Integral model for measuring unsatisfied basic needs in Colombia

Germán Méndez-Giraldo[1], Laura Palacios-Rodriguez[1], and Rafael Ropero-Lyton[1]
[1]Universidad Distrital F. J. C. Colombia, gmendez@udistrital.edu.co, lvanessarp@gmail.com, raedula96@gmail.com

Abstract* - This paper works on the concept of unsatisfied basic needs as a system for measuring the level of country's development index. In the case of Colombia and other Latin American countries, is mainly based on the determination of the level of poverty, inequality, and quality of life of the population. This research achieved the characterization of the variables that measure unsatisfied basic needs in order to make a prediction through data mining. The data analyzed was taken from official studies as the national census; it is recognized that the measurement of quality of life and well-being can be carried out both objectively and subjectively. However, official studies only include objective measurements, so it is important to answer the following question: how can this deficiency in the surveys be corrected? For this reason, a global satisfaction indicator was designed, it is built from the individual satisfaction components. It is concluded that for levels of housing and households the objective measurements correctly classify the instances, but for the level of people, level of satisfaction was considered as an indicator of subjective nature and that it will depend, among others, on demographic, cultural variables and in general on their individual beliefs. A categorization model based on demographic variables is proposed to calculate the global satisfaction index and with which the percentage of instances correctly classified is improved in all cases by more than 30%.

Key Words: NBI, Data Mining, Quality of life, Well-being, Inequality.

I. Introduction

In recent decades, populations worldwide have undergone a socioeconomic transformation that has led to a reduction in poverty, however, there are still cultural and economic gaps that reflect the existing inequality [1 and 2]. On the other hand, the issue of well-being and quality of life has acquired greater importance; for this reason, nations have established improving their levels of quality of life and social well-being as one of their main challenges, all of them with many related research [3]. Problems such as unemployment, forced displacement, drug trafficking, poverty, social exclusion, and armed conflict, are simultaneously results and causes of the inequality, which have had a strong impact on the economic and social development of Colombia, causing a constant decline in the well-being and citizens' quality of life [4, 5 and 6].

Currently, the measurement of the level of development in Latin American countries is based mainly on determining the level of poverty, inequality, and quality of life of the population, associated with different indicators. The information collected for this work includes the UBN (Unsatisfied Basic Needs) as a composite indicator in addition to each of its simple indicators. In Colombia, the UBN indicator according to DANE (National Department of Statistics in Spanish) is defined by five simple indicators, among: inadequate housing, housing with critical overcrowding, housing with inadequate public services, housing with high economic dependency, and housing with school-age children who do not they attend school.

Data presented by the DANE continue to show a high percentage of total population that is below the poverty line, those who manage to overcome this limit are in a high-risk situation. These data showed that there are a higher percentage of occurrence in terms of poverty rates at the rural level, vulnerability of women, people with a lower level of studies, and people with informal jobs [7]. In addition, variables such as well-being, defined by [8] as the level of satisfaction not only individual but also collective, in terms of their needs, including a wide range, from the most vital to the most superfluous. It is important to guarantee people access in terms of services, culture, etc., since they are means to achieve a reduction in poverty rates. The poverty is conceptualized as the deprivation of access to goods, services and income that allow a level of well-being accepted as a minimum in a society [9].

Limited availability of studies and applications related to quality of life and well-being present a series of drawbacks because they do not allow to obtain a comprehensive view of the main variables that make up social well-being and, therefore, do not offer solutions to improve. The answers obtained from the National Quality of Life Survey (NQLS) most of the time have a cultural context immersed, which implies that many times they depend on interviewee's social context and the individual and collective perception that he has, for this reason, many times the answers are not an accurate reflection of reality. This research seeks to characterize the variables that measure unsatisfied basic needs to make a prediction through data mining of satisfaction and conditions of households in Colombia. Finally, it is important to explain that DANE defines UBN only by 5 different simple indicators, and according with values reported by interviewed, it can be stated whether it is in an unsatisfied necessities state, and if it is in poor or in misery. These indicators are:

- Inadequate housing, measures at the infrastructure level: The situation of deprivation occurs in homeless people, homes without walls, or with walls made of zinc, fabric, canvas, cardboard, cans, waste, plastics or semi-permanent or perishable material, or houses with dirt floors.
- Homes with inadequate services: This indicator directly expresses the lack of access to minimum vital and sanitary conditions, that is, homes without sanitary facilities or that lack an aqueduct and are supplied with water directly from sources, tank trucks or rainwater.

