Leszek A. Maciaszek
Maurice D. Mulvenna
Martina Ziefle (Eds.)

Communications in Computer and Information Science 1856

Information and Communication Technologies for Ageing Well and e-Health

7th International Conference, ICT4AWE 2021
Virtual Event, April 24–26, 2021, and 8th International Conference, ICT4AWE 2022, Virtual Event, April 23–25, 2022
Revised Selected Papers

Springer

Communications in Computer and Information Science 1856

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Leszek A. Maciaszek · Maurice D. Mulvenna ·
Martina Ziefle
Editors

Information and Communication Technologies for Ageing Well and e-Health

7th International Conference, ICT4AWE 2021
Virtual Event, April 24–26, 2021, and 8th International Conference, ICT4AWE 2022, Virtual Event, April 23–25, 2022
Revised Selected Papers

Springer

Editors
Leszek A. Maciaszek
Wrocław University of Economics Institute of Business Informatics
Wrocław, Poland

Macquarie University
Sydney, Australia

Martina Ziefle
RWTH Aachen University
Aachen, Germany

Maurice D. Mulvenna
University of Ulster
Newtownabbey, UK

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-37495-1 ISBN 978-3-031-37496-8 (eBook)
https://doi.org/10.1007/978-3-031-37496-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The present book includes extended and revised versions of a set of selected papers from the International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2021 and 2022), which were exceptionally held as online events, due to COVID-19.

ICT4AWE 2021 received 33 paper submissions from 23 countries, of which 15% were included in this book. ICT4AWE 2022 received 50 paper submissions from 24 countries, of which 32% were included in this book.

The papers were selected by the event chairs and their selection is based on a number of quality criteria that include the classifications and comments provided by the program committee members, the session chairs' assessment and also the program chairs' general assessment of all papers included in the technical program. The authors of selected papers were then invited to submit a revised and extended version of their papers having at least 30% innovative material.

The International Conference on Information and Communication Technologies for Ageing Well and e-Health fosters close exchange and collaboration in the field of digital assistive technologies in the health care domain. The conference aims to be a meeting point for those that study age and health-related quality of life and apply information and communication technologies to help people stay healthier, more independent and active at work or in their community. ICT4AWE facilitates the exchange of information and dissemination of best practices, innovation and technical improvements in the fields of age and health care, education, psychology, social coordination and ambient assisted living. From e-Health to intelligent systems and ICT devices, the conference is a vibrant discussion and collaboration platform for all those that work in research and development or in companies involved in promoting the quality of life and well-being of people, by providing room for research and industrial presentations, demos and project descriptions.

The papers selected to be included in this book contribute to the understanding of relevant trends of current research on Information and Communication Technologies for Ageing Well and e-Health, including: Decision Support Systems, Home Care Monitoring Systems, Systems to Encourage Healthy Lifestyles, Diagnostic Support, Acceptance of Ambient Technology, Internet of Things and Smart Devices for Independent Living, Mobile Health Monitoring, Psychological Dimension of Aging Population, Privacy and Ethical Issues, Open Data and Health Information, Inclusive Design, Human Communication and Behavioural Studies of Senior Citizens, Health Information Systems, Game-Based Approaches towards Health Awareness, Electronic Health Records, Case Studies in eHealth and Autonomy and Active Ageing.

We would like to thank all the authors for their contributions and also the reviewers who have helped to ensure the quality of this publication.

April 2022

Leszek Maciaszek
Maurice Mulvenna
Martina Ziefle

Organization

Conference Chair

Leszek Maciaszek	Wrocław University of Economics Institute of Business Informatics, Poland and Macquarie University, Australia

Program Co-chairs

Martina Ziefle	RTWH Aachen University, Germany
Maurice Mulvenna	Ulster University, UK

Program Committee

Served in 2021

Amit Nanavati	IBM Research, India
Andrea Vitaletti	Sapienza University of Rome, Italy
Annica Kristoffersson	Mälardalen University, Sweden
Antonio Lanatà	University of Florence, Italy
Asta Thoroddsen	University of Iceland, Iceland
Bert-Jan van Beijnum	University of Twente, The Netherlands
Dympna O'Sullivan	TU Dublin, Ireland
Javier Gomez	Universidad Autónoma de Madrid, Spain
Kathleen McCoy	University of Delaware, USA
Mehdi Adda	Université du Québec à Rimouski, Canada
Philipp Brauner	RWTH Aachen University, Germany
Roberto Hornero	University of Valladolid, Spain
Sami Shaban	United Arab Emirates University, UAE
Sandra Baldassarri	University of Zaragoza, Spain
Xiaomu Zhou	Northeastern University, USA

Served in 2022

Andrea Tales	Swansea University, UK
Bo Xie	University of Texas at Austin, USA
Gillian Cameron	Inspire Workplaces & Ulster University, UK
Judith Sixsmith	University of Dundee, UK
Michael O'Grady	University College Dublin, Ireland
Rémi Bastide	University of Toulouse, France

Served in 2021 and 2022

Åsa Smedberg	Stockholm University, Sweden
Andreas Schrader	Universität zu Lübeck, Germany
David Fuschi	Coventry University, UK
Eila Järvenpää	Aalto University, Finland
Elvis Mazzoni	University of Bologna, Italy
Evi Zouganeli	Oslo Metropolitan University, Norway
Georg Duftschmid	Medical University of Vienna, Austria
George Xylomenos	Athens University of Economics and Business, Greece
Heikki Lyytinen	University of Jyväskylä, Finland
Jaakko Hakulinen	Tampere University, Finland
Jane Bringolf	Centre for Universal Design Australia Ltd, Australia
Jitae Shin	Sungkyunkwan University, South Korea
Josep Silva	Universitat Politècnica de València, Spain
Karsten Berns	TU Kaiserslautern, Germany
Laurent Billonnet	University of Limoges, France
Marco Porta	Università degli Studi di Pavia, Italy
Mario Ciampi	National Research Council of Italy, Italy
Marko Perisa	University of Zagreb, Croatia
Maurice Mars	University of KwaZulu-Natal, South Africa
Meng Wong	National Institute of Education, Singapore
Mikel Larrea	Universidad del País Vasco, Spain
Oh-Soon Shin	Soongsil University, South Korea
Paula Forbes	University of Abertay, UK
Raul Montoliu	Universidad Jaume I, Spain
René Meier	Lucerne University of Applied Sciences, Switzerland
Renato Bulcão-Neto	Federal University of Goiás, Brazil
Stefano Federici	University of Perugia, Italy
Susanna Spinsante	Università Politecnica delle Marche, Italy

Taro Sugihara	Tokyo Institute of Technology, Japan
Telmo Silva	University of Aveiro, Portugal
Ulrich Reimer	Eastern Switzerland University of Applied Sciences, Switzerland
Yao Chang	Chung Yuan Christian University, Taiwan, Republic of China

Additional Reviewers

Served in 2022

Edith Maier	Eastern Switzerland University of Applied Sciences, Switzerland
Ernesto Veiga	Federal University of Goiás, Brazil

