SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2022
ORGANIZATION AND MANAGEMENT SERIES NO. 167

PREFERENCE-DRIVEN CLASSIFICATION FOR COOPERATION AND RESOLVING DECISION-MAKER – AI CONFLICTS

Krzysztof KANIA[1*], Jan KOZAK[2]

[1] University of Economics in Katowice; krzysztof.kania@ue.katowice.pl, ORCID: 0000-0002-7123-6972
[2] University of Economics in Katowice; jan.kozak@ue.katowice.pl, ORCID: 0000-0002-2128-6998
* Correspondence author

Purpose: The aim of the paper is to present sources and effects of decision-maker – AI conflicts and how a modified measure of quality of classification algorithms can be used in resolving them.
Design/methodology/approach: Identification of sources and effects of decision-maker – AI conflict was based on literature research while presentation of modified quality measure was based on numerical experiment.
Findings: Using modified measure can be used to reflect user's preferences and can become a base for building flexible human-AI communication. It also helps in lowering barriers of involving AI into decision processes.
Research limitations/implications: The scope of the research was limited to the classification task, which is only one of the tasks performed by AI algorithms.
Practical implications: Results of article can be used in building AI application for every classification problem where cooperation between AI and decision-maker is needed.
Originality/value: The research enhances understanding an influence of particular AI solutions on further working in business environment and present possibilities of using modified quality measure in improving decision-maker – AI cooperation.

Keywords: AI in decision-making, classification, user preferences.

Category of the paper: Research paper.

1. Introduction

Artificial Intelligence (AI) tools are rapidly entering various areas of life, science and business. In many cases, they have enabled tasks that previously could not be solved within reasonable limits due to complexity or data volume. AI changes almost every aspect of human work and AI applications appear in a variety of devices and systems – from smartphones to city control systems and medicine. Sometimes this happens gradually, almost imperceptibly, by augmenting previously known functions or supporting everyday activities, sometimes in a spectacular way such as beating the world champion in an advanced logic game. AI changes our behavior, our work, our communication, our ways of solving tasks. Inevitably, AI enters also areas of decision-making that were previously the exclusive domain of humans. This creates inescapable conflicts between humans and AI.

The purpose of this paper is to present the sources and consequences of the decision-maker – AI conflict and how they can be mitigated by using a modified measure of classification quality that incorporates user preferences in its design.

The paper is organized as follows: the next section presents fundamental problems of human-AI cooperation, their sources and the ways proposed in the literature to overcome them. Then, it focuses on classification issues as one of the basic tasks occurring in decision-making processes. Classical measures of classification quality are presented, as well as the modified measure, that allows taking into account the preferences of decision-makers. In the next section, the performance of modified quality measure is presented using two datasets as an example. The consequences of the introduction of the measure for the possibility of solving conflicts and some problems of human-AI cooperation are presented in the discussion. The paper concludes with a summary and directions for future work.

2. Sources and consequences of human-AI conflicts

Conflict situations in decision-making processes have always been present. Different opinions, views or preferences, different management styles, different levels of risk appetite, different temperaments of decision-makers or levels of understanding and interpretation of data – these are the sources of conflicts that have been known for a long time (Schwenk, 1990). The situation changed when the computer system, which is an active and permanent element of the decision-making process, became a party to the conflict. In the case of classical IT systems, the computer was only the provider of the data or results of the various models, and the responsibility of the system was limited to delivering reliable and timely data. The problem of decision-maker-computer cooperation was exacerbated when the system began to play an expert role in the process. This is what happens with AI. The main source of problems with the use of AI in decision-making processes is the delegation of powers and the issue of responsibility for the decisions made (Ferreira, Monteiro, 2021). It is impossible to explain anything to the computer, nor convince it of anything. The computer cannot be moved and made softer, although it can be ignored or turned off. It has no self-interest, although it operates according to rules written by a man with a certain level of knowledge and specific beliefs.

