Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 1 PAGE: 307-313 YEAR: 2024
DOI:10.17798/bitlisfen.1395411

Symbolic Aggregate Approximation-Based Clustering of Monthly Natural Gas Consumption

Mehmet Eren NALİCİ[1], İsmet SÖYLEMEZ[1*], Ramazan ÜNLÜ[1]

[1]*Abdullah Gül University, Faculty of Engineering, Industrial Engineering Department, Kayseri Türkiye*

(ORCID: 0000-0002-7954-6916) (ORCID: 0000-0002-8253-9389) (ORCID: 0000-0002-1201-195X)

Keywords: Machine Learning, Symbolic Aggregate Approximation, Clustering, Energy

Abstract

Natural gas is an indispensable non-renewable energy source for many countries. It is used in many different areas such as heating and kitchen appliances in homes, and heat treatment and electricity generation in industry. Natural gas is an essential component of the transportation sector, providing a cleaner alternative to traditional fuels in vehicles and fleets. Moreover, natural gas plays a vital role in boosting energy efficiency through the development of combined heat and power systems. These systems produce electricity and useful heat concurrently. As nations move towards more sustainable energy solutions, natural gas has gained prominence as a transitional fuel. This is due to its lower carbon emissions when compared to coal and oil, thus making it an essential component of the global energy framework. In this study, monthly natural gas consumption data of 28 different European countries between 2014 and 2022 are used. Symbolic Aggregate Approximation method is used to analyze the data. Analyses are made with different numbers of segments and numbers of alphabet sizes, and alphabet vectors of each country are created. These letter vectors are used in hierarchical clustering and dendrogram graphs are created. Furthermore, the elbow method is used to determine the appropriate number of clusters. Clusters of countries are created according to the determined number of clusters. In addition, it is interpreted according to the consumption trends of the countries in the determined clusters.

1. Introduction

For many nations, natural gas is an essential non-renewable energy source. It is utilized in a wide range of applications, including heat treatment and power production in industry, as well as heating and cooking appliances in residences. Because it offers a more environmentally friendly option to conventional fuels for automobiles and fleets, natural gas is a vital component of the transportation industry. Its adaptability also extends to the manufacture of petrochemicals, where it serves as a feedstock to produce several necessary goods including chemicals, plastics, and fertilizers. Furthermore, natural gas is essential for developing combined heat and power (CHP) systems, which increase energy efficiency. These systems simultaneously provide usable heat and power. Natural gas has become more well-known as a transitional fuel as countries shift to more environmentally friendly energy sources. Although energy from environmentally damaging sources has been utilized for a long time, there has been a shift in trend toward lower carbon energy sources. In this situation, using natural gas has become a viable substitute for using coal [1]. Natural gas plays an important role in reducing contaminants and climate change because it emits 50% less carbon dioxide (CO2) than coal and 30% less than oil. The International Gas Union cites three main reasons for this, including the fact that it is more affordable than

*Corresponding author: ismet.soylemez@agu.edu.tr
Received:24.11.2023 , Accepted:08.03.2024

other energy sources based on fossil fuels, easier to deliver and install, and a sustainable resource. Considering these circumstances, several nations strive to enhance their natural gas production technologies, including the utilization of non-traditional production techniques, to augment the physical network components within the gas supply chain [2].

In modern economies, energy resources are becoming increasingly scarce. Therefore, it is important to obtain energy more cheaply and use it wisely. The need to use energy more wisely has arisen due to the rapid increase in energy consumption, which has occurred regardless of the world's population growth. This increase is also impacting the use of natural gas in particular [3]. The demand for natural gas has been rising globally since 2005, reaching a peak in 2019 before the coronavirus outbreak. Global natural gas consumption reached 3.9 million metric tons in 2018, up 4.6% over the previous year. The switch from coal to natural gas in the US and China, as well as the huge demand from these two nations, were the main causes of this growth. Natural gas, which has a lower carbon footprint than other non-renewable energy sources like coal and oil, is being promoted by laws in other nations as well. Specifically, compared to coal and oil, natural gas produces 20 to 50% less CO2, making it a greener energy source [4].

