Uncovering Decision-making Process of Cost-sensitive Tree-based Classifiers using the Adaptation of TreeSHAP[**]

Marija Kopanja[1,2,*,†], Sanja Brdar[1,†] and Stefan Hačko[1,2,†]

[1]BioSense Institute, Novi Sad, Serbia
[2]Faculty of Sciences, University of Novi Sad, Serbia

Abstract
The cost-sensitive decision tree (CSDT) method is a modification of the decision tree (DT) algorithm with incorporated misclassification costs into the learning process. Cost-sensitive tree-based classifiers are well suited for tackling class imbalance. In general, tree-based classifiers are characterized by a nice graphical representation of the model, which can be used in explaining the classifier's decision-making process. However, the depth of the tree can be a limiting factor in comprehending the model. An additional obstacle in comprehending the decision-making process of the CSDT classifiers is the implementation of the method. The TreeSHAP method, a variation of the SHAP methodology for the exact calculation of SHAP values for tree-based models, can facilitate the explanation of (deep) tree-based models. However, the current implementation of the TreeSHAP method is limited to only several tree-based models, excluding the cost-sensitive tree-based classifiers. The aim of this paper is to introduce a cost-sensitive tree explanation method based on the TreeSHAP method and analyze insights into the decision-making process of the CSDT classifiers compared to DT classifiers.

Keywords
Explainable artificial intelligence, Tree SHAP, Cost-sensitive learning, Tree-based classifiers, Cost-sensitive decision tree

1. Introduction

Cost-sensitive (CS) learning is a subgroup of machine learning (ML) classification algorithms that are able to cope with samples with different misclassification costs [1], [2], [3], [4], [5], [6]. The misclassification cost of a sample is a function of the actual and predicted class, represented as a two-dimensional cost matrix. Two types of cost matrices can be distinguished, class-

Late-breaking work, Demos and Doctoral Consortium, colocated with The 1st World Conference on eXplainable Artificial Intelligence: July 26–28, 2023, Lisbon, Portugal
[*]Corresponding author.
[**]Supported by ANTARES project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement SGA-CSA No. 739570 under FPA No. 664387, https://doi.org/10.3030/739570 and Ministry of Education, Science and Technological Development of the Republic of Serbia, grant agreement 451-03-47/2023-01/200358.
[†]These authors contributed equally.
marija.kopanja@biosense.rs (M. Kopanja); sanja.brdar@biosense.rs (S. Brdar); stefan.hacko@biosense.rs (S. Hačko)
0000-0003-1176-5677 (M. Kopanja); 0000-0002-2259-4693 (S. Brdar); 0000-0002-1067-6333 (S. Hačko)

dependent and example-dependent cost matrices [7], [8]. A former, much stronger assumption can be used in the context of imbalanced classification [1], [5]. Namely, class imbalance in a two-class classification framework can be described as a problem where one, minority class, with a smaller number of samples, is underrepresented. The CS approach is important due to the vast amount of intrinsically imbalanced data in many domains including fraud detection [9], [10], [11], credit rating [12], [13], [14], medicine [15], [16], and others [5].

The tree-based classifiers have a nice graphical representation of the model, which facilitates understanding of the decision-making process. However, the depth of the tree can be a limiting factor in comprehending which features contribute to the classification of each individual sample. An additional obstacle for CSDT is the implementation of the method. For tree-based models there is a model-specific xAI method TreeSHAP, a method for the exact calculation of SHAP values [17], [18], [19]. The objective of our research is to introduce a CS version of TreeSHAP that will facilitate the explanation of (deep) cost-sensitive tree-based models, as the current implementation is not compatible with the CSDT algorithm.

In the literature, there is a limited amount of articles dealing solely with the interpretable ML models for imbalanced data, of which the CS classification is a special case [20], [21]. To the best of the present author's knowledge, there are no articles in the literature considering the explanation methods for CSDT.