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2021.1.1.369
ISBN: 978-958-52071-8-9 ISSN: 2414-6390

• Homes with critical overcrowding: This indicator considers homes with more than three people per room.

• Homes with high economic dependency: There are considered in this case an employed with three or more members that depend on him, and whose head of household has passed at least two years of primary education.

• Households with school-age children who do not attend school: This is considers households with at least one child between 6 and 12 years old, a relative of the head of the household who does not attend a formal education center.

The DANE defines a household in poverty conditions, when it has at least one of these indicators, and in a poverty condition if they express two or more simple indicators of the NBI.

II. Material and Methods

The scope of the research was the establishment and recognition of the variables within NQLS to describe the conditions of poverty, level of satisfaction and quality of life of people; variables related to poverty, inequality, education, health, physical characteristics of the dwellings, average occupation, public services in the homes and proximity to different services in the neighborhoods or communities of Colombia were considered. The sets of variables under study are divided into three different study areas as following: a housing study area, household study area, and a people study area.

A mixed methodology is used, based and developed by the Knowledge Acquisition and Representation Group through Expert Systems and Simulation (ARCOSES in Spanish) of Francisco José de Caldas District University and the Knowledge Discovery in databases (KDD) methodology. This hybrid methodology is divided into two basic stages. The First Stage or Collection of knowledge corresponds to the acquisition of knowledge of the system to find variables that affect the quality of life and unsatisfied basic needs in Colombia. The second one, or KDD proposed by Fayyd in 1996, [10] and referred to in detail in [11] is composed of five steps: Selection, preprocessing, transformation, data mining and evaluation and interpretation. Figure 1 presents the diagram of the methodology used in this research.

The NQLS information is representative for the whole country, as it was applied to 24 main cities with their metropolitan areas and the rest of secondary cities known as municipalities as well as rural centers and other Colombian dispersed population. For the application of this methodology, firstly, it was necessary to identify the main variables that affect social well-being and quality of life in Colombia; the concept of quality of life was simplified to the quantitative measurement of factors such as income and physical conditions of the environment and housing, through numerical indicators such as GDP, UBN, Gini Index, HDI (Human development Index), among others. The WHO in 1994 proposed a definition for quality of life as the perception of the individual about his position in life, in the context of the culture and value system in which somebody lives, and in relation to his objectives, expectations, standards and interests, [12].

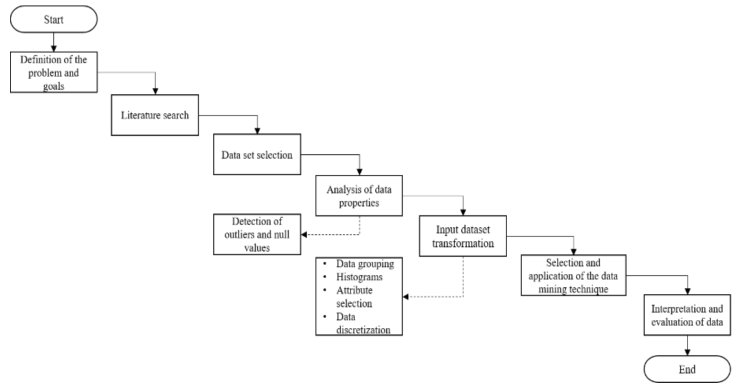

Fig. 1 Methodology used

It is necessary to determine two types of variables: The output or goals variables, these variables are predicted, calculated, or inferred, and the predictor or independent variables, which are used to process data. In literature there is not consensus about which main dimensions should be analyzed, although different authors list a basic series of dimensions that could be considered in research on well-being and quality of life. These dimensions include income, employment, nutrition, health, education, housing, social interactions and security; these dimensions generate a consensus around well-being and quality of life, [13]. For the Colombian case, DANE uses two different indices with which it measures poverty in the country, the first is the multidimensional poverty index (MPI) where satisfaction is evaluated around five dimensions: educational possibilities, conditions of childhood and youth, health, work, access to home public services and housing conditions; The second measure is an indirect method that assesses the ability of households to purchase goods and services, the monetary poverty index is calculated as the percentage of the population with incomes below minimum monthly income, this income let to cover basic needs; in other words, those necessary to acquire a basket of goods (food and non-food) that allow an adequate standard of living, [14 and 15].