The examination of predicting natural gas usage was undertaken across various scopes, encompassing global, national, gas distribution system, commercial, residential, and individual customer levels. Diverse datasets, including economic indicators, weather information, historical energy and natural gas consumption records, software simulation data, household survey data, and additional factors like days of the week, were utilized to construct forecasting models. The forecast durations spanned from a few hours to several decades ahead [5]. There are different studies in the literature regarding the estimation of natural gas consumption. Different methods have been used for natural gas consumption such as ARIMA [3,6,7], Neural Networks [6,7], Exponential Smoothing [8] and Holtz-Winter Method [8]. Moreover, artificial intelligence methods such as LASSO [9], Support Vector Regressor [9] is used for natural gas consumption prediction.

The report uses monthly data from 2014 to 2022 on natural gas usage from 28 different European nations. To analyze the data, the Symbolic Aggregate Approximation (SAX) approach is applied. Alphabet vectors for every nation are produced, and analyses are conducted using different numbers of segments and different alphabet sizes. Hierarchical clustering and the creation of dendrogram graphs require these alphabet vectors. Moreover, the elbow approach is employed to establish the proper number of clusters. Based on the number of clusters that are found, groups of nations are formed. Furthermore, the interpretation is based on the consumption patterns of the nation's inside the identified clusters.

2. Material and Method

In this section, general structure of the data will be given. also, method information will be mentioned.

2.1. Data

In this data set, monthly natural gas consumption data of 28 countries between 2014 and 2022 are used. The unit of data is determined as million cubic meters [10].

2.1. Symbolic Aggregate Approximation (SAX)

For time series clustering, the Symbolic Aggregate Approximation (SAX) approach is applied. The SAX method traditionally employs alphabetical symbols to represent and store time-series data. It is well-known for its ability to effectively represent high-dimensional time series data while maintaining the characteristics of the original data points [11]. This discretization procedure is unique because it makes use of a representation that lies between the raw time series and the symbolic characters. Initially, the normalized data is converted into a Piecewise Aggregate Approximation (PAA) model, which is a discrete string representation [12-13].

The data transformation method known as Piecewise Aggregate Approximation (PAA) reduces the data's dimensionality. PAA divides the data into segments of equal length and creates an approximation of the original sequence that is piecewise-constant, presenting the average value of each segment [14]. The SAX approach is an extension of PAA that provides a symbolic representation of time data. To reduce dimensionality, the approach calculates segment PAA values, normalizes the time series using the Z-score, and replaces the segment mean value with a symbolic code. The code is derived from a preset table of discretization intervals with mean values. The table is

Table 1. Example of Monthly Consumption of Countries in 2022

COUNTRY	2022-01	2022-02	2022-03	2022-04	2022-05	2022-06	2022-07	2022-08	2022-09	2022-10	2022-11	2022-12
Belgium	2,208	1,684	1,540	1,272	1,000	906	895	942	977	1,136	1,339	1,860
Bulgaria	384	317	349	203	202	178	187	136	140	143	221	276
Czechia	1,123	892	902	688	394	340	309	308	391	527	761	973
Denmark	301	260	267	207	160	140	114	118	126	156	200	288
Germany	12,108	9,884	9,444	6,967	4,325	3,491	3,679	3,018	4,261	5,039	7,101	9,942
Estonia	60	50	45	31	21	15	11	13	16	19	30	43
Ireland	518	403	464	461	428	416	443	445	397	380	415	519
Greece	581	503	645	303	349	427	533	485	327	246	357	467
Spain	3,741	3,098	3,047	2,403	2,329	2,607	2,779	2,538	2,525	2,507	2,507	2,525
France	5,921	4,420	4,153	3,160	2,040	1,727	1,744	1,595	1,991	2,261	3,425	5,016
Croatia	376	300	268	194	132	127	151	143	138	203	235	254
Italy	9,733	7,672	7,970	5,268	4,199	4,232	4,443	3,810	4,034	4,235	5,578	7,372
Latvia	137	103	98	61	41	19	18	41	38	27	102	158
Lithuania	198	157	198	152	141	145	95	55	90	92	112	214
Luxembourg	95	75	69	55	34	27	25	21	31	39	53	70
Hungary	1,474	1,240	1,200	848	433	440	398	345	529	545	915	1,182
Netherlands	4,492	3,601	3,423	2,818	2,223	2,040	1,848	1,743	1,822	2,208	2,759	4,015
Austria	1,156	970	1,001	705	467	374	348	318	471	559	798	963
Poland	2,390	2,071	2,076	1,798	1,332	1,148	1,103	989	1,050	1,357	1,895	2,230
Portugal	536	450	485	415	453	489	503	479	461	449	475	377
Romania	1,591	1,263	1,247	729	576	490	455	435	493	637	965	1,309
Slovenia	116	93	96	72	53	48	46	42	48	56	79	91
Slovakia	715	561	534	409	227	196	192	154	189	313	451	588
Finland	179	150	162	96	103	94	85	101	75	69	78	89
Sweden	102	82	60	55	50	37	53	54	43	55	54	82
Norway	686	618	676	640	587	1,072	922	253	916	746	902	1,148
North Macedonia	37	40	45	7	4	3	13	23	32	5	34	33
Türkiye	7,137	6,138	6,951	3,698	3,055	2,946	2,767	3,507	3,130	3,269	2,674	5,768

