

A Framework to Predict Fine-Grained Sustainable Consumption Behavior Levels of Individuals

Short Paper

Arielle Moro

Information Management Institute
University of Neuchâtel
A.L. Breguet 2, 2000 Neuchâtel
Switzerland
arielle.moro@unine.ch

Adrian Holzer

Information Management Institute
University of Neuchâtel
A.L. Breguet 2, 2000 Neuchâtel
Switzerland
adrian.holzer@unine.ch

Abstract

In order to encourage individuals to adopt more sustainable behaviors, it is crucial to know their current levels of consumption in specific domains (e.g., mobility) before exposing them to personalized incentives. Although various theoretical models exist, there is currently no technological solution that automatically estimates individual's consumption sustainability levels. This short paper aims at addressing this gap and presents the design of a framework that enables to estimate these levels based on multiple features (e.g., demographics). It also presents a preliminary validation of a part of the framework through two empirical comparative studies related to the mobility consumption domain. These studies evaluate the performance of six classifiers using a large-scale survey of approximately 3000 representative individuals living in Switzerland. The results highlight that the gradient boosting trees and the multinomial logistic regression models are promising, and accommodation, habits and demographic variables are the most decisive features to estimate mobility behaviors.

Keywords: Green IS, sustainable consumption behavior, data analytics

Introduction

In the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) published in 2014, experts unquestionably confirmed the impact of Human activities on the evolution of the climate. They also highlighted that it is still possible to mitigate the impact of global warming if we are able to reduce the greenhouse gas emissions in the near future. Consequently, individuals must reduce their carbon footprint by adopting more sustainable consumption behaviors. However, moving towards a more sustainable behavior is not obvious and there might be no one-size-fits-all solution for all individuals. More specifically, individuals do not have the same sustainability level for all consumption domains (e.g., mobility, housing, food, clothing) and thus might need to adjust their efforts on particular domains. For example, an individual, who buys all her food in bulk stores to reduce her waste but uses a private car every day, has a higher food consumption sustainability level compared to the one related to mobility.

Consequently, she needs appropriate positive incentives in order to reduce her carbon footprint in the mobility domain.

However, even though several theoretical models aim at understanding and describing sustainable consumption behavior of individuals (Kollmuss and Agyeman 2002; Geiger and al. 2018), there is no technological solution to understand and predict the levels of sustainability of an individual in a specific consumption domain. Such technological solutions could help to personalize digital interventions, such as recommender systems or gamification features that would aim at raising awareness, offering green alternatives and innovative product experiences. The two key research questions of our research work are the following: Can we design a framework that enables to produce an estimate view of the sustainable consumption behavior of an individual? Can we predict the sustainability level of a specific behavior consumption of an individual (in the mobility context for example)?

In this short paper, in order to address this issue, we make the following contributions: first we introduce a novel framework that enables to predict the different sustainability levels of the consumption behavior of an individual. Second, as a first proof-of-concept, we present an instantiation of one part of the framework with two comparative analyses of six machine learning models. These analyses focus on consumption aspects of one of the most important sources of pollution, namely mobility, which accounts for almost 25% of greenhouse gas emissions in the EU (see the following report https://ec.europa.eu/clima/policies/transport_en). In Switzerland, transport represents the highest sector of emissions, i.e., 36% of the total energy consumption (<https://www.sccer-mobility.ch/aboutus/nutshell/>). Furthermore, energy consumption is mainly related to fuel for private cars and travels by plane for the mobility sector in Switzerland (see the following report: <https://www.bafu.admin.ch/bafu/en/home/documentation/reports/environmental-report-2018.html>). We also highlight the features that have the most influence on the chosen predictions. Third, we use a large-scale dataset called Swiss Household Energy Demand Survey (SHEDS) (www.sccer-crest.ch/research/swiss-household-energy-demand-survey-sheds) and collected by the Competence Center for Research in Energy, Society and Transition (SCCER CREST). It contains 15'000 survey responses of representative individuals living in Switzerland over three years. In terms of use, such a framework could be integrated into Green recommendation platforms in order to make individuals aware of their current consumption behavior and then to help them moving towards more sustainable behavior per domain. In addition, our framework could help urban planners to design smart cities and therefore to help citizens to move towards sustainable behaviors.