Invited Speakers

2021

Eling D. de Bruin	ETH Zürich, Switzerland and Karolinska Institute, Sweden
Bo Xie	University of Texas at Austin, USA
Alex Mihailidis	University of Toronto, Canada
Sunil Agrawal	Columbia University, USA

2022

Sabine Koch	Karolinska Institutet, Sweden
Arlene J. Astell	University of Toronto, Canada
Ray Jones	University of Plymouth, UK
Jesse Hoey	University of Waterloo, Canada

Contents

Rescuing Relevant Features from Active Aging Surveys: A Data Mining Perspective

Juan-Fernando Lima[1], Priscila Cedillo[2], María-Inés Acosta-Urigüen[1,3], Marcos Orellana[1(✉)], and Alexandra Bueno-Pacheco[1]

[1] Laboratorio de Investigación y Desarrollo en Informática - LIDI, Universidad del Azuay, Cuenca, Ecuador
{flima,macosta,marore}@uazuay.edu.ec

[2] Departamento de Ciencias de la Computación, Universidad de Cuenca, Cuenca, Ecuador
priscila.cedillo@ucuenca.edu.ec

[3] Facultad de Informática, Universidad Nacional de La Plata, Bueno Aires, Argentina
maria.acostau@info.unlp.edu.ar

Abstract. Within the psychological currents, several proposals on active aging have been defined, conceptualizing it as a perspective or differentiated way of aging satisfactorily. These proposals generate indicators that assess the level of physical health, psychological well-being and adequate social and spiritual adaptation. The indicators are quantified based on active ageing surveys whose are differs for each ageing proposal and collects different features of active aging such as: health, cognition, activity, affection, fitness, and satisfaction levels. This methodology is focused on rescuing the relevant factors (features) facilitating the interpretation of the data, avoiding the non-required characteristics. The methodology proposes a set of data mining techniques for different types of data that could be present in the forms of active aging, and seeking to make a concrete proposal, the features of an active aging survey are evaluated, determining a subset of features where their weights make them more relevant in data collection, and finally, this methodology is positively evaluated as a model of acceptance by geriatric psychologists.

Keywords: Active ageing · Data mining · Features relevance · TAM · Methodology

1 Introduction

When old age concept was analyzed, until not long ago it was considered to be only related to illness, memory problems, senility, dementia, poverty and depression [35]. The World Health Organization, in 2015 presented the World Report on Aging and Health, which covers aging as the sum of several changes; biologically, aging is associated with the accumulation of a wide variety of molecular and cellular damages that gradually reduce physiological reserves, increase the risk of many diseases, and generally decrease the individual's capacity [59]. In addition, it defines Active or Healthy Aging as the process of promoting and maintaining the functional capacity that allows

L. A. Maciaszek et al. (Eds.): ICT4AWE 2021/2022, CCIS 1856, pp. 138–156, 2023.
https://doi.org/10.1007/978-3-031-37496-8_8

well-being in old age, quantifying it through variables that include personal characteristics, health characteristics, genetic inheritance, functional capacity and intrinsic capacity of being human; as well as variables of the environment in which the person develops, such as identity, relationships, the possibility of enjoyment, autonomy, security and the potential for personal growth [5,59].

The World Health Organization (WHO) considers that there are three main periods of healthy aging: a relatively high and stable capacity, a diminished capacity, and a significant loss of capacity. It also identifies that these there are different approaches to quantify the active ageing, but all of them keep the same purpose of promoting and maintaining the human intrinsic capacity, so that people with reduced functional capacity can continue to carry out activities that are important to them [59]. Trajectory that does not depend on chronological age and that is not uniform among the individuals of a population.

Worldwide, the concept of active aging has gained interest; for example, the European Union has focused on issues related to a transition from the perception that older adults are only recipients of a retirement to an active orientation in which they are active subjects at the family, community, work, educational level, etc. [57] Another example is observed in China, Japan, and South Korea; they have proposed an integrated, composite index for measuring the contribution of older people to the society, to their communities and their families, considering the percentage of older population aged 60 or more. This index is compound of metrics that identify the staying in labor market longer, the contribution or participation in the community, society and family activities, the capacity of living independently, and the promotion of individuals' capacities [56].

Although identifying active aging in a people is relevant, for governments, it is even more relevant to generate population indicators that allow them to build policies with a view of improving the reality of this segment of the population. In this context, data mining techniques is useful to quantify, tabulate and obtain results oriented to active aging, where 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 [40,41,54]. Prediction, detection of outliers (anomaly detection), clustering and decision making are some of the most used techniques in the data mining field [38,47].

According to the above descriptions, active ageing definitions such as by the WHO, European, and Oriental have some similar characteristics among them, but also differs on multiples definitions and variables to measure the active ageing, this obstruct the way to create a general framework to evaluate the active ageing surveys designed by psychologists basing on the definitions that they believe proper. Thus, it is required to propose a data mining methodology to evaluate variables collected by active ageing surveys. The perspectives of data scientists are base to communicate the relevant techniques and tools to face and overcome the most important barriers for each data analysis stage. A previous work presented the first steps to discover insights among active ageing variables and the method was embodied as a business process notation [2]. Based on it, this paper presents the methodology for surveys guided by a software engineering specification, deep into the data mining techniques, and shown an instance of the method analyzing variables of an active ageing survey.

The paper is organized as follows: In Sect. 2, in form of background, the active ageing definitions are presented allowing to give context to readers. In Sect. 3, The methodology to face the evaluation of active ageing variable is shown. In Sect. 4, the evaluation by psychologists of this method is presented. In Sect. 5, the conclusions and viable applications of this method are shown.

2 Related Work

While the concept of active ageing is analyzed, the WHO highlight that old ageism refers to creating stereotypes, discrimination and prejudice against old people [59]. Since at the end of the 60s of the 20th century, various sociological perspectives have been proposed as theoretical objectives focused on distinguish the "good" and "bad" ways of aging. However, models of aging based on exact psychological indicators have now been developed, although these models have a great common basis, each of them represents a different perspective of aging successfully, below the most relevant perspectives are shown.

2.1 Successful Aging

This concept presents a model based that measure the capacities of elderly taking into account the absence and risk reduction of disease and disability; the high functional capacities, and active engagement with life. This term was introduced by Robert J. Havighurst (1961) [39] in order to classify aging patterns in two groups, normal and pathological capacities; and according to authors such as [23] or [45], it is necessary to distinguished heterogeneity within elderly groups, each of them having different stereotypes, experiences and expectations. Thus, the Rowe and Kahn model (1997) establishes three key indicators to characterize successful aging related to the probability of suffering diseases or disabilities, the function of cognitive and physical capacities and the engagement to life. The model of successful cognitive ageing compares an elderly individual to other groups, the approach compares the cognitive performance of an specific person with data obtained of the same chronological age and establishes an index on relative homogeneity in young adulthood but heterogeneity thereafter [35].