On the one hand, there are purely physical or technical problems (control), where AI consultation with a human, e.g. due to the short time required to make a decision, is redundant or even harmful. In the case of physical phenomena in limited and closed world the influence of the will of the decision-maker is small. When the key factors of the application of AI are physical and technical one they are relatively easy to define and measure. The more the result of decision-making process or the use of the result of AI modules depends on the human being, the more important the subjective factors become. Therefore, at the other end, there are problems that people find impossible to objectify and treat them as their exclusive domain. As long as AI is used to support simple technical activities in devices that facilitate everyday functioning (e.g. object recognition in a camera lens), these issues do not matter much. Their importance increases when decisions are made that affect the health, life or personal rights of others or have a social impact. In such cases, ethical, legal and economic questions arise immediately (Rodrigues, 2020; Zhang, Chen, Xu, 2022; Formosa, Rogers, Bankins, Griep, Richards, 2022; Čartolovni, Tomičić, Lazić Mosler, 2022). Human-AI problems and conflicts begin when the degree of automation of AI activities exceeds the threshold acceptable to the decision-maker (Mackeprang, Müller-Birn, Stauss, 2019).

The main reasons for difficulties of human-AI cooperation are: poor understanding of the meaning of the results (low interpretability), the inability to obtain a simple explanation of the results (black-box problem, low transparency), the inability to indicate to the algorithm the personal preferences of the decision-maker, no influence on the way the algorithm works or lack of confidence in the correctness of the result (Dafoe et al., 2020), (Leyer, Schneider, 2021). The situation was additionally aggravated by the spread of Big Data technology, where it become obvious that a human cannot beat a computer in the tasks of analyzing huge datasets.

The ways in which human and AI make decisions are completely different. The activity of AI is guided by the past, while human activity by the future. AI is based strictly on data, human has the ability to think abstractly, formulate a vision, predictive intuition, holistic view, the ability to take into account new factors unknown to the machine, creativity, the ability to improvise and quickly assess a new situation, operating in conditions of discontinuity, etc. Good cooperation between the decision-maker and AI consists in maintaining a balance between the capabilities of the computer and the use of natural human abilities (Jarrahi, 2018).

AI tools will remain useless if the knowledge they discover is not included in the decision-making process (Mikalef, Krogstie, 2019). If decision-makers are unwilling to cooperate or - despite their will – they are unable to interpret, justify, understand or apply the result delivered by AI, all efforts to implement it will be wasted (Janssen, van der Voort, Wahyudi, 2017), (Duan, Edwards, Dwivedi, 2019). When the human-AI conflict cannot be resolved, and AI prevails, the result of the lack of proper cooperation is a reduction in the creativity of decision-makers, laziness, resignation from the principles of ethics and morality, which leads to frustration and resignation. When humans prevail, the result is the rejection of AI tools and a waste of its potential. The imbalance between the analytical capabilities of a machine and the ability to synthesize that humans are capable of has negative consequences. Too little AI contribution wastes the discovered knowledge, while too much, marginalizes and degrades human.

Therefore, various patterns of human-AI cooperation are analyzed (Leyer, Dootson, Oberländer, Kowalkiewicz, 2020). Effective use of the opportunities offered by AI in decision-making processes requires meeting several conditions. These include (Mikalef, Krogstie, 2019), (Duan et al., 2019):

- the ability of the decision-maker to work with these tools,
- the ability to interpret the obtained results,
- embedding AI tools into decision-making processes and using its results in the subsequent steps of the these processes.

Lowering the barriers to introducing AI into decision-making processes and resolving conflict situations can be carried out in two ways: from the management side and from the IT side. From the management side through (Langer, Landers, 2021):

- raising awareness and training of decision-makers in human-computer cooperation,
- building systems with an appropriate level of AI use,
- allowing a person an appropriate scope of competence and freedom in making decisions,

while from the IT side, among others by (Wang et al., 2020; Gunning et al., 2019; Samek, Montavon, Vedaldi, Hansen, Müller, 2019):

- bringing AI tools closer to the decision-makers (similarly to data reporting and visualization tools), so that they can be operated them on their own, without the intermediation or help of an IT specialist,
- building the convenient AI systems interfaces,
- making AI algorithms more flexible so that decision-makers can influence their operation,
- expansion of AI towards interpretable and explainable intelligence (eXplainable AI - XAI).

There is no single recipe for overcoming all problems. They are solved in small areas, in terms of particular tools or algorithms.

