Multidiszciplináris tudományok, 13. kötet, 2. sz. (2023), pp. 289-296 **https://doi.org/10.35925/j.multi.2023.2.25**

CAN WE PREDICT THE CRITICALITY NOW BETTER THAN IN 2017?

Kitti Fodor
assistant lecturer, University of Miskolc, Institute of Economic Theory and Methodology
3515 Miskolc-Egyetemváros, e-mail: kitti.fodor@uni-miskolc.hu

Beatrix Varga
associate professor, University of Miskolc, Institute of Economic Theory and Methodology
3515 Miskolc-Egyetemváros, e-mail: beatrix.varga@uni-miskolc.hu

Abstract
Critical raw materials are part of our daily lives, despite what we may not think. For example, they are found in cars, but they are used in many other areas. Our aim in this study was to examine how the critical raw materials sector has changed over the last 6 years. Our analysis was carried out using logistic regression. We observed that in 2023, we can more accurately categorise raw materials using the SR and EI parameters than in 2017. The model was able to correctly categorise the raw materials with an accuracy of 96.6%.

Keywords: *critical raw materials, logistic regression, change, classification*

1. Introduction

There are some objects that we can hardly imagine our lives without, such as mobile phones, laptops, cars, televisions, etc. You would not think that these products require raw materials that are considered critical raw materials. In technological innovation, critical raw materials play a key role, as they are essential raw materials for some innovations. But what are critical raw materials?

"Critical Raw Materials (CRMs) are those raw materials which are economically and strategically important for the European economy, but have a high-risk associated with their supply. Used in environmental technologies, consumer electronics, health, steel-making, defence, space exploration, and aviation, these materials are not only 'critical' for key industry sectors and future applications, but also for the sustainable functioning of the European economy." (CRM Alliance)

Critical raw materials have been under EU review for more than a decade, and every few years the EU produces an up-to-date list of critical raw materials.

2. CRM and EU

The EU has been monitoring critical raw materials for many years. However, the CRM list was created for a number of purposes, including to encourage the mining and recycling of CRMs within the EU and to make countries, companies and investors aware of the potential risks of critical stocks of these raw materials.

Criticality is basically assessed along two main factors, Supply Risk (SR) and Economic Importance (EI). For both factors, there is a cut-off point above which a given raw material is considered critical. The figure below shows the raw materials studied by the EU and their categorisation.

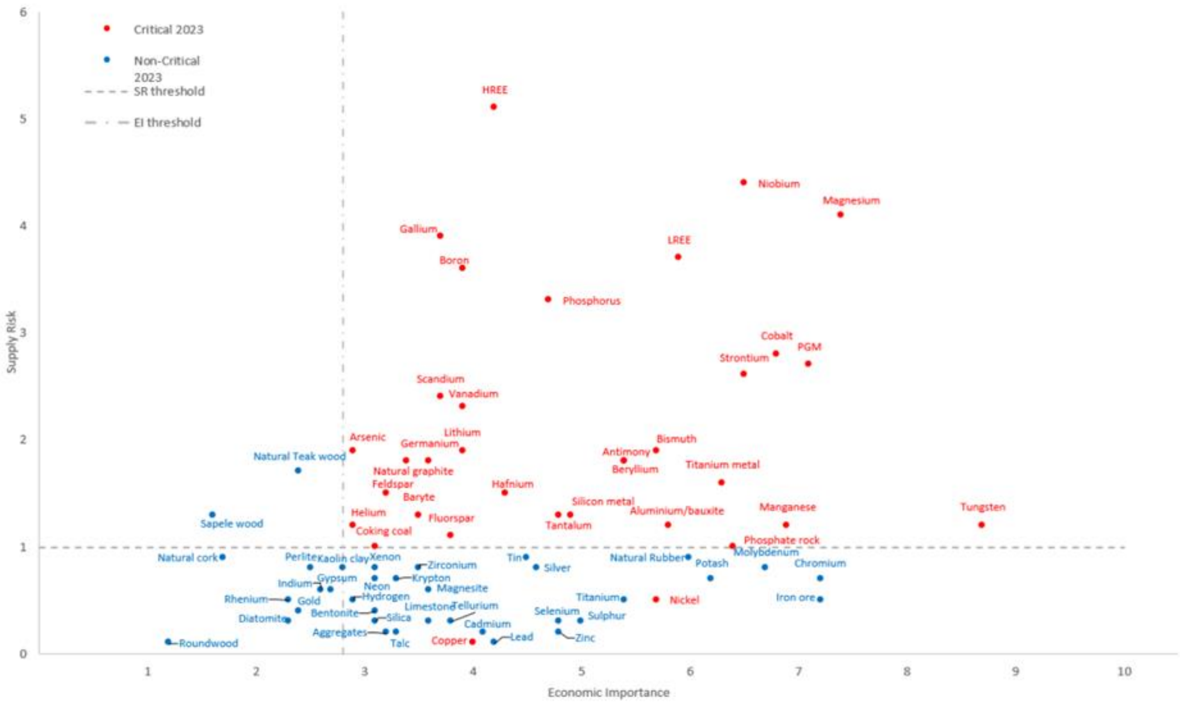

Figure 1. *Result of the 2023 EU criticality assessment*
Source: European Commission 2023

It can be observed that the threshold of SR is 1.0 and the threshold of EI is 2.8.