The short paper is organized as follows. Firstly, we detail the literature related to the main research subjects of this research work. Secondly, we present an overview of our framework as well as the methodology used to implement and evaluate it. Thirdly, we describe the preliminary analysis before we discuss its findings. Finally, we introduce the future goals and the expected contributions of this research.

Literature Review

The domains of this research work are related to both the field of modeling sustainable behavior and the field of predicting behavior. The first field enables to have an overview of existing research works done to model and understand sustainable behaviors, while the second field presents the state of the art of the machine learning models used to predict behaviors in general.

Modeling Sustainable Behavior

In order to better understand individuals' sustainable behaviors, Kollmuss and Agyeman (2002) studied economic, social and cultural factors that have an influence on pro-environmental behavior. Clark et al. (2003) did a similar analysis in the context of a green electricity program. Other authors found that

emotional criteria and individuals' awareness could have a high impact on increasing green behavior (Amel et al. 2009; Dillahunty et al. 2008). Juárez-Nàjera et al. (2010) analyzed sustainable behaviors of individuals in higher institutions with a focus on moral norms and values, in German and Mexican universities. At the city scale of Madrid, Figueroa-García et al. (2018) explored the social factors that enable to determine sustainable consumption behavior of individuals. Regarding mobility usage, Van Acker et al. (2016) studied individuals' lifestyles, how they have an influence on their mobility and, more importantly, how they can be modified in order to move towards more sustainable behaviors. About electricity consumption, Guo et al. (2018) reviewed and assessed existing works linked to residential electricity consumption, from factors to the adoption of sustainable plans. In order to better understand the key elements to model individual's sustainable behavior, Geiger and al. (2018) presented an exhaustive theoretical model, called Sustainable Consumption Behavior (SCB) cube, that takes into account two sustainable dimensions (ecological and socio-economic) and may include multiple consumption areas and various consumption phases. This is the closest model to our framework in terms of axes of consumption. However, the SCB cube is only a theoretical model unlike our framework, which is a technological solution.

Predicting Behavior

Predicting a human behavior can be studied from a dynamic (i.e., studying the evolution of the behavior) or a static (i.e., studying the behavior at a single point in time) point of view. Firstly, from a dynamic point of view, Subrahmanian and Kumar (2017) highlighted that human behaviors are constantly evolving and that the models should learn these changes automatically in order to be always performant. Pentland and Liu (1999) used a Markov chain model to predict behaviors of automobile drivers. Regarding mobility, Kulkarni et al. (2016) studied the behavior of individuals by analyzing their movements over time and the places they visited. Secondly, from a static point of view, Wei et al. (2017) proposed a framework, including several classifiers, to predict user personality based on heterogeneous information such as social media data. Other authors also presented an approach combining several classifiers based on a genetic algorithm in order to predict customers' purchase behavior (Kim et al. 2003). Kim and Yoon (2017) described a model that aims at predicting green advertising attitudes including a regression analysis to study the variables that have an influence on the responses to green advertising. Other behavioral research analyses are also relevant. Several authors used machine learning models and strategies, such as active learning, neural network, logistic regression and k-nearest neighbors to make behavior predictions in a wide variety of domains, from animal behavior to malware behavior and email reply predictions (Kabra et al. 2013; Rieck et al. 2011; Yang et al. 2017). In addition, a last research paper, written by Zhou (2017), conducted a data mining analysis about individual consumer credit default prediction in the context of e-commerce. The author compared two different ensemble approaches (bagging and boosting) and highlighted the importance of social features for this specific type of prediction. In the preliminary analysis of our framework, we use some of the models already presented in the aforementioned research papers, such as ensemble approaches, neural network and regression models.

Framework Description

The main goal of our framework is to allow creators of green information systems (Green IS, i.e., information systems designed to encourage sustainable behavior), to personalize digital intervention. To do so, the framework estimates an individual's (i.e., user) sustainability levels (output) based on her personal data (input). More formally the framework offers a function called *getSustainability(domain, user_data)*. This function requires a domain (e.g., SHORT_MIDDLE_FLIGHT_NB, MOBILITY, FOOD, HOME_WORK_TRANSPORTATION_MODE, ALL) and user information data (see Figure 1).

