Explainable machine learning for healthcare decision-making tasks

Marina Yelis[1]*, Yan Kuchin[2], Adilkhan Symagulov[2], Elena Muhamedieva[2]

[1]Satbayev University, Kazakhstan
[2]Institute of Information and Computational Technologies MES RK

**Corresponding author's e-mail: k.marina92@gmail.com*

Abstract

Recently, artificial intelligence (AI) has made great strides due to the rapid development of machine learning technologies. Despite this, there are potential risks associated with a "black box" approach to learning. Unlike classical ML methods, where model results can usually be explained, deep model learning lacks transparency, making it difficult to understand how the model made a particular decision. In this paper, we consider Explainable Machine Learning (EML) methods as applied to the construction of a decision support system in healthcare. The essence of the proposed approach is to build an explanatory machine learning model and then use the explanations to make management recommendations for healthcare organizations. To evaluate the applicability of the approach, we used the "Hospital General Information" open dataset. The results show that the explanatory machine learning model correctly ranked the key hospital performance measures. This confirms the applicability of EML as part of a decision support system for health care.

Keywords: explainable machine learning, multi-criteria decision support system, GeoAI, SHapley Additive exPlanations (SHAP), black box explanation

1 Introduction

Advances in technology, improvements in learning algorithms, and access to large amounts of data have enabled advances in machine learning (ML), leading to its widespread industrial adoption.

As a result, many real-world problems are solved today using machine learning models. Machine learning is successfully used to solve problems in medicine [1, 2], biology [3], robotics, urban agriculture [4] and industry [5, 6], agriculture [7], modeling environmental [8] and geo-environmental processes [9], communication system design [10], astronomy [11], petrographic research [12, 13], geological exploration [14], natural language processing [15, 16], etc.

However, when decisions ultimately affect people's lives (e.g., in medicine, law, finance, or defense), there is a need to understand how such decisions are provided by AI methods [17]. This is a major barrier to the practical application of machine learning.

Recently obtained results in the development of explanatory systems [18, 19] allow not only to apply them to assess the weight of machine learning model features, but also, in our opinion, to use them in the decision support process.

2 Method

Our approach, which be called as MCDSS based on the explanation of "black boxes" (MCDSS&BBE). It was discussed in [20]. MCDSS&BBE consists of the following key elements (Figure 1).

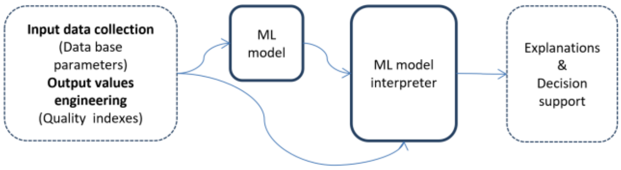
Input data collection
(Data base parameters)
Output values engineering
(Quality indexes)
ML model
ML model interpreter
Explanations & Decision support

FIGURE 1 MCDSS&BBE workflow schema

First, we collect input data and define targets, indicators that can be synthesized, or be some aggregation of external indicators.

Secondly, we build a non-linear model based on the controlled learning method (ML model), where we take into account the maximum number of available parameters.

Thirdly, we estimate the contribution of parameters to the result achieved by the model as a whole and by a separate object (using the model interpreter ML).

Fourthly, we use the results of the interpretation to develop recommendations.

The ML model considered as a way to answer the question "Why do we have this or that classification or regression result? The answer to this question for a single object is a recommendation to change input parameters to increase the values of target indicators (quality indicators). In other words, we consider the data model as a way to describe the regularities of real systems and form recommendations for their change. The SHAP interpreter [19] is used to solve the multicollinearity problem.

The application of MCDSS&BBE was described in [21].

The development of MCDSS&BBE is an element of the GeoAI Healthcare system designed to make recommendations for the management of healthcare organizations.

To assess the applicability of this approach, we used elements of MCDSS&BBE to analyze the Hospital General Information open dataset provided by Medicare.

3 Results

Dataset has the overall hospital rating, which we use as our target in the building model, and the other 7 features which affect that overall rating. To build the recommendation system we have to figure out dependencies, how each feature affects the overall score. To do so machine learning models were built and trained to predict quality features. As a result of the experiments, a CatBoost regression model was applied [22]. Then using SHAP, the model's features were ranked in terms of their weight in the process of getting the result. In Figure 2, each line corresponds to a specific feature, which is sorted by significance (SHAP value) in descending order. Each point represents the value of each indicator and its impact on the overall rating. Its position along the horizontal axis shows how negative (left) or positive (right). At the same time, the color of the dot indicates the value of the factor - the red dots have a high value of this factor (higher than the average value in the sample), and the blue color, respectively, a low value.

The overall hospital quality rating includes more than 100 indicators, divided into 7 groups or categories of indicators: Mortality, Safety of Care, Readmission, Patient Experience, Effectiveness of Care, Timeliness of Care, and Efficient Use of Medical Imaging. The data source states that a statistical model known as the hidden variable model is used. Seven different hidden variable models are used to calculate scores on 7 groups of indicators. The hospital's cumulative score is then calculated by taking the weighted average of these group scores.

