ISSN 2395-8618

Evolutive Algorithms with Restrictions Based on Decision Trees Models for the Enhancement of Satisfaction in Time Use

Mario E. Marin, Julio C. Ponce, Ángel E. Muñoz, and Alejandro Padilla

***Abstract*—A hybrid intelligent system is presented in which the parameters of trained decision trees-based models can be used to define a search space that another intelligent algorithm can utilize to optimize an objective function. This system is of value in social sciences research and industry applications dealing with datasets with categorical attributes and non-linear and complex interactions, interested in the question of what changes in the realities represented by the datapoints would bring them to a more desirable class or regression value according to a decision maker or policy. The approach has the advantage of high interpretability compared to other black box type intelligent algorithms. A case study is presented in which a dataset with 30 attributes is used to explore the less costly changes in time-use assignations to improve satisfaction in time use in academic activities.**

***Index Terms*—Interpretable machine learning, decision trees, random forest, genetic algorithms, time use, time management, satisfaction.**

I. Introduction

Machine learning (ML) models based on decision trees are one of the most interpretable ML methods widely available. Still, they are not known for their high accuracy compared with other commonly used ML techniques [1]. However, their use has grown in recent years, in part because there are efforts underway to develop ways of constructing decision trees-based models as interpretable surrogate models to explain deep learning ones [2]; but also because they happen to be a high accuracy option for certain kinds of datasets, such as those produced by measurement instruments designed to explore concepts attributable to people's decisions, and adapt well to model the non-linear and complex interactions usually found in these datasets [3,4] even when they contain categorical attributes [5]. These advantages make them a handy option to train and use ML models in social sciences and in industry applications built on datasets from surveys with good statistical design in which these models can have both high accuracy and high interpretability.

This paper presents an application of a hybrid intelligent system (HIS) to inquire into what changes people who do not evaluate themselves as satisfied with their time use in academic activities can make to their weekly routines to achieve such a state of satisfaction.

This satisfaction is an essential subjective wellbeing metric correlated to other key wellbeing components [6]. The application uses a HIS built with decision trees-based models combined with evolutive algorithms.

Still, combinations with other optimization and metaheuristic techniques can be used for other, more complex datasets [7]. Besides their interpretability, single decision tree models and random forests were chosen for this application because they have been found to be highly accurate for analyzing the datasets used [8]. It was found that combining them with genetic algorithms offers ways of presenting suggestions to a user or decision-maker to potentially change how time is used to achieve satisfaction in time use in academic activities.

II. Interpretability and Accuracy of Decision Trees-Based Models

A literature review finds many hybrid approaches with decision trees and genetic algorithms, such as combining them to enhance the first ones with better accuracies [7, 9], finding classification rules applicable to small groups of datapoints in a big dataset [10], enhancing the construction of the decision trees [11] or its pruning [12], and so on. However, the application presented in this paper is not concerned with the training, construction, or pruning process but with already trained accurate decision trees whose parameters are used to build constraints for an intelligent search algorithm and describe how the classification or regression process is done.

Interpretability in ML models has clear advantages, such as allowing researchers and analysts to directly explore how a trained model reaches a conclusion and how different or similar datapoints reach similar or different results. These advantages can serve to find research questions or to produce useful information beyond what a highly accurate black box model can offer and has already been done in academic settings [13,14]. Crucially, this interpretability can also serve to mathematically inquire into the ways in which the realities represented by the datapoints need to be changed to achieve a more desirable outcome as modeled by the trained decision tree. Questions such as finding the minimum change, or the less costly, required to bring about a datapoint to be classified in a more desirable class or produce a more desirable regression value can be answered by deploying a HIS in which the trained decision tree parameters are used to define the search space in which an intelligent algorithm looks for answers.

It is essential to mention that it is the accuracy of the ML model that backs the inferences made from it, as well as the hypotheses that could be put forward. For instance, if the leaf nodes of a classification decision tree have high impurities, it would be more challenging to describe the classification

Manuscript received 03/02/2021, accepted for publication on 20/06/2021.

The autors are with the Universidad Autónoma de Aguascalientes, Centro de Ciencias Básicas, Mexico (marin@ia.org.mx, {julio.ponce,angel.muñoz,alejandro.padilla}@edu.uaa.mx).