generated based on the observation that values for different types of normalized time-series segments follow a Gaussian distribution. To obtain the levels of discretization for the symbolic code, the Gaussian distribution of the PAA value is divided into equiprobable intervals. The number of symbols in the SAX alphabet for a given method instance is determined by the number of symbolic intervals [15]. Figure 1 shows an example of the SAX application. In this application, the number of segments is 9 and the alphabet size is 8 as SAX parameters. The blue line shows the normalized consumption data. The orange line shows the graph resulting from PAA. Since we reduced it to 9 data, it was created by taking the average of 12 consecutive months of natural gas consumption data. Finally, the PAA data was marked with the symbol of the value corresponding to the values divided into 8 equal parts of the Normal distribution and the Alphabet vector was created. The alphabet vector of each country is created according to the alphabet value corresponding to each "PAA" value. Hierarchical clustering is done according to the "Unicode" values of the letters in these vectors and dendrogram graphics are plotted. The combinations of the number of alphabets and the number of segments are directly proportional to the size of our data set, which is 108. The divisor numbers of the number 108 are determined as the number of segments, and the alphabet number is determined as all numbers between 4 to 10. Combinations between these two values can be selected as SAX parameters. In this study, 4 different combinations are considered such as “Alphabet Size=8 Number of Segment=12”, “Alphabet Size=6 Number of Segment=12”, “Alphabet Size=8 Number of Segment=27”, “Alphabet Size=6 Number of Segment=9”.