According to the latest reports in terms of multidimensional poverty, Colombia increased its percentage between 2016 and 2018 by 1.8 percentage points, while in terms of monetary poverty between 2017 and 2018 there was an increase of 0.1 points percentage. Additionally, there is a history between 2010 and 2018 of the MPI in the different areas of the country, capital cities, populated centers and dispersed rural areas, statistics show that in Colombia regions have different levels of incidence of multidimensional and monetary poverty. On the other hand, social differences implies that groups and classes have some form of differentiation that usually increases with their size. Then social inequality is the hierarchization of these differences, it is the condition by which people have unequal access to valued resources [16].

For the case of this study, the selected database is the NQLS made by DANE, which has a nationwide coverage and includes within its observations the three levels required for a complete study of well-being and life satisfaction of the population; these levels are: households, housing, and people. On the other hand, it contains all the variables required to analyze the level of incidence of the UBN in Colombia. See Table 1.

Table 1. Data analyzed

D.B.	Tables	N.O.	N.V.	T.S.
Households	Household living conditions and possession of assets	8612	94	809.528
	Household living conditions and possession of property programs)	119	9	1071
	Household living conditions and possession of assets (aliases)	1721	7	12.047
	Ownership and financing of the dwelling occupied by the household	8612	20	172.240
	Ownership and financing of the dwelling occupied by the household (alias)	1351	7	9.457
	Updated home services	8612	58	499.496
Housing	Housing data	8501	38	323.028
People	Characteristics and composition of the home	26501	39	1.033.539
	Health	26501	103	2.729.603
	Comprehensive care for children under 5 years of age	1853	56	103.768
	Education	24649	48	1.183.152
	Work force	21698	122	2.647.156
	Information and communication technologies	24649	72	1.774.728
	Child labour	2952	43	126.923

As can be seen in table 1, in some of the cases of the databases that make up the levels of households and people, there are differences with the number of observations, this is because some of them are only applied to a specific population, either due to the inherent interviewer's characteristics or because the information type on a particular variable which not everyone decides to answer. All these differences require adjustments which can be the removal of noisy or atypical data, selecting the strategies for handling unknown data, null data, duplicate data, and statistical techniques for their replacement.

According to [17], a series of techniques are available to address the problems of null or missing values, such as: 1) Discard the records with missing data, such as the Listwise Deletion that eliminates all that record and the Pairwise Deletion which only removes records that have missing data on irrelevant variables. 2) Deletion of lost data, desirable when the lost data cannot be easily imputed. 3) Impute the missing data, it can be a simple or multiple imputation. For this stage of the process, first a review of each of the tables is carried out, variable by variable, eliminating those that include information that is not relevant for the study or there are records that have many missing fields. In these cases, it cannot be determined if this is due to system error, the pollster or because the interviewed chose not to answer. It is possible to eliminate some variables that are used as keys to open a series of subsequent questions.

Subsequently, atypical data must be analyzed and treated, determining what conditions the class variable is found, this is a way to identify if there is an imbalance of the observations, using heuristics such as subsampling, which consists of training the model with an equal number of samples in each class, then it is possible discarding instances of the class bigger, in this way its balance is achieved. Other way is the oversampling, which, on the contrary, seeks to increase the number of samples in the smaller class replicating them in the database, is a method characterized by the fact that it does not lead to the loss of information, such as the well-known algorithm SMOTE (synthetic minority oversampling technique), [18].