ISSN 2395-8618

process usefully or to construct solid hypotheses from the information provided by the trained decision tree. Therefore, while an interpretable ML model with low accuracy can undoubtedly be described in its low performance, it has a low capacity to help explain the realities represented by the datasets in which it is trained or the behavior of a black box ML model it could be used to explain as a surrogate model. A useful, highly interpretable model needs to be accurate to help better describe reality and not only how the model works.

It can be argued that models based on ensembles of decision trees, such as random forests, can also be thought of as interpretable ML. A random forest model uses sets of decision trees in which specific attributes have been kept from some trees and then delivers as the result of the whole ensemble the mode or average of the individual decision trees results [3,15]. These incomplete trees can also help explain how they reach the same conclusion as the whole ensemble or not. If they do, their individual accuracies will back the inferences made from them. Various techniques can be used to select a representative tree from an ensemble [16], or it can be selected because of its usefulness as judged by an analyst or a decision-maker; in both cases, the selection should be in accordance with the ensemble's result that usually has better accuracy than single decision trees [3,15].

As mentioned, while there are various methods to train a decision tree or an ensemble of them, the focus of this paper will be on trained decision trees, leaving the training methods to the preference or needs of the analyst. In the presented examples, Python 3.9 is used with the scikit-learn 1.1.2 implementation of the models.

III. Decision Tree Branches as Search Space, a General Model

Model interpretability, while useful to describe how the realities represented by the training dataset behave, has another valuable use: the model's parameters can be used to further model how minimally different datapoints would obtain more desirable results in classification or regression. This use case has important applications in fields in which datapoints represent realities subject to modification. A general model for an optimization problem involving a single decision tree would look like this:

Parameters:

$a_1, \dots, a_n$	Attributes of a n-dimensional instance (datapoint).
$b_1, \dots, b_m$	Parameters of the trained decision tree with m nodes.
$k_1, \dots, k_n$	Cost of change for each of the n attributes.

Variables:

$x_1, \dots, x_n$	Change in each of the n attributes.

Objective function:

$$\min_{x_1,\dots,x_n} f(x_1, \dots, x_n) = \sum_{i=1}^{n} k_i |x_i|.$$

Constraints:

$$[(x_1 + a_1 > b_1) \cap (x_2 + a_2 > b_2) \cap \dots (x_s + a_s > b_s)] \cup ,$$

$$[(x_1 + a_1 < b_1) \cap (x_3 + a_3 > b_3) \cap \dots (x_t + a_t > b_t)] \cup .$$

As can be seen, the constraints would be the union of all the intersections of the inequations with operators that lead to forming paths to a desirable outcome in leaf nodes s, t, and so on; therefore, the operators used are not necessarily the original operators of the trained decision tree. The problem can also be seen as several traditional optimization problems, each with a set of straightforward constraints representing a single path down the tree towards the desired result in a leaf node, for example, node s:

$$\begin{cases} x_1 + a_1 > b_1 \\ x_2 + a_2 > b_2 \\ \quad\vdots \\ x_s + a_s > b_s \end{cases} .$$

Also, the problem can be taken as a global one in which all the paths are considered at once, and the whole trained decision tree $\mathbb{T}$ can be used as a constraint in which d delimiters the set of desired outcomes or is the desired outcome:

$$\mathbb{T}(x_1, \dots, x_n) > d \quad \text{or} \quad \mathbb{T}(x_1, \dots, x_n) < d \quad \text{or} \quad \mathbb{T}(x_1, \dots, x_n) = d.$$

Furthermore, the constraints could be relaxed by trimming parts of the intersections at the level that the impurity or error of the nodes is considered tolerable to the decision-makers or by eliminating specific paths down the decision tree forestalling them. These judgments can be made iteratively, or an optimization hybrid metamodel can be designed to consider multiple options of constraint relaxation and path elimination.

A simple example of this HIS approach can be obtained using a trained classification decision tree with the very known Iris dataset [17], as seen in Fig. 1.