Internally, the framework estimates the sustainability levels of the consumption behavior of an individual as a form of a tree structure composed of three distinct levels. The low level of this tree corresponds to the predictions of precise acts of consumption of the individual related to specific consumption domains (e.g., commuting mode, number of short/middle distance flights). The intermediary level of the tree contains the sustainability levels of various consumption domain (e.g., mobility, housing, food, clothing). Each of these levels are computed according to an aggregation of the sustainability levels of the low level of the tree related to each of these specific domains. And finally, the highest level of the tree is the global sustainability level of the individual. This last value is also obtained by aggregating all the sustainability levels of the intermediary level of the tree. The prediction tasks related to the low level of the tree are crucial in this framework because they have an impact on the computation of all other upper levels of the tree structure. These prediction tasks correspond to multi-class prediction problems with three possible pre-defined predicted classes (low, medium or high sustainability level) that reflect the polluting impact of one specific act of consumption related to a consumption domain as mentioned previously. These values will be described in detail in the description of the preliminary analysis in the next section. These predictions are based on independent variables related to personal characteristics of an individual (e.g., demographic, socio-economic, psychological variables).

To fine tune the predictions, the framework offers a function called *trainModel(pred_task, training_data)* (see Figure 1). The *pred_task* value indicates the sustainability level that we want to predict between all the sustainability levels that are possible to predict at the low level of the tree (e.g., SHORT_MIDDLE_FLIGHT, HOME_WORK_TRANSPORTATION_MODE), while the *training_data* value corresponds to all the data used to fit the model selected for this specific prediction. The selection of the model is automatically done by the framework itself.

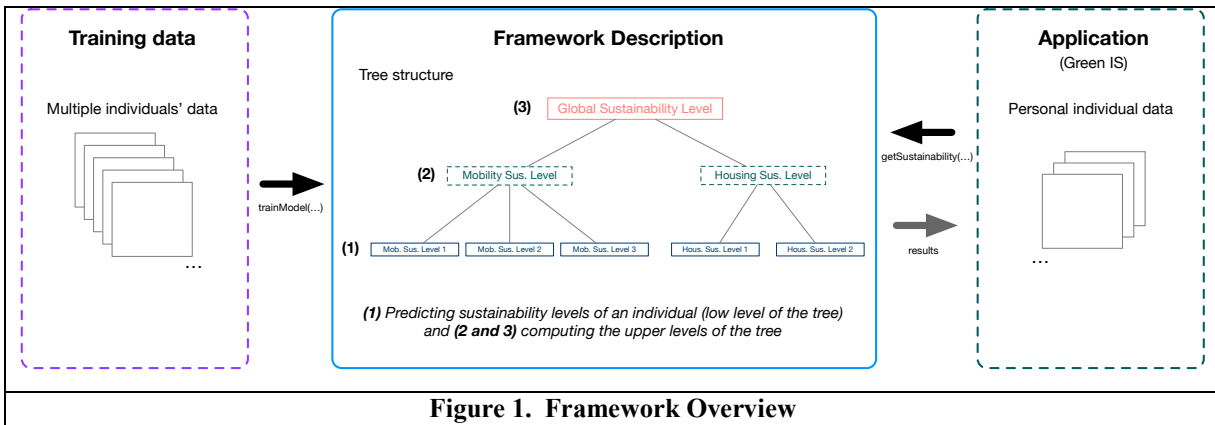


Figure 1. Framework Overview

Global Methodology

Our work follows the design science research methodology fine-tuned for Green information systems (Brendel et al. 2018). In order to carry out the first part of our framework, which consists in predicting the sustainability levels of individuals' consumption behavior, we perform several comparative analyses including different classifiers in order to find the most accurate one per act of consumption. The classifiers were selected according to their popularity of use and also their belonging to different categories (e.g., ensemble approach, deep learning approach) to ensure a maximum of diversity during the comparative analysis. To do so, we follow a data science process: explore and prepare the data, select and explore the machine learning model candidates and fine-tune them and present the findings