Over the last more than a decade, the EU has examined these raw materials 5 times, each time producing a list of raw materials identified as critical. The list has been updated in the following years:

- 2011
- 2014
- 2017
- 2020
- 2023

While on the first list was only 14 critical raw materials, this list nearly doubled by 2017, and by 2023 50 raw materials will be listed as critical. Once a raw material is on the list, it does not mean that it will be critical forever, but there are raw materials that have been on every list since 2011:

- Antimony
- Beryllium
- Cobalt
- Flourspar
- Gallium
- Germanium
- HREE
- Indium

- LREE
- Magnesium
- Natural graphite
- Niobium
- PGMs
- Tungsten (European Commision, 2023)

3. Previous research

In 2019, we conducted several studies on critical raw materials, using a variety of multivariate statistical methods. The methods used were logistic regression and cluster analysis.

Logistic regression was used to form two equations. The explanatory variables of the first model were the SR and EI parameters and categorized 5 raw materials incorrectly. And the explanatory variables of the second model were EU import reliance, end-of-life recycling input rate and major world producers share, and 8 items were misclassified. These two equations formed the complex model, in which a raw material was considered critical if both equations indicated that it was. The complex model reduced the 5 misclassifications in the initial model to 3. The following 3 raw material categorisations were incorrect:

- Hafnium
- Silicon metal
- Tantalum.

During the cluster analysis, 5 homogeneous clusters were created and named as follows:

- Innovation dependent
- Treasures of Asia
- Emerging Criticals
- Harmless
- Recyclables

For the third cluster it was found that, hogy the analysis of the EU confirms that these raw materials are becoming increasingly important. The three raw materials that could not be correctly categorised by the logistic regression were all in this cluster and became critical raw materials by 2023. (Varga et al., 2019; Varga et al., 2021)

4. Database and methodology

4.1. Database

For our study, we have created a database which is based on the research of the European Commission. 87 raw materials were included in the database, and 56.3% of these elements were categorized as critical in 2023. The database includes 7 variables, for example, the name of the raw material, supply risk, economic importance, end-of-life recycling input rate, critical classification in 2023.

4.2. Logistic regression

The aim is to classify observation units into predefined groups of the dependent variable. The model is based on the "odds", the value of which is calculated by the model. This gives the probability of being in each group. If we want to express this in a formula, we can do so as follows

$$odds_x = \frac{P_x}{1-P_x}, \rightarrow \; odds_x = \frac{P_{not\ critical|x}}{1-P_{not\ critical|x}}$$

In logistic regression, we assume that the logarithm of the odds can be defined as a linear function of the explanatory variables, which can be written as follows:

$$\ln(odds_x) = logit(P_x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

After further transformations, we arrive at the following probability of non-criticality:

$$P_x = \frac{e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\ldots+\beta_p x_p}} = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}}}{1+e^{\boldsymbol{\beta}^T \boldsymbol{x}}}$$

The classification also requires a so-called cut point value , the role of which is that if the calculated probability exceeds this value, the given raw material becomes critical in the model. This value can of course be varied, but it is advisable to set it in such a way that the loss due to misclassification is as small as possible. (Hajdu, 2003)

An important aspect in conducting the analysis is the model fit, which in the case of logistic regression can be determined by Cox&Snell R2 and Nagelkerke R2. It captures the explanatory power of the model. (Malhotra, 2008; Székelyi, Barna, 2002)

If the coefficient is considered significant, it can be said to contribute to the analysis. (Hajdu, 2003; Malhotra, 2008)

Also in the case of logistic regression, a classification matrix contains the classification results and accuracy.

As with all statistical methods, logistic regression has its disadvantages, including its high sensitivity to multicollinearity and outliers, which are therefore of high importance to manage, and its predictive ability for large sample sizes. However, it has the advantage of requiring less conditions to be fulfilled.

4.3. Evaluation of the method

There are several ways to evaluate classification models:

- Classification matrix
- ROC curve
- Gini coefficient
- Kolmogorov-Smirnov test

The Kolmogorov–Smirnov test is not one of the most widely used methods, due to the approximate nature of the method, so I did not use this assessment option in the research.

4.3.1. Classification matrix

The essence of the classification matrix, compares observed and predicted group memberships, thereby determining the proportion of cases that, overall, can be categorised correctly. *Table 1* shows an example for classification matrix.