If, by some process of domestication, the plants were modified in petal width and lengths so that they all look like the original Iris Virginica, the objective function in which x_1 stands for petal width and x_2 for petal length, would look like this:

$$\min_{x_1,x_2} f(x_1, x_2) = k_1|x_1| + k_2|x_2|.$$

Apart from feasibility constraints, the main constraints of the model would use the inequality operators that lead to the desired outcome as in:

$$(x_1 + a_1 > 0.8) \cap (x_2 + a_2 > 4.95).$$

However, if, according to a decision-maker, the domesticated plants must be like the original Iris Virginica only in petal width, then the main constrain would consider the branch with the Iris Virginia node leaf only down to Node C in the diagram in Fig. 1, even as that node has a high impurity as far as the original decision tree model goes. The relaxed constraint would look thus: $(x_1 + a_1 > 0.8)$.

ISSN 2395-8618

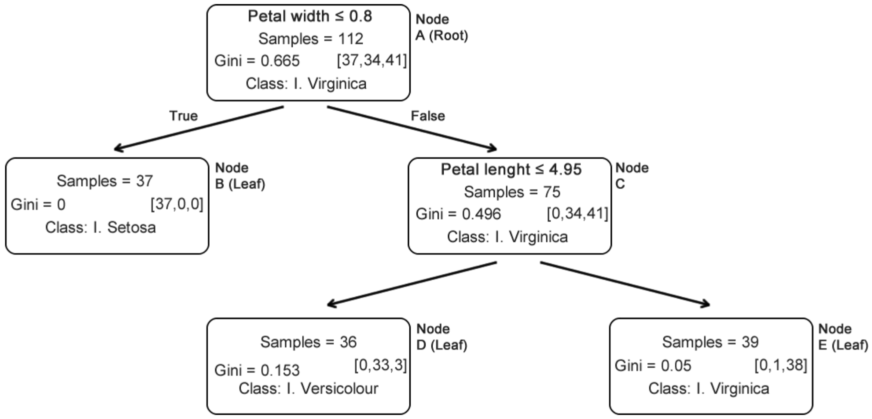

Fig. 1. A decision tree trained with attributes belonging to the Iris dataset

IV. The Problem: Satisfaction in Time Use in Academic Activities

A Mexican university is interested in having its diverse student body practice time management because these practices and the perception of being in control of their time may help their academic performance [18,19]. Preserving or enhancing satisfaction in time use in academic activities is vital in this endeavor [6]. Rather than having one-on-one mentoring, the university wants to train and test a HIS with interpretable ML that can suggest to students how they may change their weekly routines to reach a status of satisfaction on time use in their academic activities by answering a brief survey.

There are massive public datasets that contain time use assignations and sometimes satisfaction levels: time use surveys microdata. In this case, appropriate publicly available datasets with proven usefulness for the application of intelligent classification algorithms with high performance [8] would be microdata from the National Time Use Survey (ENUT) done by INEGI in both 2014 (with 42118 datapoints) and 2019 (with 71404 datapoints) in all 32 federal entities in Mexico.

The ENUT contains hundreds of weekly time use attributes, some select demographic and socioeconomic ones, and a set of attributes that indicate the presence or absence of satisfaction in time use in academic activities [20]; thus, the variable of interest would be binary. The classification problem would have two classes: a negative one for the absence of time use satisfaction in academic activities and a positive one for its presence.

Preprocessing was done to prepare the datasets for analysis and reduce their number of attributes without losing relevant information. Metadata attributes were dropped. Non-time use attributes were transformed into either 0-1 binary attributes or 1-5 integer ordinal attributes. Categorical attributes were transformed into sets of as many 0-1 binary attributes as the number of categories in the original attribute.

Time use data attributes were reduced from five attributes linked to an activity to just two attributes with integer values representing the minutes devoted to the activity during weekdays and weekends. Attributes with null values were either transformed into zeros or statistically appropriate values. Attributes that contained information that could be derived from other attributes were dropped.