(<https://books.google.ch/books?id=I6qkDAEACAAJ>). In addition, we will also identify the user data that has the most impact on the prediction tasks (i.e., low level of the tree). This will possibly allow us to reduce the amount of personal data needed as input in order to compute the targeted predictions. Two of these comparative analyses are the focus of the preliminary analysis we present in this short paper for the mobility consumption domain. After having identified the most accurate and relevant classifiers for all consumption domains (e.g., mobility, housing, food, clothing) we will target an instantiation of the entire framework, we will start creating the tree structure with its low, intermediary and global levels, which is the second part of our framework. In terms of tests, we will evaluate the prediction accuracy of the classifiers during the creation of the first part of the framework and then the accuracy of the intermediary and global levels of the tree of the framework in a second time. For these two evaluations, we will use an empirical dataset that contains a sample of representative individuals living in Switzerland that is fully described in the next section.

Preliminary Analysis and Results

Our preliminary analysis is focused on the mobility consumption domain. In this domain, we explore two prediction tasks: the prediction of sustainability levels related to the *home-work transportation mode* and the other one linked to the *average of short/middle flight number per year*. This preliminary analysis aims at assessing six classifiers, identifying the most accurate classifier and the relevant features that enable to predict the targeted sustainability levels of consumption behavior of an individual. We first detail the dataset, the selected features of this dataset for the prediction tasks and the possible predicted values also extracted from this dataset. Then, we present the classifiers and their selected parameters for the two types of prediction. Finally, we expose and discuss the obtained results.

Dataset and Features

The preliminary empirical analysis we realized in this research paper uses a dataset called Swiss Household Energy Demand Survey (SHEDS) (www.sccer-crest.ch/research/swiss-household-energy-demand-survey-sheds). This dataset has been developed as part of the research agenda of the Competence Center for Research in Energy, Society and Transition (SCCER CREST). The collection period of this dataset is five years from 2016 to 2020 (one survey per year). Every yearly wave of the dataset contains the survey answers of about 5000 representative respondents living in Switzerland. Some respondents could participate in all or several survey waves. It contains respondents' anonymized answers of surveys about Swiss Households' energy consumption related behavior. The surveys are focused on three axes of energy consumption behavior: electricity, housing and mobility (Weber et al. 2017). In the surveys, the respondents also gave additional information related to demographic, socio-economic and psychological questions. For this paper, we only used the most recent, complete and refined survey of 2018, which is focused on mobility. From this dataset, we extracted 119 features belonging to seven features groups as partially detailed below.

- Demographic and socio-economic features: age, gender, household type (single person, couple with/without children...), homeplace zip code, work place zip code, household monthly gross income, etc.;
- Accommodation features: accommodation type (flat in a building with less than 5 flats, ...), size of the living area, distance to the grocery store, home-work distance, etc.;
- Psychological attitudes features: average of the environmentally friendly feeling, average of the feelings related to the environment and potential environmental change, etc.;

- Social performance features: frequency of visiting art exhibitions or galleries, money spent per person for a really good dinner, etc.;
- Habits/routine features: level of importance of having your own car, taking long showers, living close to place of work, etc.;
- Social context features: overall life satisfaction level, being or not an active member of a club or an association, etc.;
- Energy literacy features: response (following a scale) to the following statements: CO₂ emissions play a crucial role in global warming, coal is a renewable energy resource, etc.

Predicted Values

As mentioned in the previous section, we define the predicted values as three possible targeted values: low, medium and high sustainability levels that correspond to high, medium and low polluting impacts respectively (see Table 1 below).