To facilitate data processing, on some occasion's variables were grouped by type of information. In case of the "Housing data" database, the variables could be divided into three sections: Variables related to the physical conditions of the house, including type of house and material of the floors, ceilings, and walls; variables related to public services in homes, focusing on basic services (electricity, sewerage, aqueduct, and garbage collection) and finally, variables related to the physical and environmental conditions outside. For the "People data", two types of subdivisions were made, the first by thematic and the second by population ranges. Additionally, in the creation of the variable "satisfaction", considering that this is a multidimensional variable, five types of satisfactions included in the original survey were used. However, to perform a calculation, a different weight was given to each of the variables, to see what impact, it had in satisfaction component; and its specific contribution to estimate a specific metric of well-being and to value the performance of the algorithms in each of these scenarios. In the case of the "Household data", even though the household section was made up of several tables, these were grouped into one that contained the total information of this level, since for the construction of this model it was used as a variable of classification the total incidence of the NBI in each one of the observations.

Subsequently, the data adjustment and its projection are carried out, using methods of reduction of dimensions or transformation to reduce the effective number of variables under consideration or to find invariant representations of data. For this reason, other data adaptation tasks are necessary, typically, elimination of redundant attributes, the creation or combination of existing attributes, normalization of attributes or the transformation of these to facilitate the processing. For example, the variable "Life Satisfaction; Satisfaction with: Income, health, safety and job satisfaction", a Likert scale was made, converting the numerical responses into a 5-point scale. With values such as: Very High: 9 -10, High: 7 -8, Medium: 5 - 6, Low: 3 - 4 and Very Low: 0-2. This process of discretizing the continuous numerical variables in the different databases allows the creation of ranges within which the possible values that the variable can take can be accommodated, this allows reducing the possible number of states, this process is of special help to improve the performance of the models applied in the data mining (DM) process.

These techniques can simplify a table both in a database horizontally or vertically. Techniques such as aggregations, data comprehension, histograms, segmentation, discretization, among others, are used; these techniques allow selecting a data mining method, at this stage it is necessary to decide which is the most appropriate technique, thus it is necessary to determine the type of information to be extracted from the data. DM methods refer to the types of functions provided by the tools used to look for patterns in the data, such as deciding which models and parameters may be appropriate and matching a particular DM method to the general criteria of the KDD process, [19]; these can be of classification, regression, clustering, or summarization, [20]. To choose algorithms and subsequently carry them out, some techniques are used that serve as guidelines to do this. The first depends on which data mining methodology is chosen and the second, where the choice is conditioned only to the characteristics of the learning problem and the data, [21]. For the last techniques are used the algorithms Zero R, J48, Random Forest, Random Tree, Naive Bayes, SMO and IBK. Each algorithm was run 15 times on the data base and each experiment was carried out using a different seed, to find the standard deviation to apply each algorithm on the training data base, which allows selecting the technique that best fits in each case. All of the above allows to know possible deviation of each model may have during its evaluation stage.

The training process was carried out using a partition in the database, where the WEKA software used 70% of the observations to carry out the training and the construction of the model, and the remaining 30% to carry out its evaluation.

At the beginning of the validation process, it was found that for household variables these classification techniques were very good, as will be seen in the next section, but in the case of people variables it did not work adequately. Recognizing that for the second case there is a component of subjectivity under which, not only the NQLS is based on subjectivity due to the way in which information is captured for certain variables, but at the same time the components of the selected global satisfaction depend on the person interviewed while their classifying as variable. For this, an analysis is also carried out on this database compiled at the level of people in different ways. The construction of the model was proposed, where each of the evaluations of satisfaction in their different components were made by population ranges.

The proposed model, equation 1, which is considered as a contribution to the measurements made institutionally by DANE, since individuals have different degrees of perception of quality of life and therefore well-being is based on characteristics own from each one of them; for this it is recognized that through the available information there are three main characteristics such as gender (g), age (a) and level of education (s). The output variable is Global Satisfaction (Gs). This is represented by:

$$\begin{aligned} Gs = {} & wl^{s}_{g,a} \times Ls^{s}_{g,a} + wi^{s}_{g,a} \times Li^{s}_{g,a} \\ & + wh^{s}_{g,a} \times Hs^{s}_{g,a} + ws^{s}_{g,a} \times Ss^{s}_{g,a} \\ & + wj^{s}_{g,a} \times Js^{s}_{g,a} \qquad (1) \end{aligned}$$

Where:

Gs: It is the global satisfaction index

$wl^{s}_{g,a}$: Is the weight given to the life satisfaction index based on gender, age, and educational level

$Ls^{s}_{g,a}$: Is the life satisfaction index based on gender, age, and educational level

