UDC 330.303.4:338.2 DOI 10.33111/mise.105.4

Житкевич О.В.,
кандидат економічних наук, докторант,
КНЕУ імені Вадима Гетьмана

Zhytkevych O.V.,
PhD (Economics), Doctoral candidate,
KNEU named after Vadym Hetman

ПОЕТАПНЕ МОДЕЛЮВАННЯ ВИКИДІВ CO_2 ІЗ ВИКОРИСТАННЯМ САМООРГАНІЗАЦІЙНИХ КАРТ КОХОНЕНА

STEP BY STEP MODELING OF CO_2 EMISSIONS USING KOHONEN SELF-ORGANIZING MAPS

Анотація. *Стаття присвячена виявленню закономірностей у динаміці декарбонізації національних економік на основі методу самоорганізаційних карт (SOM). Дослідження полягає у виявленні типології країн світу в контексті генерації та споживання енергії, а також аналізі змін у структурі їхніх енергетичних профілів у 2013–2022 роках. Було розроблено та апробовано емпіричну модель оцінки потенціалу декарбонізації за допомогою самоорганізаційних карт. Методологія дослідження включає шість етапів. На першому етапі автори зібрали та систематизували дані з відкритих джерел, зокрема Світового банку та щорічника EnerData, які включають макроекономічні, енергетичні та екологічні показники. Другий етап передбачав попередню обробку даних – виключення пропусків, ідентифікацію викидів та перетворення їх в єдиний формат «країна-рік-показник-значення». На третьому етапі було сформовано узгоджений багатовимірний набір даних з 14 змінними (із забезпеченням відсутності мультиколінеарності), виконано нормалізацію даних. На четвертому етапі побудовано та навчено SOM із гексагональною топологією, оптимальну структуру якої було визначено за показниками Quantization Error і Topographic Error. П'ятий етап присвячений кластеризації вагових векторів нейронів за допомогою агломеративного підходу, де оптимальну кількість кластерів визначено за метриками Silhouette Score, Davies–Bouldin та Calinski-Harabasz. На заключному етапі результати були інтерпретовані, країни були типологізовані за енергетичними профілями (лідери, промислові, ресурсоорієнтовані тощо) та проаналізовані динаміки їхнього руху між кластерами в часі. Отримані результати дозволили виявити структурні відмінності в моделях декарбонізації та визначити ключові фактори, які забезпечили успіх енергетичного переходу країни на глобальному рівні.*

Ключові слова: *декарбонізація, самоорганізаційні карти, кластеризація, викиди CO_2, сценарне моделювання, багатовимірний аналіз.*

Abstract. *The article is devoted to identifying patterns in the dynamics of decarbonization of national economies based on the methods of self-organizing maps (SOM). The study consists in revealing a typology of countries of the world in the context of energy generation and consumption, as well as analyzing*

changes in the structure of their energy profiles in 2013–2022. An empirical model for assessing the decarbonization potential using self-organizing maps was developed and tested. The research methodology includes six stages. At the first stage, the authors collected and systematized data from open sources, in particular the World Bank and the EnerData yearbook, which include macroeconomic, energy and environmental indicators. The second stage involved pre-processing of data – eliminating omissions, identifying outliers and converting them into a single format "country-year-indicator-value". At the third stage, a coordinated multidimensional data set with 14 variables was formed (ensuring the absence of multicollinearity), and data normalization was performed. In the fourth stage, a SOM with hexagonal topology was built and trained, the optimal structure of which was defined using quantization and topographic error indicators. The fifth stage is devoted ti the clustering of neuron weight vectors using an agglomerative approach, where the optimal number of clusters was determined using the Silhouette Score, Davies-Bouldin, and Calinski-Harabasz indices. In the final stage, the results were interpreted, countries were typified by energy profiles (leaders, industrial, resource-oriented, etc.) and the dynamics of their movement between clusters over time were analyzed. The results obtained allowed us to identify structural differences in decarbonization models and identify key factors that ensured the success of the country's energy transition at the global level.
Keywords: *decarbonization, self-organizing maps, clustering, CO_2 emissions, scenario modeling, multivariate analysis.*

General statement of the problem. Climate transformation and energy decarbonization are not just an ecological but also an economic necessity for states striving for sustainable development and international integration. In Ukraine, which is in a state of war, constant active reconstruction and reforms after the start of large-scale Russian aggression, this challenge is gaining particular attention. In 2025, Ukraine has already taken a number of steps towards adapting to the climate standards of the European Union. The National Energy and Climate Plan 2025–2030 was adopted and implemented, a report on its first stage of implementation was submitted, and initiatives to modernize the energy resource structure with an emphasis on decarbonization were launched [1, 2].

Despite significant infrastructure destruction and temporary instability of the energy system, Ukraine demonstrates significant potential resources for reducing emissions. In particular, studies show that rooftop solar photovoltaic installations alone have the potential to generate ~238.8 GW and ~290 TWh/year [3]. In addition, strategic decarbonization scenarios, which are already included in government and analytical documents, envisage a transformation of the energy sector by 2050 with the prospect of gradually covering the entire economy [4].

Determining the decarbonization potential (the maximum possible reduction in greenhouse gas emissions, taking into account technological, resource, institutional, social and financial constraints) is critically important for each country. And for Ukraine, it is extremely

important for several reasons. This allows for the formulation of realistic targets that are consistent with Ukraine's climate commitments under the Paris Agreement and EU standards. Also, such a substantiated potential can be used as an argument in negotiations with international donors, investors and in the transformation of the "green recovery" model. Taking into account such potential by sector (energy, industry, transport, buildings, land use, etc.) will prevent ineffective measures that may cause social or economic tensions. This is especially important in the context of European integration.

In accordance with the terms of the associated partnership and the prospects of full integration with the EU, Ukraine must gradually align its climate and energy policy with the norms of the European Green Deal [4, 5]. In addition, cooperation with the EU in the field of energy and climate is being strengthened through programs such as REPowerEU, which are aimed at accelerating the "green" transformation and reducing dependence on fossil sources [6]. Thus, an accurate assessment of the decarbonization potential is not only a tool for scientific analysis, but also a strategic mechanism that provides Ukraine with the technical basis for building climate excellence on the path to the European Union.

Since the analysis of decarbonization potential faces a number of methodological and practical challenges. Data characterizing the energy and economic profile of countries are multidimensional and heterogeneous. The relationships between individual indicators are often nonlinear, which complicates the application of classical regression and index models. In this context, modern machine learning methods, in particular self-organizing Kohonen maps, demonstrate high efficiency in visualization, dimensionality reduction and clustering of multidimensional data.

