to be analyzed. According to the Research of Adult Learning and Development, to manage these age groups, a path is to transform ages to standardize in groups as Childhood, Adolescence, Emerging adulthood, Average adulthood, and Late adulthood, according to the Research of Adult Learning and Development (Smith & Nancy DeFrates-Densch, 2008). In the same way, Ethnicity can be categorized according to Nerenz et al. (2009).

3.3.4 Data Cleaning

In data containing words or phrases, it is necessary to identify words with values that will not contribute to the data mining process. Instead, it will make the process difficult. The typical stage includes frequently repeated words in the writings, such as articles or propositions (Moreira & Namen, 2018).

3.4 Modelling

Due to the diversity of data types, choosing the appropriate technique to guarantee that data is processed and showing the proper variables that characterize active aging is the main interest of this stage. Several approaches determine the influence of variables as "Select by weights" that selects only those whose weights satisfy a criterion concerning the input weights (Malik & Mishra, 2014). However, we seek to divide a group of interest into subgroups (clusters). Therefore, an unsupervised method is a proper approach.

The application of unsupervised methods has been used in similar works as (Pal & Pal, 2013) where evidence the predictive value of different measures of cognition, based on clusters was found that girls with the high socioeconomic status trend to higher academic achievement in science stream, and boys with low socioeconomic status had trends higher academic achievement in general. Clustering on social domains has been widely used as an unsupervised technique for human activity recognition (Ariza Colpas et al., 2020).

Clustering is essential for the Knowledge Discovery tool because the goodness of data is proportional to the fulfilment of its purpose. The technique creates groups mutually exclusive based on three possible conditions: a) Defining the maximum cluster distance and minimizing it, b) Compute the sum of averages of the distances and minimizing it, and c) Compute the total cluster distance and minimize it (Pandove et al., 2018).

Once the problem is apparent, and it is clear how clustering can solve it, it is necessary to select the proper technique among the numerous approaches such as Hierarchical, K-means, Random sampling, Randomize search, Condensation based, Density-Based, Grid-based, Probabilistic Model-based, and Clustering Graphs and Network Data (Pandove et al., 2018). The requirement of this work defines two segments: proper or no proper values for active aging and based on this. K-means application is proposed to include the number of groups before executing the technique.

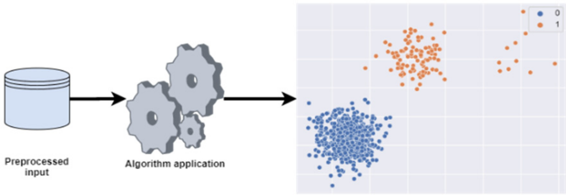

Figure 3: Expected results after clustering.

According to (Kandasamy et al., 2020), the behavior of the responses could be better understood since it ensures that the most significant disadvantage of the Likert scale is losing information and introducing distortion to data. Therefore, applying clusters to segment groups could be advantageous to find these variables in the same group of people. The cluster method that sticks to the data set of our characteristic for effective aging is a proprietary method described based on clustering to find the relationship of the variables.

3.5 Evaluation

Due to the application of an unsupervised technique done in the above section, the evaluation of the results must be verified to verify what kind of variables define active aging. Being clustering the method used to analyze data, the proper way to show how grouped data is related is through multidimensional graphs. Generally, 2D or 3D dimensional (Kandogan, 2001) display the groups. Therefore, an analysis of the clusters is required to evaluate the goodness of the clustering results.

Visual Analytics: The novel strategy of incorporating human capabilities to describe the behavior. The popular visualization methods used are: Bar charts, Line charts, Pie charts, and Scatter plots. Scatter plot is the standard method to describe the clustering application, including 2D or 3D axes on the cartesian, the human can evaluate according to the dispersion of data. However, this is hard to assess and validate if clusters are very closed (Chen et al., 2015).

4 CASE STUDY

This section presents the case study, with all the activities proposed by Runeson et al. (2012). The results of this evaluation represent an essential input for data engineers; who analyze factors associated with active aging in the proposed methodology.

Experts in active aging evaluate the inputs used in the proposed methodology. It follows the method proposed by Runeson et al. (2012). The activities to be tracked are 1) design, 2) preparation for data collection, 3) collecting evidence, 4) analysis of collected data and reporting, and 5) threats of validity analysis.

4.1 Design

A degree of agreement among the experts is obtained from this evaluation regarding the variables considered for healthy aging. Besides, this case study aims to evaluate health personnel's perceptions regarding the usefulness of the results derived from artificial intelligence in the characterization of healthy aging.