As has been mentioned, the model's accuracy is paramount to the validity of its conclusions. Accordingly, a performance review was run using a 5-fold cross-validation regime for two samples from different years of the ENUT using five different ML models, two of which are decision trees-based. This review was implemented in Python 3.9 using the scikit-learn 1.1.2 implementation of the classifiers with default parameters except for the following ones:

- **Random forest (RF):** 1500 trees, a maximum of 100 levels; Gini criterion.
- **Support vector machines (SVM):** Radial basis function (RBF) kernel, with a decision function one vs. one shaped.
- **Deep artificial neural network (NN):** Feedforward architecture with 1000 hidden layers with as many artificial neurons in their input and hidden layers as attributes in the preprocessed dataset and a one-element output layer. An L_2 penalty of 10^{-5}, a rectified linear unit function (RELU), and an adam stochastic gradient-based optimizer shuffling samples in each iteration were used.
- **Logistic regression (LR):** Multinomial, with a Broyden–Fletcher–Goldfarb–Shanno (LBFGS) optimization algorithm and an L_2 penalty term.
- **Decision trees:** Default optimized version of the Classification and Regression Tree (CART) algorithm.

The results of this performance review are shown in Table 1. As shown in Table 1, deep neural networks perform better than a single decision tree model for one of the samples; random forest models have the highest performance in both samples. In both cases, a balanced training dataset was used; therefore, a zero-rule classifier would have a 0.5 accuracy. An accuracy higher than 0.9 can be reasonably judged adequate by the decision-makers. Thus, single decision trees ML models can be used for this application.

The results in Table 1 suggest that a decision tree trained with either of the surveys' microdata will have enough accuracy to both explain to students how their time use patterns lead to satisfaction or dissatisfaction in time use in academic activities and to power an optimization model to offer suggestions to students that the model finds currently dissatisfied with their time use in such settings. However, a practical problem arises. The original ENUT surveys contain hundreds of questions, and their use as-is is impractical.

ISSN 2395-8618

TABLE I
PERFORMANCE REVIEW OF ML MODELS TRAINED WITH BALANCED DATASETS FROM BOTH SAMPLES OF THE ENUT MICRODATA

Sample	Classifier	*Accuracy*	*Precision*	*F1 Score*	C. Time [a]
2014's sample (440 attributes)	Random Forest	**0.9495**	**0.9521**	**0.9494**	18.05s
	Decision Tree	0.9232	0.9232	0.9232	**0.91s**
	SVM	0.4977	0.4973	0.4874	46.57s
	Deep Neural Network	0.8634	0.8758	0.8612	61.45s
	Logistic regression	0.8952	0.8981	0.8950	18.78s
2019's sample (432 attributes)	Random Forest	**0.9583**	**0.9607**	**0.9583**	32.66s
	Decision Tree	0.9241	0.9243	0.9241	1.64s
	SVM	0.9128	0.9136	0.9128	37.65s
	Deep Neural Network	0.9376	0.9378	0.9375	171.18s
	Logistic regression	0.9177	0.9182	0.9177	**0.94s**

[a] *As run in a laptop computer with an Intel Core i5-9300H CPU @ 2.40GHz*

TABLE II
SUBSET OF ATTRIBUTES FROM THE ENUT WITH HIGH ACCURACY USING DECISION TREE-BASED ML MODELS. IN ITALICS, THOSE THAT ARE NOT TIME USE RELATED [21]

Attribute	
Number of rooms in the house you live in	Commuting to and from school (Weekdays)
Number of people living in your house	Cooking and preparing drinks (Weekdays)
Age	Serving meals, washing dishes (Weekdays)
Civil status (Married/Unmarried)	Cleaning house interior (Weekdays)
Exclusive dedication to studying (Yes/No)	Cleaning house interior (Weekends)
Goes to school (people aged 5 to 24 years)	Folding and storing clothes (Weekdays)
Work (Weekdays)	Securing doors/windows (Weekdays)
Sleep (including naps) (Weekdays)	Workouts or sports (Weekdays)
Sleep (including naps) (Weekends)	Watching TV (Weekdays)
Eating meals (Weekdays)	E-mail, social networks, chat (Weekdays)
Eating meals (Weekends)	E-mail, social networks, chat (Weekends)
Grooming, personal hygiene (Weekdays)	Consult or read on the web (Weekdays)
Grooming, personal hygiene (Weekends)	Classes (Weekends)
Classes (Weekdays)	Homework or study at home (Weekdays)
Homework or study at home (Weekdays)	Commuting to and from school (Weekends)

However, a briefer instrument has already been validated with a 0.9639 accuracy for a random forest model [21] and 0.9034 accuracy for a single decision tree model that only requires 30 questions related to the attributes in Table 2, of which 24 are time use attributes for which there can be potentially actionable suggestions.