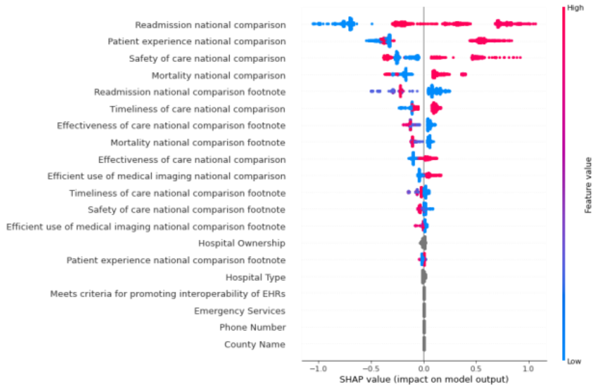

FIGURE 2. The influence of factors on the feature “overall hospital rating”.

According to Figures 2, we can see the influence of individual features on the results of the regression model. If we compare the results with the main indicators mentioned above, we see that the 4 main indicators have a strong influence on the hospital overall rating. In this way, we show that it is possible to build recommendation systems to improve the quality of hospital services by applying machine-learning model.

4 Conclusion

The explanatory system built based on a machine learning regression model correctly ranked the main parameters of hospitals. Consequently, the influence of these factors is correctly reflected in the built model and can be applied to assess the quality of a medical institution.

The results confirm the hypothesis that Models of machine learning (explainable ML - EML) can be applied to support decision-making on the management of health care facilities.

The main limitations of this approach are:

- Difficulties in interpreting the results of explanations by applied specialists. In other words, the results obtained with EML still need further interpretation for end users.
- If the parameters under consideration have a complex semantic interpretation or do not have it at all, the use of current EML models makes no sense.
- Objectives of further research:
- Use the proposed scheme of EML application in other subject areas.
- Development of automatic or semi-automatic systems to explain the obtained results.

5 Acknowledgments

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. ИРН AP09259587). The authors would like to thank prof. Ravil Mukhamediev (SU) for the consultations and help with the paper preparation, prof. Elena Zaitseva and prof. Vitaly Levashenko from Dept. of Informatics, University of Žilina for their valuable feedback and cross-country collaborations.

References

[1] Cruz J, Wishart D *Applications of Machine Learning in Cancer Prediction and Prognosis Cancer Informatics* 2006 pp **59–77**

[2] Miotto R, Wang F, Wang S, Jiang X, Dudley J *Deep learning for healthcare: review, opportunities and challenges: Briefings in bioinformatics* 2017 pp **1236-1246**

[3] Ballester Pedro, John BO Mitchell *A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking: Bioinformatics* 26 V 9 2010 pp **1169-1175**

[4] Mahdavinejad M, Mohammadreza R, Mohammadamin B, Peyman A, Payam B, Amit S *Machine learning for Internet of Things data analysis: A survey. Digital Communications and Networks* 4 no 3 2018 pp **161-175**

[5] Charles F, Worden K *Structural health monitoring:* a machine learning perspective 2012 pp **66**

[6] Lai J, Qiu J, Feng Z, Chen J, Fan H *Prediction of soil deformation in tunnelling using artificial neural networks:* Computational Intelligence and Neuroscience 2016

[7] Liakos K, Busato P, Moshou D, Pearson S, Bochtis D Machine learning in agriculture: A review: *Sensors* 2018

[8] Recknagel F *Application Of macine Learning To Ecological Modelling*: Ecological Modelling V 146 2001 pp *303-310*

[9] Tatarinov V, Manevich A, Losev I *System approach to geodynamic zoning on the basis of artificial neural networks:* Mining Sciences and Technology 2018 pp **14-25**

[10] Charles C, Hecker J, Stuntebeck E, Shea T *Applications of machine learning to cognitive radio networks:* Wireless Communications 2007 pp **47-52**

[11] Ball N, Brunner R, *Data mining and machine learning in astronomy*: Journal of Modern Physics 2010 pp **1049-1106**

[12] Muhamediyev R, Amirgaliev E, Iskakov S, Kuchin Y, Muhamedyeva E *Integration of Results of Recognition Algorithms at the Uranium Deposits*: Journal of ACIII 2014 pp **347-352**

[13] Kuchin Y, Mukhamediev R, Yakunin K *One method of generating synthetic data to assess the upper limit of machine learning algorithms performance*: Cogent Engineering 2020

[14] Chen Y, Wu W *Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data Geochemistry*: Exploration, Environment, Analysis 2017 pp **231-238**

[15] Hirschberg J, Manning C *Advances in natural language processing*: Science 2015 pp **261-266**

[16] Goldberg Y *A primer on neural network models for natural language processing:* Journal of Artificial Intelligence Research 2016 pp **345-420**

[17] Arrieta A, a D´ıaz-Rodr´ıguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F Explainable *Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible* AI: Information Fusion 2020 pp **82-115**

[18] Ribeiro M, Guestrin S *Local Interpretable Model-Agnostic Explanations (LIME): An Introduction A technique to explain the predictions of any machine learning classifier* 2016

[19] Lundberg S, Lee S A unified approach to interpreting model predictions: *Advances in Neural Information Processing Systems* 2017 pp **4765-4774**

[20] Muhamedyev R, Yakunin K, Kuchin Y, Symagulov A, Murzakhmetov S, Abdurazakov A *The use machine learning interpreter for the development of decision support system:* The 18th International Scientific Conference Information Technologies And Management 2020 pp **19-20**

[21] Muhamedyev R, Yakunin K, Kuchin Y, Symagulov A, Buldybayev S, Murzahmetov S, Abdurazakov A *The use of machine learning "black boxes" explanation systems to improve the quality of school education:* Cogent Engineering 2020

[22] CatBoostRegressor documentation, **E-resource**: https://catboost.ai/docs/concepts/python-reference_catboostregressor.html#python-reference_catboostregressor__purpose