Consumption area	Predicted Value and Distribution	Raw Value
TM - Home-Work Transportation Mode (workers only) (3005 survey respondents)	High Sustainable Consumption Behavior Level (656 survey respondents)	Slow Mobility (bike, foot...) / Working from Home
	Medium Sustainable Consumption Behavior Level (1204 survey respondents)	Public Transportation / Car Sharing
	Low Sustainable Consumption Behavior Level (1145 survey respondents)	Private Car / Motorbike
SMF - Short/Middle Flight Number (average number per year) (3204 survey respondents)	High Sustainable Consumption Behavior Level (1251 survey respondents)	0
	Medium Sustainable Consumption Behavior Level (1357 survey respondents)	1 or 2
	Low Sustainable Consumption Behavior Level (596 survey respondents)	More than 2 (not included)

Preliminary Results

For this preliminary analysis, we perform two comparative studies including state of the art classifiers in order to select the most accurate models. The two studies include the use of six supervised approaches: one basic approach (i.e., decision tree), two ensemble approaches (i.e., random forests, gradient boosting trees), one deep learning approach (Multi-Layer Perceptron, i.e., MLP) and two other approaches (i.e., multinomial logistic regression, support vector machine). Comparing a wide range of classifiers allows us to highlight the most accurate classifier for each prediction task we want to perform.

Classifier	Parameter and Range	Selected Values (TM) / (SMF)
Decision Tree (random_state = False)	max_depth = [2, 10]	3 / 2

Random Forests (max_features = auto, random_state = False)	max_depth = [2, 10]	9 / 8
	n_estimators = (10, 50, 100, 250, 500)	500 / 50
Gradient Boosting Trees (max_features = auto, random_state = False)	max_depth = [2, 10]	3 / 4
	n_estimators = (10, 50, 100, 250, 500)	50 / 250
	learning_rate = (1e-5, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0)	0.1 / 0.1
Multinomial Logistic Regression (multi_class = multinomial, solver = newton-cg)	max_iter = (100, 500, 1000, 2000, 5000, 7000, 10000)	100 / 100
	C = (0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0, 1e5, 1e6)	1 / 0.01
Support Vector Machine (kernel = rbf)	C = (0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0, 1e5, 1e6, 1e7, 1e8, 1e9, 1e10)	1e6 / 1e5
	gamma = (1e-9, 1e-8, 1e-7, 1e-6, 1e-5, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 'auto')	1e-6 / 1e-8
Neural Network (MLP) (solver = adam, activation = tanh, max_iter = 700)	hidden_layers = ([500], [1000], [500, 500], [1000, 1000])	[1000, 1000] / [500, 500]
	learning_rate_init = (1e-6, 1e-5, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0)	0.01 / 1e-6
	alpha = (1e-6, 1e-5, 0.00001, 0.0001, 0.001, 0.01)	1e-5 / 1e-6

We used the Python programming language and the open-source *Scikit-Learn* Python library to set and evaluate the machine learning models (<https://scikit-learn.org>). It is crucial to mention that we transformed categorical features into multiple binary features through one-hot-encoding and normalized each of these features (mean value of 0 and standard deviation of 1). In order to select appropriate classifier parameters, we used a grid search approach using ranges of parameters for each classifier with a stratified 5 folds cross validation in order to preserve the percentage of samples of each class (parameters are exposed in Table 2 above with Transportation Mode (TM) and Short-Middle Fights (SMF)). When the classifiers were properly configured, we performed a stratified 10 folds cross validation to obtain the final results discussed hereafter. The training and test datasets of each fold were automatically generated in the k-folds cross validation process used during our comparative analyses.

Results and Discussion

Firstly, we computed the average of F1 scores for the two comparative analyses (TM and SMF). The F1 score is the weighted average of precision and recall measures and enables to properly evaluate the performance of classifiers in the case of uneven class distribution. Figure 2 shows that the gradient boosting trees approach is a most accurate for the first analysis, while for the second analysis, the gradient boosting trees, the multinomial logistic regression and the multi-layer perceptron are the best classifiers. The results of these classifiers indeed maximize the micro, macro and weighted values. In addition, these results increase the predictions of a random classification, which would result in a 33% accuracy by 50%-100%.