V. SPECIFIC MODEL FOR TIME USE SATISFACTION IN ACADEMIC ACTIVITIES

A specific model for an optimization problem testing the decision tree, trained with the 2019's ENUT time use attributes in Table 2, to get users from unsatisfied to satisfied with time use in academic activities with minimum change to weekly routines is:

Parameters:

$a_1, \ldots, a_{24}$	Time use attributes of the instance (users, or respondents).
$b_1, \ldots, b_{1951}$	Parameters of the decision tree's 1951 decision nodes.
$k_i = 1, \forall\, i \in [1,24]$	Cost of change for each of the 24 attributes is unitary.

Variables:

$x_1, \ldots, x_{24} \in [-60,60]$	Change in each of the n time use attributes (minutes).

Objective function:

$$\min_{x_1,\ldots,x_{24}} f(x_1, \ldots, x_{24}) = \sum_{i=1}^{24} |x_i|.$$

Constraints:

$$[(x_1 + a_1 > b_1) \cap (x_2 + a_2 > b_2) \cap \ldots (x_s + a_s > b_s)] \cup,$$

$$[(x_1 + a_1 > b_1) \cap (x_3 + a_3 > b_3) \cap \ldots (x_t + a_t > b_t)] \cup,$$

$$\text{Or: } \mathbb{T}(x_1, \ldots, x_{24}) > 0,$$

where $\mathbb{T}$ is the complete trained decision tree, this constraint is equivalent to the whole set of intersections for the unrelaxed constraints created with the parameters of the decision tree.

ISSN 2395-8618

TABLE III

SUGGESTIONS FOR CHANGE IN TIME USE FOR A RANDOMLY SELECTED SET OF RESPONDENTS TO THE 2014 ENUT SURVEY WHO TAKE CLASSES AND HAVE NO SATISFACTION WITH TIME USE IN ACADEMIC ACTIVITIES. SUGGESTIONS ARE FOR WEEKDAYS ONLY UNLESS OTHERWISE NOTED

Change recommended	**Freq.**	**Change recommended**	**Freq.**
Less cooking or preparing drinks	326	More commuting to school	1
More cleaning house's interior [a]	75	More cleaning house's interior	1
Less watching TV or video streaming	2	Less folding clothes	1
Less sleeping	1	Less sports and workouts	1
More grooming, hygiene, bathroom	1	More email, chat, social networks [b]	1

[a] Activity only during weekends [b] Activity during both weekdays and weekends

An output value of 1 for the decision tree means the presence of satisfaction in time use in academic activities, while a value of 0 means its absence.

As can be inferred from the above model, instances with 24 attributes can have a vast number of combinations of changes considering that time attributes can be changed to go up or down by different amounts. Also, with 1951 nodes, hundreds of intersections represent possible ways from the root of the decision tree to a leaf node that concludes with the instance being classified as with satisfaction with time use in academic activities. This complexity calls for a metaheuristic method, with evolutionary approaches being suitable. In this case, genetic algorithms were used.

For this problem, a genetic algorithm was used with a population of 200, probabilities of 0.8 for recombination and 0.02 for mutation, and 200 generations. The selection was random, and by contest. Recombination was done erasing the values of an instance's vector from a randomly selected point and appending to the severed vector the missing part from another randomly selected instance. The mutation was done by assigning a new random value to a randomly selected variable in the instance selected for mutation. All random processes were done using uniformly distributed random variables. The initial population comprises instances with values randomly chosen from a range of minus 60 to 60 weekly minutes, just as other random values assigned.

VI. RESULTS AND DISCUSSION

For 400 randomly selected ENUT respondents with no satisfaction in time use in academic activities from the 2014 ENUT sample, with the decision tree trained with the 2019 ENUT sample, in 383 cases the only change needed to achieve satisfaction in time use in academic activities is to add time assigned to taking classes. For the other 17 cases, it was only necessary to add time to doing homework, practice, or study at home; the time needed for the changes is just 1 or 2 weekly minutes, respectively. These small changes indicate that the decision tree is trained very tightly around the time use patterns, though as it is being tested with a sample from another year, overfitting is less of an issue. Given these results, the simplest option is to suggest to the user to increase the time used for these activities by a reasonable amount.