2.2 Optimal Aging

The concept of optimal ageing can be described in the manner an elderly spends his life and the opportunities he has including not only aspect related to the biological, medical and socioeconomic environment, but also including a form of life that can be optimized and cultivated [46]. It measures satisfaction despite the presence of adverse medical conditions that can alter the life of an older adult [10]. The concept of optimal ageing includes the decision of an individual to optimize different stages of his life, depending on his goals, his projection or his environment. The concept includes, in addition to the level of health, the mental health and the social functioning as major elements that determines plasticity in aging. In [3], where the selection, optimization and compensation are mandatory to adjust proper goals and objectives related to that age [6]. The concept

includes the analysis of aspect related to the effect of physical activity, nutrition, and social as elements to determine the physical functioning, the mental/cognitive functioning, the active social functioning, and an overall health indicator [37]. Brummel-Smith considered the social environment and support system as elements that influences the process in adaptation to changes and also prevents individuals from pathologies caused by stress [9].

2.3 Active Aging

The WHO in 2002 raised the concept of active aging. This proposal, from the beginning, included health, participation and safety as basic components, although later this work was taken up again and learning was included as a fourth component of this concept [8]. In this way, active aging has been defined as "... the process of optimizing opportunities for health, participation and safety in order to improve the quality of life as people age" [28], proposal that can be approached from a public policy perspective. The first pillar, health, refers to preventing diseases and promoting healthy habits that promote well-being on a physical, mental and social level [28]. The pillar of physical level depends on peace, security, food, income, ecosystem, and other variables that constitute the individual's system [28]. Active aging has a third pillar, the safety, since insecurity of any kind could have harmful effects on health and well-being. At the individual level, the risks of illness, death, unemployment for long periods; food insecurity, economic and cultural insecurity [8]. The latest and newest pillar in active aging is the learning. This is seen as a resource that improves the ability to remain healthy, by allowing different knowledge and skills to be acquired and updated to maintain people's capacities and also to better reinforce personal safety [28].

2.4 Positive Aging: European Model

Fernández - Ballesteros points out that this concept integrates what has been defined as optimal, successful, active, productive, healthy aging, and other nomenclatures that express, in short, an aging that entails the integral well-being of the subject. The author points out that three main aspects are grouped under all these concepts, a) successful aging is multidimensional, since it encompasses biophysical, emotional, cognitive and social aspects; b) they cannot be reduced to purely biological domains in terms of health and disease, nor to subjective conditions; and c) there are no great intercultural differences in this construct. Thus, it is determined, through exploratory factor analysis, that there are 5 factors that affect this type of aging, and they are: a) health, cognition, activity, affect, and fitness [20].

2.5 New Proposals for Successful Aging

A recent proposal states that successful aging is determined by four aspects: physical, psychological, social and pleasant activities [34]. The psychological factor refers to the existence of a lower stress level, as well as a generalized sense of psychological well-being, which implies self-acceptance, positive relationships with others, autonomy, mastery of the environment, purpose in life and personal growth [49]. Regarding

the social factor, social support (availability of people that one can count on), religion and commitment to life are mentioned as essential points that help people face the adversities of life [51]. Pleasant activities include actions like exercise or vacations. In order to evaluate this factor, questions were asked that inquired about the frequency with which the person traveled outside the city or the frequency with which they carried out exercise. This model was tested using structural equations, and had a very good fit showing that these factors make up the successful aging construct [34].

2.6 Psychosocial Models of Successful Aging

Carver and Buchanan's proposal seeks to highlight non-biomedical aspects that are considered in other models. This would allow to speak of a good or optimal aging even in those older adults who have some disease. Their approach considers, as a main point, leaving aside biomedical constructs so that more older adults can be included. In this case, several aspects need to be considered: commitment to life, optimism or positive attitude, resilience. In addition, the factor of spirituality or religiosity is included, and can be beneficial. It is understood as the fact of getting involved in contemplative or altruistic activities. Finally, self-efficacy and self-esteem are indicated as significant in the model, as well as gerotranscendence, which refers to the older adult being able to see his life in retrospect and give it meaning and purpose, as well as understand it coherently [13].

In Sweden another model established that successful aging has to do with a subjective and retrospective assessment of overall quality of life. From this point, they affirm then that the commitment with life is the main factor to have a good aging [52]. This concept is based on the theory of activity, specifically related to the participation in pleasant or leisure activities. In this sense, a person who presents higher levels of activity in this type of dynamism will be opposing the so-called lack of productivity that comes with retirement, thus guaranteeing their successful aging. In [26] a model for activity domain includes culture-entertainment, productive-personal growth, physical-outdoors, recreation-expressive, friendship, and formal group. Those items remark the importance of participation and the commitment that they generate. As it has been analyzed, these models in general have something in common: they all seek the integral well-being of people. They show the interest in the deep human being, beyond the only visible and objective, it considers the person in all its facets.

3 Methodology

Based on the Software & Systems Process Engineering Metamodel (SPEM) specification, the proposed methodology is an instance of the six stages of the Cross-Industry Standard Process for Data Mining (CRISP-DM) [58], and it is shown in Fig. 1. The active ageing context is fundamental providing: roles, concepts, and artefacts such as codebooks and the raw datasets. Several datasets has to be transformed to ordinal types, therefore, data dictionaries are essential to accomplish this task. And after that, the model can be build according the data quality, the feature selection model has to support different data types.

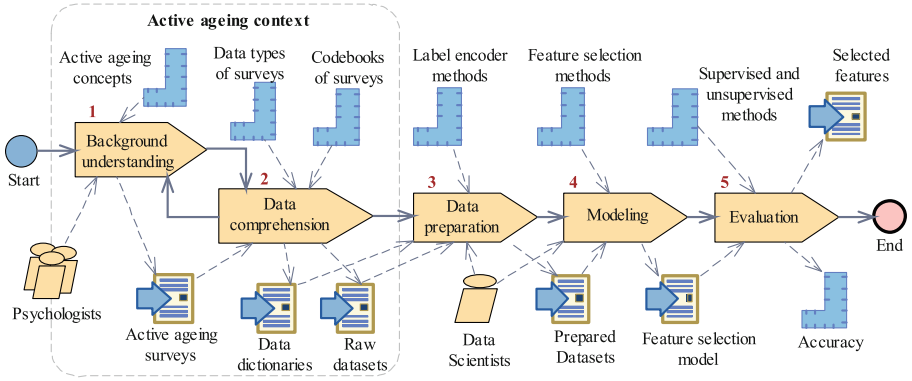

Fig. 1. Methodology for select the relevant features among active ageing surveys.

3.1 Data Comprehension

The different approaches focused on evaluate the active ageing seen in the above section collect several variables and data types making hard the interpretation if collected data, it inclusive for only one ageing survey. In this context, grateful to the web page of Institute for Research on Ageing it is possible to confirm the several approaches and waves for ageing analysis, the institute collected and opened multiple datasets seeking to improve the research on this knowledge domain [27].