The evaluation's objectives and scope were established with Goal Question Metric (GQM) approach proposed by Basili et al. (Basili et al., 1994), who offers a paradigm that defines the evaluations' scope and objectives. The proposed GQM scheme follows the scheme of a) The evaluation analyses the inputs for the methodology proposed by the data engineers; b) What is the purpose of the objective measures the agreement among health experts on the variables considered in the characterization of active aging regarding the usefulness of the information resulting from the methodology; c) From the point of view of Clinical psychologists; and d) In the health context where this study is carried out.

In this context, the research questions are: What variables are considered important for characterizing active aging for two health personnel, and what the perception of health personnel on the usefulness of a methodology that allows selecting active aging variables collected utilizing artificial intelligence is.

According to Runeson et al. (Runeson et al., 2012) recommendations, this case study method is holistic-multiple, and the units of analysis are presented in Figure 4.

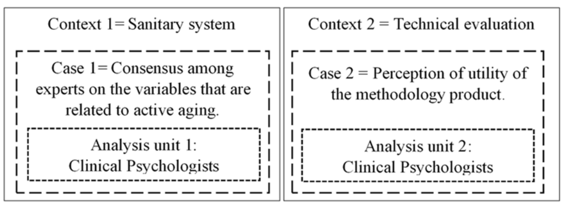

Figure 4: Holistic-multiple method.

4.2 Preparation for Data Collection

Two surveys have been designed to achieve the objectives of this case study. For Context 1, a form (see https://n9.cl/iwle) has been made based on Lak, Rashidghalam, Myint, and Bradaran (Lak et al., 2020), who propose a list of active aging characteristics based on the study of related work. The purpose of this first form is to reach a consensus of experts in the gerontological health area of the variables that should be considered for the characterization of healthy aging through data mining.

For Context 2, a form has been designed based on the technology assessment model (TAM) proposed by Davis (Davis, 1985). On this occasion, only the constructs of the Perceived Utility (PU) and the Intent to Use (UIT) in the future are analysed, specifically of the final products of the methodology. Data mining experts have designed the form with its respective explanation. As shown in the following URL: https://n9.cl/qcmb4, this questionnaire uses a 5-point Likert scale.

4.3 Collecting Evidence

Both questionnaires were presented to two Clinical psychologists with experience in the gerontological area.

4.4 Data Analysis and Results Reporting

By analyzing the results, it is found that they allow answering the case study questions. In Case 1, Fleiss' Kappa is used. It is a statistical measure for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings to several items or classifying items; the action is scored between 0 and 1 (0 means low agreement, and one refers to a high deal). Fleiss' Kappa is used to validate the process of inclusion/exclusion of variables presented in Appendix 1 (see https://n9.cl/iwle). Finally, the selection of each reviewer was checked, and the discrepancies were resolved with consensus.

For the two raters, Fleiss's Kappa for agreement on inclusion in the active aging characteristics was 0.82. Landis and Koch (Landis & Koch 1977) provide a table to interpret the values, and values between 0.81 and 1.00 are considered almost perfect. Furthermore, Fig. 5 presents the degree of agreement by dimension. The open public space, housing, and cultural environment reaches a fair deal, the social climate qualifies a substantial agreement, and the rest obtains an almost perfect agreement.

The average of the responses obtained for the two TAM constructs analysed was calculated (see Fig. 6). It is concluded that clinical psychologists mention that this technological contribution can reduce the time and effort of characterizing active aging in older adults. Also, the participants recall that it is a valuable input since it will allow an excellent characterization of the study variable to develop future primary, secondary, or tertiary intervention plans with the elderly.

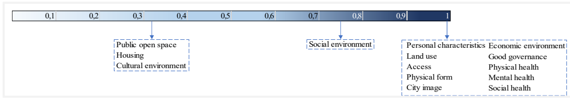

Figure 5: Fleiss Kappa measures per dimension.

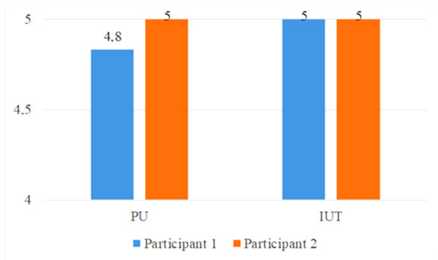

Figure 6: Results of the case study - Clinical psychologist perceptions.

4.5 Threats of Validity

Validity is measured analysing four categories: construct, internal, and external validity, and reliability.