The classification table is one of the simplest ways to measure the performance of models.

Table 1. *Classification matrix*

		Predicted		
		Not critical	Critical	
Observed	Not critical	TN	FP	Specificity
	Critical	FN	TP	Sensitivity
				Accuracy

Source: Own editing, based Quantitative Statistical methods – Logistic regression

4.3.2. ROC curve

The curve is presented in a square of unit sides, with sensitivity on one axis and 1-specificity on the other. The curve connects the sensitivity and 1-specificity values for the different cut points. The upper left corner symbolises the perfect classification. The different curves can be compared by the area under the curve, which is used in the literature as AUC. If the AUC is around 80-90%, it is considered to be outstanding. The following picture shows some examples.

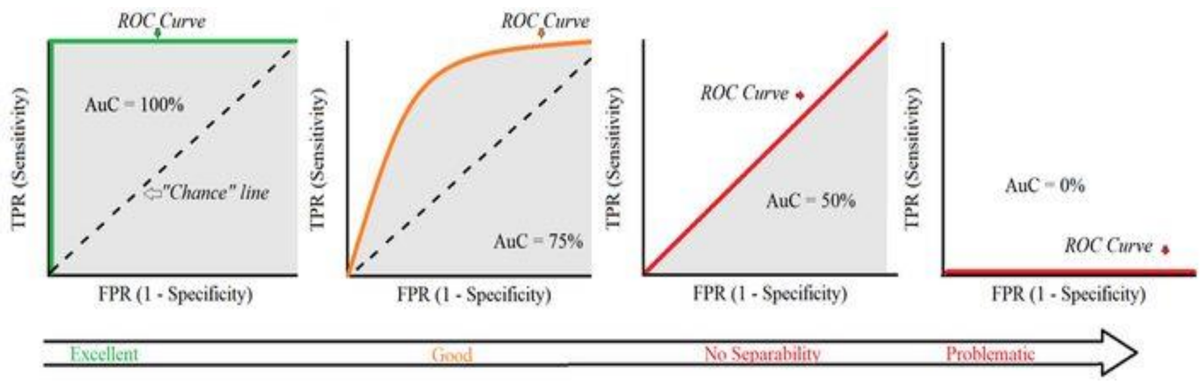

Figure 2. *Some possible ROC curves*
Source: https://www.datasciencecentral.com/roc-curve-explained-in-one-picture/

4.3.3. Gini coefficient

There are several ways to determine the indicator, the Gini coefficient can be determined from the AUC value calculated for the ROC curve using the formula:

$$Gini = 2\,(AUC - 0{,}5)$$

For the different Gini coefficient (AR) values, the following interpretations have been used:
– AR < 60%: wrong model

– 60% < AR < 70%: correct model
– 70% < AR < 80%: good model
– 80% < AR < 90%: best predictive model
– 90% < AR < 100%: up to 97%, it is possible to have an exceptionally good model, but it is worth revisiting the results in this case (Engelman et al, 2003; Olawale, 2020)

5. Empirical research

In this research, we examined the accuracy with which raw materials can be categorised along the two main parameters. The dependent variable was whether the raw material was critical in 2023 and the explanatory variables were SR and EI. The Omnibus test ($p < 0.001$) and the Hosmer and Lemeshow goodness-of-fit test ($p = 0.922$) showed a reliable model with a good fit. The generated model has high explanatory power (Nailkerke R2 = 90.4%).

The model equation can be written in the following form:

$$P_{(critical)} = \frac{e^{0,00001+1,00644x_1+1,08114x_2}}{1+e^{0,00001+1,00644x_1+1,08114x_2}}$$

where,
x_1: EI*100
x_2: SR*100

The performance of the model was evaluated on 3 criteria. First, the classification matrix was examined. For the classification matrix, we chose a value smaller than the default value of 0.5, which was determined by the Youden statistic, so the cut point chosen was 0.464.

Table 2. *Classification matrix*

Observed		Predicted		
		C2023		Percentage Correct
		0	1	
C2023	0	37	1	97.4
	1	2	47	95.9
Overall Percentage				96.6
a. The cut value is ,464				

Source: Own editing, based Quantitative Statistical methods – Logistic regression

Based on the classification matrix, we can conclude that the model has a high classification accuracy, 3 raw materials could not be categorised properly, so overall we got a more favourable result now than when we examined the 2017 raw material list.

The second aspect was the ROC curve and the area under the curve.

The *Figure 3* shows that the curve approaches the upper left corner, the AUC value is 99.1%, which is considered to be outstanding.

The last aspect was the Gini coefficient. A similar conclusion can be drawn for the Gini coefficient, where a value above 70% indicates a very strong model, in this case the value is 98.2%.