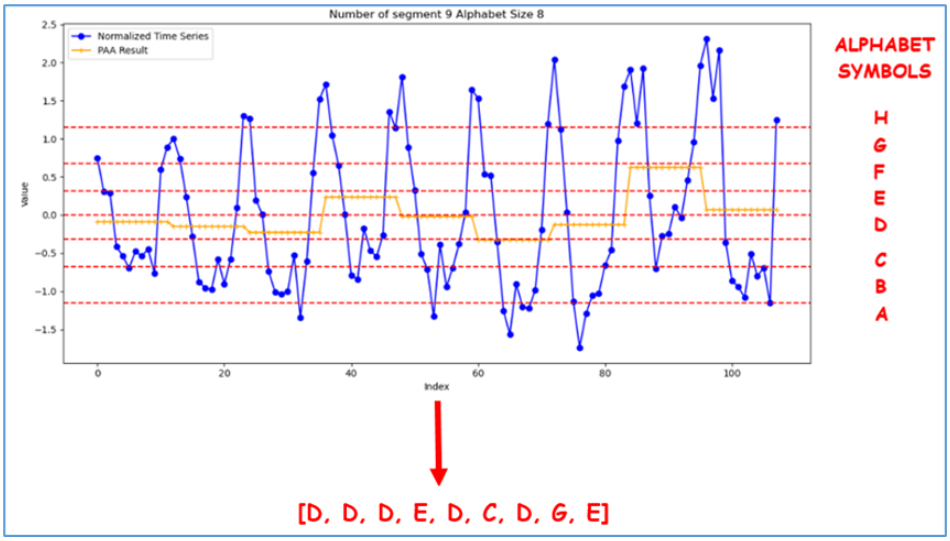
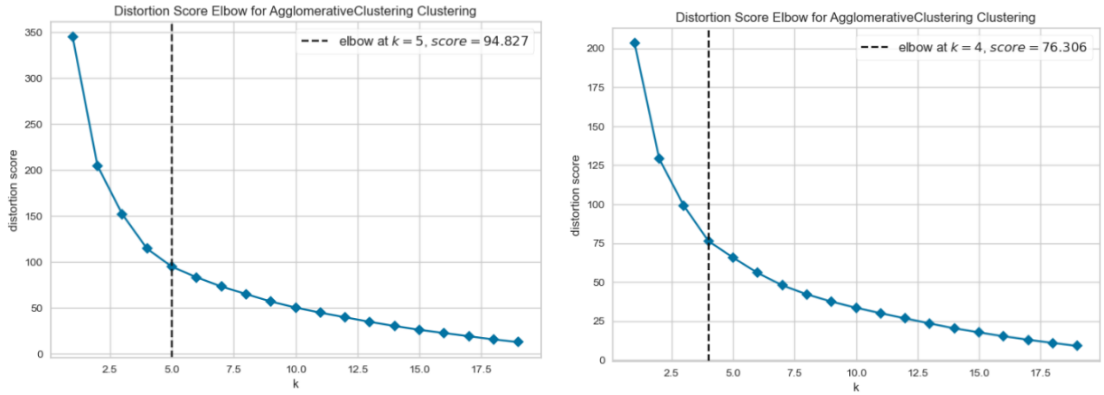


Figure 1. Example of PAA Application for Türkiye Monthly Consumption with Alphabet Size = 8 Number of Segment = 9

