K-Means clustering method results.

Using the K-Means clustering method (with K=3, selected taking into account the results of the Laktia method, as well as the recommendation to define 3 groups according to the previous classification method), three groups of countries were identified that are strongly correlated with the three levels of efficiency used in the classification model:

1) Cluster 1 (18 countries). Countries in this cluster have high indicators of governance, innovation and education. The highest is sdg9_eurd (2.56), the highest is sdg16_cpi (78.94), the lowest is sdg8_unemp (5.22), the highest is sdg4_adult (20.17).

2) Cluster 2 (16 countries). Countries in this cluster have average indicators. High income levels and relatively low inequality, but moderate investment in innovation. Average sdg8_income (21422.94), low sdg10_gini (29.38), moderate sdg16_cpi (56.06).

3) Cluster 0 (8 countries). The countries included in this cluster have the lowest governance and innovation indicators, the highest inequality and the worst social indicators. The lowest sdg9_eurd (0.75), the lowest sdg16_cpi (43.12), the highest sdg8_unemp (9.88), the highest sdg10_gini (35.00).2.

The fact that using K-Means three groups were found that perfectly coincide with the geographical and economic levels confirms the validity of the initial Random Forest classification.

The most noticeable differences between the clusters are observed in the indicators sdg16_cpi (Quality of Governance) and sdg9_eurd (R&D Expenditure). This confirms the conclusion that these features are strong predictors of sustainable performance, as was revealed in the previous analysis using Feature Importance methods. Identification of the High Risk Cluster (Cluster 0): This cluster is clearly defined by the combination of low CPI and low innovation spending. Countries in this cluster (including Bulgaria, Greece, Romania and Turkey) need the most targeted investments from the Recovery Fund, aimed primarily at improving the quality of governance and economic sustainability. The clustering results provide independent confirmation of the following recommendations - EU funding should be differentiated. That is, for countries classified as Cluster 1 (Leaders), support for R&D and advanced technologies is needed. For countries classified as Cluster 0 (High Risk), funding for governance reforms and the fight against corruption is needed, as this is their biggest difference from the other groups. For cluster 2, it is recommended to prioritize financing of large infrastructure projects aimed at decarbonizing industry and accelerating the transition to renewable energy (especially wind and solar generation) in order to reduce the environmental footprint, which is medium (sdg13_co2gcp - 5.61).

Comparing the results.

Comparing the results of clustering and data classification, it can be noted that there is a high level of such overlap.

In particular, Cluster 1 includes most of the countries of Western and Northern Europe (Austria, Germany, France, Sweden, Norway, the Netherlands), which are usually classified as Leaders (Level 2) in SDG reports.

Cluster 0 unites the countries that have the lowest indicators on critical indicators (CPI, R&D). This group includes Bulgaria, Romania, Serbia, Turkey and others, which can be attributed to Level 0 (Needs Improvement) in the controlled classification.

Cluster 2 is an intermediate group and includes countries from Southern and Central Europe (Spain, Portugal, Poland, Czech Republic), which are usually in Level 1 (Good Performers). This cluster reflects countries that have made significant progress, but do not have the high innovation and governance indicators characteristic of Cluster 1.

Thus, the level of agreement between the results of K-Means Clustering and Random Forest Classification is extremely high and confirms the reliability of the developed models.

Thus, it is possible to relatively clearly separate three levels of progress of countries according to all indicators. Economic and geographical groups (North/West, Central/South, East/Candidates) naturally cluster, which adds to the interpretability of the results.

Critical features, such as CPI and R&D, are the main factors that determine which group a country belongs to, which is directly consistent with Feature Importance Analysis from Random Forest. Thus, it can be argued that the ways to improve the development levels of countries identified by the Random Forest model are based on an objective assessment confirmed by the obtained data structures.