From the 400 tested instances, only 9 were respondents taking classes, for an average of 1773.33 weekly minutes, and these were part of the 383 cases in which more classes were suggested.

All 17 cases in which more time to do homework, practice, or study at home were recommended for respondents who did not take a single weekly minute of classes. So, a second run was made with a sample of 400 randomly selected respondents from the 2014 ENUT sample with no satisfaction in time use in academic activities and a non-zero value in minutes assigned to taking classes, a subset of respondents which would be of more interest to the decision-makers at the university.

As was the case with the first run, the total changes of time use assignations suggested are for a small number of weekly minutes, from 5 to 7, in practice merely indicating that the time used for the activity must be increased or decreased. Still, the suggestions for this subset have more variety. The most frequent suggestions are shown in Table 3.

As seen in Table 3, most of the suggested changes in time use assignations are unrelated to academic activities, excluding commuting to school. Except for suggestions for less time used in cooking or preparing drinks and more time used in cleaning the interior of one's house, all other suggestions seem to be for specific cases. While the HIS recommends more than one change to a few respondents, satisfaction with time use in academic activities is just a single change away for most users.

The main themes are that students' satisfaction with time use in academic activities would benefit the most from using less time preparing food and having more time to clean their houses. Why is that so, and how these two non-academic activities interact with academic ones can be a subject of further research: is it because excessive academic activities do not allow time to clean even on weekends and the degraded domestic environment reflects poorly in the academic activities time use evaluation? Is it that preparing food is time-consuming and reduces time to be satisfactorily spent studying?

Even with 30 attributes, the tested trained decision tree has 1951 nodes. Still, as seen in Fig. 2(a) and Fig. 2(b), it can be gleaned that the leftmost part of the decision tree is dominated by unsatisfied respondents who do not take classes and that exceptionally become satisfied. On the other hand, the rightmost parts of the decision tree are dominated by satisfied respondents who take classes, where a small portion of respondents end up unsatisfied, as seen in Fig. 2(c).

The dissatisfaction of those who do not take classes continues persists, particularly if they also do not do some studying at home, which was also relevant in the first run of the genetic algorithm. However, for people who attend classes, the vast majority are satisfied with their time assignations to academic activities, particularly those who sleep more than a bare minimum of 35 hours weekly, as seen in Fig. 2(c).