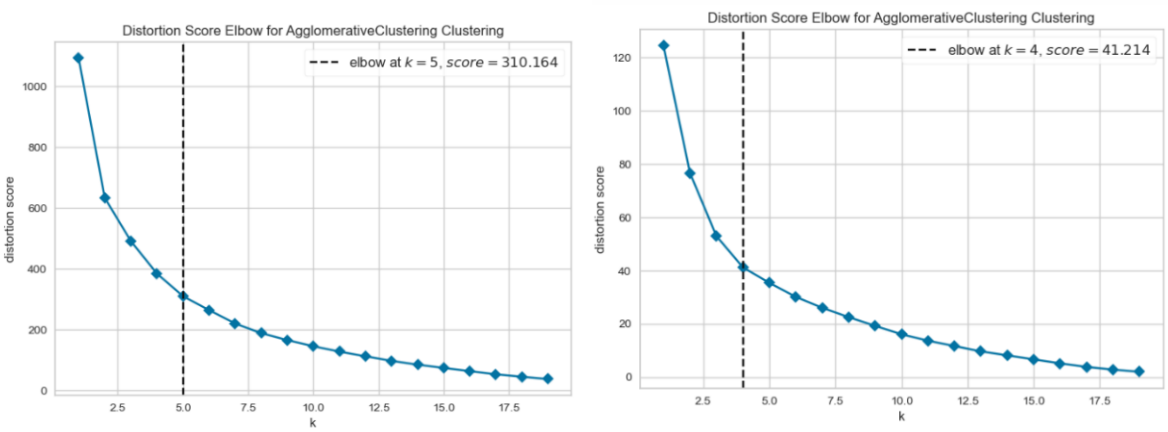
3. Results and Discussion

Dendrogram graphs are plotted for 4 different segment number and alphabet number combinations. Moreover, the "Elbow Method" is used to determine the ideal number of clusters. In Figure 2, the results of the elbow methods applied for Number of Segments 12 and 12, Alphabet Sizes 6 and 8 respectively. Moreover, in Figure 3, the results of the elbow methods applied for Number of Segments 27 and 9, Alphabet Sizes 8 and 6 respectively. According to these results, the appropriate cluster numbers are determined as 5, 4, 5, and 4, correspondingly. Dendrogram graphs are plotted (see figure 4 and 5) for the segment number and alphabet number combinations applied in the elbow method, and the elements in the clusters are determined according to these dendrograms. When the dendrogram graphs are analyzed, the clusters formed according to the number of clusters determined are stated in "Table 2". The results obtained for the case where "Number of Segment = 12" and "Alphabet Size = 8" are as follows. Ireland and Portugal are in the first cluster. Greece Spain and North Macedonia are clustered at the second group. While Denmark, Estonia, Latvia, Romania, Lithuania, Netherlands, Luxemburg, and Finland are grouped together, Türkiye, Sweden, Poland, Slovakia, Belgium, Italy, Hungary, Slovenia, Austria, Czechia, France, Croatia, Bulgaria, and Germany are assigned together. Norway has not acted in concert with any country. More details are given in the next chapter.



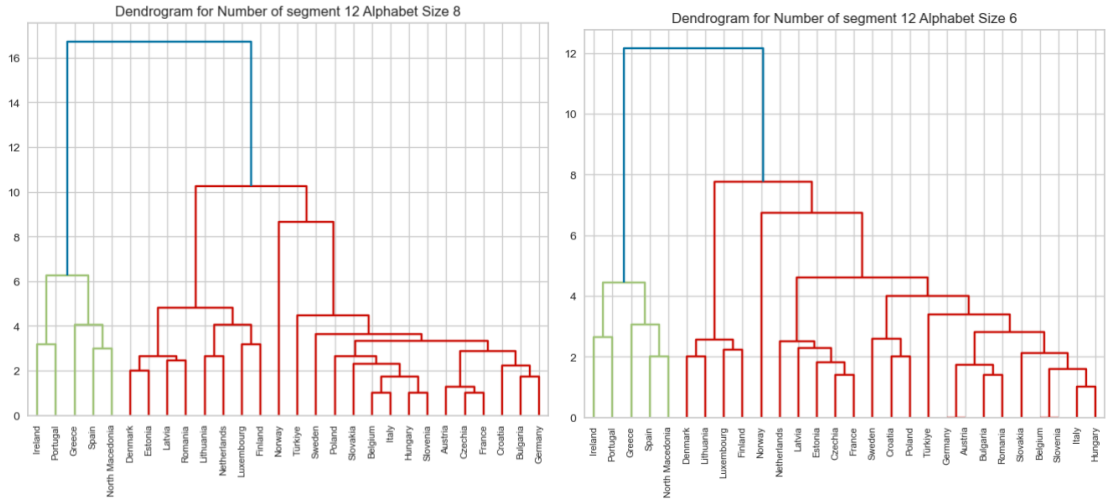
(A) (B)

Figure 2. Application of Elbow Method for Number of Segment = 12 Alphabet Size = 6 and 8



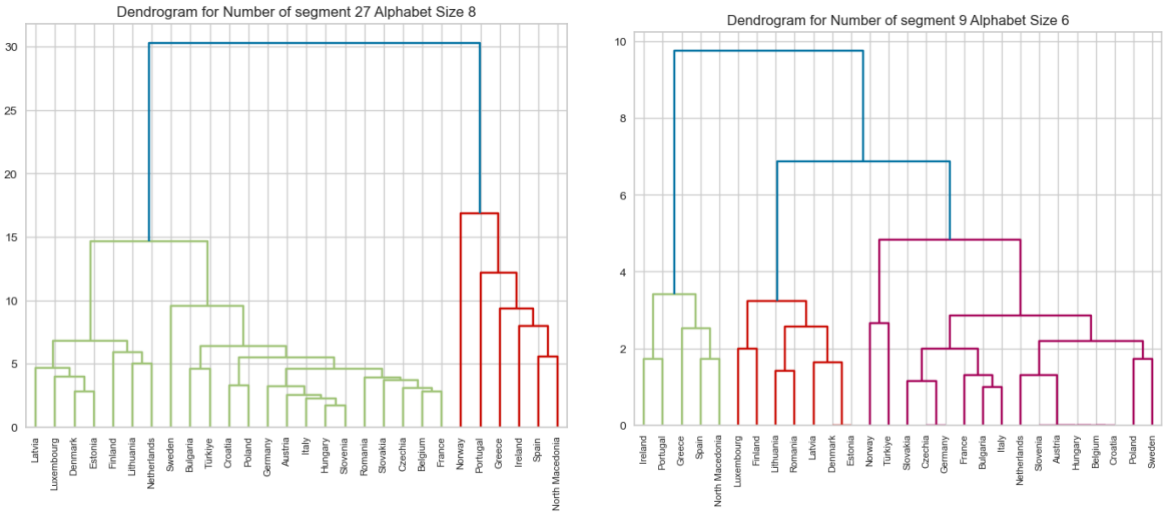
(A) (B)