To address this propose, one survey collected in the study by the Advanced Cognitive Training for Independent and Vital Elderly carried out in United States since 1999 until 2001 [55] is used as an input example to detect the relevant features and discard the others without lose the weight of the input data for a machine learning algorithm. In order, from the 43 collected forms in the study, seeking proof but also simplify this whole methodology, we selected the data of the DS0027 folder related to the form 618-Medication Audit. This dataset is conformed by 38 features and 5156 records by people aged between 65 and 94 years. Each feature is a medicament, the description of each feature is stored in the respective codebook, and the referent value indicates how many the person forgot to take the medicine.

3.2 Data Preparation

In a data mining process the data preparation stage wastes more than the 50% of the whole process, this is caused by one or multiple techniques are applied until to get a proper dataset to be used as input in a data mining algorithm. The processes inside of the data preparation stage commonly are: normalization, label encoder, missing values treatment, data discretization, dimensional reduction; in addition, surveys require extra techniques due to the text existence, in these cases, techniques such as: lower case, quantities parsing, punctuation treatments, stop words, stemming, and lemmatization are indispensable prior to extract relevant information from the active ageing forms collected in textual mode. This subsection aims to describe the more relevant

data preparation techniques among the listed, and according to the sample form apply the required tasks.

Dealing with Missing Values. In the latest years, across several academic domains data is always collected in different types, however, no matter the type of data, it could be numerical or not, discrete or continuous, but perhaps missing values have been unavoidable in all of these domains. From this point, several approaches to deal with missing data have been developed, among these it is possible to highlight: assurance of data quality, imputation, and dropping tuples (based on its percentage) [18].

In cases where features are numerical or time series-based then the forecasting methods work well, but in case of surveys such as active ageing, the features tend to be categorical, or ordinal types complicating much more the execution of imputation approaches. Works such as Swindell [54] shows how to deal with continuous and categorical features, proposing to fill in missing data with the "Imputation approach" method. The missing data are imputed based on the average value between the $k = 20$, but the credibility of the imputed data is low. In psychology domain, the choice about missing data relapses on the opinion of experts and on the amount of data collected [18], 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 [41].

Transforming by Data Discretizacion. According to [41], a challenge to face in surveys is the standard for peoples' ages transformation, who are part of the group to be analyzed. Ages are indispensable across studies for healthy. The Research of Adult Learning and Development Handobbok address the guidelines for treatment of these age groups [53]. A simple way is to transform ages focused on standardize the age in groups as: childhood, adolescence, emerging adulthood, average adulthood, and late adulthood. In complement to it, the ordinal encoders can give support to these age groups making possible to include in numerical data mining algorithms [61]. In the same way, Ethnicity can be categorized [42] but label encoders can be used instead of ordinal encoders [61]. The encoders are trained to allow transform nominal to numerical types, but also these algorithms execute the inverse task facilitating to trace the data discretization and transformation.

Dealing with Likert Scales. In the same way, satisfaction and comfortable feedback levels are part of active ageing surveys, commonly these levels are based on the Likert scale. Both ways to deal with this scale are present, on the one hand, ordinal encoders can transform and make these appropriate for a data mining process, but on the other hand, special algorithms for deal with this scale are available increasing the power of data mining process improving the goodness in the results [30].

Dealing with Texts. In data containing words or phrases, it is necessary to identify words that will contribute or not to a mining process [38]. Text mining is considered a huge topic inside data knowledge retrieval, perhaps to develop a Natural Language Processing (NLP) perspective over surveys can fit as a specific papers to address the

treatment. However, in this point, it is presented a overview of the main techniques for a text mining process. In an initial description, we highlight that it has the same six steps than a data mining process [58]. According to [24], the text preparation includes the following techniques:

Nonalphabetic Characters Removal. The preprcessing technique allows to remove numbers, punctuation, special characters. However, depending of the domain understanding, number can be transformed to words to avoid loss information. For example, A question in survey related to number of pill administration per patient.

Text Conversion. Lowercase and uppercase conversions are used to standardize the inputs, and reduces the vocabulary, therefore, it decreases the number of dimensions. It is necessary highlight that inside NLP each word is a dimension (feature).

Spelling Correction. Due to manual writing could contain wrong words or personal styles, the spelling is used for correct these issues. Also, this technique helps to increase the statistical power and avoid the capture of styles.

Contractions and Abbreviations. It uses a dictionary to transform the abbreviations (both formal and informal) and get the complete words.

Stop Words. It is a list of word contains common words to be removed without value in the analysis, for example, the sentence connectors.

Lemmatization and Stemming. These techniques are focused on get a root among words wrote in different tenses. The first one uses a dictionary to get the simple root of a word, for example: "$wrote$"=> "$write$". On the other hand, stemming seeks to truncate the word through rules, for example: "$speaking$"=> "$speak$", however, it does not consider the tenses, for example, "$funnier$"=> "$funnier$".

3.3 Modeling

Reveal the features whose characterize the active aging is the main interest of this work, several feature selection approaches could be take into account to resolve the objective, however, only a select group of approaches have support for the multiple data types collected in the related surveys of active ageing how is shown in the above. The most simple approach is determine the influence of variables using the algorithm "Select by weights" that selects only those whose weights satisfy a criterion concerning the input weights [36].

Clustering. The selection or removal of features depends of the input data, therefore, in some cases it is need to apply filters or a deeper filter analysis for discovering some groups of individuals with similar patterns, hence, when this requirement floated out, the clustering technique is justly focused on divide data onto subgroups of interest, incorporating a unsupervised approach as part of the analysis. The application of unsupervised methods has been used in similar works focused on data segmentation, the work of [43] evidences the predictive value of different measures of cognition, clustering 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. The application of clustering over social

domains has been widely used as an unsupervised technique for human activity recognition [4], this leads to take it as part of a possible segmentation case.

When it is clear how clustering will solve it, it is necessary to select the proper technique among the multiple algorithms such as: hierarchical, k-means, random sampling, randomize search, condensation based, densityBased, grid-based, probabilistic model-based, clustering graphs, or network data [44]. The goodness of data is proportional to the fulfilment of its purpose, normally, 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 [44]. The problem defines two segments: proper or no proper values for active aging. Therefore, k-means algorithm is proposed to generate the two segments [30], clustering gives another advantage related to the responses of likert scale due to its nature of losing information and data distortion.

Feature Selection and Dimensional Reduction. The above points aim to create a validated and structured input to be used in this stage. However, a concern is data with redundant variables [25], and also the unstructured data collected from interviews or tests such as cognitive diagnostics or clinical treatments are also an issue to overcome [11], in both cases, high dimensional treatment is the gap to overcome, therefore, two ways are visible to accomplish to task.

Feature Extraction. In this method algorithms are focused on transforming a set of features (variables) to a new dataset, with low dimensions using mathematical functions, some of them can revert the process, but there is some loss in precision of the output values. These techniques are widely used in the neuropsychological domain as a dimensionality reduction to deal with data collected from digital machines as function of Magnetic Resonance Images [19,29], or Electroencephalograms images (EEG) [1,7,17]. The principal component analysis (PCA) or its variants is the common method to perform this task.

