White-Box or Black-Box Decision Tree Algorithms: Which to Use in Education?
Boris Delibašić, Member, IEEE, Milan Vukićević, Miloš Jovanović, and Milija Suknović

Abstract—University students are usually taught data mining through black-box data mining algorithms, which hide the algorithm’s details from the user and optionally allow parameter adjustment. This minimizes the effort required to use these algorithms. On the other hand, white-box algorithms reveal the algorithm’s structure, allowing users to assemble algorithms from algorithm building blocks. This paper provides a comparison between students’ acceptance of both black-box and white-box decision tree algorithms. For these purposes, the technology acceptance model is used. The model is extended with perceived understanding and the influence it has on acceptance of decision tree algorithms. An experiment was conducted with 118 senior management students who were divided into two groups—one working with black-box, and the other with white-box algorithms—and their cognitive styles were analyzed. The results of how cognitive styles affect the perceived understanding of students when using decision tree algorithms with different levels of algorithm transparency are reported here.

Index Terms—Algorithms, decision support systems, decision trees, open-source software, student experiments.

I. INTRODUCTION

THE INCREASING use of data mining in industry has created a need for good quality courses on this topic for undergraduate and graduate students [1]. Data mining courses are usually focused on teaching the whole data mining process[2] or on the application and implementation of data mining algorithms [3]. Popular data mining books [4], [5] concentrate on presenting black-box algorithms, without identifying algorithm building blocks for white-box algorithms. This paper examines the difference between students’ acceptance of black-box and white-box decision tree algorithms.

The experiment was performed on the open-source data-mining platform Rapid Miner [6], which includes black-box (BB) decision tree algorithms. For the white-box (WB) approach, a plug-in for Rapid Miner was used [7].

The BB approach enables users to use predefined algorithms to set parameters and retrieve models that help them detect regularities (patterns) in data. This facilitates use since the algorithm’s details remain hidden from the user.

The WB approach reveals more algorithm details than the BB approach and, in addition, allows users to assemble algorithms by putting algorithm building blocks together. In [8], it was shown that WB decision tree algorithms can outperform famous BB algorithms on a variety of datasets. However, this higher algorithm transparency increases the complexity of the system due to the increased number of user options. In the same paper, several BB decision tree algorithms were analyzed, and a generic decision tree structure was proposed that can reproduce famous BB decision trees and allow the creation of even more new WB algorithms by exchanging parts of BB algorithms. The same authors propose a generic structure that consists of subproblems that must be solved to make an algorithm work (e.g., to decide what evaluation measure should be used for splits evaluation). Many algorithmic components, also known as reusable components (RCs), are available for solving the subproblems (e.g., gain ratio or gini index can be used for the evaluating split subproblem). By combining reusable components (algorithm building blocks) through subproblems, a plethora of algorithms can be designed.

II. BACKGROUND

The students’ acceptance of decision tree algorithms as described in this paper was modeled with the popular Technology Acceptance Model (TAM) [9], as well an extended TAM model [10]. TAM includes five constructs: perceived usefulness (PU), perceived ease of use (PEOU), attitude toward use (ATU), intention to use (ITU), and actual system use. The key determinants of technology acceptance are the belief that the computer system will help improve job performance (PU) and the belief that the computer is easy to use (PEOU). These two determinants are considered to be the basis for evaluating the attitudes toward using particular computer systems and ultimately generating the intention to use. The intention to use then leads to actual end-user behavior.

In addition to testing PU, PEOU, ATU, and ITU, perceived understanding (PUND) was used to extend the original TAM model because [11] showed that PUND had a positive impact on user’s acceptance, and also because BB and WB approaches vary in the level of openness, with WB considered being the more open.

Users’ interaction with decision trees was previously researched in [12], where the accuracy, response time, and confidence of users working with comprehensible models (decision tables, decision trees, and rule-based predictive models) were analyzed. In [13], an interactive decision tree classifier was used to show that experts could be involved interactively in building decision tree models. A pilot study in [14] investigated how users interacted with machine learning systems.
The openness of systems has always been an interesting research topic. The author of [15] investigated marketing decision support systems, analyzing the influence that openness has on mental model quality, experience, decision confidence, and intensity of use. It was shown that openness decreases the reliance effect, but does not influence the decision makers’ evaluation of their decision. In [11], it is shown that transparency (openness) has a positive impact on user’s trust and user’s acceptance of a content-based art recommender.