Figure 3. Application of Elbow Method for Number of Segment = 27 and 9 Alphabet Size = 8 and 6



(A) (B)

Figure 4. Dendrograms for Number of Segment = 12 Alphabet Size = 8 and 6



(A) (B)

Figure 5. Dendrograms for Number of Segment = 27 and 9 & Alphabet Size = 8 and 6

Table 2. Country Clustering in Natural Gas Consumption: Applying the Elbow Method to Segment & Alphabet Analysis

Cluster	Number of Segment = 12 Alphabet Size =8	Number of Segment = 12 Alphabet Size = 6	Number of Segment = 27 Alphabet Size = 8	Number of Segment = 9 Alphabet Size =6
1	Ireland, Portugal	Ireland, Portugal, Greece, Spain, N. Macedonia	Greece, Ireland, Spain, N. Macedonia, Portugal	Sweden, Poland, Croatia, Belgium, Hungary, Austria, Slovenia, Netherlands, Italy, Bulgaria, France, Germany, Czechia, Slovakia
2	Greece Spain, N. Macedonia	Denmark, Lithuania, Luxemburg, Finland	Portugal	Türkiye, Norway
3	Denmark, Estonia, Latvia, Romania, Lithuania, Netherlands, Luxemburg, Finland	Norway	Norway	Estonia, Denmark, Latvia, Romania, Lithuania, Finland, Luxemburg,
4	Norway	Netherlands, Latvia, Estonia, Czechia, France, Sweden, Croatia, Poland, Türkiye, Germany, Austria, Bulgaria, Romania, Slovakia, Belgium, Slovenia, Italy, Hungary	France, Belgium, Czechia, Slovakia, Romania, Slovenia, Hungary, Italy, Austria, Germany, Poland, Croatia, Türkiye, Bulgaria, Sweden	N. Macedonia, Spain, Greece, Portugal, Ireland
5	Türkiye, Sweden, Poland, Slovakia, Belgium, Italy, Hungary, Slovenia, Austria, Czechia, France, Croatia, Bulgaria, Germany	-	Netherlands, Lithuania, Finland, Estonia, Denmark, Luxemburg, Latvia	-

4. Conclusion and Suggestions

Natural gas could be an essential resource for countries around the world. As countries' policies regarding carbon footprint increase, natural gas usage may also increase. In this study, "Symbolic Aggregate Approximation" method is used for made to cluster the natural gas consumption trends of 28 countries. The tests are made on 4 different SAX combinations and countries are tried to be clustered according to the results of the "Elbow method". When the results are analyzed, the consumption trends of the countries might be in 4 groups. Norway can be considered as cluster 1. This country tends to act alone in different tests. The reason for this may be that it has a significantly different natural gas consumption pattern than other countries. The fact that Norway ranks 2nd in Europe [10] in natural gas production and uses these resources efficiently may be another indicator that affects this. The second cluster can be considered as Southern European countries and these can be considered as the "Portugal, Greece, Spain, N. Macedonia" cluster. Warm climatic conditions can reduce natural gas consumption, especially in countries such as Spain and Greece. The impact of the tourism sector may be significant in these countries. Economic growth and energy efficiency policies can influence consumption trends. Furthermore, the reason why countries such as "Ireland" and "Portugal" act together in some tests may be that both countries have small populations, and their energy demands are generally lower. The 3rd cluster can be considered as the cluster of Northern European countries such as Denmark, Estonia, Latvia, Romania, Lithuania, Netherlands, Luxemburg, Finland. Northern European countries may have cold climates, so heating needs may be higher. Economic growth and industrial activities can also affect natural gas