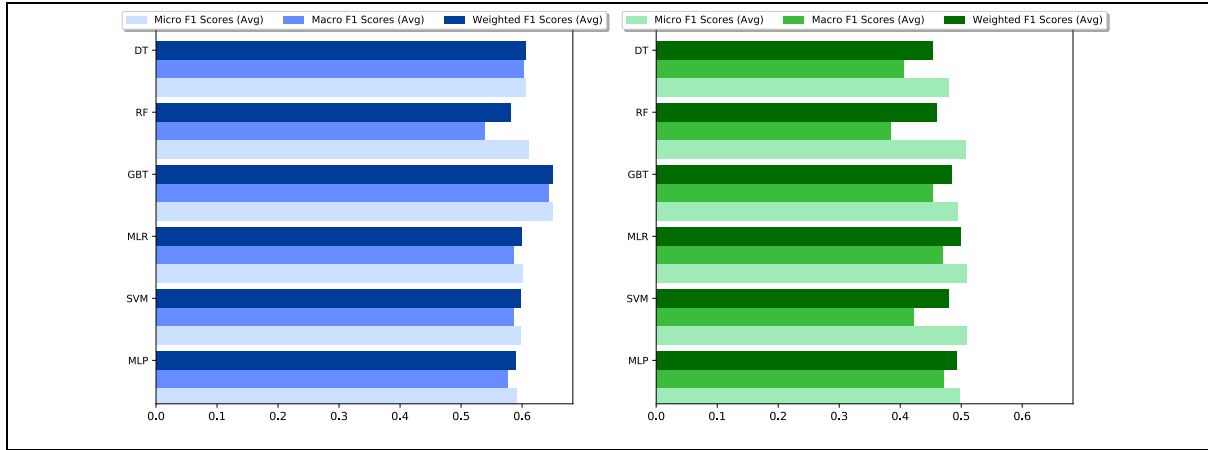


Figure 2. Average of F1 scores
(Left: Home-Work Transportation Mode Level / Right: Short-Middle Flight Number Level)
(DT: Decision Tree, RF: Random Forests, GBT: Gradient Boosting Trees, MLR: Multinomial Logistic Regression, SVM: Support Vector Machine and MLP: Multi-Layer Perceptron (Neural Network))

Then, we explored in more detail the result accuracy for each targeted class by exploring the average confusion matrix of each approach in Figure 3. The confusion matrix is a precise set of indicators of the accuracy performance of a predictive model for multi-class prediction problems. The confusion matrix of an ideal model would have values of 1 on the top left to bottom right diagonal, as every instance of a class would be predicted by the model and would thus have values of 0 elsewhere as no instance of a class would be misclassified. The results of the confusion matrices confirm the relevance of the gradient boosting tree approach for the first analysis. However, the results of the multinomial logistic regression classifier are slightly better than those of the gradient boosting trees for the second analysis. We can also notice that it is more difficult to predict all the classes of the second comparative analysis (SMF). Finally, we extracted the 20 most relevant features used to predict these two types of sustainability levels by using the gradient boosting trees approach fine-tuned with the relevant parameters for the two prediction goals (see Table 2). We can observe that several features of the first analysis are very powerful to predict the first sustainability levels compared to the second levels in Figure 4. For the first analysis (TM), accommodation and habits features are very relevant to predict it and may reflect the wealth of the household. Regarding the second analysis (SMF), accommodation, demographic, habits, psychological and social performance features are in the top ten of the most useful features. Interestingly, energy literature features are never highlighted, while the distance between regions (home-work) is indicated as the best feature for both analyses.