TAM has been widely used in modeling the acceptance of management information systems in education. Students’ acceptance of ERP systems was discussed in [16], the acceptance of technology in education was modeled in [17], and an e-collaboration technology acceptance model among management students was proposed in [18].

Currently, there are no acceptance models for data mining education available in the literature; this paper proposes acceptance models for both BB and WB decision trees.

### III. Experimental Setting

The 118 participants in the experiment were senior-year business administration students who had taken a course in Business Intelligence (BI). The 14-week course consisted of 2 h of lectures and 2 h of practical exercises per week and had been offered, in a similar form, over an eight-year period. In chronological order, the topics covered were the following: decision support systems, collaboration systems, data warehousing, data mining, case-based reasoning, and knowledge management. Each topic lasted for two weeks, except for data warehousing and data mining, which each took three weeks. The experiment was performed in the weeks following the data mining part of the course. It was noticed that this particular part of the course was hard for students to grasp, as they had difficulty with data preprocessing, choosing the right algorithm and setting its parameters, and analyzing the models produced from algorithms.

The students were set the task of searching for the most accurate decision tree algorithm in 15 trials. All students worked with the Hayes Roth dataset available from UCI [19]; in both the BB and WB approach, the accuracy of the best algorithm was the same. The chosen dataset was fairly simple (160 instances, four categorical attributes, and one class attribute with three classes) and was chosen because it is easy for students to understand, produces fairly different decision trees for different algorithm parameters, and has a mixed attribute type.

Before the experiment started, students working with the BB approach were shown how to use the C4.5 decision tree; those working with the WB approach were shown how to design an algorithm of their own. Students subsequently received user manuals that helped them to understand the algorithm parameters and algorithm building blocks.

Students working with BB algorithms (BB students) could choose between three algorithms (C4.5, CART, and CHAID) and set their parameters. Students working with WB algorithms (WB students) had to design algorithms and set the parameters of the algorithm building blocks through the WB decision tree designer interface shown in Fig. 1.

### IV. Results and Evaluation

Each button on the left-hand panel represents a subproblem. When a subproblem is selected, the upper central panel shows the RCs available for solving it and allows users to select and save one or more of these. The lower central panel shows the parameters for the selected RCs.

The right-hand panel documents the designed generic tree algorithm (the selected RCs and their parameters).

The top panel contains options for creating new WB algorithms, saving current ones, or opening existing ones.

The questionnaire contained in total 21 items measured on a 1–7-point Likert-like scale. The questionnaire and user manuals used for the experiment are available at whibo.fon.bg.ac.rs/joomla/images/stories/downloads/the-whibo-experiment.pdf. PUND was measured with items adapted from [16], where PU, PEUS, ATU, and ITU was measured with items adapted from TAM [9].

Of the BB students, 91.38% found the most accurate algorithm after the 15 trials, as compared to only 80% of the WB students. This suggests that WB algorithms are more complex to use.
Table I shows students’ scores in terms of the factors of the acceptance model. No significant differences were found in any of the factors, which is surprising given that WB decision trees are more complex to use than BB decision trees. However, as shown by [20], users are willing to use more complex models if they understand their potential benefits.

Students’ cognitive styles were analyzed, and their effect on the PUND scores reported. Cognitive style was tested using the popular MBTI questionnaire [21] that includes 95 items with four bipolar scales: Extraversion-Introversion, Sensing-Intuition, Thinking-Feeling, and Judging-Perception. Introverts are oriented primarily to internal hints, and extroverts to external. The sensing type pays most attention to what she/he can sense, and the intuitive type relies mostly on inner feelings. Someone with a thinking style makes assessments based on logical impersonal processes, while the feeling type makes assessments based on personal, subjective processes. Judgers try to structure and use models that allow decisions to be made quickly, while perceivers keep options open with no concern for deadlines. A student might fall entirely within one of these cognitive aspect categories or straddle several categories.

BB students who were neither sensing nor intuitive (a total of nine students) achieved a PUND of 3.28 (1.37), while WB students with the same characteristics (13 students) achieved a PUND of 4.62 (0.98). The difference was significant on a 0.05 level \((F = 7.146, \text{Sig.} = 0.015)\). Eta square for this finding is 0.26, which is considered as a large effect size. Therefore, students who are neither sensing nor intuitive should use WB.