3. Classification in decision-making processes

AI techniques and tools cover a wide variety of tasks. One of them is classification of objects which refers to the problem of determining the class (group) (one of at least two) into which the analyzed case should be placed. The classifier algorithm (in short, the classifier), on the basis of the training data (previously known cases) about the class to which the given cases has been placed, builds the classifier (the classifier learning phase), according to which it is possible to classify current or future cases (the phase of using the classifier in a specific situation).

Classification is related to a myriad of decision problems, such as credit scoring, identifying employees or customers who want to leave the company (churn analysis), customer segmentation, medical diagnosis, etc. Developers of AI algorithms try to objectify the evaluation of classifier quality, and the basic tool for this evaluation is the confusion matrix (error matrix). It is defined as a matrix of combinations of actual classifications and those predicted by the algorithm. For binary (two class) classification, it can be represented as a matrix (Figure 1).

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Predicted class</td></tr>
<tr><td>A</td><td>¬A</td></tr>
<tr><td rowspan="2">Actual class</td><td>A</td><td>Number of cases A classified properly as A
(true positive – TP)</td><td>Number of cases A classified improperly as ¬A
(False negative – FN)</td></tr>
<tr><td>¬A</td><td>Number of cases ¬A classified improperly as A
(false positive – FP)</td><td>Number of cases ¬A classified properly as ¬A
(true negative – TN)</td></tr>
</table>

Figure 1. Confusion matrix for binary classification.

Source: (Kohavi, Provost, 1998).

At first glance, the best classifier is the one in which the ratio of the number of correctly classified cases (actual A classified as A and actual ¬A classified as ¬A) to the total number of classified cases is as high as possible. This measure of classification quality is referred to as *accuracy* (formula 1).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Striving for higher accuracy means that we want to correctly classify as many cases of A as A and ¬A as ¬A. This is the most commonly used measure to evaluate the quality of classifiers. However, in practice, this measure is found to be insufficient (Gilli, Schumann, 2015), (Campagner, Sconfienza, Cabitza, 2020) and other measures are also used that may be more useful in specific situations.

The first of these (*precision*) is the ratio of the number of instances of A correctly classified as A to all instances classified as A (formula 2).

$$precision = \frac{TP}{TP + FP} \tag{2}$$

The pursuit of higher precision means that we do not want to classify cases ¬A as A, even at the expense of omitting some A cases.

The second measure (*recall*) is the ratio of the number of A cases correctly classified as A to all A cases (formula 3).

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Striving for higher recall means that we want to classify as many cases A as A, even at the expense of classifying cases ¬A as A.

Two next measures are *negative predictive value (npr)* (formula 4) and *true negative rate (tnr)* (formula 5). They are similar to the two previous measures but relate to the false cases.

$$npr = \frac{TN}{TN + FN} \tag{4}$$

$$tnr = \frac{TN}{TP + FN} \tag{5}$$

The meaning of these measures is difficult for average users to intuitively understand and apply. In everyday decision-making, they operate with terms they understand more, such as "I would like to take into account more cases when ...", whereas using raw algorithms they need know and interpret specific values.

Another obstacle to the proper use of AI is the inflexibility of the algorithms, which puts decision-makers in a situation where they can only accept the result delivered by the AI or - after taking into account their own knowledge - reject it. Meanwhile, decision-makers like to work online, conducting what-if simulation so that they can quickly test the effects of changes of the introduced parameters before making a final decision. However, as long as only a single-measure maximized classifier is available, this will not be possible. We illustrate the dilemmas facing the decision-makers with two examples.

In the case of credit risk assessment, a bank may identify as much more serious the risk of granting a bad loan than the risk of rejecting an application that, despite doubts, would be repaid correctly. In other words, it is more important to avoid the case of granting credit to a person/company that will have trouble paying it back than to lose the benefit of granting credit to a person/company that will pay it back without a problem. The simplest strategy would be to reject all applications because then the risk of granting bad credit disappears altogether. However, such a strategy is disadvantageous from the point of view of the bank's revenues, so the best for the decision maker would be such a classifier that rejects all doubtful applications, even at the expense of rejecting some reliable customers, with their number being as small as possible (maximize precision). In turn, when the bank is more inclined to risk, the classifier according to which the bank will be more willing to grant a loan will be more useful, even at the cost of granting it to a certain number of entities that will not be able to pay it back (greater recall).