Construct validity focuses on the relation between the theory behind the study and the practical experience. Ensure that the operational measures studied represented what the researchers had planned to investigate and what they were investigating. For Context 1, a previous work by Sutano was used, who proposed a list of variables derived from a rigorous systematic review. Also, the Fleiss Kappa nominal scale was used (Fleiss, 1971). It provides the degree of experts agree to not fall into subjectivities. Furthermore, in Context 2, a validated questionnaire was used whose Cronbach's alpha is ideal. Thus, constructs were interpreted in the same way by the researcher and the interviewees.

Internal validity depends on how the participants are selected. In this study, educational and professional experience and the participant's knowledge about the data mining field could influence the responses and perceptions when using the proposed solution. To mitigate this threat, the selected participants have a similar professional profile.

Due to the COVID-19 pandemic, the external validity considered that the access to numerous groups was restricted. Thus, selecting the sample of individuals who participated was made at convenience; for this reason, the results have to be analysed carefully because they are not generalizable to the population.

From the two forms design to the analysis of results, the reliability considers how the evidence chain was carried out to respect the data's literality. Moreover, the qualitative responses were quantified using a Likert scale to avoid introducing interpretation bias or, failing that, the participants mentioned textually.

5 CONCLUSIONS AND FURTHER WORK

This paper allows the determination of variables' values to determine active aging. Data mining allows identifying among variables that are strongly associated with the topic. Data recollected from different sources in the psychological tests as mental, physical, social, policy health, and personal behavior (Fernandez-Ballesteros, 2011), these variables are matched with variables of models proposed by the WHO and Neuropsychologists (Nayak, Buys, & Lovie-Kitchin, 2006).

The multiple models for measuring and evaluating individuals' active aging have allowed creating this framework to identify the appropriate methods of active aging. Moreover, the proper techniques to analyze them into each data mining process: LikertSvm for Likert scale values, listwise deletion for missing values, standardized data discretization for sociodemographic variables according to their categorization in health care.

Then, according to the input data and literature, a clustering technique is proper to evaluate the groups of active aging in the next stage of modeling. However, the data scientists have to perform a performance evaluation of clusters using metrics or visual analytic analysis to get the best precision in splitting groups.

Due to it not being a standardized technique o evaluation to identify active aging, it is impossible to classify the people in the two groups who have successful aging among those who do not. Thus, the number of variables to consider is significant, so determining which are related is essential to open a path to active aging.

After implementing the proposed framework and getting results, we seek to report results after applying different data science techniques.

ACKNOWLEDGMENTS

The authors wish to thank the Vice-Rector for Research of the University of Azuay for the financial and academic support and all the staff of the Laboratory for Research and Development in Informatics (LIDI), and the Department of Computer Science of Universidad de Cuenca.

REFERENCES

Ariza Colpas, P., Vicario, E., De-La-Hoz-Franco, E., Pineres-Melo, M., Oviedo-Carrascal, A., & Patara, F. (2020). Unsupervised Human Activity Recognition Using the Clustering Approach: A Review. *Sensors*, *20*(9), 2702. https://doi.org/10.3390/s20092702

Baltes, P. B., & Baltes, M. M. (2010). 成功衰老的心理学观点：有偿的选择性优化模型. In *Successful aging*.

Basili, V. R., Caldiera, G., & Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of Software Engineering*, *2*, 528–532. https://doi.org/10.1.1.104.8626

Brummel-Smith, K. (2007). *Optimal aging, part II: Evidence-based practical steps to achieve it*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., & others. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS Inc*, *9*, 13.

Chen, W., Guo, F., & Wang, F. (2015). A Survey of Traffic Data Visualization. *IEEE Transactions on Intelligent Transportation Systems*, *16*(6), 2970–2984. https://doi.org/10.1109/TITS.2015.2436897

Davis, F. (1985). A technology acceptance model for empirically testing new end-user information systems: Theory and results [Massachusetts Institute of Technology]. In *Massachusetts Institute of Technology*. https://doi.org/oclc/56932490

Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.

Fabila Echauri, A. M., Minami, H., & Izquierdo Sandoval, M. J. (2013). La Escala de Likert en la evaluación docente : acercamiento a sus características y principios metodológicos. *Textos y Contextos*, 31–40.

Fernandez-Ballesteros, R. (2011). Positive ageing. Objective, subjective, and combined outcomes. *EJAP (Test)*, *7*(1), 22–30.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619

Gachet Páez, D., de Buenaga Rodríguez, M., Puertas Sánz, E., Villalba, M. T., & Muñoz Gil, R. (2018). Healthy and wellbeing activities' promotion using a Big Data approach. *Health Informatics Journal*, *24*(2), 125–135. https://doi.org/10.1177/1460458216660754

Kandasamy, I., Kandasamy, W. B. V., Obbineni, J. M., & Smarandache, F. (2020). Indeterminate Likert scale: feedback based on neutrosophy, its distance measures and clustering algorithm. *Soft Computing*, *24*(10), 7459–7468.