2. Methodology

Given a cost matrix, either class-dependent or sample-dependent, the CSDT method uses a greedy binary splitting procedure for feature space stratification by using a cost-sensitive splitting criterion that is expressed as expected misclassification cost reduction. The cost-based impurity measure is the minimal cost of labeling a given node to a particular class. More details about the method can be found in [13].

Different methods have been proposed for calculating SHAP values [22]. One model-specific method is TreeSHAP [17], a method that computes exact SHAP values in polynomial time for tree-based models. The method is helpful for understanding the global tree model structure based on many local explanations. This is especially important for deep tree-based models, as the depth of the tree is a limiting factor in comprehending the model's decision process. For the CSDT classifiers, another limiting factor is the impractical graphical representation of the model. Extracting decision rules for individual samples is hard even for shallow CSDT models. On the other hand, applying the TreeSHAP algorithm was impossible due to its incompatibility with CSDT algorithm.

The TreeSHAP uses a tree-based model structure to compute SHAP values exactly. The CSDT uses a cost matrix in the tree-building process for making a decision on whether the sample will be classified as positive or negative, depending on the number of samples from each class in the node and cost matrix. Due to the latter fact, the cost matrix needs to be taken into consideration by the TreeSHAP algorithm.

To create the CSTreeSHAP method as a subclass of TreeSHAP, we developed several recursive functions for extracting required information from the CS tree model, such as sequential ID numbers of left and right nodes. The cost matrix is not used explicitly due to its sample-dependent nature for the dataset used in our research, rather attributes of the existing CS tree object are adapted such as information about the number of samples in each class per node and the corresponding part of the cost matrix. This is a crucial part since the prediction in each terminal node is determined based on the number of samples in the node and their costs in the cost matrix. Recall, each terminal node in the CSDT model is allocated to the least costly class. The available implementation of TreeSHAP uses a special tree structure of a tree-based model, which CSDT does not have, at least not in the required form. However, the CSDT tree model can be described and therefore recreated using the mentioned list of information. Extraction of the information for classic DT models is straightforward due to its implementation in scikit-learn library [23]. The code of our CSDTreeSHAP implementation will be available on GitHub within the next extended version of the paper.

3. Experiments

The aim of our experimental work was to analyze feature importance retrieved from the obtained local explanations using CSTreeSHAP and TreeSHAP methods for CSDT and DT models, respectively. By providing insight into the decision-making process of different ML models it is possible to make evidence-based decisions about models that would be preferable.

We used a well-known CS1 dataset [24], from the credit scoring domain, where the minority class of risky clients has a higher misclassification cost (i.e. the clients that are more likely to default on a loan) [25], [12], [26], [27], [8]. More information about the dataset and the cost matrix can be found in Costcla library [1] and in the article [12], respectively.

3.1. Experimental setup

For partitioning the dataset we used a 10-fold cross-validation procedure. Measures used for evaluation include precision and recall per class, $F1$ score and relative cost reduction measure (RCR) from studies [12], [28], [8], which is the most relevant measure in CS framework. Due to space limitations, results for mentioned measures are not reported.

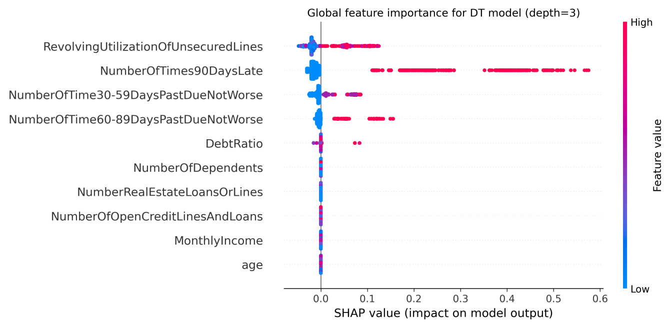
In the explaining phase, obtained SHAP values per sample are averaged for each feature on the whole dataset and represented in the global feature importance plot. The tree depth is varied, ranging from 2 to 10, i.e. from shallow to deeper tree models, as the objective was to investigate whether DT and CSDT models trained on the same data make decisions based on similar indicators and which model would be preferred by domain experts as an end user.