In the opposite situation are diagnosticians, for whom the safest strategy is to refer all patients for further tests to be sure not to miss any case of disease (maximize recall). The problem arises when there are limits (cost, time) of tests and performing tests for all patients is not possible and/or the test itself is not indifferent to the patient's health. Then it is better not to test some patients than to test a healthy one. In such a situation, the best classifier for diagnosticians would be the one, which would focus on the patients who are actually ill and omit the healthy ones, even agreeing to omit some actually ill patients (maximize precision).

As can be noticed, this classifier will be different for different levels of limits, diagnostician preferences, and patient risk propensity.

Assessing classifier quality is a complex task. To solve it, many measures are used (Tharwat, 2018). One of the problem is that classical measures of classification quality refer only to the confusion matrix and are related directly to the numerical outcome of the classification. In many human independent problems it is very necessary, but in decision-making contexts it only exacerbates human-AI conflicts because decision-makers may argue that AI does not take into accounts additional preferences and as a result its solutions are detached from reality. Situation is even more complicated in the case of multi-class classification, when decision-makers want to set up preferences for many classes or each class separately. The main problem is that in the classical measures described above, there is no place to indicate decision-makers' preferences for individual classes (Kozak, Kania, Juszczuk, Mitręga, 2021).

The proposed contribution to solving the problem of human-AI collaboration is the introduction of a modified classifier quality measure that allows taking into account the user's preferences (Kozak, Probierz, Kania, Juszczuk, 2022). This makes it possible to control the choice of classifier such that the resulting classifier best reflects these preferences (formula 6).

$$preference\_driven_{\vec{\kappa}} = \frac{1}{c}\sum_{i=1}^{c}(\kappa_i * precision_i + (1-\kappa_i) * recall_i) \tag{6}$$

where:
κ_i – is the value of preference parameter for i-th class, while $\kappa_i \in [0,1]$,
c – is the number of classes.

For two classes using measures (4) and (5) *preference driven measure* (pd_2) could take the form of formula (7):

$$pd_2 = (\kappa_1 * prec + (1-\kappa_1) * rec + \kappa_2 * npr + (1-\kappa_2) * tnr)\ /2 \tag{7}$$

The value of the parameters κ_i (vector of κ values or κ vector, for short) reflects the decision-maker's preference for balancing the expectation of classification precision with the desire to avoid missing interesting cases. A higher value of κ_i for particular class, indicates an expectation of a classifier that will identify fewer instances (cases) of A, but with a higher probability that they will indeed be A, while a lower value of κ_i indicates a desire to identify as many instances of A as possible, even if by doing so the classifier will identify as A, a number of instances ¬A.

Once the decision maker has indicated a value of κ vector , a new classifier for the decision maker will be indicated to use in the decision process. Since this operation is fast, this activity can be iterative - the decision maker, after analyzing the effects of the classifier, can adjust the values of the κ_i parameters and run the simulation again. In this way, a learning loop will be realized in practice.

4. Numerical experiment

The purpose of the experiment is to investigate how changing the parameters in κ vector that describe the preferences of the decision-maker affect the choice of classifier. The experiment was conducted on two data sets: the first reflects the situation of the diagnostician described above, the second - the grantor in a bank.

4.1. Data description and experiment design

The first dataset (Breast Cancer Wisconsin (Diagnostic) Data Set) (https://archive.ics.uci.edu/ml/datasets/breast...) contains 569 cases of images with 32 attributes classified as “cancer/no cancer” and it is related to the case of diagnostician described above. The second dataset (German Credit Data Set) (https://archive.ics.uci.edu/ml/datasets/statlog...) contains data about 1000 cases of people classified as good or bad credit risk described by 20 attributes and it is related to the case of decision-maker in a bank.

Five ML algorithms have been selected for the experiments, and the results of which have a form known in management sciences and provide a high degree of diversity in the approach to the classification task. Three of them classify using decision trees: CART, C4.5, and Random Tree, and the other two generate decision rules: Decision Table and Rido. All selected algorithms are available in the free WEKA package. With its help, both sample sets were classified. As a result, for each of the data sets, five different classifiers were obtained, which can be assessed using classical measures of classification quality and the proposed measure taking into account the preferences of the decision-maker. The aim of the experiment is to test whether, when using a preference-driven measure, the classifier assessment depends on the values of κ vector given by the decision-maker.