Feature Selection. This approach selects only the relevant variables from a dataset. Unlike feature extraction which transform data, this method keeps the understanding of the variables. This approach requires a better data pre-processing, and finally, an evaluation process to ensure the quality of the output features. The evolution of feature selection has granted multiple vias or methods such as: Filter Methods. A ranking mechanism is used to grade the features (variables), and based on a value as a threshold, the features are removed. The methods are based on relevance and redundancy. The common methods are: variance, information measure, correlation measure, distance measure, consistence measure, fuzzy set theory, and rough set theory [14]. Wrapped Methods, this method uses the predictor performance as an objective function to evaluate the variable subset, suboptimal subsets are found by employing search algorithms, the algorithms start with an empty dataset and then add features (or by full dataset and removing features) until to obtain the maximum objective function [60]. Evolutionary Computation Paradigms, some authors of consider the evolutionary computation into the wrapped methods, however, a more recent study [60] details the importance of split these methods, the heuristic search evaluates different subsets until it optimizes the objective function, multiple subsets

are created by searching in a searchspace, the used algorithms by this method are genetic algorithms, genetic programming, particle swarm optimization, ant colony optimization, and differential evolution. The feature selection process is shown in Fig. 2.

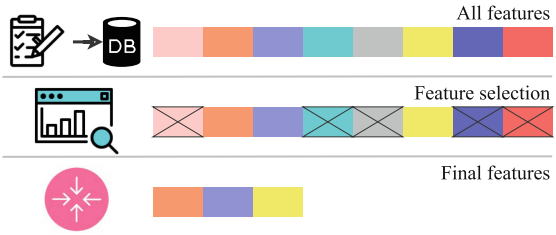

Fig. 2. The overview of feature selection for active aging forms.

Case of Rescue Relevant Features. Using as input the data of the survey 618 explained in Sect. 3.1, The objective is to evaluate relevance among features (X axes), therefore, the feature selection has to be applied. As preprocessing operations: the AID, OTHERMED, and NUMMEDS features were dropped, and the VISIT feature was transformed using an ordinal encoder and set itself as evaluation predictor known also as a target.

In order, in the modeling stage, to evaluate the weight of each feature, a filter method was used due to the absence of a target feature, how we have seen the wrapped method uses a predictor to evaluate each variable. The chi-square test evaluates the relationship among the features, its main advantage is the support for nominal, ordinal, and numeric values. The test compares the features against the normal distribution, the distribution is described as $x^2 = \Sigma Zi^2$, where $z1, z2, .., z(n)$ are standard normal variables. In feature selection the test aims to select the features which are highly dependent on the response. For null hypotheses (H0) the two variables are independent, and for alternative hypotheses (H1) two variables are not independent. The python package for machine learning provides an API for feature selection, and it can be used by the method $SelectKBest(chi2, k = n)$ simplifying the analysis. In Fig. 3 is possible see the $p - value$, the null hypotheses is rejected if the value falls in the error region (alpha from 0 to 0.05). In this survey, 12 features are independents due to these fall into the alpha range, and the rest of features are not independent.

In this case, one of the surveys was analyzed, but the rest of these can to follow the guidelines described in the above subsection. Therefore, evaluate the relevance among: successful, optimal, active, positive, psychological proposals described in Sect. 2.

3.4 Model Evaluation

The main advantage models based on a predictor analysis is itself validation due to its precision values, as we have seen, the p-values are fundamental for evaluating the

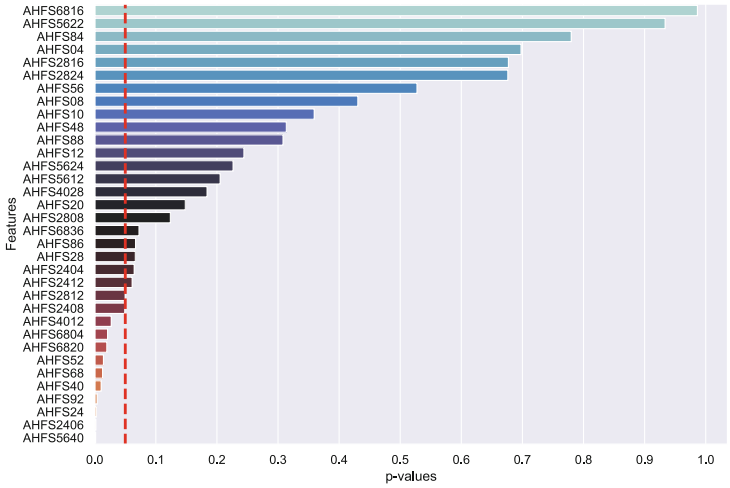

Fig. 3. Rescuing relevant features from the Form 618-Medication Audit, $alpha = 0.05$.

weight among features. On the other hand, If clustering was used prior the feature selection, a simple way is verify the grouped data is related is through multidimensional graphs. Generally, 2D or 3D dimensional plots are used to display the groups [31], the visual analytics is a recent strategy in which the human capabilities are used to interpret the data behavior. The interpretation is executed over popular visualization methods as: Bar charts, Line charts, Pie charts, and Scatter plots. In this case, the scatter plot is a standard method to describe a clustering application, while the human can evaluate according to the dispersion of data. However, this method is hard to assess if clusters are very closed [15]. In this case a mathematical indicator is required to improve the verification of goodness. Two indicators are proposed by the literature, the elbow analysis and the silhouette indicator are viable to increase the confidence of the generated groups [50].

4 Evaluating the Feature Selection of Surveys

In the same way as the previous work [2], This section presents a case study seeking to evaluate the acceptance of this methodology. In this context, Runeson in 2012 proposed a set of activities to evaluate a product [48], 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. The results of this evaluation represent an essential knowledge for data engineers who work in factors associated with active aging in the proposed methodology. To achieve this evaluation experts in active aging evaluate the inputs used in the proposed methodology according to Runeson.

4.1 Designing the Evaluation

Seeking to obtain a degree of agreement among the experts, this evaluation analyzes regarding the variables in a survey of healthy aging. Besides, this case study is focused

on evaluate the perceptions among health personnel regarding the usefulness of the results derived from the stages of the data method.

The evaluation and the scope were addressed by the Goal-Question-Metric (GQM) approach proposed by Basili et al. in 1994 [12]. The GQM scheme is structured as follows: a) The analysis of the inputs for the methodology proposed by the data engineers, b) What is the purpose of the objective measures the agreement among active ageing experts against each activity regarding the usefulness of the information resulting from the methodology, c) From the point of psychologists, and d) In the health context where this study is carried out.

Analyze. The evaluation analyzes the inputs for the methodology proposed by the data engineers.
With the Purpose of. reach a degree of agreement among health experts, and evaluate their perceptions regarding the usefulness of the information resulting from the methodology.
From the Point of View of. Clinical psychologists.
In the Context of. a real execution of feature evaluation.

And, in this context, the research questions are:

1. What stage(s) are not avoidable in the health domain?.
2. What the perception of health personnel on the usefulness of a methodology that allows selecting active aging variables collected utilizing data mining techniques?.

Based on Runeson et al. [48] recommendations, this case study method is holistic-multiple, and the units of analysis are presented in Fig. 4.