3.2. Results

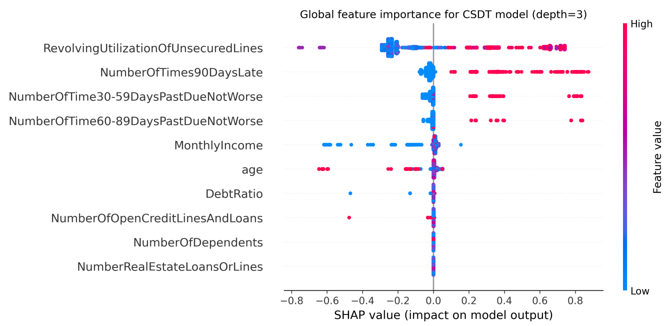
The cross-validated results for varying depths of a tree model are analyzed. The results for CSDT models coincide with those reported in [8]. Given RCR as a performance measure, the model of choice would be the CSDT model, which achieves the highest RCR. There is consistently good performance of CSDT in terms of RCR at all depths compared to DT. Global feature importance plot using summary plots (Fig.1) shows the top-rated features in descending order according to their importance, i.e. mean absolute SHAP values.

[1]A Python module for cost-sensitive machine learning. https://pypi.org/project/costcla/

Due to space limitations, not all plots are presented. For the dataset, the results are shown in figure Fig.1. for depth 3, as there is a slight increase in RCR for CSDT until the tree depth reaches 5. Notably, the top features for DT and CSDT coincide, with slight perturbation in their order. However, the CSDT model is characterized by a larger number of splitting features compared to DT for the same depth, meaning its harder to comprehend the most important features for decisions the model makes, especially due to the lack of code implementation that could facilitate visualization of the CSDT tree model represented as a dictionary object.



(a) Global feature importance for DT model.



(b) Global feature importance for CSDT model.

Figure 1: Global feature importance for CS1 dataset with parameter depth = 3.

4. Concluding remarks and future work

The presented article introduces the CSTreeSHAP method as an xAI tool for better understanding the driving forces behind the decision-making process of CS classifiers, with a focus on CSDT models. Except for being well-suited for CS problems, the potential of the tool and CSDT classifiers can be seen in a large number of imbalanced data on which the methodology could be applied. The usefulness of the explanation method is demonstrated for different datasets at different tree depths. Having the ability to compare explanations of different tree models at varying depths enables well-grounded model selection. Directions for future work include extending the application of the CSTreeSHAP method on ensemble methods with cost-sensitive tree models as base learners and other (ensemble) cost-sensitive models.

References

[1] A. Ali, S. M. Shamsuddin, A. Ralescu, Classification with class imbalance problem: A review 7 (2015) 176–204.

[2] B. A. Draper, C. E. Brodley, P. E. Utgoff, Goal-directed classification using linear machine decision trees, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1994) 888–893. doi:10.1109/34.310684.

[3] P. Kumar, R. Bhatnagar, K. Gaur, A. Bhatnagar, Classification of imbalanced data: Review of methods and applications, IOP Conference Series: Materials Science and Engineering 1099 (2021) 012077. doi:10.1088/1757-899X/1099/1/012077.

[4] I. Mienye, Y. Sun, Performance analysis of cost-sensitive learning methods with application to imbalanced medical data, Informatics in Medicine Unlocked 25 (2021). doi:10.1016/j.imu.2021.100690.

[5] S. Yanmin, A. Wong, S. Mohamed, Classification of imbalanced data: a review, International Journal of Pattern Recognition and Artificial Intelligence 23 (2011). doi:10.1142/S0218001409007326.

[6] C. Elkan, The foundations of cost-sensitive learning, in: Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 2, 2001, p. 973–978.

[7] M. Lazaro, A. Figueiras-Vidal, A Bayes Risk Minimization Machine for Example-Dependent Cost Classification, IEEE transactions on cybernetics 51 (2021) 3524–3534. doi:10.1109/TCYB.2019.2913572.