In this article, we propose a methodological approach to the quantitative assessment of the decarbonization potential of Ukraine. Special attention is paid to the construction and preparation of a dataset for modeling CO_2 emissions using self-organizing Kohonen maps. **The aim of the study** is to develop and empirically test a model for assessing decarbonization potential using SOM. To achieve the aim, the following **tasks** were formulated:

1. To form a multidimensional database that includes a set of energy, environmental and macroeconomic indicators using official statistical sources.

2. To carry out data preprocessing, including format unification, elimination of omissions and emissions, as well as normalization of indicators to ensure comparability.

3. To build a consistent database with a defined set of key variables, supplemented by derived indicators, and to perform multicollinearity reduction based on the results of correlation analysis.

4. To develop and train a Kohonen self-organizing map (SOM) model with appropriate topology parameters and to assess its quality using standard metrics.

5. To implement clustering of SOM neuron weight vectors with testing of different variants of the number of clusters and determination of the optimal partition using integral criteria.

6. Interpret the obtained clusters from the standpoint of decarbonization potential, determine their key characteristics and trace the dynamics of changes in country affiliation over time.

Analysis of recent studies and publications. The scientific literature offers various approaches to assessing decarbonization potential, including index methods, regression models and scenario analysis, among others. Index methods, such as the Energy Sustainability Index and the Carbon Intensity Index, allow us to assess the efficiency of energy systems in the context of reducing CO_2 emissions. In the article [7] proposed a new decarbonization index that allows us to assess the progress of countries towards achieving carbon neutrality. This index is an important tool for analyzing the effectiveness of decarbonization policies and strategies. However, we believe that such methods often simplify the multidimensional nature of the problem, which can lead to underestimation or overestimation of the potential of certain countries.

The article [8] examines the impact of macroeconomic stabilization components on decarbonization and energy efficiency in the five largest greenhouse gas emitting countries of the European Union (France, Germany, Italy, Poland and Spain) for the period from 1990 to 2020. Correlation analysis and linear regression (OLS and SUR) methods are used to assess the statistical significance of the impact of macroeconomic factors on energy efficiency and CO_2 emission reduction. The results show that macroeconomic stabilization components have different impacts on decarbonization and energy efficiency, which highlights the need for macroeconomic and environmental policies to be aligned [8]. However, regression models that analyze the relationship between indicators (e.g. macroeconomic) and CO_2 emissions also have limitations, as they do not always take into account the complex nonlinear relationships between these variables.

Scholars analyze various paths to achieving carbon neutrality in the US energy system at near-optimal costs. The study uses modeling to

identify variability in decarbonization scenarios, which allows for a better understanding of possible technology portfolios and their interrelationships [9]. Scenario analysis allows for modeling different development options, but its accuracy depends on the quality of the input data and the assumptions used.

For example, authors [10] used data-driven approach for optimizing district heating networks using source-load mapping, focusing on Stockholm as a case study. The authors analyze the limitations of traditional clustering methods, such as k-means and hierarchical clustering, in the context of complex, nonlinear relationships between variables. The authors note that these methods may inadequately reflect complex data structures, leading to inaccurate classifications, especially in conditions of high spatial heterogeneity and nonlinear dependencies. They propose alternative approaches that take these complexities into account, including methods that integrate nonlinear correlations and spatial variability.

Clustering approaches such as k-means and hierarchical clustering allow for grouping countries according to similar characteristics, which is useful for assessing decarbonization potential. However, these methods have limitations, in particular, they may not account for complex nonlinear relationships between variables [11].

Dimensionality reduction methods, such as principal component analysis (PCA), are widely used tools for processing multidimensional data in the context of assessing the decarbonization potential of countries. However, as noted in [12], PCA is a linear method that may not adequately capture complex nonlinear relationships between variables such as CO_2 emissions, energy consumption and economic indicators. This limitation may affect the accuracy of classifying countries according to their decarbonization potential. To overcome these limitations, the authors propose combining PCA with other methods, such as t-SNE, which allows for the detection of more complex patterns in the data.

Self-organizing maps, proposed by Teuvo Kohonen, are widely used for visualizing high-dimensional data and detecting clusters. In the energy sector, SOMs have been applied to classify energy consumers, forecast loads, and analyze energy efficiency. In particular, authors [13] examined the application of Kohonen self-organizing maps to assess the energy sustainability of countries. The authors use 28 indicators, such as CO_2 emissions per capita, the share of renewable energy in the energy balance, and others, to create a two-dimensional map that allows for the classification of countries by their level of energy sustainability. This study demonstrates the effectiveness of SOMs in visualizing and

analyzing complex relationships between various energy and economic indicators. However, the application of SOM specifically to assess the decarbonization potential of countries remains an understudied area, which makes the research innovative.

Existing methods for quantifying decarbonization potential are mainly based on regression approaches or aggregated index models. Existing methods for quantifying decarbonization potential are mainly based on regression approaches or aggregated index models. However, such tools do not always adequately reflect the complexity of the interaction of energy, economic and social factors. The high dimensionality of the data and their significant variability cause information loss, which reduces the accuracy of the estimates and the predictive ability of the models. At the same time, the task of identifying groups of countries with similar characteristics arises, which allows for the formation of more targeted political and economic strategies to achieve climate goals.

The main material of the study. To achieve the set goals, a comprehensive qualitative analysis of twelve clustering methods was conducted, covering both traditional statistical approaches and modern machine learning algorithms. Each method was evaluated according to a number of specialized criteria of effectiveness, advantages and limitations. As a result of experimental evaluation, the method of self-organized Kohonen maps was selected for the study. Unlike other clustering algorithms, SOM allows you to unambiguously determine the level of development of the object of analysis (in this case, countries by decarbonization potential), since the most and least developed countries are located in opposite sections of the Kohonen map. Accordingly, the distance of the country from the cluster of the most developed countries reflects the effectiveness of implementing the policy of transition to renewable energy. Analysis of changes in the positions of countries on the SOM map in dynamics allows you to systematically track progress in the implementation of these policies over time. Analytical environment and tools used: Python 3.9 – data preparation, processing and transformation; NumPy/Pandas – structured data management; Scikit-learn – standardization and normalization of indicators; MiniSom – implementation of the Kohonen self-organizing map method for cluster analysis; Deductor Studio Academic – graphical visualization of the SOM map; Matplotlib/Seaborn – construction of heat maps and graphical display of indicator profiles and Jupyter Notebook – integrated and interactive execution of the entire analytical process [14].