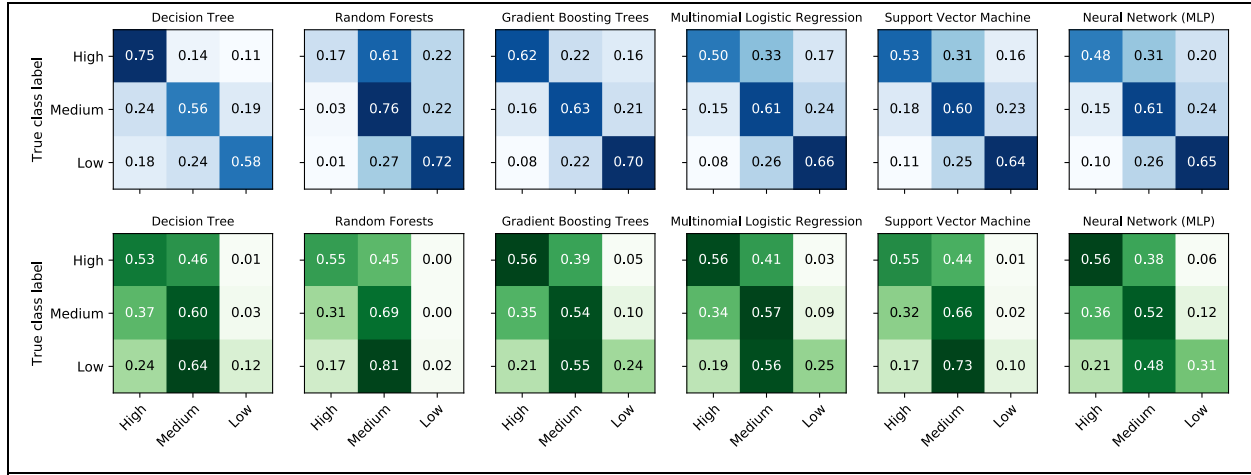


Figure 3. Normalized Confusion Matrices
 (Left: Home-Work Transportation Mode Level / Right: Short-Middle Flight Number Level)
 (High/Medium/Low: High/Medium/Low Sustainability Levels)

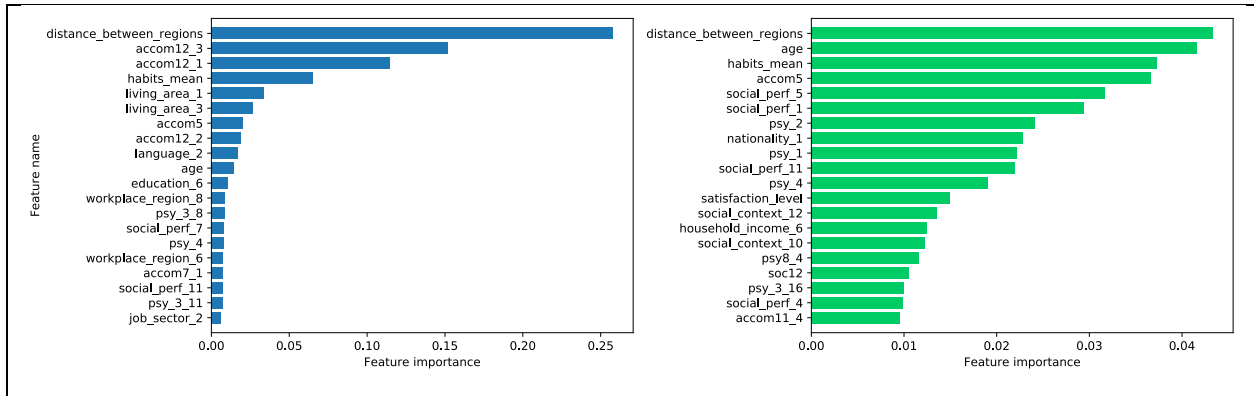


Figure 4. 20 Most Important Features
 (Left: Home-Work Transportation Mode Level / Right: Short-Middle Flight Number Level)

Future Work and Expected Contributions

In this paper, we presented an outline of a novel framework that allows Green IS to automatically obtain the sustainability levels of an individual. For a preliminary analysis, we performed two comparative analyses in order to highlight the most accurate classifiers and the most useful features to predict the sustainability levels of two acts of consumption related to the mobility domain. This analysis highlighted that it is possible to use classifiers to predict sustainability levels of individuals. Secondly, according to the results, the gradient boosting trees ensemble approach and the multinomial logistic regression approach seem to be the most promising models in terms of accuracy. Future work will aim at confirming these findings with other comparative analyses and helping to automatically adapt the selection of the classifier in the framework. We are currently designing a full instantiation of the framework including other consumption domains (e.g., housing) with their corresponding prediction tasks (e.g., average temperature of a living room). Then, we will have to devise the aggregation strategies for the upper levels of the tree by possibly taking into account the reliability level of the predictions. In order to evaluate the full instantiation of the framework, we will also use the SHEDS dataset. This research work will not only provide a technological solution for the research community and the industry but will also contribute to a real societal improvement in order to reach the sustainable world that we need in the close future.