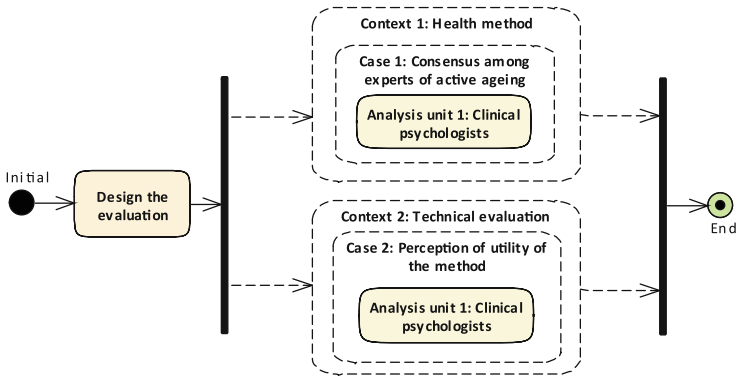

Fig. 4. Process for evaluation the acceptance of this work.

4.2 Preparation for Data Collection

Two surveys were designed to achieve the objectives of this case study. For Context 1, a form was developed based on Lak, Rashidghalam, Myint, and Bradaran [32] (see

https://n9.cl/iwle), 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.

Next, seeking to evaluate the Context 2, a form was created based on the technology assessment model (TAM) proposed by Davis [16]. However, to limit its application, we only analyze the constructs of the Perceived Utility (PU) and the Intent to Use (UIT) in the future, therefore, the final product of the methodology is analyzed. Hence, data mining experts designed the form with its respective explanation based on the likert scale, it is shown in the following URL: https://n9.cl/qcmb4.

4.3 Data Analysis and Reporting of Results

Both questionnaires were explained, then were filled by two Clinical psychologists with experience in the gerontological domain, they are experts due to their experience both in academic degree and also in practical field.

Next while the results were interpreted, it was found that in Case 1, Fleiss' Kappa statistical measure used for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings (likert) 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 also is used to validate the process of inclusion/exclusion of variables presented in Appendix 1 of the form (see https://n9.cl/iwle). Finally, the selections of each reviewer was checked and some discrepancies were resolved with consensus among they.

For the both two raters, the Fleiss's Kappa for agreement on inclusion in the active aging resulted 0.83, Landis and Koch [33] provide a table to evaluate and interpret the resulting values, thus, values between 0.81 and 1.00 are considered almost perfect. The average of the responses obtained for the two TAM constructs analysed was calculated (see Fig. 5). It is concluded that clinical psychologists mention that this technological contribution can reduce the time and effort to evaluate the relevance among features. Also, the experts recall that it is a valuable input since it will allow an excellent characterization of the study variable to develop intervention plans in different levels.

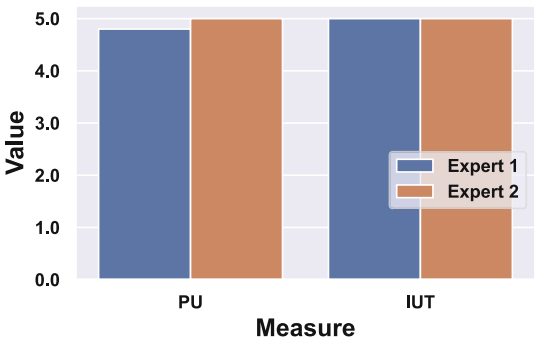

Fig. 5. Results of the case study - Perceptions of the clinical psychologists.

In comparison with the previous work, the results of the evaluation of this TAM does not change, the psychologists maintain their responses. It could be due to the simplicity to fill the forms and the experience of the raters. The values obtained in the previous works of 0.82 and 0.83 in this work do not change the results of this case study.

4.4 Threats of Validity

The evaluation of threats to validity in a case of study is a critical task focused on secure the quality of empirical studies in Software Engineering (SE), this includes the systems and methods for data treatment and analysis due to these are part of information systems. Threats to validity normally is understood by four categories: construct, internal, and external validity, and conclusions [62].

Construct Validity: focuses on identify the correct use of operational measures for the concepts being studied. In context 1, the work by Sutano [22] as a list of active ageing features collected in a systematic literature review. In order to compare the degree of experts agree seeking to not fall into personal subjectivities the Fleiss Kappa scale was used. For context 2, the questionnaire proposed and validated by the Cronbach's alpha is ideal. Thus, constructs were interpreted in the same way by the researcher and the interviewees.

Internal Validity: is focused on establish a causal relationship, whereby certain conditions are believed to lead to other conditions, in this case, how the participants were selected. The knowledge and experience of participants about data mining could influence the responses and perceptions when using the proposed solution. To avoid this threat, the participants have a similar professional profile.

The External Validity: defines the domain to which the findings can be generalized. in this case, 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 due to the specific requirements of each data mining technique.

The Conclusion Threat: demonstrates that the operations of the case of study such as the data collection procedure can be repeated, with the same results, this threat is avoid due to the use of general forms, the first one based on literate, and the second one validated by the Cronbach's alpha value. Moreover, the qualitative responses were quantified using a Likert scale to avoid introducing interpretation bias.

5 Conclusions and Further Work

This paper in determine the relevance among features of surveys providing by active ageing evaluation. In perspective of data scientists, the data mining techniques are fundamental to achieve a proper data evaluation identifying among features are strongly associated with the topic discarding the others. Data recollected from different sources in the psychological tests as mental, physical, social, policy health, and personal behavior [21], these variables commonly are collected based by models proposed by the international organizations and according to the neuropsychologists perceptions [41].

The multiple data mining techniques for measuring variables have allowed creating this method to rescue the most relevant variables among different data types, therefore,

evaluating the relevance of these it is possible to focused on the proper variables for active aging. Moreover, the proper techniques to analyze them into each data mining process: Likert scale values treatment, listwise deletion or missing values imputation for missing values dealing, standardized data discretization of continuous values, and ordinal and label encoders for sociodemographic variables.

Then, according to the objective evaluation, a clustering technique could be used as part of the analysis to split data in subgroups and then evaluate the features weight. However, the use of clustering has to perform a performance evaluation of groups using indicators or visual analytics to get the best precision in the generated groups, after getting the results, we seek to report results seeking to apply different data mining techniques to evaluate the features. The critical premise is to evidence the existence of a target feature as predictor. According to the application case shown in Sect. 3.3.3, this predictor allow to compare among features and rescue the most relevant. On the other hand, when there is not a predictor, feature extraction techniques such as PCA can be used although the original structure of features has to change. If the data analysts want to maintain the original values among features, filter methods are candidates to execute this task, the information measures will be generated by techniques such as correlation or variance among the popular techniques.

In our application over the survey 618-Medication Audit, we detect 12 features are independent among them, rescuing this features as the relevant to be interpreted by clinical psychologists and gerontology experts. The evaluation of this method as a TAM is positive among the experts. Further efforts are mainly centered on evaluate features between the feature selection and the feature extraction techniques. The perspectives to create this method are fundamental and were well received by experts who do not part of the data science domain, this paper grants several tools and theoretical applications of techniques over survey collected by active ageing evaluations. Based on these evidences, it will be possible to detect the most relevant features among the surveys and contrast them with the base proposal and its principles of ageing to verify the application of the active proposal.