The first stage of the study was the collection and preparation of data for further cluster analysis of countries according to their decarbonization profiles. A systematic review of a wide range of databases and information resources of international organizations was conducted. Taking into account the relevance of the available data for the task, their completeness by indicators, years and countries, the World Bank databases and EnerData reports were selected as the main sources. Based on these sources, a sample of 14 indicators for the period 2013–2022 for 40 countries was formed.

In particular, the following indicators were selected from the World Bank database [15]: GDP per capita growth rate (in percent), urbanization rate (in percent of total population), total energy consumption and energy intensity of GDP (energy consumption per 1000 USD of GDP). The EnerData Yearbook reports [16] provide data on the trade balance by major energy sources (coal, petroleum products, natural gas, electricity), domestic consumption of these resources, the share of renewable sources in electricity production and the average CO_2 emission factor for each country.

It is fundamentally important that the analysis did not use ready-made indicators of CO_2 emissions in their pure form. Instead, an aggregated set of factors was formed that allows for a more detailed and accurate description of the structure of energy consumption and the economic context of each country, providing a more correct assessment of the decarbonization potential.

To retrieve data from the World Bank database, the official World Bank Application Programming Interface (API) for data collection was used, which provides systematic access to current statistical indicators and allows them to be integrated into analytical research models. Example code fragment to demonstrate the data retrieval procedure [14]:

```
import wbdata
import pandas as pd
import datetime
# Select indicators
indicators = {
 'SP.URB.TOTL.IN.ZS': 'Urban population %',
 'NY.GDP.PCAP.KD.ZG': 'GDP per capita growth %'
}
# Country selection
countries = ["USA", "CHN", "UKR", "NOR", "SWE", "IND", ...] # ISO-codes of 40 nations
# Time
```

data_date = (datetime.datetime(2013, 1, 1), datetime.datetime(2022, 12, 31))

 # Data receiving

wb_data = wbdata.get_dataframe(indicators, country=countries, data_date=data_date, convert_date=False)

 # Transformation

wb_data = wb_data.reset_index().pivot(index=['country', 'date'], columns='indicator', values='value')

Regarding the data loading and pre-processing from EnerData Yearbook, since EnerData does not provide a direct application programming interface (API), the data was obtained by manual loading and further structuring. The data from EnerData Yearbook was converted from Excel to CSV format to ensure compatibility with analytical models. The data was then pre-processed and standardized, bringing them to a single structure: "Country | Year | Indicator | Value", using tabular data processing procedures, which allows integrating the information into subsequent stages of the analytical study. Example of a code fragment to demonstrate the data collection procedure [14]:

import pandas as pd

 # Data

enerdata_df = pd.read_csv('enerdata.csv')

 # Transformation

enerdata_pivot = enerdata_df.pivot_table(index=['Country', 'Year'], columns='Indicator', values='Value')

Handling missing data is important because missing data is a fairly common phenomenon in real databases. To ensure analytical validity and avoid systematic error, the database was limited to 40 countries, leaving only those countries where the proportion of missing values did not exceed 10 % of the total number of indicators. Despite this, even among the selected countries, values were missing for individual indicators both for individual years and systematically. The following three approaches were used to address the problem of missing data, described below.

First approach is the systematic missing data, in cases where data for a certain indicator are completely missing for a country (e.g., the electricity trade balance for Kazakhstan), the group aggregation method was used, namely, the average value of the indicator was calculated for the group of countries to which the country with missing data belongs (e.g., EU, G7, CIS). Missing values were filled in according to the calculated average values, ensuring data consistency and minimizing

the impact of missing data on subsequent analysis. Example of a code fragment to demonstrate the procedure [14]:

```
# Replacing gaps with average values by groups (for Kazakhstan – CIS)
cis_mean_trade_balance =
enerdata_pivot[enerdata_pivot['Region'] == 'CIS']['Electricity
Trade Balance'].mean()
enerdata_pivot.loc[
 (enerdata_pivot['Country'] == 'Kazakhstan') &
 (enerdata_pivot['Electricity Trade Balance'].isna()),
  'Electricity Trade Balance'
] = cis_mean_trade_balance
```

Sometimes incomplete data were observed for individual years (for example, the absence of a value for 2015 for a certain country), so the second approach was utilized. Methods that take into account the nature of the dynamics of indicators were used to handle such omissions, so for indicators with slow changes over time (for example, the level of urbanization or the structure of energy consumption), linear interpolation between available values was used, which allows preserving the continuity and trends of the indicator. For indicators with high variability over time (for example, the CO_2 emission factor), omissions were replaced by average values for the corresponding group of countries, which ensures that the impact of extreme fluctuations is minimized and the data is maintained. Example of a code fragment to demonstrate the procedure [14]:

```
df['Urban population  %'] = df['Urban
population  %'].interpolate(method='linear')
```

Also, sometimes anomalous values were found in the data (for example, the CO_2 emission factor for Norway according to the WDI in one year was sharply different by a factor of five from neighboring years). A combined approach was used to handle such anomalies. First, an automatic analysis of the presence of emissions was performed using the Interquartile Range (IQR) method. Values identified as potential anomalies were manually checked and corrected if measurement errors or data entry errors were detected. In case the anomalous value reflected real changes in the indicator (for example, a sharp decrease or increase in emissions due to specific economic or energy events), it was left unchanged, ensuring the reliability of the data dynamics. Example code fragment to demonstrate the procedure [14]:

```
Q1 = df['CO2 emission factor'].quantile(0.25)
```

Q3 = df['CO2 emission factor'].quantile(0.75)
IQR = Q3 – Q1
*outliers = df[(df['CO2 emission factor'] < (Q1 – 1.5 * IQR)) | (df['CO2 emission factor'] > (Q3 + 1.5 * IQR))]*

As a result of data cleaning and transformation, an initial dataset was formed, which included 400 records corresponding to a combination of 40 countries and a 10-year observation period. Each record contained 14 normalized indicators, including: indicators of total energy consumption and energy intensity of GDP, trade balances by main types of energy carriers (coal, petroleum products, natural gas, electricity), the share of renewable sources in total electricity production, the CO_2 emission factor, as well as demographic and economic indicators, in particular the level of urbanization and GDP per capita growth rates. The structured nature of this dataset ensured its suitability for further use in the Kohonen self-organizing maps model, which allowed for detailed visualization of the dynamics and profiles of all 40 countries during the analyzed decade. After the stage of collecting primary data from sources, a heterogeneous dataset was formed, in which the values of different indicators differed in order of magnitude and units of measurement. In particular, GDP growth rates were expressed in percentages, energy consumption in terajoules, and CO_2 emissions in tons per megawatt-hour. This discrepancy creates a risk of distorting the clustering results, since during training the Kohonen neural map, the algorithm optimizes the distances between feature vectors based on the Euclidean metric.