References

- Amel, E. L., Manning, C. M. and Scott, B. A. 2009. "Mindfulness and sustainable behavior: Pondering attention and awareness as means for increasing green behavior". *Ecopsychology*, 1(1):14–25.
- Brendel, A. B., Zapadka, P., & Kolbe, L. M. 2018. "Design Science Research in Green IS: Analyzing the Past to Guide Future Research". In *Proceedings of the 26th European Conference on Information Systems (ECIS)*, Portsmouth, UK.
- Clark, C. F., Kotchen, M.J. and Moore, M. R. 2003. "Internal and external influences on pro-environmental behavior: Participation in a green electricity program". *Journal of environmental psychology*, 23(3):237–246.
- Dillahunt, T., Becker, G., Mankoff, J. and Kraut, R. 2008. "Motivating environmentally sustainable behavior changes with a virtual polar bear". In *Pervasive 2008 Workshop Proceedings*, volume 8, pages 58–62.
- Figuroa-García, E., García-Machado, J. and Pérez-Bustamante Yábar, D. 2018. "Modeling the social factors that determine sustainable consumption behavior in the community of madrid". *Sustainability*, 10(8):2811.
- Geiger, S. M., Fischer, D. and Schrader U. 2018. "Measuring what matters in sustainable consumption: an integrative framework for the selection of relevant behaviors". *Sustainable Development*, 26(1):18–33.
- Guo, Z., Zhou, K., Zhang, C., Lu, X., Chen, W. and Yang, S. 2018. "Residential electricity consumption behavior: Influencing factors, related theories and intervention strategies". *Renewable and Sustainable Energy Reviews*, 81:399–412.
- Juárez-Nàjera, M., Rivera-Martínez, J. G., and Hafkamp, W. A. 2010. "An explorative socio-psychological model for determining sustainable behavior: Pilot study in german and mexican universities". *Journal of Cleaner Production*, 18(7):686–694.
- Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. and Branson K. 2013. "Jaaba: interactive machine learning for automatic annotation of animal behavior". *Nature methods*, 10(1):64.
- Kim, E., Kim, W., and Lee, Y. 2003. "Combination of multiple classifiers for the customer's purchase behavior prediction". *Decision Support Systems*, 34(2):167–175.
- Kim, Y. J., and Yoon, H. J. 2017. "Predicting green advertising attitude and behavioral intention in south korea". *Social Behavior and Personality: an international journal*, 45(8):1345–1364.
- Kollmuss, A. and Agyeman, J. 2002. "Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior?". *Environmental education research*, 8(3):239–260.
- Kulkarni, V., Moro, A., and Garbinato, B. 2016. "Mobidict: a mobility prediction system leveraging realtime location data streams". In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming*, page 8. ACM.
- Pentland, A. and Liu., A. 1999. "Modeling and prediction of human behavior". *Neural computation*, 11(1):229–242.
- Rieck, K., Trinius, P., Willems, C., and Holz T. 2011. "Automatic analysis of malware behavior using machine learning". *Journal of Computer Security*, 19(4):639–668.
- Subrahmanian, V. and Kumar, S. 2017. "Predicting Human Behavior: The Next Frontiers". *Science*, 355(6324):489-489.
- Van Acker, V., Goodwin, P. and Witlox, F. 2016. "Key research themes on travel behavior, lifestyle, and sustainable urban mobility". *International journal of sustainable transportation*, 10(1):25–32.
- Weber, S., Burger, P., Farsi, M., Martinez-Cruz, A. L., Puntiroli, M., Schubert, I. and Volland, B. 2017. "Swiss Household Energy Demand Survey (SHEDS): Objectives, design, and implementation". *SCCER CREST Working Paper WP2 - 2017/04*.
- Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., Rui, Y. and Ma W.-Y. 2017. "Beyond the words: Predicting user personality from heterogeneous information". In *Proceedings of the tenth ACM international conference on web search and data mining*, pages 305–314. ACM.

- Yang, L., Dumais, S. T., Bennett, P. N., and Awadallah, A. H. 2017. "Characterizing and predicting enterprise email reply behavior". In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 235–244. ACM.
- Zhou, J. 2017. "Data Mining for Individual Consumer Credit Default Prediction under E-commerce Context: A Comparative Study". In Proceedings of the 38th International Conference on Information Systems (ICIS), South Korea.