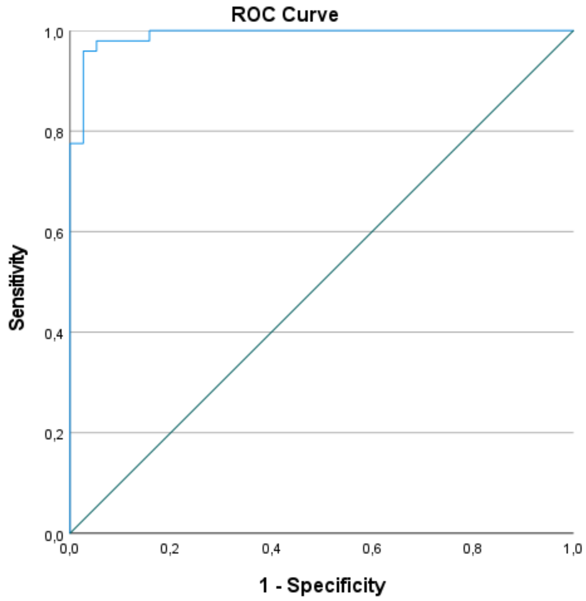

Figure 3.

6. Summary

Overall, the number of critical raw materials has increased since 2017. Based on the 2017 values, 5 raw materials were incorrectly categorised using the main parameters. However, for 2023, only 3 raw materials were incorrectly categorised. Based on the logistic regression analysis, it can be concluded that the change in SR value has a greater effect on the criticality. Previous years and the current research have confirmed that this is still an area worth looking at, because it is a changeable area and it is difficult to predict what the future holds.

References

[1] CriticEl. *Kritikus Elemek*. **http://kritikuselemek.uni-miskolc.hu/**

[2] *CRM Alliance*. **http://criticalrawmaterials.org/**

[3] Engelmann, B., Hayden, E., Tasche, D. (2003). *Testing rating accuracy*. 2003 január. Credit Risk. **www.risk.net**

[4] European Commission (2017). *Methodology for establishing the EU list of critical raw materials* (Publications Office of the EU, Luxembourg). **https://publications.europa.eu/en/publication-detail/-/publication/2d43b7e2-66ac-11e7-b2f2-01aa75ed71a1**

[5] European Commission (2017). *Study on the review of the list of Critical Raw Materials – Criticality Assessment* (Publications Office of the EU, Luxembourg). **https://publications.europa.eu/en/publication-detail/-/publication/08fdab5f-9766-11e7-b92d-01aa75ed71a1/language-en**

[6] European Commission (2017). *Study on the review of the list of Critical Raw Materials – Critical Raw Materials Factsheets* (Publications Office of the EU, Luxembourg). **https://publications.europa.eu/en/publication-detail/-/publication/7345e3e8-98fc-11e7-b92d-01aa75ed71a1/language-en**

[7] European Commission (2017). *Study on the review of the list of Critical Raw Materials – Non-critical Raw Materials Factsheets* (Publications Office of the EU, Luxembourg). **https://publications.europa.eu/en/publication-detail/-/publication/6f1e28a7-98fb-11e7-b92d-01aa75ed71a1/language-en**

[8] European Commission. *Critical Raw Materials.* **https://ec.europa.eu/growth/sectors/raw-materials/specific-interest/critical_en**

[9] European Commision (2023). Study on the Critical Raw Materials for the EU. **https://op.europa.eu/en/publication-detail/-/publication/57318397-fdd4-11ed-a05c-01aa75ed71a1**

[10] Hajdu O. (2003). *Többváltozós statisztikai számítások*. Budapest: Központi Statisztikai Hivatal.

[11] Malhotra, N. K. (2008). *Marketingkutatás*. Budapest: Akadémiai Kiadó.

[12] Ojo Olawale. *The CAP curves.* **https://waleblaq.medium.com/the-cap-curves-the-cumulative-accuracy-profile-58a141e01fae**

[13] Székelyi M., Barna I. (2002). *Túlélőkészlet az SPSS-hez.* Budapest: Typotex Kiadó.

[14] Varga, B., Fodor, K. (2019b). Kann die logistische Regression zur Klassifizierung kritischer Rohstoffe verwendet werden? In: People and their values in the society. pp. 15–25. **https://www.irisro.org/selc2019/02VargaBeatrix-FodorKitti-nemet.pdf**

[15] Varga, B., Fodor, K. (2021). Technological Innovation and the Critical Raw Material Stock. In: *Data Analysis and Rationality in a Complex World.* pp. 321–329. **http://doi.org/10.1007/978-3-030-60104-1_35**

[16] Varga B., Szilágyi R. (2011). *Kvantitatív információképzési technikák.* **https://www.tankonyvtar.hu/hu/tartalom/tamop425/0049_08_kvantitativ_informaciokepzesi_technikak/3268/index.html**