$wi^{s}_{g,a}$: Is the weight given to the income satisfaction index according to gender, age, and educational level

$Li^{s}_{g,a}$: Is the income satisfaction index based on gender, age, and educational level

$wh^{s}_{g,a}$: Is the weight given to the health satisfaction index based on gender, age, and educational level

$Hs^{s}_{g,a}$: Is the health satisfaction index based on gender, age, and educational level

$ws^{s}_{g,a}$: Is the weight given to the safety satisfaction index based on gender, age, and educational level

$Ss^{s}_{g,a}$: Is the safety satisfaction index based on gender, age, and educational level

$wj^{s}_{g,a}$: Is the weight given to the job satisfaction index based on gender, age, and educational level

$Js^{s}_{g,a}$: Is the job satisfaction index based on gender, age, and educational level

The values for each characteristic are presented in Table 2 and constitute the selection values for the category of membership and subsequently carry out the classification process.

Table 2. Levels of each selection characteristic

Characteristic	Number of Levels	Level description
Gender	2	– g(1): Man – g(2): Woman
Age	6	– a(1): Early childhood: 0 -5 years – a(2): Childhood: 6 - 11 years – a(3): Adolescence: 12 - 18 years – a(4): Youth: 19 - 26 years – a(5): Adulthood: 27 - 59 years – a(6): Old age: 60 years or more
Scholarship	6	– sc(1): None – sc(2): Preschool – sc(3): Primary – sc(4): High school – sc(5): University – sc(6): Technical / Technological

III. Results

In the first instance, it is recognized that the measurement of quality of life and well-being can be carried out both on an objective and subjective way, however, the governmental institutions decided to measure it only objectively. This proposal is to address the issue of unsatisfied basic needs and well-being, measured through the global satisfaction indicator. For the levels of housing and households, unsatisfied basic needs were considered as an indicator of well-being, which is verifiable and evident with reality, thus it is constituted as an objective measurement. On the other hand, for the level of people, the level of satisfaction was considered as an indicator, which is based on the perception that the individual has of their situation, so it is considered as an indicator of a subjective nature and it depends, among others, of demographic, cultural and general variables of own individual beliefs.

Figure 2 presents the structure of the KDD process for the construction of the model with the databases, this allows to show each of the stages that were followed to obtain the results of the algorithms applied to the databases.

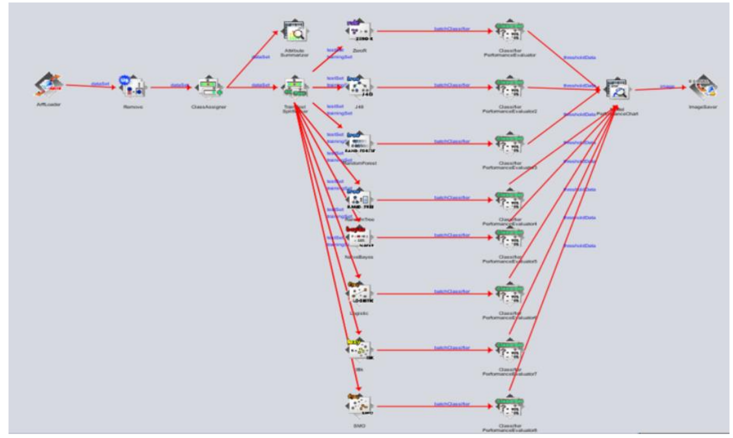

Fig. 2 KDD flow chart for the construction of the classification model

Next, in table 3 is the summary of the information of the algorithms with the best performance, these are the better models to predict the category of a complete observation given some input variables.