ISSN 2395-8618

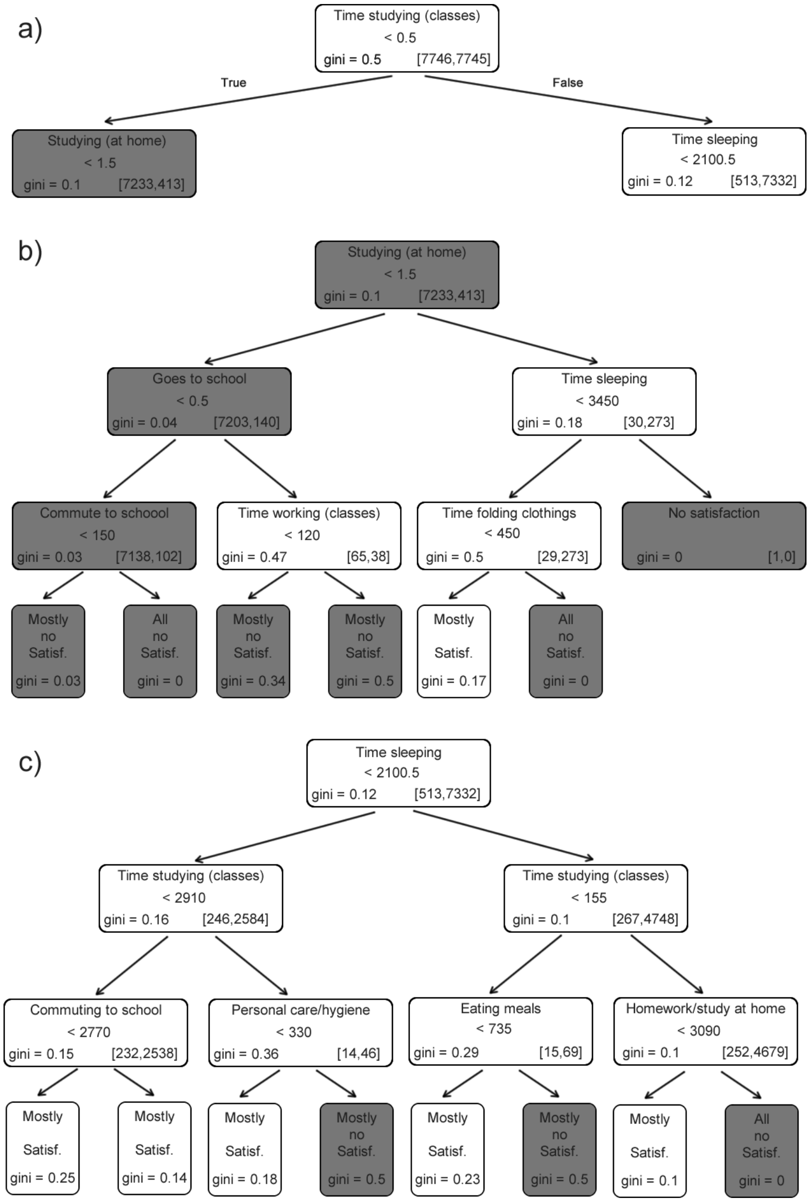

Fig. 2. Diagram of the root and upper part (a), leftmost part (b), and rightmost part (c) of the decision tree trained with the 2019's ENUT. Gray nodes are dominated by unsatisfaction

For those who do take classes and sleep more than 35 hours weekly, dissatisfaction can come in the form of unbalanced assignations to activities such as cooking, cleaning, commuting, meals, personal hygiene, and a myriad of other activities finetuned to more particular cases in the lower regions of the tree not shown in Fig. 2. For each of these cases, further hypotheses can be explored.

Fortunately for the Mexican university of this case study, the results of the HIS tested in this paper show that most people that take at least one class will be satisfied with their time use on academic activities.

At the same time, dissatisfaction in this area runs rampant for those who study in other ways or do not study at all, even if they attend school for other activities.

On the other hand, the results supply evidence that the suggestions to bring unsatisfied students to satisfaction on time use on academic activities could be few and impacting activities mostly outside the campus.

VII. Conclusions and Future Work

Interpretable ML models, and hybrid approaches to them, are not only for data mining practitioners. In many fields, particularly in the social sciences, these ML models may have the accuracy to become a tool of research for both exploratory and confirmatory studies. These models can also be purposeful tools to be used in industry and organizational settings, provided they are trained with datasets obtained from high-quality measurement instruments and can be validated with datasets from different samples.

The HIS presented in this work combining genetic algorithms and decision tree classifiers for the absence or presence of satisfaction in time use in academic activities, can

ISSN 2395-8618

be used to offer actionable suggestions as part of an intelligent time management system that considers the satisfaction of its users as a key criterion. This HIS can even provide support for organizational or public policy changes to facilitate time use changes leading to satisfaction in time use, which is linked to other wellbeing components [6]; this is several steps beyond simply teaching or using time management.

For the application presented in this paper, its results can lead to future research as cases can be made for the hypotheses that the university's students could be well served by policies such as providing affordable meal plans to release students from using too much of their time preparing food, and by avoiding excessive time assigned for academic activities themselves to leave time for other basic activities that people can consider important such as cleaning one's domestic environment.

REFERENCES

[1]. R. Caruana and A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms," *Proceedings of the 23rd International Conference on Machine Learning*, Association for Computing Machinery, pp. 161–168, 2006. DOI:10.1145/1143844.1143865.

[2]. A. Blanco-Justicia and J. Domingo-Ferrer, "Machine Learning Explainability Through Comprehensible Decision Trees," *Machine Learning and Knowledge Extraction*, pp. 15–26. Springer International Publishing, Cham, 2019. DOI:10.1007/978-3-030-29726-8_2.

[3]. A. Joshi, "Decision Trees. In: Machine Learning and Artificial Intelligence," pp. 53–63, *Springer International Publishing*, Cham 2020. DOI:10.1007/978-3-030-26622-6_6.

[4]. A. Dobra, "Decision Tree Classification," in Liu Ling and Özsu, M.T. (*eds.) Encyclopedia of Database Systems.* Springer US, pp. 765–769, 2009. DOI:10.1007/978-0-387-39940-9_554.