In the absence of scaling, the dominance of features with large absolute values can lead to the loss of information about other, no less important indicators. Given this, data preparation for SOM included several critical stages:

1. Construction of derived features (feature engineering) – the formation of new indicators that generalize or refine the economic and energy profile of the country.
2. Scaling and normalization of features (feature scaling) to ensure comparability of variables.
3. Checking correlations between features to eliminate multicollinearity.
4. Selection of features according to their significance for the model.
5. Formation of the final feature matrix for further training of SOM.

Construction of derived features. Based on the primary indicators of the World Bank and EnerData databases, a number of new aggregated variables were created that reflect more comprehensive

characteristics of the country's energy profile: Net trade balance (by energy type) – calculated as the difference between the volumes of exports and imports of each type of energy resources, and Energy structure coefficient – an integral indicator that combines the share of renewable energy sources in electricity production and the energy intensity of GDP. This indicator characterizes the level of diversification of the energy balance and the efficiency of energy use. Example of a code fragment to demonstrate the aggregation of the Net Trade Balance [14]:

```
df_scaled['Net Coal Balance'] = df_scaled['Coal Export'] – df_scaled['Coal Import']
df_scaled['Net Oil Balance'] = df_scaled['Oil Export'] – df_scaled['Oil Import']
df_scaled['Net Gas Balance'] = df_scaled['Gas Export'] – df_scaled['Gas Import']
```

Scaling (Normalization & Scaling). To bring the feature values to a comparable scale, the Min-Max Normalization procedure was used, with all variables being brought to the range [0,1]. This approach is justified by two main aspects: the sensitivity of the SOM algorithm to the feature scales – since the Kohonen map training is based on Euclidean distances between feature vectors and neuron parameters, the dominance of variables with large absolute values can lead to the loss of information about less scaled features; and also by preserving the relative proportions within each indicator – minimax normalization allows maintaining the relationship between observations for one variable, which is important for the correct interpretation of clustering results and assessment of the decarbonization potential of countries. An example of a code fragment to demonstrate the procedure [14]:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
scaled_features = scaler.fit_transform(df[feature_columns])
df_scaled = pd.DataFrame(scaled_features, columns=feature_columns)
```

It should be noted that in a number of analytical tasks, the use of minimax normalization may not be effective enough. In such cases, it is advisable to use Z-standardization, which normalizes the deviation of each value from the mean by dividing by the standard deviation. This procedure provides a more balanced data set and often improves the efficiency of clustering in cases where other data preparation methods are insufficient.

Correlation check. To avoid duplication of features and reduce multicollinearity in the input data set, an analysis of the correlation matrix was performed [14]:

```
import seaborn as sns
import matplotlib.pyplot as plt
corr = df_scaled.corr()
plt.figure(figsize=(12,10))
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.show()
```

The conclusions on the correlation of features are based on the fact that features with a correlation exceeding 0.95 were either combined or replaced by aggregate factors to reduce data redundancy. At the same time, some highly correlated indicators, such as Total Energy Consumption and Energy Intensity, were left in the analysis, since they reflect the impact on energy consumption and decarbonization processes in different aspects and thus carry unique information for modeling. Selection of features. Based on the analysis, the features with the most significant impact on the country's energy profile and its decarbonization potential were selected. It was decided to limit ourselves to no more than two dozen variables, since an increase in the number of features reduces the individual contribution of each predictor to the modeling result. As a result of forming the final feature matrix, 14 key indicators were left, among which the following are worth highlighting: GDP growth per capita (%), share of urbanized population (%), total energy consumption; GDP energy intensity (energy consumption per unit of GDP); CO_2 emission factor (tons of CO_2 per MWh); trade balance by energy type: coal, oil, gas, electricity; share of renewable sources in electricity production; domestic energy consumption by type: coal, oil, gas, electricity.

Construction of the final feature matrix for the self-organizing Kohonen map. Before providing data for training the self-organizing Kohonen map, an additional assessment of the quality of the input dataset was carried out. In particular, the distributions of values for each feature were analyzed to exclude the presence of sharp outliers or flat distributions that could distort the topology of the map during training. As a result, a final feature matrix was formed, which is characterized by: 400 observations (40 countries × 10 years) and 14 normalized, informative features, balanced in scale and without redundancy.

This structured dataset provides a robust basis for efficient training and further analysis of the self-organizing Kohonen map, allowing for clear visualization of the dynamics and profiles of countries regarding

their decarbonization potential over a decade. Example code snippet to demonstrate the final matrix [14]:

X = df_scaled.values # final feature matrix for submission to SOM