Table 3. Summary of metrics for the selected models

Table Name	Selected Algorithm	Instances ranked well (%)	ROC (%)	PRC
Data to housing - Exterior	SMO	99.49	1	0,99
Home services	Random Forest	81.8	0,85	0,84
Gender – Full Table g(1) y g(2)	SMO	93.28	0,97	0,9
Gender: g(1)	SMO	91.08	0,96	0,87
Gender g(2)	SMO	93.03	0,97	0,9
Age: a(3)	SMO	91.68	0,94	0,88
Age: a(4)	SMO	89.85	0,94	0,86
Age: a(5)	SMO	92.61	0,97	0,9
Age: a(6)	SMO	92.44	0,97	0,89
Scholarship: sc(1)	SMO	88.33	0,96	0,84
Scholarship: sc(2)	SMO	66.67	0,75	0,56
Scholarship: sc(3)	SMO	91.49	0,97	0,88
Scholarship: sc(4)	SMO	92.18	0,96	0,89
Scholarship: sc(5)	SMO	91.77	0,94	0,88
Scholarship: sc(6)	SMO	88.33	0,92	0,84

ROC: Receiver Operating Characteristic PRC: Precision ReCalls

With these probabilities it is possible to generate a ROC curve; This is obtained by varying minimum probability value necessary for a certain record to be associated with a certain class. For this work, cost is associated with the rate of false positives (x-axis), and benefit is associated with the rate of true positives or accepted instances (y-axis). Figure 3 shows the ROC curves of the models obtained after applying the algorithms already mentioned (ZeroR, J48, Random Forest, Random Tree, Naive Bayes, SMO and IBK). As can be seen, the curve that is closest to one (1) is by the SMO algorithm.

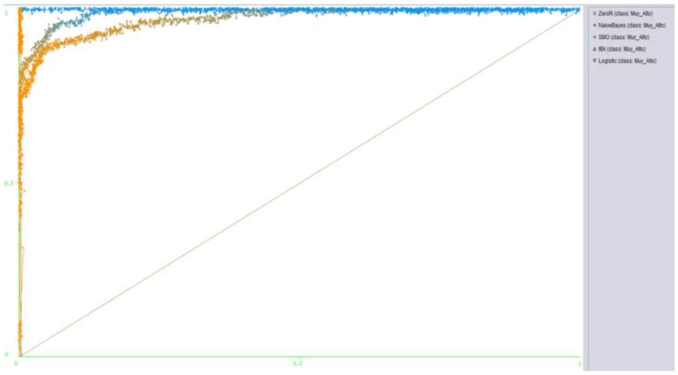

Fig. 3 ROC curve of the models obtained from the table "Housing data - Exterior" using the algorithms Naive Bayes, IBK and SMO. Software: WEKA, Knowledge Flow GUI.

All models built at people level where all the components are included using the proposed model, (see equation 1) measured the level of satisfaction for each demographic class (gender, age, and level of education), this model generates more robust classification results (Table 5) than if it does not use (Table 4). It is important to say that results obtained by the traditional way or where there are no instances of demographic categorization and are based only characteristics given from database. It is found that the number of instances correctly classified is significantly reduced in this traditionally way, and the ROC and PCR values have a value close to 0.5, which indicates that there is a random behavior at the time of classifying, or that the algorithm cannot find relationships between the variables to have a separation capacity between classes.

Table 4. Results of the classification process without demographic categorization

Variables Name	No demographic components of categorization		
	Well-classified instances (%)	ROC (%)	PCR
Overall satisfaction	50.02	0,66	0,47
Gender: g(1)	59.26	0,74	0,52
Gender: g(2)	55.01	0,69	0,49
Age: a(3)	59.52	0,56	0,5
Age: a(4)	54.36	0,61	0,48
Age: a(5)	57.81	0,73	0,52
Age: a(6)	62.33	0,81	0,55
Scholarship: sc(1)	57	0,77	0,51
Scholarship: sc(2)	57.79	0,78	0,52
Scholarship: sc(3)	53.81	0,69	0,49
Scholarship: sc(4)	61.05	0,57	0,51
Scholarship: sc(5)	52.18	0,55	0,46

Table 5. Results of the classification process with demographic categorization

Variables Name	Demographic components of categorization		
	Well-classified instances (%)	ROC (%)	PCR
Overall satisfaction	93.28	0,97	0,9
Gender: g(1)	91.08	0,96	0,87
Gender: g(2)	93.03	0,97	0,9
Age: a(3)	91.68	0,94	0,88
Age: a(4)	89.85	0,94	0,86
Age: a(5)	92.61	0,97	0,9
Age: a(6)	92.44	0,97	0,89
Scholarship: sc(1)	88.33	0,96	0,84
Scholarship: sc(2)	91.49	0,97	0,88
Scholarship: sc(3)	92.18	0,96	0,89
Scholarship: sc(4)	91.77	0,94	0,88
Scholarship: sc(5)	88.33	0,92	0,84

IV. Discussion

The results of "Household Data" reach a degree of correctly classified instances of 99.49% and ROC and CRP values of 1 and 0.99, which confirms that these variables are of an objective nature, that is, they are evidenced and allow determining the satisfaction of the easy and correct way. This facilitates the implementation of strategies to improve the level of quality of life of households, allowing and facilitating access to public services, improving the conditions of the room and the increase of goods and equipment that improve the living standards of the community.