Kandogan, E. (2001). Visualizing Multi-Dimensional Clusters, Trends, and Outliers Using Star Coordinates. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 107–116. https://doi.org/10.1145/502512.502530

Koufakou, A., Gosselin, J., & Guo, D. (2016). Using data mining to extract knowledge from student evaluation comments in undergraduate courses. *2016 International Joint Conference on Neural Networks (IJCNN)*, 3138–3142.

Lak, A., Rashidghalam, P., Myint, P. K., & Bradaran, H. R. (2020). Comprehensive 5P framework for active aging using the ecological approach: An iterative systematic review. *BMC Public Health*, *20*(1), 1–22. https://doi.org/10.1186/s12889-019-8136-8

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*. https://doi.org/10.2307/2529310

Lupien, S. J., & Wan, N. (2004). Successful ageing: From cell to self. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1449), 1413–1426. https://doi.org/10.1098/rstb.2004.1516

Malik, H., & Mishra, S. (2014). Feature selection using RapidMiner and classification through probabilistic neural network for fault diagnostics of power transformer. *2014 Annual IEEE India Conference (INDICON)*, 1–6.

Moreira, L. B., & Namen, A. A. (2018). A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Computer Methods and Programs in Biomedicine*, *165*, 139–149. https://doi.org/10.1016/j.cmpb.2018.08.016

Nadler, J. D., Damis, L. F., & Richardson, E. D. (1997). *Psychosocial Aspects of Aging*.

Namakforoosh, M. N. (2002). *Metodología de la Investigación*.

Nayak, R., Buys, L., & Lovie-Kitchin, J. (2006). Data Mining in Conceptualising Active Ageing. In J. Li, S. J. Simoff, P. J. Kennedy, P. Christen, & G. J. Williams (Eds.), *Data Mining and Analytics 2006: Proceedings of the Fifth Australasian Data Mining Conference* (pp. 39–46). Australian Computer Society Inc.

Nayak, R., Buys, L., & Lovie-Kitchins, J. (2006). *Data Mining in Conceptualising Active Ageing BT - Fifth Australasian Data Mining Conference (AusDM2006)*. *61*, 39–45. http://crpit.com/confpapers/CRPITV61 Nayak.pdf

Nerenz, D. R., McFadden, B., & Ulmer, C. (2009). *Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement*. National Academies Press.

Pal, A. K., & Pal, S. (2013). Evaluation of teacher's performance: a data mining approach. *International Journal of Computer Science and Mobile Computing*, *2*(12), 359–369.

Pandove, D., Goel, S., & Rani, R. (2018). Systematic Review of Clustering High-Dimensional and Large Datasets. *ACM Trans. Knowl. Discov. Data*, *12*(2). https://doi.org/10.1145/3132088

Runeson, P., Höst, M., Rainer, A., & Regnell, B. (2012). Case Study Research in Software Engineering: Guidelines and Examples. In *Case Study Research in Software Engineering: Guidelines and Examples*. https://doi.org/10.1002/9781118181034

Smith, M. C., & Nancy DeFrates-Densch, A. E. (2008). *Handbook of Research on Adult Learning and Development*. Taylor & Francis.

Stowe, J. D., & Cooney, T. M. (2015). Examining rowe and kahn's concept of successful aging: Importance of taking a life course perspective. *Gerontologist*, *55*(1), 43–50. https://doi.org/10.1093/geront/gnu055

Swindell, W. R., Ensrud, K. E., Cawthon, P. M., Cauley, J. A., Cummings, S. R., & Miller, R. A. (2010). Indicators of "healthy aging" in older women (65-69 years of age). A data-mining approach based on prediction of long-term survival. *BMC Geriatrics*, *10*, 1–24. https://doi.org/10.1186/1471-2318-10-55

Um, J., Zaidi, A., & Choi, S. J. (2019). Active Ageing Index in Korea – Comparison with China and EU countries. *Asian Social Work and Policy Review*, *13*(1), 87–99. https://doi.org/10.1111/aswp.12159

Walker, A., & Maltby, T. (2012). Active ageing: A strategic policy solution to demographic ageing in the European Union. *International Journal of Social Welfare*, *21*(SUPPL.1), 1–10. https://doi.org/10.1111/j.1468-2397.2012.00871.x

World Health Organization. (2015). *Informe mundial sobre el envejecimienot y la salud*.