Acknowledgements. 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

1. Abdulhay, E., et al.: Computer-aided autism diagnosis via second-order difference plot area applied to EEG empirical mode decomposition. Neural Comput. Appl. **32**(15), 10947–10956 (2020). https://doi.org/10.1007/s00521-018-3738-0
2. Acosta-Urigüen., M., Cedillo., P., Orellana., M., Bueno., A., Lima., J., Prado., D.: Finding insights between active aging variables: towards a data mining approach. In: Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE, pp. 268–275. INSTICC, SciTePress (2022). https://doi.org/10.5220/0011068100003188

3. Aldwin, C.M., Spiro, A., Park, C.L.: Five - health, behavior, and optimal aging: a life span developmental perspective. In: Birren, J.E., Schaie, K.W., Abeles, R.P., Gatz, M., Salthouse, T.A. (eds.) Handbook of the Psychology of Aging (Sixth Edition), pp. 85–104. Academic Press, Burlington (2006). https://doi.org/10.1016/B978-012101264-9/50008-2. https://www.sciencedirect.com/science/article/pii/B9780121012649500082
4. Ariza Colpas, P., Vicario, E., De-La-Hoz-Franco, E., Pineres-Melo, M., Oviedo-Carrascal, A., Patara, F.: Unsupervised human activity recognition using the clustering approach: a review. Sensors **20**(9) (2020). https://doi.org/10.3390/s20092702. https://www.mdpi.com/1424-8220/20/9/2702
5. Baldassar, L., Atkins, M.: Healthy active ageing rapid evidence review, heart foundation walkwise, pp. 978–1 (2020). https://doi.org/10.26182/zdh2-ej22
6. Baltes, P.B., Baltes, M.M.: Psychological perspectives on successful aging: the model of selective optimization with compensation. Successful aging: Perspectives from the behavioral sciences, pp. 1–34. Cambridge University Press, New York (1990). https://doi.org/10.1017/CBO9780511665684.003
7. Basterrech, S., Krömer, P.: A nature-inspired biomarker for mental concentration using a single-channel EEG. Neural Comput. Appl. **32**(12), 7941–7956 (2020). https://doi.org/10.1007/s00521-019-04574-2
8. Brazil, I.L.C.: Active ageing: A policy framework in response to the longevity revolution (2015)
9. Brummel-Smith, K.: Optimal aging, part ii: evidence-based practical steps to achieve it. Ann. Long Term Care **15**(12), 32 (2007)
10. Brummel-Smith, K.: Optimal aging, part i: demographics and definitions. Ann. Long Term Care **15**(11), 26 (2007)
11. Buendía, F., Gayoso-Cabada, J., Juanes-Méndez, J.A., Sierra, J.L.: Transforming unstructured clinical free-text corpora into reconfigurable medical digital collections. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp. 519–522 (2019). https://doi.org/10.1109/CBMS.2019.00105
12. Caldiera, V.R.B.G., Rombach, H.D.: The goal question metric approach. Encycl. Softw. Eng., 528–532 (1994)
13. Carver, L.F., Buchanan, D.: Successful aging: considering non-biomedical constructs. Clin. Interv. Aging **11**, 1623–1630 (2016)
14. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014). https://doi.org/10.1016/j.compeleceng.2013.11.024, https://www.sciencedirect.com/science/article/pii/S0045790613003066. 40th-year commemorative issue
15. Chen, W., Guo, F., Wang, F.Y.: A survey of traffic data visualization. IEEE Trans. Intell. Transp. Syst. **16**(6), 2970–2984 (2015). https://doi.org/10.1109/TITS.2015.2436897
16. Davis, F.D.: A technology acceptance model for empirically testing new end-user information systems : theory and results (1985)
17. Devipriya, A., Nagarajan, N.: A novel method of segmentation and classification for meditation in health care systems. J. Med. Syst. **42**(10), 193 (2018). https://doi.org/10.1007/s10916-018-1062-y
18. Enders, C.K.: Applied Missing Data Analysis. Guilford Press, New York (2010)
19. Estella, F., Delgado-Márquez, B.L., Rojas, P., Valenzuela, O., Roman, B.S., Rojas, I.: Advanced system for automously classify brain MRI in neurodegenerative disease. In: 2012 International Conference on Multimedia Computing and Systems, pp. 250–255 (2012). https://doi.org/10.1109/ICMCS.2012.6320281
20. Fernández-Ballesteros, R.: Envejecimiento saludable. In: Congreso sobre envejecimiento. La investigación en España, pp. 9–11 (2011)
21. Fernández-Ballesteros, R.: Positive ageing: objective, subjective, and combined outcomes. E-J. Appl. Psychol. **7**(1), 22–30 (2011)

22. Fleiss, J.L.: Measuring nominal scale agreement among many raters. Psychol. Bull. **76**(5), 378–382 (1971). https://doi.org/10.1037/h0031619
23. Gutiérrez, M., Desfilis, E.S., Zacarés, J.J.: Envejecimiento óptimo: perspectivas desde la psicología del desarrollo. Promolibro (2006)
24. Hickman, L., Thapa, S., Tay, L., Cao, M., Srinivasan, P.: Text preprocessing for text mining in organizational research: review and recommendations. Organ. Res. Methods **25**(1), 114–146 (2022). https://doi.org/10.1177/1094428120971683
25. Houari, R., Bounceur, A., Kechadi, M.T., Tari, A.K., Euler, R.: Dimensionality reduction in data mining: a copula approach. Expert Syst. Appl. **64**, 247–260 (2016). https://doi.org/10.1016/j.eswa.2016.07.041, https://www.sciencedirect.com/science/article/pii/S0957417416303888
26. Hutchinson, S.L., Nimrod, G.: Leisure as a resource for successful aging by older adults with chronic health conditions. Int. J. Aging Hum. Dev. **74**(1), 41–65 (2012). https://doi.org/10.2190/AG.74.1.c. PMID: 22696843
27. Institute for Research on Ageing, McMaster University: Open access datasets from aging studies (2022). https://mira.mcmaster.ca/research/open-access-datasets-from-aging-studies. Accessed 15 Aug 2022
28. Kalache, A., Gatti, A.: Active ageing: a policy framework. Adv. Gerontology = Uspekhi Gerontologii **11**, 7–18 (2003). http://europepmc.org/abstract/MED/12820516
29. Kale, V.V., Hamde, S.T., Holambe, R.S.: Brain disease diagnosis using local binary pattern and steerable pyramid. Int. J. Multimed. Inf. Retrieval **8**(3), 155–165 (2019). https://doi.org/10.1007/s13735-019-00174-x
30. Kandasamy, I., Kandasamy, W.B.V., Obbineni, J.M., Smarandache, F.: Indeterminate likert scale: feedback based on neutrosophy, its distance measures and clustering algorithm. Soft. Comput. **24**(10), 7459–7468 (2020). https://doi.org/10.1007/s00500-019-04372-x
31. Kandogan, E.: Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2001, pp. 107–116. Association for Computing Machinery, New York (2001). https://doi.org/10.1145/502512.502530
32. Lak, A., Rashidghalam, P., Myint, P.K., Baradaran, H.R.: Comprehensive 5p framework for active aging using the ecological approach: an iterative systematic review. BMC Public Health **20**(1), 33 (2020). https://doi.org/10.1186/s12889-019-8136-8
33. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159–174 (1977). http://www.jstor.org/stable/2529310
34. Lee, P.L., Lan, W., Yen, T.W.: Aging successfully: a four-factor model. Educ. Gerontol. **37**(3), 210–227 (2011). https://doi.org/10.1080/03601277.2010.487759
35. Lupien, S.J., Wan, N.: Successful ageing: from cell to self. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. **359**(1449), 1413–1426 (2004)
36. Malik, H., Mishra, S.: Feature selection using rapidminer and classification through probabilistic neural network for fault diagnostics of power transformer. In: 2014 Annual IEEE India Conference (INDICON), pp. 1–6 (2014). https://doi.org/10.1109/INDICON.2014.7030427
37. McREYNOLDS, J.L., Rossen, E.K.: Importance of physical activity, nutrition, and social support for optimal aging. Clin. Nurse Spec. **18**(4) (2004). https://journals.lww.com/cnsjournal/Fulltext/2004/07000/Importance_of_Physical_Activity,_Nutrition,_and.11.aspx
38. Moreira, L.B., Namen, A.A.: A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. Comput. Methods Programs Biomed. **165**, 139–149 (2018). https://doi.org/10.1016/j.cmpb.2018.08.016. https://www.sciencedirect.com/science/article/pii/S0169260718307569