When analyzing the results, without using the proposed model, gave on average the values of 56.7% of the well-classified instances and values of ROC and CRP of 0.68 and 0.5, respectively; while with the use of the classification model by demographic categories, the values are higher and on averages are 91.3% of well-classified instances and ROC and CRP levels of 0.96 and 0.88, respectively. This validates the hypothesis of the existence of subjectivity in the national survey. The results show that the people model with the perspective of population ranges, is more robust and it is possible to get more information on the behavior of these variables and levels of satisfaction. For this reason, if the analysis is carried out from this approach, it can be found that depending on the population range and giving greater weight to a certain component of satisfaction, besides a better fit is found in certain ranges. For example, in the case of the population range of old age, the best classification percentage is when the scenario is used in which the component with the greatest weight of global satisfaction is with their specific satisfaction in health, and for people between adolescence and youth ranges, the best classification is when their specific satisfaction is based on increasing their income. As an example, Table 6 shows the measurement of satisfaction by age group at each level of education.

Table 5. Average of Instances correctly classified with the SMO algorithm of the people table divided by population range l

Table	a(1)	a(2)	a(3)	a(4)	a(5)	a(6)
Total Education	93%	93%	92%	93%	92%	92%
Technique	87%	88%	87%	86%	86%	86%
Higher	89%	90%	88%	89%	89%	88%
High school	92%	92%	92%	92%	91%	92%
Primary	91%	92%	91%	91%	91%	91%
Preschool	77%	77%	64%	77%	83%	77%
None	86%	85%	86%	86%	83%	86%

Adolescence	90%	90%	90%	90%	89%	88%
Adulthood	92%	92%	92%	92%	91%	92%
Youth	90%	87%	88%	89%	88%	87%
Old age	90%	91%	91%	92%	91%	91%
Woman	92%	92%	92%	92%	91%	92%
Man	91%	92%	92%	92%	91%	91%

In conclusion, it can be said that by increasing the degree of instances correctly classified for the measurement of satisfaction in any index such as quality of life, satisfaction with income, or with social security, health, or job satisfaction, as well as the overall satisfaction index. All this allows to improve reliability for decision-makers who want to define public policies, strategies or simply government actions. In this way, they will be more assertive if it is considered that the population handles different expectations according to their age, gender, and age, this allows for focusing efforts on these actions by the State to improve the well-being and quality of life of their fellow citizens.

As previously said, it is recognized that the measurement of quality of life and well-being can be carried out both objectively and subjectively, and the housing table is not alien to this situation, the variables that compose it can be divided into these two forms. The objective variables are those related to the physical conditions of the home and those related to public services, focusing on basic services such as: Electricity, Sewerage, Aqueduct, and garbage collection. The analysis of overall satisfaction is correlated with the physical and environmental conditions from the outside and which provide a subjective view, because it is related to the perception that respondents have about their environment.

References

[1] Manuscript Templates for Conference Proceedings, IEEE. http://www.ieee.org/conferences_events/conferences/publishing/templates.html Amara, S., Idrisb, I. P., & Anisd, A. (2020). Exploring the link between income inequality, poverty reduction and economic growth: An ASEAN perspective. International Journal of Innovation, Creativity and Change, 11(2), 24-41.

[2] Roy, J., Tscharket, P., Waisman, H., Abdul Halim, S., Antwi-Agyei, P., Dasgupta, P., ... & Pinho, P. F. (2018). Sustainable development, poverty eradication and reducing inequalities.

[3] Uduporuwa, R. J. M. Sustainable City Development is Possible? A Review of Challenges and Key Practices towards Urban Development in Developing Countries.