After preparing the normalized dataset, a key stage of the research was conducted – training a Kohonen self-organizing map (SOM). This method allows to detect hidden patterns in multidimensional data and visualize them in the form of a two-dimensional map, where objects with similar characteristics (in this case, countries) are located close to each other. The tools used were: Python MiniSom – a compact and efficient library for implementing Kohonen SOM and visualization – Matplotlib and Seaborn for building graphical representations, as well as Deductor Studio Academic for interactive display of U-Matrix heat maps.

An important stage of the modeling is determining the size of the map and its training parameters. For this study, a map with a size of 16×12 neurons was chosen, which corresponds to a total number of 192 neurons. This choice is justified by the empirical rule for calculating the optimal number of neurons [17]:

$$M \approx 5 \times \sqrt{N} \quad (1)$$

where M is the total number of neurons and N is the number of features in the dataset. Since in our case $N = 400$ (40 countries × 10 years), the optimal range of the number of neurons is 100–250.

To provide a more natural local connection between features on the map, a hexagonal topology of neurons was chosen, which promotes a smooth distribution of countries and facilitates the interpretation of clusters. Example code fragment to demonstrate the procedure [14]:

from minisom import MiniSom

som = MiniSom (x=16, y=12, input_len=14, sigma=1.0, learning_rate=0.5,

neighborhood_function='gaussian', random_seed=42)

The additional model parameters are input_len = 14 – the number of input features; sigma = 1.0 – the initial radius of the neuron's influence area; learning_rate = 0.5 – the initial learning rate, which gradually decreases over the epochs and neighborhood_function = 'gaussian' – a neighborhood function with a smooth weakening of the influence of the input vector on neurons distant from the winning neuron (Best Matching Unit, BMU).

The model was trained in random batch learning mode for 5500 epochs. The use of a random order of input vectors is justified by the fact that it allows you to avoid the phenomenon of "freezing" the map, when fixed sequences can lead to a loss of flexibility in forming the SOM topology.

In each epoch, a data vector corresponding to a certain country in a certain year was randomly selected. After selecting an object, the weights of the winning neuron and its neighbors were updated according to a defined learning function, which ensures gradual smoothing of the map and formation of coherent clusters. Example of a code fragment to demonstrate the procedure [14]:

som.random_weights_init(X) # Initializing neuron weights with random values from the dataset

som.train_random(X, num_iteration=5500) # Learning

During map training, the following metrics were used to assess the quality of training: Quantization Error (QE) — a measure of the correspondence between data objects and winning neurons, which is defined as the average distance between the input vector and the weight vector of the nearest neuron (QE allows us to assess how well the map reflects the original data), and Topographic Error (TE) — a metric that assesses the ability of the SOM to preserve the topological structure of the data (it is calculated as the fraction of points in the input set for which the two nearest BMUs (the best and the second best) are not adjacent on the map; TE provides information about the correctness of the spatial arrangement of clusters on the SOM). To demonstrate the training control procedure, use the following code fragment [14]:

qe = som.quantization_error(X)

te = som.topographic_error(X)

print(f'Quantization Error: {qe}, Topographic Error: {te}')

Our target Quantization Error (QE) < 0.2 and Topographic Error (TE) < 0.1 indicators indicate that SOM adequately reflects hidden patterns in the data and correctly preserves the topological structure of the neighborhood between objects.

SOM does not perform explicit clustering of data, but projects multidimensional data into a two-dimensional space while preserving the topology. Therefore, to isolate individual groups of objects, additional clustering of neurons based on their weight vectors was performed using Agglomerative Clustering. In parallel, optimization of the number of clusters was performed, in which the formed Kohonen map was sequentially divided into different numbers of clusters in the

range from 3 to 8. This stage is critically important, since incorrect determination of the number of clusters can lead to excessive generalization of groups or, conversely, loss of interpretability with excessive detail.

The accuracy of the segmentation was assessed using three standard criteria: the silhouette coefficient, the Davis–Boldin index, and the Kalinsky–Kharabash index, as well as an integral indicator based on them. As a result, six clusters were selected that most relevantly segmented countries according to decarbonization profiles, allowing to distinguish groups of leaders, followers, and problem countries.

The procedure for constructing the Kohonen map and subsequent clustering included four steps:

Step 1. Visual assessment via U-Matrix. After SOM optimization, a U-Matrix (Unified Distance Matrix) was constructed, a heat map where the distances between the weight vectors of neighboring neurons are displayed in color. Areas with low distance reflect densely grouped neurons that are potential clusters, while areas with high distance values outline the boundaries between separate groups. The U-Matrix analysis revealed six distinct zones that visually appeared as distinct "islands" on the map and served as the basis for further interpretation of country clusters based on their decarbonization profiles, which was confirmed by quantitative metrics. Example code snippet to demonstrate the procedure [14]:

```
from pylab import bone, pcolor, colorbar, plot, show
bone()
u_matrix = som.distance_map().T # Distances between neurons (U-Matrix)
pcolor(u_matrix) # Visualizing map
colorbar() # Color legends
show()
```

Step 2. Agglomerative clustering based on neuron weight vectors. Each SOM neuron is represented by a weight vector consisting of 14 elements, according to the number of input features. For further clustering, we used all 192 neuron vectors and applied agglomerative hierarchical clustering. This approach allows us to group neurons based on the similarity of their weight vectors, which reflect the multidimensional characteristics of the countries' decarbonization profiles. The result is the formation of clearly defined groups of neurons, each of which corresponds to a certain segment of countries in terms of the level and characteristics of decarbonization potential. To demonstrate the clustering procedure, an example code was used that

illustrates the agglomeration process and the subsequent assignment of clusters to map neurons [14]:

from scipy.cluster.hierarchy import linkage, fcluster

from scipy.spatial.distance import pdist

weights = som.get_weights().reshape(-1, 14) # Extract all neuron weights

Z = linkage(weights, method='ward') # Hierarchical clustering of weight vectors

Step 3. Validation of clusters using quality metrics. To determine the optimal number of clusters (K), three generally accepted metrics for assessing the quality of clustering and an integral criterion based on them were used [18]:

1. Silhouette coefficient (SC) – varies from -1 to 1 and characterizes the degree of similarity of objects to their own cluster compared to other clusters (the closer to 1, the better the segmentation).

2. Calinski-Harabasz index (CHI) – evaluates the ratio of intercluster dispersion to intracluster dispersion (higher values indicate high clustering quality).

3. Davies-Bouldin index (DBI) – determines the ratio of intracluster dispersion to the distance between clusters; lower values indicate better segmentation quality.

When testing for different values of K, the following results were obtained (Table 1).

Table 1

CLUSTERIZATION QUALITY ASSESSMENT METRICS

K	SC	DBI	CHI
3	0.33	1.09	227.37
4	0.25	1.09	182.87
5	0.34	0.37	261.31
6	0.44	0.90	290.43
7	0.34	0.87	248.23
8	0.26	1.00	215.51

Source: Developed by author

According to two metrics — SC and CHI — optimal segmentation is observed at K = 6. DBI indicates K = 5 as the best option. Thus, the

integrative assessment demonstrates that K = 6 provides the optimal balance between internal homogeneity and cluster separation, which confirmed the choice of six groups for the final model.

Step 4. Semantic validation of clusters. Semantic validation is an additional evaluation stage for unsupervised clustering tasks and involves analyzing the logical consistency of the resulting groups with real economic and energy profiles of countries.

To do this, the average values of all indicators were calculated for each cluster, which allowed to form a generalized profile of a "typical country" within the group. Comparison of the profiles of different clusters allowed to identify the key characteristics, which are described below. Leaders – decarbonization centers (Canada, New Zealand, Norway, Sweden, etc.) – countries with a high share of renewable energy and low energy intensity. Followers (UK, EU countries, developed Latin American countries) – show progress in decarbonization, but are partially dependent on traditional energy sources. Resource-oriented economies (Nigeria, UAE, Saudi Arabia, Ukraine, etc.) – high energy intensity of GDP and dependence on fossil fuels. Industrial-oriented countries (Australia, India, Germany, Poland, Turkey, Japan) – significant energy intensity, high CO2 emissions, low share of renewables. Hydrocarbon-oriented countries (Russia, USA) – large volumes of oil and gas production and domestic consumption. Large industrial countries (China, USA) – largest total energy consumption, significant CO2 emissions, but significant investments in renewable energy.

Since each country is represented by ten entries (one per year), some countries may have changed cluster affiliation in different years according to changes in the energy profile.

The choice of six clusters provided a balanced and interpretable structure, avoiding excessive fragmentation or artificial aggregation of heterogeneous countries. The resulting groups became the basis for further analysis of the dynamics of the movement of countries between clusters, which allows assessing progress or stagnation in the implementation of decarbonization policies of national economies [19].

Thus, the joint use of clustering quality metrics and semantic verification allowed to provide a reliable and economically understandable segmentation of countries on the Kohonen SOM map.

The first and foremost, SOM is a visualization tool that effectively allows you to explore patterns in complex multidimensional data and present them in a visual form. As part of this study, we have identified three levels of visual analytics:

1. Heat maps for individual features. During SOM training, each neuron acquired weight parameters that were formed based on the objects (countries) assigned to it or its neighbors. Each weight parameter corresponds to one of 14 selected features.

For each parameter, heat maps are built on all SOM neurons, which allows you to visually identify areas with high and low values of the corresponding characteristic. The intensity of the feature on the map is displayed as a color gradient – from blue (minimum values) to red (maximum values), with a corresponding color scale for accurate determination of the value.

Such maps make it easy to identify patterns, spatial clustering, and features of energy consumption and decarbonization profiles of different countries, providing a visual and intuitive representation of multidimensional data (Fig. 1).

Heat maps have become an effective tool for exploring spatial patterns of key indicators across countries, tracking the distribution of country profiles on the map, and identifying local patterns that are difficult to detect using traditional tabular data analysis, especially when there are a large number of characteristics.

2. Visualizing the positions of individual countries on the SOM map allowed us to assess their location within clusters. This made it possible to identify countries that are in the "core" of clusters, that is, have a clearly expressed typical decarbonization profile. At the same time, countries located on the borders between clusters were identified, which indicates the presence of hybrid profiles — a combination of characteristics from different groups or a gradual transition from one decarbonization strategy to another.

This approach allowed us not only to highlight clear leaders and outsiders, but also to identify countries that are in the process of transforming their energy profile, providing a deeper and more detailed understanding of the dynamics of decarbonization at the global level.

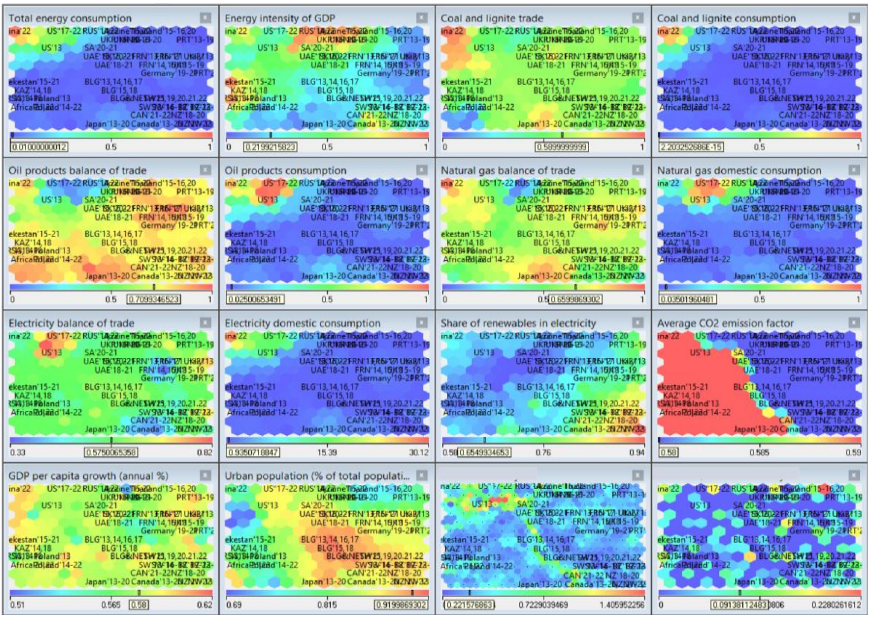

Fig. 1. Heat maps for exploring spatial patterns
of key indicators across countries [19]

3. Trajectories of countries on the Kohonen map — reflecting changes in the country profile over time. One of the key advantages of Self-Organizing Maps is the ability to visualize the trajectories of objects in a multidimensional feature space over time, which allows us to track how countries changed their decarbonization profiles over a 10-year period (Fig. 2). To do this, for each annual data set and each country, a winning neuron (Best Matching Unit, BMU) was determined that most accurately reflected the country profile in the corresponding year. Based on the BMU sequence, the trajectory of the country's movement on the Kohonen map was constructed, which allowed us to visually assess the dynamics of changes: the country could approach the clusters of leaders, remain in its previous position, or move away from the target decarbonization profiles. This approach allows us not only to analyze the state of the country in a specific year, but also to track trends and progress in the implementation of decarbonization policies, which provides a deep understanding of the dynamics of the transition to sustainable energy. Example of a code fragment to demonstrate the procedure [14]:

plt.figure(figsize=(10,8))