39. Nassir, S., Leong, T.W., Robertson, T.: Positive ageing: elements and factors for design. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, OzCHI 2015, pp. 264–268. Association for Computing Machinery, New York (2015). https://doi.org/10.1145/2838739.2838796
40. Nayak, R., Buys, L., Lovie-Kitchin, J.: Influencing factors in achieving active ageing. In: Sixth IEEE International Conference on Data Mining - Workshops (ICDMW 2006), pp. 858–862 (2006). https://doi.org/10.1109/ICDMW.2006.100
41. Nayak, R., Buys, L., Lovie-Kitchin, J.: Data mining in conceptualising active ageing. In: Li, J., Simoff, S.J., Kennedy, P.J., Christen, P., Williams, G.J. (eds.) Data Mining and Analytics 2006: Proceedings of the Fifth Australasian Data Mining Conference, pp. 39–46. Australian Computer Society, Australia (2006). https://eprints.qut.edu.au/14011/
42. Nerenz, D., McFadden, B., Ulmer, C.: Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. National Academies Press (2009). https://books.google.com.ec/books?id=JDOYmzQSUNsC
43. Pal, A.K., Pal, S.: Evaluation of teacher's performance: a data mining approach. Int. J. Comput. Sci. Mob. Comput. **2**(12), 359–369 (2013)
44. Pandove, D., Goel, S., Rani, R.: Systematic review of clustering high-dimensional and large datasets. ACM Trans. Knowl. Discov. Data **12**(2) (2018). https://doi.org/10.1145/3132088
45. Posada, F.V., Tur, M.C.T., Resano, C.S., Osuna, M.J.: Bienestar, adaptación y envejecimiento: cuando la estabilidad significa cambio. Rev. Multidiscip. Gerontol. **13**(3), 152–162 (2003)
46. Poscia, A., et al.: Workplace health promotion for older workers: a systematic literature review. BMC Health Serv. Res. **16**(5), 329 (2016). https://doi.org/10.1186/s12913-016-1518-z
47. Páez, D.G., de Buenaga Rodríguez, M., Sánz, E.P., Villalba, M.T., Gil, R.M.: Healthy and wellbeing activities' promotion using a big data approach. Health Inf. J. **24**(2), 125–135 (2018). https://doi.org/10.1177/1460458216660754
48. Runeson, P., Host, M., Rainer, A., Regnell, B.: Case Study Research in Software Engineering: Guidelines and Examples, 1st edn. Wiley Publishing, Hoboken (2012)
49. Ryff, C.D.: Beyond ponce de Leon and life satisfaction: new directions in quest of successful ageing. Int. J. Behav. Dev. **12**(1), 35–55 (1989). https://doi.org/10.1177/016502548901200102
50. SAPUTRA, D.M., SAPUTRA, D., OSWARI, L.D.: Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. In: Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019), pp. 341–346. Atlantis Press (2020). https://doi.org/10.2991/aisr.k.200424.051
51. Seeman, T.E., Lusignolo, T.M., Albert, M., Berkman, L.: Social relationships, social support, and patterns of cognitive aging in healthy, high-functioning older adults: Macarthur studies of successful aging. Health Psychol. **20**(4), 243–255 (2001). https://doi.org/10.1037/0278-6133.20.4.243
52. Silverstein, M., Parker, M.G.: Leisure activities and quality of life among the oldest old in Sweden. Res. Aging **24**(5), 528–547 (2002). https://doi.org/10.1177/0164027502245003
53. Smith, M., DeFrates-Densch, N.: Handbook of Research on Adult Learning and Development. Routledge, Milton Park (2009). https://books.google.com.ec/books?id=HrWslN2zgL4C
54. Swindell, W.R., et al.: 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), 55 (2010). https://doi.org/10.1186/1471-2318-10-55
55. Tennstedt, S., et al.: Advanced cognitive training for independent and vital elderly (active), United States, 1999–2001 (2010). https://doi.org/10.3886/ICPSR04248.v3

56. Um, J., Zaidi, A., Choi, S.J.: Active ageing index in Korea - comparison with China and EU countries. Asian Soc. Work Policy Rev. **13**(1), 87–99 (2019). https://doi.org/10.1111/aswp.12159
57. Walker, A., Maltby, T.: Active ageing: a strategic policy solution to demographic ageing in the European union. Int. J. Soc. Welf. **21**(s1), S117–S130 (2012). https://doi.org/10.1111/j.1468-2397.2012.00871.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2397.2012.00871.x
58. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, vol. 1, pp. 29–39. Manchester (2000)
59. World Health Organization: World report on ageing and health. Technical report, World Health Organization (2015). Accessed 01 Aug 2022
60. Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. IEEE Trans. Evol. Comput. **20**(4), 606–626 (2016). https://doi.org/10.1109/TEVC.2015.2504420
61. Zhang, H., Zheng, G., Xu, J., Yao, X.: Research on the construction and realization of data pipeline in machine learning regression prediction. Math. Prob. Eng. **2022**, 7924335 (2022). https://doi.org/10.1155/2022/7924335
62. Zhou, X., Jin, Y., Zhang, H., Li, S., Huang, X.: A map of threats to validity of systematic literature reviews in software engineering. In: 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), pp. 153–160 (2016). https://doi.org/10.1109/APSEC.2016.031