Of the 60 WB students, 41 were judging, 14 perceiving, and 5 were neither judging nor perceiving. Judging students achieved a PUND of 4.35 (std. dev. 1.17), while perceiving students had an average of 5.29 (1.01). The difference was significant on a 0.05 level \((F = 7.017, \text{Sig.} = 0.011)\). Eta square for this finding is 0.12, which can be considered as a medium effect size. Perceiving WB students will achieve significantly better PUND than will judging students, a difference not seen with BB students or when comparing the WB and BB approach.

A partial least square (PLS) analysis was used for proposing models that could predict the influences for students’ acceptance of BB and WB decision trees. PLS was a convenient choice for this study as the sample size was relatively small [10]. For the analysis, SmartPLS software was used [22]. The proposed models and their quality criteria are shown in Fig. 2 and Tables II and III.

In Tables II and III, the original number of items is shown for each factor, as well as the remaining items (only items that had loadings higher than 0.8 were kept, which guarantees convergent validity). Cronbach’s alpha is greater than 0.7 for almost all factors, except for PUND, but this is acceptable given that Cronbach’s alpha can underestimate reliability in factors with a small number of items. Composite reliability is greater than 0.8 for all factors. The average variance extracted (AVE) being greater than 0.5 for all factors means that they all have a significant effect, which suggests the validity and discrimination of the model is satisfactory. In both models, the Fornell–Larcker criterion is satisfied (square root of AVE is larger than any interfactor correlations), which indicates sound discriminant validity.

For the BB model, the coefficient of determination \(R^2\) of the dependent variable ITU is 0.449, showing a moderate strength effect, whereas for the WB model, the \(R^2\) of the dependent variable is 0.703, which can be considered as a substantial strength effect. \(R^2\) shows the percentage of variance explained in the model. These figures indicate that a good prediction can be made of ITU, which is itself a good predictor of user acceptance.

Fig. 2 shows models for BB and WB decision tree acceptance. Arrows show path coefficients (with t-values and significance levels). The original TAM model is shown inside the bracketed rectangle.

Based on the models, several observations can be made. In BB decision trees, PUND had a positive impact on PU, PEOU, and ITU. PU had positive impact on ATU, and ATU had positive impact on ITU. Students who perceived that they had a better understanding were more likely to find the system more useful and easy to use and to accept it. It is interesting that PEOU had no influence on ATU and that PU had no influence on ITU. PUND, however, played a major role in acceptance of BB decision trees.

In WB decision trees, PUND had a positive impact only on PEOU. PEOU had a positive impact on PU, and PU had a positive impact on ATU and ITU. ATU had a positive impact on ITU. Here, students with greater understanding found it easier to use. However, students who found WB decision trees useful were more likely to accept the systems.

V. CONCLUSION

This paper proposed models for BB and WB decision tree acceptance among management students. As data mining algorithms are usually the hardest topic for students in a BI course, an analysis was made of whether more open algorithms would
TABLE II
QUALITY CRITERIA OF BB ACCEPTANCE MODEL

<table>
<thead>
<tr>
<th>Factor</th>
<th>Original number of items</th>
<th>Remaining Items (Loadings)</th>
<th>Cronbach’s alpha</th>
<th>Composite Reliability</th>
<th>AVE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUND 2</td>
<td>PUND1 (0.89)</td>
<td></td>
<td>0.728</td>
<td>0.88</td>
<td>0.785</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PUND2 (0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU 6</td>
<td>PU1 (0.89)</td>
<td></td>
<td>0.89</td>
<td>0.924</td>
<td>0.751</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>PU2 (0.86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU4 (0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU5 (0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU 6</td>
<td>PEOU1 (0.92)</td>
<td></td>
<td>0.789</td>
<td>0.904</td>
<td>0.825</td>
<td>0.369</td>
</tr>
<tr>
<td></td>
<td>PEOU5 (0.89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATU 5</td>
<td>ATU1 (0.91)</td>
<td></td>
<td>0.893</td>
<td>0.926</td>
<td>0.758</td>
<td>0.747</td>
</tr>
<tr>
<td></td>
<td>ATU2 (0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATU4 (0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATU5 (0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITU 2</td>
<td>ITU1 (0.94)</td>
<td></td>
<td>0.74</td>
<td>0.878</td>
<td>0.784</td>
<td>0.449</td>
</tr>
<tr>
<td></td>
<td>ITU2 (0.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III
QUALITY CRITERIA OF WB ACCEPTANCE MODEL