4.2. Experiment results

Table 1 contains results of the experiment for the Breast Cancer Wisconsin (Diagnostic) Data Set (the best classifiers in terms of a specific measure are bolded). Contrary to the classical classification measures, which always return one value, the results for proposed preference-driven measure are different for different values in the κ vector. Table 1 shows numerical results for classical measures (accuracy, precision, recall) and the three sample κ vectors (p-d[default], p-d[0.7, 0.1], p-d[0.7, 0.1], where default values reflect the ratio of number of cases in each class to the number of all cases), while the full results are shown in Figure 2. Since the set of values w of the parameter κ_i is continuous, the evaluation results form surfaces.

Table 1.
Classifier evaluation results based on different measures - Breast Cancer Wisconsin (Diagnostic) Data Set

	CART	**C4.5**	**Random Tree**	**Decision Table**	**Rido**
Accuracy	0.9342	**0.9442**	0.9242	0.9285	0.9399
Recall	0.9432	0.9541	**0.9672**	0.9651	0.9476
Precision	0.9558	**0.9604**	0.9210	0.9286	0.9602
p-d[default]	0.9304	**0.9398**	0.9050	0.9120	0.9367
p-d[0.0,0.7]	0.9223	0.9357	0.9359	**0.9363**	0.9286
p-d[0.5,1.0]	0.9221	0.9356	**0.9377**	0.9375	0.9284

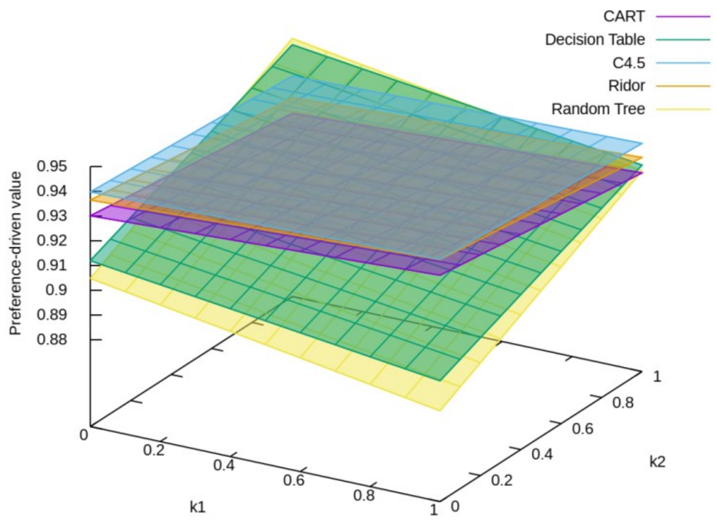

Figure 2. Solution space for Breast Cancer Wisconsin (Diagnostic) Data Set.

Figure 2 shows the evaluation for each classifier depending on κ_1 and κ_2 i.e. the exact values of the preference vector. As can be seen, the whole solution space changes the choice of classifier depending on the preference, e.g. for κ_1 close to 0.0 and κ_2 close to 1.0 the best classifiers will be Decision Table and Random Tree, while for κ_1 close to 1.0 and κ_2 close to 0.0 the best will be C4.5 and Ridor. It is also worth mentioning that for this dataset the accuracy measure indicates that the best classifier is built using the C4.5 algorithm, the same is true for the precision measure. However, the recall measure indicates the Random Tree algorithm as the best.

In the case of German Credit Data Set, the situation is slightly different (Table 2 and Figure 3). In this case, the CART algorithm turns out to be the best for each sample κ vector. However, it is worth noting, the change between the Random Tree, Decision Table and Ridor algorithms in subsequent positions in the ranking, where for κ_1 from about 0.5 to 1.0 and κ_2 from 0.0 to about 0.2 the classifier built using the Random Tree algorithm is the best rated, and for values of κ_1 about 1.0 and κ_2 about 1.0 Random Tree is the worst rated.