[4] Palacio, J., Maya-Jariego, I., Blanco, A., Amar, J., & Sabatier, C. (2017). Quality of life and Health in displaced communities affected by the armed conflict in Colombia. In Quality of Life in Communities of Latin Countries (pp. 167-184). Springer, Cham.

[5] Pulido-Velásquez, M., & Alegría-Castellanos, A. (2017). An analysis of the Colombian Civil Conflict: Synthetic Control Approach.

[6] Giusta, G., Pastor, M., & Benner, C. (2018). Post-conflict Colombia: Inclusive Economies and urban development.

[7] Soto Iguarán, C. (2016). Lucha contra la pobreza y extensión de la cobertura social en Colombia. Opera, (18), 35–59. https://doi.org/10.18601/16578651.n18.04

[8] Ayvar, F., Navarro, J., & Giménez, V. (2015). El bienestar social en América Latina, 1990-2014: Un análisis DEA a partir de las dimensiones del desarrollo humano. Revista Nicolaita de Estudios Económicos, 10(2), 7–28.

[9] Benvin, E., Rivera, E., & Tromben, V. (2016). Propuesta de un indicador de bienestar multidimensional de uso del tiempo y condiciones de vida aplicado a Colombia, el Ecuador, México y el Uruguay. Cepal Review, 2016(118), 115–137. https://doi.org/10.18356/94FD7D0B-EN

[10] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27–34. https://doi.org/10.1145/240455.240464

[11] Brodley, C., Lane, T., & Stough, T. (1999). Knowledge Discovery and Data Mining. American Scientist, 87(1), 54. https://doi.org/10.1511/1999.16.807

[12] Guevara, H., Dominguez, A., Ortunio, M., Padrón, D., & Cardozo, R. (2010). Percepción de la calidad de vida desde los pricipios de la complejidad. Revista Cubana de Salud Pública, 36(4), 357–364. Retrieved from https://www.redalyc.org/articulo.oa?id=21416138011&idp=1&cid=1753836

[13] Alkire, S. (2016). The Capability Approach and Well-Being Measurement for Public Policy. In M. D. Adler & M. Fleurbaey (Eds.), The Oxford Handbook of Well-Being and Public Policy (p. 32). https://doi.org/10.1093/oxfordhb/9780199325818.013.18

[14] DANE. (2019a). Boletín Técnico Pobreza Monetaria en Colombia Año 2018. Retrieved from https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2018/bt_pobreza_monetaria_18.pdf

[15] DANE. (2019b). Boletín Técnico Pobreza Multidimensional en Colombia Año 2018. Retrieved from https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2018/bt_pobreza_multidimensional_18.pdf

[16] Álvarez Rivadulla, M. J. (2018). ¿Por qué preocuparnos por la desigualdad en Colombia? Retrieved November 7, 2019, from Universidad de los Andes website: https://uniandes.edu.co/es/noticias/desarrollo-regional/por-que-preocuparnos-por-la-desigualdad-en-colombia

[17] Farhangfar, A., Kurgan, L., & Dy, J. (2008). Impact of imputation of missing values on classification error for discrete data. Pattern Recognition, 41(12), 3692–3705. https://doi.org/10.1016/j.patcog.2008.05.019

[18] Wallace, B. C., Small, K., Brodley, C. E., & Trikalinos, T. A. (2011). Class Imbalance, Redux. 2011 IEEE 11th International Conference on Data Mining, 754–763. https://doi.org/10.1109/ICDM.2011.33

[19] Seng, J.-L., & Chen, T. C. (2010). An analytic approach to select data mining for business decision. Expert Systems with Applications, 37(12), 8042–8057. https://doi.org/10.1016/j.eswa.2010.05.083

[20] Brodley, C., Lane, T., & Stough, T. (1999). Knowledge Discovery and Data Mining. American Scientist, 87(1), 54. https://doi.org/10.1511/1999.16.807

[21] Hilario, M., Kalousis, A., Nguyen, P., & Woznica, A. (2009). A data mining ontology for algorithm selection and meta-mining. In V. Podpečan, N. Lavrač, J. de Bruin, & J. Kok (Eds.), ECML PKDD 2009 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Third Generation Data Mining: Towards Service-Oriented Knowledge Discovery, SoKD 2009 (pp. 76–87). Bled.