demands. The last group can be Türkiye, Sweden, Poland, Slovakia, Belgium, Italy, Hungary, Slovenia, Austria, Czechia, France, Croatia, Bulgaria, Germany. The countries that make up this cluster generally have large economies and diverse climatic conditions. The level of industrialization, economic growth and population density can affect natural gas demands. Factors such as energy policies, energy security and foreign dependency may also shape this cluster. Given the fact that energy is produced with scarce resources, especially non-renewable ones, accurate forecasting allows planning for the required amount of energy production.

For the future research papers, this study also shows that the SAX method can be applied for different data sets. A comparative study can also be carried out with different forecasting time series methods.

Contributions of the authors

Each author contributed equally to the article.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study has complied with research and publication ethic

References

[1] T. S. Adebayo, M. T. Kartal, and S. Ullah, "Role of hydroelectricity and natural gas consumption on environmental sustainability in the United States: Evidence from novel time-frequency approaches," *Journal of Environmental Management*, vol. 328, no. 116987, p.116987, 2023.

[2] M. O. Turan, T. Flamand, "Optimizing investment and transportation decisions for the European natural gas supply chain," *Applied Energy*, vol. 337, p. 120859, 2023.

[3] S. Yildiz, "Doğal Gaz Tüketim Tahmini," *Sosyal Ve Beşeri Bilimler Dergisi*, vol. 5, no. 1, 440-452, 2013.

[4] M. S. Shaari, T. B. Majekodunmi, N. F. Zainal, N. H. Harun, A. R Ridzuan, "The linkage between natural gas consumption and industrial output: New evidence based on time series analysis," *Energy*, vol. 284, no. 1, p. 129395, 2023.

[5] B. Soldo, "Forecasting natural gas consumption," *Applied Energy*, vol. 92, pp. 26–37, 2012.

[6] O. Kaynar, S. Taştan, F. Demirkoparan, "Yapay sinir ağlari ile doğalgaz tüketim tahmini," *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 25, pp. 463-474, 2012.

[7] O. Çoban, C. C. Özcan, "Sektörel Açidan Enerjinin Artan Önemi: Konya İli İçin Bir Doğalgaz Talep Tahmini Denemesi," *Sosyal Ekonomik Araştırmalar Dergisi*, vol. 11, no. 22, pp. 85-106, 2011.

[8] K. Oruç, K., &Ş. Çelik, "Isparta İli İçin Doğal Gaz Talep Tahmini," *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 22, no. 1, pp.31-42, 2017

[9] Y. Hou, Q. Wang, & T. Tan, "A robust stacking model for predicting oil and natural gas consumption in China. Energy Sources," *Part B: Economics, Planning, and Policy*, vol. 19, no. 1, 2024.

[10] EUROSTAT. (2023). DATABASE. Europa.eu. URL: https://ec.europa.eu/eurostat/ databrowser /view/nr g_cb_gasm/default/table?lang=en

[11] D. H. Yang, & Y. S. Kang, "Distance- and Momentum-Based Symbolic Aggregate Approximation for Highly Imbalanced Classification," *Sensors*, vol. 22, no.14, pp. 5095–5095, 2022.

[12] J. Lin, E. Keogh, S. Lonardi, & B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," *in Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery - DMKD '03*, 2003

[13] Lin, J., Keogh, E., Wei, L., & Lonardi, S. "Experiencing SAX: a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp.107–144, 2007.

[14] A. Roques, & A. Zhao, "Association Rules Discovery of Deviant Events in Multivariate Time Series: An Analysis and Implementation of the SAX-ARM Algorithm," *Image Processing on Line*, 12, pp. 604–624, 2022

[15] J. W Earnest, "Sum of Gaussian Feature-Based Symbolic Representations of Eddy Current Defect Signatures," *Research in Nondestructive Evaluation*, vol. 34, no. 3-4, pp. 136–153, 2023.