[5]. G. Origel-Rivas, E. Rendón-Lara, I. M. Abundez-Barrera and R. Alejo-Eleuterio, "Redes neuronales artificiales y árboles de decisión para la clasificación con datos categóricos," *Research in Computing Science*, vol. 149, no. 8, pp. 541–554, 2020.

[6]. I. Boniwell, "Satisfaction with time use and its relationship with subjective well-being," 2006. DOI:10.21954/ou.ro.0000e971.

[7]. J. Monalisa, R. K. Behera and D. Satchidananda, "Hybrid Decision Tree for Machine Learning: A Big Data Perspective," Dehuri Satchidananda and Chen, Y.-W. (ed.) *Advances in Machine Learning for Big Data Analysis*, Springer Nature Singapore, Singapore, pp. 223–239, 2022. DOI:10.1007/978-981-16-8930-7_9.

[8]. M. E. Marin and J. C. Ponce, "Intelligent Time Use Suggestions for Wellbeing Enhancement," *Research in Computer Science*, vol. 149, no. 10, pp. 101–114, 2020.

[9]. Z. Fu, B. L. Golden, S. Lele, S. Raghavan and E. A. Wasil, "A genetic algorithm-based approach for building accurate decision trees". *Informs Journal on Computing*, vol. 15, no. 1, pp. 3–22, 2003. DOI:10.1287/ijoc.15.1.3.15152.

[10]. D. R. Carvalho and A. A. Freitas, "A Hybrid Decision Tree/Genetic Algorithm Method for Data Mining", *Information Sciences*, vol. 163, no. 1–3, pp. 13–35, 2004. DOI:10.1016/j.ins.2003.03.013.

[11]. S.-H. Cha and C. Tappert, "A Genetic Algorithm for Constructing Compact Binary Decision Trees," *Journal of Pattern Recognition Research*, vol. 4, no. 1, pp. 1–13, 2009. DOI:10.13176/11.44.

[12]. J. Chen, X. Wang and J. Zhai, "Pruning decision tree using genetic algorithms," *International Conference on Artificial Intelligence and Computational Intelligence*, pp. 244–248, 2009. DOI:10.1109/AICI.2009.351.

[13]. E. H. Thomas and N. Galambos, "What Satisfies Students? Mining Student-Opinion Data with Regression and Decision Tree Analysis," *Research in Higher Education*, vol. 45, pp. 251–259, 2004. DOI:10.1023/B:RIHE.0000019589.79439.6e.

[14]. M. del Pilar, M. Bucio and G. Gutiérrez-Carreón, "Decision Tree-Based Model to Determine Student's Dropout Factors in a Mexican Higher Education Institution," *Research in Computer Science*, vol. 151, no. 1, pp. 35–41, 2022.

[15]. V. K. Ayyadevara, "Random Forest," in *Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R*, Apress, Berkeley, pp. 105–116, 2018.

[16]. A. I. Weinberg and M. Last, "Selecting a representative decision tree from an ensemble of decision-tree models for fast big data classification," *Journal Big Data*, vol. 6, no. 23, 2019. DOI: 10.1186/s40537-019-0186-3.

[17]. D. Dua and C. Graff, "*UCI Machine Learning Repository*" 2017.

[18]. T. H. Macan, C. Shahani, R. L. Dipboye and A. P. Phillips, "College Students" Time Management: Correlations with Academic Performance and Stress," *Journal of Educational Psychology*, vol. 82, no. 4, pp. 760–768, 1990. DOI:10.1037/0022-0663.82.4.760.

[19]. B. Aeon and H. Aguinis, "It's About Time: New Perspectives and Insights on Time Management," *Academy of Management Perspectives*, vol. 31, no. 4, pp. 309–330, 2017. DOI:10.5465/amp.2016.0166.

[20]. INEGI, "*Encuesta Nacional sobre Uso del Tiempo 2019, Diseño Conceptual*", Instituto Nacional de Estadística y Geografía, Aguascalientes, 2020.

[21]. M. E. Marin, J. C. Ponce, Á. E. Muñoz and A. Padilla, "Peligros y ventajas del diseño de instrumentos de medición auxiliado con estrategias de selección de atributos automáticas", *Research in Computing Science*, vol. 150, no. 5, 2021.