```
    plt.title('Trajectory of Ukraine (2013-2022)')
    coords = []
    for year in range(2013, 2023):
    country_year_data = df[(df['Country'] == 'Ukraine') & (df['Year'] == year)].iloc[0][feature_columns].values
    winner = som.winner(country_year_data)
    coords.append(winner)
    # draw the trajectory of movement
    for i in range(len(coords)-1):
    plt.arrow(coords[i][0]+0.5, coords[i][1]+0.5,
    coords[i+1][0]-coords[i][0], coords[i+1][1]-coords[i][1],
    head_width=0.3, head_length=0.3, fc='blue', ec='blue')
    plt.pcolor(som.distance_map().T, cmap='Greys', alpha=0.3)
    plt.show()
```

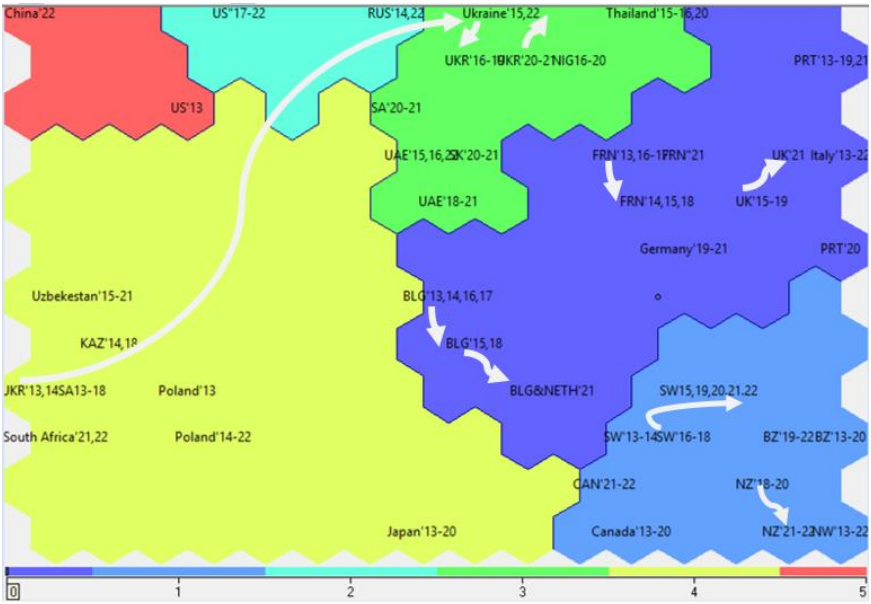

Fig. 2. Trajectory of movement for some counties on the Kohonen map [19]

This visualization allows not only to assess the static state of countries, but also to analyze the dynamics of their development towards decarbonization. During the study, several characteristic patterns of behavior were identified, described below.

Decarbonization leaders – countries that have traditionally actively implemented emission reduction policies – maintained high positions throughout the entire period under study. All countries included in the leader cluster (lower right corner of the map) constantly remained in it. Countries-candidates for leadership (follower cluster, upper right corner of the map) also remained in their cluster, but a gradual shift towards leaders was observed, which indicates real progress in the transformation of energy infrastructure and the reduction of CO_2 emissions.

Countries with high energy consumption in the clusters on the left side of the map remained stable. The exception is the USA, which moved from the far left cluster (next to China) to the "hydrocarbon countries" cluster together with Russia. This shift occurred not due to the deterioration of US policy, but due to the outpacing industrial development of China, against the background of which the US position in relative indicators shifted.

Countries in the central clusters of the map demonstrated less stability. Thus, the United Kingdom and some Latin American countries moved from the cluster of industrial countries (lower left corner of the map) to the cluster of followers, improving their positions in the context of decarbonization. Ukraine in 2015 made a sharp jump from the cluster of industrial countries to the cluster of "resource-oriented economies" in the upper central part of the map, after which it remained within this cluster, moving between different neurons.

This approach allows not only to track the progress of countries in decarbonization, but also to identify transformational patterns characteristic of different groups of economies.

At the initial stage of the study, the academic platform Deductor Studio was used to work with U-Matrix. It allows for interactive map analysis and tagging without the need for programming. At the same time, the platform has a number of limitations: visualization customization options are limited, which makes it difficult to configure more complex scenarios or specific data display designs. In addition, Deductor Studio does not allow for integrating visualizations into a web environment, which limits the possibilities for further use of the results in dashboards or analytical web applications. To overcome these limitations, the visualization of clustering results was transferred to a web dashboard format based on Plotly Dash. This approach provided an interactive interface that allows the user to: select a specific country and track its trajectory on the SOM map over time; filter data by year or cluster; and analyze heat maps with the ability to dynamically configure the display of factors.

The implementation process included the following steps:

1. Generating coordinates of objects (countries) on the SOM map in Python and importing the results into the visualization environment.

2. Building heat maps using built-in visualization tools and custom components.

3. Displaying trajectories of countries' movements in the form of scatter/line graphs to demonstrate dynamics across neurons.

4. Adding interactive filters that allow the user to adjust the visibility of objects by year, country, and cluster.

The results of applying this approach allowed: identifying "hot zones" of key decarbonization factors and localizing regions with high or low progress; visualizing trajectories of countries' movements in time dynamics, which became the basis for assessing the effectiveness of national energy policies; and involving not only technical specialists, but also business analysts in working with the model due to interactivity and a clear interface.

After building the Self-Organizing Map, identifying six clusters and visualizing them, the next step was to understand the significance of these clusters for national decarbonization strategies and apply them in scenario modeling (what-if analysis). The goal was to assess how changing individual indicators of the country's energy and economic structure affects its position on the cluster map and a possible move to a more progressive profile.

The scenario modeling methodology was based on modifying the country's profile, so for each country, the current vector of characteristic indicators was taken and its targeted change was carried out. For example, the share of renewable energy in electricity production was increased by 10 % or the energy intensity of GDP was reduced by 20 %. This methodology also involved running through SOM, the modified profile was passed through the SOM.winner() function, which made it possible to determine whether the winning neuron had changed, and accordingly, whether the country had moved to another cluster. The final stage was the analysis of the results, where the modeling results were stored in the form of an interactive matrix of changes, which allowed answering the question: "What specific parameters need to be changed in the country's energy structure so that it moves to the desired cluster (for example, the group of decarbonization leaders)?" The average values of the indicators of the leading countries of the corresponding cluster were used as a guideline.

This approach has transformed clustering into a practical tool for scenario analysis, allowing to identify key structural changes needed to

achieve a country's target profile in the context of decarbonization. Example code fragment to demonstrate the procedure [14]:

modified_profile = original_profile.copy()

modified_profile['Share of Renewables'] += 0.1 # +10 % to the share of renewable energy

*modified_profile['Energy Intensity'] *= 0.8 # -20 % from energy intensity*

new_cluster = som.winner(modified_profile.values)

Scenario modeling based on clustering results allows for the formulation of specific recommendations for the strategic development of different groups of countries.

For example, for countries in the cluster conventionally designated as "resource-oriented economies" (which includes, in particular, Ukraine in recent years), the key factor in increasing their global positioning in the field of decarbonization is investment in renewable energy. Only if the share of renewable sources increases significantly do these countries have the potential to move to the "follower" or "leader" clusters (towards the lower right corner of the SOM map). It is important to note that the leader cluster includes countries with developed green energy, including powerful resource exporters such as Canada and Norway.

As for large industrial countries (USA, China), scenario modeling showed that even a significant increase in the share of renewable energy without a simultaneous decrease in energy intensity will not allow them to move to more "green" clusters. This indicates the need for a comprehensive approach to energy conservation and increasing production efficiency. A similar situation is observed for russia, which is located in the cluster of large hydrocarbon countries, but borders the cluster of resource-based economies, demonstrating the characteristics of an intermediate profile.