Table 2.
Classifier evaluation results based on different measures - German Credit Data- Data Set

	CART	**C4.5**	**Random Tree**	**Decision Table**	**Rido**
accuracy	**0.7520**	0.7350	0.7010	0.7010	0.7120
recall	0.8814	0.8729	0.8114	**0.8900**	0.8671
precision	**0.7890**	0.7764	0.7728	0.7373	0.7569
p-d[default]	**0.6588**	0.6346	0.6226	0.5580	0.5970
p-d[0.7,0.1]	**0.6418**	0.6178	0.6168	0.5337	0.5790
p-d[1.0,1.0]	**0.7041**	0.6793	0.6373	0.6203	0.6436

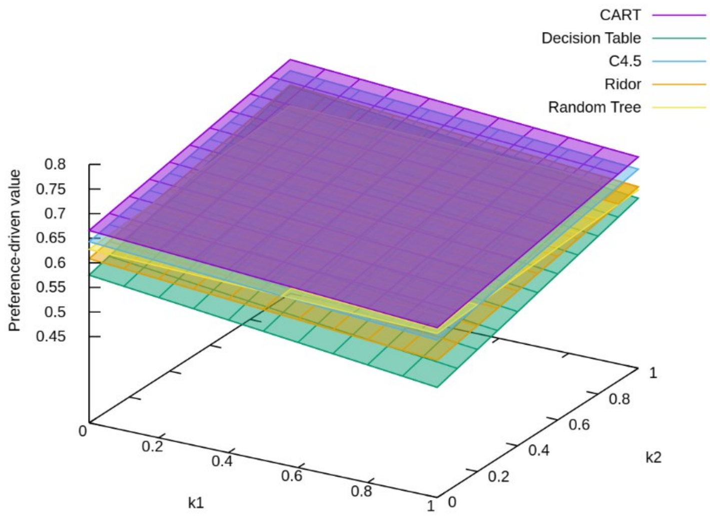

Figure 3. Solution space for German Credit Data Set.

5. Discussion

The results of the experiment showed that the best classifier depends on the parameters provided by the decision-maker (in the case of breast cancer for p-d[default] it is C4.5, for p-d[0.7, 0.1] it is Decision Table and for p-d[0.7, 0.1] it is Random Tree, while in the case of credit risk it is CART for all three sample κ vectors). By choosing a specific classifier, the decision-maker has the ability to flexibly react and take into account the current decision-making context. The use of the measure in the proposed form will contribute to increasing the flexibility of ML algorithms, and thus facilitate building the foundations of interactive and iterative human-AI cooperation. It also will help to reduce the risk of falling into the potential pitfalls of introducing AI (Buschek, Mecke, Lehmann, Dang, 2021):

- lack of interaction – by introducing learning loop and real dialog,
- conflict of territory and agony of choice – the algorithm derives classifier that takes into account the expectations of the user,
- conflict of creation & responsibility – man and AI are co-creators of the solution, and the final decision-maker is indicated by man.

One of the principles of the human-AI collaboration environment is the use of hybrid techniques involving the user in the learning process (Wu et al., 2021). It is possible at any stage of the process; from data preparation to supporting the construction of a training model. Efforts to improve the influence of users on the construction of classifiers and include them in the learning loop are undertaken in many scientific studies. Examples of suggestions can be found in (Ware, Frank, Holmes, Hall, Witten, 2001) or (Talbot, Lee, Kapoor, Tan, 2009). In the first case, the classifier is actually built by the user himself using a graphical interface, and in the second, the classifier is selected from a predefined set of classifiers. In the solution proposed in the article, it is possible to choose the classifier on-line – immediately after the parameters are set by the user. In this way, it is possible to engage humans interactively in the learning loop, which has many advantages. On the one hand, it is possible to obtain and use additional knowledge in the system that can be used to improve AI, on the other hand, the decision-maker slowly gets used to the presence of AI and gradually incorporates it into his decision-making process (Hoi, Sahoo, Lu, Zhao, 2021). However, the condition for such action is real, not mocked, interactivity and the ability of the system to respond to changing user preferences.

