New Concept to Multi-Criteria Model Automatization - Machine Learning Based Approach

Jarosław Wątróbski
University of Szczecin;
National Institute of Telecommunications
Szczecin; Warsaw, Poland   *jaroslaw.watrobski@usz.edu.pl*

Aleksandra Bączkiewicz
University of Szczecin
Szczecin, Poland   *aleksandra.baczkiewicz@usz.edu.pl*

Iga Rudawska
University of Szczecin
Szczecin, Poland   *iga.rudawska@usz.edu.pl*

Abstract

The commonness of information systems based on machine learning (ML) models and multi-criteria decision analysis (MCDA) methods is growing due to the increasing dimensions of data required for processing. This paper presents a hybrid framework combining the MCDA method with ML models to predict the rankings of countries considering the fulfillment of Sustainable Development Goal 7 based on the identified preferences of decision-makers. The results proved that the proposed approach can be regarded as a functional tool for multi-criteria assessment in the case of inaccessibility of experts' knowledge. The proposed approach enables to mitigate the shortcomings of MCDA methods arising from the necessity to engage decision-makers.

Keywords: Decision support systems, Machine learning, Ranking prediction, Multi-Layer Perceptron Regressor, Multi-criteria decision analysis

1. Supplementary material

1.1. Criteria model

Among indicators of SDG 7 included by us in the proposed model are eight criteria with the goal of maximization: $\{C_1$ - Primary energy consumption in Tonnes of oil equivalent (TOE) per capita, C_2 - Final energy consumption in TOE per capita, C_3 - Final energy consumption in households per capita in Kilogram of oil equivalent (KGOE), C_4 - Energy productivity in Euro per KGOE, C_5 - Share of renewable energy sources (RES) in gross final energy consumption in %, C_6 - Share of RES in transport in gross final energy consumption in %, C_7 - Share of RES in electricity in gross final energy consumption in %, C_8 - Share of RES in heating and cooling in gross final energy consumption in %}. The framework includes also two criteria with the goal of minimization: $\{C_9$ - Energy import dependency by products in %, C_{10} - Population unable to keep home adequately warm in %}. Mentioned attributes C_1-C_{10} denotes features of the input dataset. The data for these indicators for years 2013-2022 was collected from the Eurostat database [1] (accessed on 24 March 2024). We have used dataset EU policies $\rightarrow$ Sustainable development indicators $\rightarrow$ Goal 7 - Affordable and clean energy.

The fragment of the training dataset before preprocessing, namely normalization, is visualized in Table 1. The fragment of the training dataset after the normalization procedure is

---

[1] https://ec.europa.eu/eurostat/data/database

displayed in Table 2.

Table 1. Fragment of training dataset before preprocessing and target variable (Score).

Symbol	Country	Year	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	Score
A_1	Belgium	2013	4.36	3.29	818	5.97	7.671	5.088	12.547	7.639	77.777	5.8	0.4258
A_2	Bulgaria	2013	2.27	1.21	308	2.28	18.898	5.886	18.682	29.234	38.313	44.9	0.1742
A_3	Czechia	2013	3.87	2.3	691	3.67	13.927	6.445	12.781	17.706	27.58	6.2	0.4141
A_4	Denmark	2013	3.17	2.51	799	13.19	27.173	6.456	43.084	34.679	12.312	3.8	0.5674
A_5	Germany	2013	3.82	2.74	775	8.07	13.757	7.299	25.284	13.408	62.411	5.3	0.4693
...	...	...	...	...	...	...	...	...	...	...	...	...	...
A_{29}	Norway	2021	4.86	3.64	864	12.81	74.018	20.66	113.642	32.274	-616.598	1.4	0.6986

Table 2. Fragment of preprocessed dataset and target variable (Score).

Country symbol	Year	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	Score
A_1	2013	0.1654	0.2831	0.6081	0.3639	0.0594	0.3121	0.1042	0.0588	0.0431	0.8886	0.4258
A_2	2013	0.0437	0.0154	0.1265	0.0388	0.2191	0.3658	0.1624	0.4077	0.1075	0.0000	0.1742
A_3	2013	0.1369	0.1557	0.4882	0.1612	0.1484	0.4034	0.1064	0.2214	0.1250	0.8795	0.4141
A_4	2013	0.0961	0.1828	0.5902	1.0000	0.3369	0.4041	0.3941	0.4957	0.1500	0.9341	0.5674
A_5	2013	0.1340	0.2124	0.5675	0.5489	0.1460	0.4608	0.2251	0.1520	0.0682	0.9000	0.4693
...	...	...	...	...	...	...	...	...	...	...	...	...
A_{29}	2021	0.2346	0.3453	0.5695	0.4638	0.8411	0.6709	1.0000	0.2961	1.0000	0.9824	0.6986

1.2.   Research results

Table 3 includes values of performance evaluation metrics such as MAE, MSE, and R^2 for investigated ML models. The results show that the performance of examined ML models is high and comparable.

Table 3. Values of performance evaluation metrics achieved by examined ML regression models.

Test dataset size, Regression model	MAE	MSE	R^2
105 samples MLP	0.0088	0.0001	0.9896
105 samples LR	0.0085	0.0001	0.9909
105 samples KRR	0.0184	0.0005	0.9618
105 samples GBR	0.0100	0.0002	0.9873
105 samples BRR	0.0085	0.0001	0.9909
30 samples MLP	0.0257	0.0007	0.9617
30 samples LR	0.0252	0.0007	0.9649
30 samples KRR	0.0321	0.0012	0.9380
30 samples GBR	0.0191	0.0005	0.9734
30 samples BRR	0.0254	0.0007	0.9644

The Weighted Spearman rank correlation coefficient was employed to compare the convergence of predicted rankings with factual reference ranking and among themselves. The high values of this coefficient that approach 1 prove high convergence of compared rankings. The obtained correlation values are displayed in Figure 1.
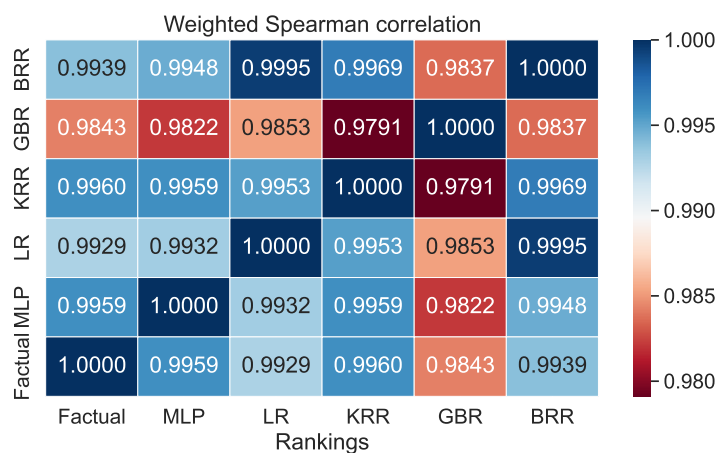
Fig. 1. Correlation of factual ranking in 2022 with rankings predicted by examined ML regression models.

Acknowledgements