<table>
<thead>
<tr>
<th>Factor</th>
<th>Original number of items</th>
<th>Remaining Items (Loadings)</th>
<th>Cronbach’s alpha</th>
<th>Composite Reliability</th>
<th>AVE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUND 2</td>
<td>PUND1 (0.84)</td>
<td></td>
<td>0.68</td>
<td>0.862</td>
<td>0.758</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PUND2 (0.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU 6</td>
<td>PU1 (0.8)</td>
<td></td>
<td>0.857</td>
<td>0.902</td>
<td>0.698</td>
<td>0.302</td>
</tr>
<tr>
<td></td>
<td>PU3 (0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU5 (0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU6 (0.86)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU 6</td>
<td>PEOU1 (0.90)</td>
<td></td>
<td>0.773</td>
<td>0.898</td>
<td>0.815</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>PEOU3 (0.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATU 5</td>
<td>ATU1 (0.94)</td>
<td></td>
<td>0.861</td>
<td>0.915</td>
<td>0.783</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>ATU2 (0.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATU5 (0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITU 2</td>
<td>ITU1 (0.87)</td>
<td></td>
<td>0.739</td>
<td>0.884</td>
<td>0.792</td>
<td>0.703</td>
</tr>
<tr>
<td></td>
<td>ITU2 (0.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

improve their acceptance by students. It is shown that perceived understanding had a positive impact on students accepting BB decision trees. While PUND did not play a major role in accepting WB systems, PU had a positive impact on accepting WB decision trees. It seems that the BB and WB approaches are complementary, as students tend to accept BB decision trees if they perceive to understand the algorithm structure. On the other hand, students tend to accept the WB system if they find it useful.

The recommendation for lecturers teaching BB decision trees would be to make a greater effort to explain the algorithm structure better to students, perhaps by using WB decision trees as a tool. When using WB trees, students should be carefully taught how to use the system and shown the benefits such a system could provide.

This paper has shown how PUND depends on a student’s cognitive style. This finding might suggest dividing students into groups for learning (by the BB or WB approach). Lecturers
should recommend the WB approach to students who are neither sensing nor intuitive, and they can expect that judging and perceiving types of students will probably have a significantly different level of PUND if using the WB approach.

Although WB decision trees are more complex, students perceived that these algorithms were easy to use; this probably has to do with the findings in [20], where it was shown that users are willing to use more complex models if they understand their benefits.

This study has several limitations. Experiments were performed on business major students; it would be more interesting to do the experiment with computer science/engineering undergraduate students (since the WhiBo platform offers an easily extendable structure for the design and implementation of new algorithms [7]). Another limitation is that, in the TAM model, only PUND was analyzed as an external factor. The influence of MBTI factors on PUND was also analyzed without taking into account other potential confounding factors.

REFERENCES


Boris Delibašić (M’12) received the Ph.D. degree in business decision making from the University of Belgrade, Belgrade, Serbia, in 2007. He is currently an Assistant Professor with the Faculty of Organizational Sciences, School of Business Administration, University of Belgrade. He teaches business intelligence and business decision making. Since 2011 he has been Assistant to the Board of the European Working Group on decision support systems. His research interests are in data mining algorithm design, data warehousing, and decision support systems.

Dr. Delibašić received a post-doctoral Fulbright scholarship at the Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA, in 2011.

Milan Vukičević received the M.Sc. degree in business decision making from the University of Belgrade, Belgrade, Serbia, in 2009. He is a Teaching Assistant with the Faculty of Organizational Sciences, School of Business Administration, University of Belgrade, where he teaches business intelligence and business decision making. He has been active since 2009 in the Rapid Miner open-source data mining platform community. His research interests are in data mining algorithm design, data warehousing, and software engineering.

Milica Suknović received the Ph.D. degree in business decision making from the University of Belgrade, Belgrade, Serbia, in 2003. He is a Full Professor with the Faculty of Organizational Sciences, School of Business Administration, University of Belgrade, where he teaches business intelligence and business decision making. Currently, he is Vice-Dean for Education with the Faculty of Organizational Sciences. He is part of the Editorial Board of the Yugoslav Journal of Operations Research. His research interests are in data warehousing, group decision support systems, and group decision making and negotiations.