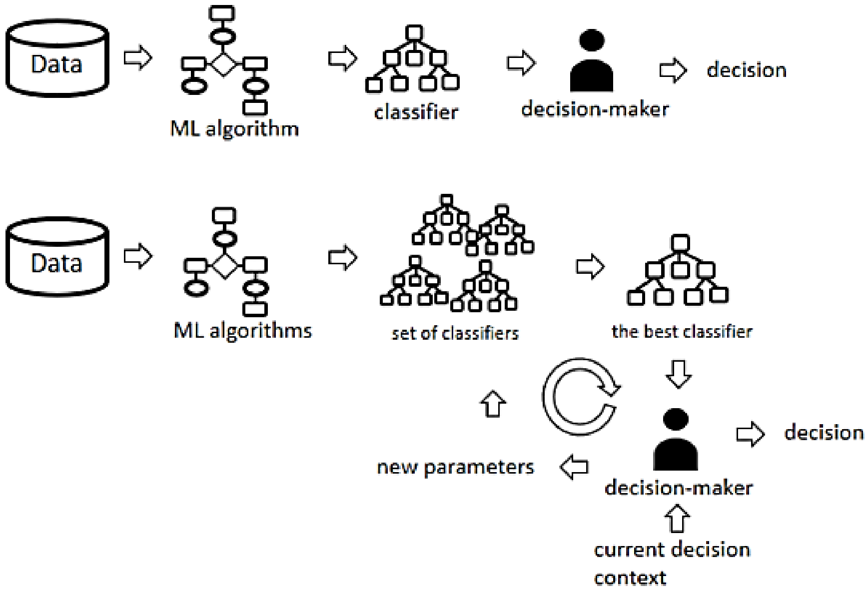

Figure 4. Difference in decision-making process after introducing parametrized classification measure.

Introducing the described vector of κ parameters to the decision procedure and the possibility of checking their effects leads to the construction of a learning loop (Figure 4) related to model training, the purpose of which is to build an appropriate classifier. Including the user in building AI solutions speeds up the learning process, improves its results and the end result is more credible for other people (Amershi, Cakmak, Knox, Kulesza, 2014).

6. Conclusions and future works

This paper presents the causes of human-AI conflicts and the problems of building a collaborative environment. The results show that an environment using a modified measure of classification quality does indeed take preferences into account and changes the classifier finally used. It is also shown that using a measure that takes preferences into account can help overcome some of the causes of these conflicts and build such an environment.

The use of the measure itself is automatic, but on the organizational side, it seems necessary to train decision-makers so that they can reflect complex decision contexts with parameters. On the IT side, it is also necessary to prepare an appropriate interface for the decision-makers, so that they can flexibly set the parameters of the algorithm for choosing the classifier without knowing the concepts strictly related to machine learning.

References

1. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105-120. https://doi.org/10.1609/aimag.v35i4.2513.
2. Buschek, D., Mecke, L., Lehmann, F., Dang, H. (2021). Nine Potential Pitfalls when Designing Human-AI Co-Creative Systems. *CEUR Workshop Proceedings*, *2903*. College Station. USA.
3. Campagner, A., Sconfienza, L., Cabitza, F. (2020). H-Accuracy, an alternative metric to assess classification models in medicine. *Studies in Health Technology and Informatics*, *270*, 242-246. https://doi.org/10.3233/SHTI200159.
4. Čartolovni, A., Tomičić, A., Lazić Mosler, E. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review. *International Journal of Medical Informatics*, *161*(March). https://doi.org/10.1016/j.ijmedinf.2022.104738.
5. Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K.R., Leibo, J.Z., … Graepel, T. (2020). *Open Problems in Cooperative AI,* August 2019. Retrieved from: http://arxiv.org/abs/2012.08630.
6. Duan, Y., Edwards, J.S., Dwivedi, Y.K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, *48*(February), 63-71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021.
7. Ferreira, J., Monteiro, M. (2021). The human-AI relationship in decision-making:

AI explanation to support people on justifying their decisions. *Transparency and Explanations in Smart Systems (TExSS) Workshop, Held in Conjunction with ACM Intelligent User Interfaces (IUI)*. Retrieved from: https://arxiv.org/ftp/arxiv/papers/2102/2102.05460.pdf.