Similarly, individual recommendations can be determined for each country depending on the current state and targets in the field of decarbonization. Thus, dynamic scenario modeling allows not only to segment countries by energy profiles, but also to identify the most sensitive parameters that affect the improvement of positions in global decarbonization. This forms a practical basis for the development of targeted strategies for the development of the energy sector of national economies and support the adoption of informed policy decisions.

Conclusions and prospects for further research. The study of the decarbonization potential of countries around the world demonstrated the effectiveness of using multidimensional clustering methods for

analyzing energy profiles. In particular, the use of self-organizing maps made it possible to identify hidden patterns in the data structure and visualize them in two-dimensional space, which ensures the preservation of the topology of connections between countries.

The accuracy of segmentation was assessed based on three standard criteria – the Silhouette coefficient, the Davies-Bouldin and Calinsky-Harabasz indices, as well as by conducting a qualitative analysis of semantic correspondence (to what extent such segmentation is logical for studying decarbonization processes). As a result of the analysis, six clusters were identified that reflect different types of energy and decarbonization profiles: leaders in the field of renewable energy, followers, resource-based economies, industrial countries, large hydrocarbon producers and large industrial states. The clusters demonstrated significant internal homogeneity and clear demarcation between groups, as confirmed by quantitative indicators of clustering quality and semantic assessment of the logical correspondence of the resulting groups to the real economic and energy characteristics of the countries.

The study of the dynamics of the positions of countries over time showed that some countries demonstrate gradual progress towards more "green" profiles, while others remain stable in their cluster positions or change them due to transformations of the energy structure. The analysis allowed to identify the most important characteristics that affect the position of the country in the global context of decarbonization, and to formulate scientifically based recommendations on targeted areas for improving the energy profile.

It is noted that the clustering methodology is not a predictive model for accurately determining future CO_2 emission indicators. Further research should include the integration of clustering with predictive models that can take into account the individual characteristics of each cluster. This will increase the accuracy of assessments and the effectiveness of strategic energy resource management at the national and global levels.

Further research involves the use of fuzzy clustering, which will allow assessing the degree of belonging of each country to each of the selected clusters regardless of the main distribution of clusters. This approach makes it possible to build ensembles of forecast models in which the influence of each model is determined by the degree of belonging of the country to the corresponding cluster. This provides increased adaptability and accuracy of forecasts regarding the development of the energy profile of countries.

Also, it is planned to integrate the study with current open sources of data on trade and consumption of energy carriers, which will allow

the formation of analytical tools to support strategic decision-making at the national and international levels in real time.

References

1. Ministry of Economy of Ukraine. (2025, September 9). *Ukraine advances in implementing the National Energy and Climate Plan*. Government of Ukraine. https://mev.gov.ua/en/news/ukraine-advances-implementing-national-energy-and-climate-plan-olha-yukhymchuk

2. DiXi Group. (2025, March). *Ukraine submitted its first integrated report on NECP together with EU member states*. DiXi Group. https://dixigroup.org/en/ukraine-submitted-its-first-integrated-report-on-necp-together-with-eu-member-states/

3. Winkler, C., Dabrock, K., Kapustyan, S., Hart, C., Heinrichs, H., Weinand, J. M., Linßen, J., & Stolten, D. (2024). *High-resolution rooftop-PV potential assessment for a resilient energy system in Ukraine*. arXiv. https://arxiv.org/abs/2412.06937

4. Stockholm Environment Institute. (2025, July 1). *Green transition report for Ukraine shows paths to green recovery and EU integration*. *Green Agenda*. https://green-agenda.org/en/ukraine/news/green-transition-report-ukraine

5. European Policy Centre. (2024). *Greener, better, stronger together: Why cooperation in renewable energy should be a priority for EU-Ukraine relations*. https://www.epc.eu/publication/greener-better-stronger-together-why-cooperation-in-renewable-energy-should-be-a-priority-for-eu-ukraine-relations/

6. European Commission. (2025). *REPowerEU*. Retrieved from October 4, 2025, from https://commission.europa.eu/topics/energy/repowereu_en

7. Qi, Y., Lu, J., & Liu, T. (2024). *Measuring energy transition away from fossil fuels: A new index. Renewable and Sustainable Energy Reviews, 200,* Article 114546. https://doi.org/10.1016/j.rser.2024.114546

8. Misztal, A., Kowalska, M., Fajczak-Kowalska, A., & Strunecký, O. (2021). *Energy efficiency and decarbonization in the context of macroeconomic stabilization in the European Union. Energies, 14*(16), Article 5197. https://doi.org/10.3390/en14165197

9. Sinha, A., Verdolini, E., & Tavoni, M. (2024). *Diverse decarbonization pathways under near cost-optimal conditions*. Nature Communications, 15(1), 1-12. https://www.nature.com/articles/s41467-024-52433-z

10. Shahcheraghian, A., et al. (2025). K-means and agglomerative clustering for source-load mapping in district heating networks. *ScienceDirect*. https://www.sciencedirect.com/science/article/pii/S2590174524003386

11. Rattle, I., Gailani, A., & Taylor, P. (2025). *Towards a typology of industrial decarbonisation initiatives*. Wiley Online Library. https://rgs-ibg.onlinelibrary.wiley.com/doi/pdf/10.1002/geo2.70000

12. Jiménez-Preciado, A. L., Cruz-Aké, S., & Venegas-Martínez, F. (2024). Identification of patterns in CO_2 emissions among 208 countries: K-means clustering combined with PCA and non-linear t-SNE visualization. *Mathematics, 12*(16), Article 2591. https://doi.org/10.3390/math12162591

13. Vlaović, Ž., Stepanov, B. L., & Anđelković, A. S. (2023). Mapping energy sustainability using the Kohonen self-organizing maps. *Journal of Cleaner Production, 412*, Article 137351. https://doi.org/10.1016/j.jclepro.2023.137351

14. Matviychuk, A. (2025, October 4). Як ми будували модель скорочення викидів CO_2 за допомогою SOM (Kohonen Self-Organizing Maps). *DOU*. https://dou.ua/forums/topic/55663/?utm_source=other&utm_medium=email&utm_campaign=03102025

15. The World Bank. (2023). *World development indicators* [dataset]. https://datacatalog.worldbank.org/search/dataset/0037712

16. EnerData. (2023). *World Energy & Climate Statistics – Yearbook 2023* [dataset]. https://yearbook.enerdata.net/total-energy/world-consumption-statistics.html

17. Heaton, J. (2008). *Introduction to neural networks for Java* (2nd ed.). Heaton Research, Inc.

18. Poznyak, S., & Kolyada, Y. (2023). Comparative analysis of the effectiveness of dimensionality reduction algorithms and clustering methods on the problem of modeling economic growth. *Neuro-Fuzzy Modeling Techniques in Economics, 12*, 67–110. https://doi.org/10.33111/nfmte.2023.067

19. Matviychuk, A., Zhytkevych, O., & Osadcha, N. (2024). Modeling carbon dioxide emissions reduction. *Energy Reports, 12*, 1876–1887. https://doi.org/10.1016/j.egyr.2024.08.004