[8] J. Mediavilla-Relaño, M. Lázaro, A. R. Figueiras-Vidal, Imbalance example-dependent cost classification: A bayesian based method, Expert Systems with Applications 213 (2023) 118909. doi:10.1016/j.eswa.2022.118909.

[9] B. A. Correa, A. Stojanovic, D. Aouada, B. Ottersten, Cost Sensitive Credit Card Fraud Detection Using Bayes Minimum Risk, in: 2013 12th International Conference on Machine Learning and Applications, 2013, pp. 333–338. doi:10.1109/ICMLA.2013.68.

[10] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access 7 (2019) 93010–93022.

[11] L. Zhang, D. Zhang, Evolutionary cost-sensitive extreme learning machine, IEEE transactions on neural networks and learning systems 28 (2016) 3045–3060.

[12] B. A. Correa, D. Aouada, B. Ottersten, Example-Dependent Cost-Sensitive Logistic Regression for Credit Scoring, in: 2014 13th International Conference on Machine Learning and Applications, 2014, pp. 263–269. doi:10.1109/ICMLA.2014.48.

[13] B. A. Correa, D. Aouada, B. Ottersten, Example-dependent cost-sensitive decision trees, Expert Systems with Applications 42 (2015) 6609–6619. doi:10.1016/j.eswa.2015.04.042.

[14] N. Khalili, M. Rastegar, Optimal cost-sensitive credit scoring using a new hybrid performance metric, Expert Systems with Applications 213 (2023) 119232. doi:10.1016/j.eswa.2022.119232.

[15] D. Ibomoiye, S. Yanxia, Performance analysis of cost-sensitive learning methods with application to imbalanced medical data, Informatics in Medicine Unlocked 25 (2021) 100690. doi:10.1016/j.imu.2021.100690.

[16] X. Wan, J. Liu, W. Cheung, T. Tong, Learning to improve medical decision making from imbalanced data without a priori cost, BMC medical informatics and decision making 14 (2014) 1–9.

[17] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S. Lee, From local explanations to global understanding with explainable ai for trees, Nature Machine Intelligence 2 (2020) 2522–5839.

[18] S. Hart, Shapley Value, in: J. Eatwell, M. Milgate, P. Newman (Eds.), Game Theory, Palgrave Macmillan UK, London, 1989, pp. 210–216.

[19] R. Mitchell, E. Frank, G. Holmes, Gputreeshap: Massively parallel exact calculation of shap scores for tree ensembles, 2020. doi:10.48550/ARXIV.2010.13972.

[20] D. Dablain, C. Bellinger, W. Aha, N. Chawla, B. Krawczyk, Understanding Imbalanced Data: XAI & Interpretable ML Framework, ResarchGate Preprint (2022). doi:10.13140/RG.2.2.14645.96489.

[21] Y. Gao, Y. Zhu, Y. Zhao, Dealing with imbalanced data for interpretable defect prediction, Information and Software Technology 151 (2022). doi:10.1016/j.infsof.2022.107016.

[22] S. Lundberg, S. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 4765–4774.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[24] W. C. Credit Fusion, Give me some credit, 2011. URL: https://kaggle.com/competitions/GiveMeSomeCredit.

[25] D. Boughaci, A. Alkhawaldeh, A new variable selection method applied to credit scoring, Algorithmic Finance 7 (2018) 43–52. doi:10.3233/AF-180227.

[26] S. Liu, R. Wang, Y. Han, Research on personal credit evaluation based on machine learning algorithm, in: 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT), 2021, pp. 48–52. doi:10.1109/ISCIPT53667.2021.00016.

[27] A. Markov, Z. Seleznyova, V. Lapshin, Credit scoring methods: Latest trends and points to consider, The Journal of Finance and Data Science 8 (2022) 180–201. doi:https://doi.org/10.1016/j.jfds.2022.07.002.

[28] B. A. Correa, A. Stojanovic, D. Aouada, B. Ottersten, Improving credit card fraud detection with calibrated probabilities, 2014. doi:10.1137/1.9781611973440.78.