8. Formosa, P., Rogers, W., Bankins, S., Griep, Y., Richards, D. (2022). Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. *Computers in Human Behavior*, *133*(*March*), 107296. https://doi.org/10.1016/j.chb.2022.107296.
9. Gilli, M., Schumann, E. (2015). Accuracy and Precision in Finance. *SSRN Electronic Journal*, *September*. https://doi.org/10.2139/ssrn.2698114.
10. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z. (2019). XAI-Explainable artificial intelligence. *Science Robotics*, *4*(*37*), 0-1. https://doi.org/10.1126/scirobotics.aay7120.
11. Hoi, S.C.H., Sahoo, D., Lu, J., Zhao, P. (2021). Online learning: A comprehensive survey. *Neurocomputing*, *459*, 249-289. https://doi.org/10.1016/j.neucom.2021.04.112.
12. Janssen, M., van der Voort, H., Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, *70*, 338-345. https://doi.org/10.1016/j.jbusres.2016.08.007.
13. Jarrahi, M.H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577-586. https://doi.org/10.1016/j.bushor.2018.03.007.
14. Kohavi, R., Provost, F. (1998). Glossary of Terms. *Machine Learning*, *30*, 271-274. https://doi.org/https://doi.org/10.1023/A:101718182689.
15. Kozak, J., Kania, K., Juszczuk, P., Mitręga, M. (2021). Swarm intelligence goal-oriented approach to data-driven innovation in customer churn management. *International Journal of Information Management*, *60*(*November 2020*). https://doi.org/10.1016/j.ijinfomgt.2021.102357.
16. Kozak, J., Probierz, B., Kania, K., Juszczuk, P. (2022). Preference-Driven Classification Measure. *Entropy*, *24*(4), 531. https://doi.org/10.3390/e24040531.
17. Langer, M., Landers, R.N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, *123(October 2020),* 106878. https://doi.org/10.1016/j.chb.2021.106878.
18. Leyer, M., Dootson, P., Oberländer, A.M., Kowalkiewicz, M. (2020). Decision-making with artificial intelligence: Towards a novel conceptualization of patterns. *Proceedings of the 24th Pacific Asia Conference on Information Systems: Information Systems (IS) for the Future, PACIS 2020* (Heathmann 2018), 1-11. Dubai, UAE.
19. Leyer, M., Schneider, S. (2021). Decision augmentation and automation with artificial intelligence: Threat or opportunity for managers? *Business Horizons*.

https://doi.org/10.1016/j.bushor.2021.02.026.

20. Mackeprang, M., Müller-Birn, C., Stauss, M. (2019). Discovering the sweet spot of human—computer configurations: A case study in information extraction. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). https://doi.org/10.1145/3359297.
21. Mikalef, P., Krogstie, J. (2019). Investigating the data science skill gap: An empirical analysis. *IEEE Global Engineering Education Conference, EDUCON*, *April*, 1275-1284. https://doi.org/10.1109/EDUCON.2019.8725066.
22. Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*(*August*), 100005. https://doi.org/10.1016/j.jrt.2020.100005.
23. Samek, W., Montavon, G., Vedaldi, A., Hansen, L., Müller, K. (2019). *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. M. Samek, G. Montavon, A. Vedaldi, L.K. Hansen, K.-R. Müller (eds.). https://doi.org/10.1007/978-3-030-28954-6.
24. Schwenk, C. (1990). Conflict in organizational decisomaking: an exploratory study of its effects in for-profit and not-for-profit organizations. *Management Science*, *36*(4), 436-448. Retrieved from: http://dx.doi.org/10.1016/j.jaci.2012.05.050.
25. Talbot, J., Lee, B., Kapoor, A., Tan, D.S. (2009). EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1283-1292.
26. Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168-192. https://doi.org/10.1016/j.aci.2018.08.003.
27. Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., Wang, Q. (2020). *From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people.* Conference on Human Factors in Computing Systems - Proceedings, 1-6. https://doi.org/10.1145/3334480.3381069.
28. Ware, M., Frank, E., Holmes, G., Hall, M., Witten, I.H. (2001). Machine Learning: letting users build classifiers. *International Journal of Human-Computer Studies*, *55*(3), 281-292.
29. Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., He, L. (2021). A Survey of Human-in-the-loop for Machine Learning. *ArXiv*, *4*(48). Retrieved from: http://arxiv.org/abs/2108.00941.
30. Zhang, Z., Chen, Z., Xu, L. (2022). Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. *Journal of Experimental Social Psychology*, *101*(*March*), 104327. https://doi.org/10.1016/j.jesp.2022.104327.