Prediction Of Oil Consumption and Oil Access of Countries in The European Union Region with Machine Learning

Murat Beken
Computer Engineering Department
Bolu Abant Izzet Baysal University
Bolu, Turkey
murat.beken@ibu.edu.tr

Onder Eyecıoglu
Computer Engineering Department
Bolu Abant Izzet Baysal University
Bolu, Turkey
onder.eyecioglu@ibu.edu.tr

Nursac Kurt
Software Engineering Department
Istınye University
Istanbul, Turkey
nursac.kurt@istinye.edu.tr

Abstract— Depending on the pandemic process, there are some problems around the world. One of them is the problem of fuel consumption and access to fuel. As a result of breakdowns in supply chains, accessibility and consumption processes for oil have changed. Accordingly, it is of great importance that future oil production, consumption, and access to oil can be predicted by some methods. Artificial intelligence stands out as a tool that can be used in this prediction. In current studies, artificial intelligence is often used for predictive purposes.
In this study, it is tried to predict the future change in oil consumption and access to oil in the European Union region and candidate countries. Decision trees, Naive Bayes, Support vector machines, K nearest neighbor (KNN), and Ensemble Boosted trees were used as methods from supervised and unsupervised machine learning approaches. Depending on the different test parameters of the methods used, the estimation successes were observed and the results were reported.

Keywords—Oil consumption; time-series estimation; artificial intelligence; machine learning

I. INTRODUCTION

European Union countries are calculated according to the share of energy obtained from renewable sources in gross final consumption using various mathematical models [1]. They tried to perform according to the development levels of the countries. A study focused on energy studies per capita in 19 Eurozone countries to achieve the correction targets in EU countries [2]. In addition, there are numerous studies describing the relationship between energy consumption and several macroeconomic indicators such as: (EU-28 countries [3], and new EU Member States [4]).

Sadorsky has defined two empirical models to evaluate the relationship between renewable energy consumption and income in emerging markets [5]. Sadorsky accepted a cointegration relationship and showed that real growth of per capita income resulted in a positive and significant relationship with per capita renewable energy consumption. He confirmed that if real income per capita increases by 1% in the long run, renewable energy consumption per capita in emerging markets will increase by 3.5%. Many studies are focusing on the relationship between renewable energy consumption and various macroeconomic variables from different perspectives and methodologies [6-7]. In a recent article, Duro et al. (2010) A Thil index that breaks down inequality into per capita energy consumption broken down into explanatory factors. Although differences in energy consumption are the most important factor in explaining per capita energy inequality, reducing energy intensity inequality is likely to result in energy per capita energy consumption among OECD countries between 1980 and 2006. shows that it plays an extraordinary role in reducing inequality.

In this study, the annual oil consumption of the EU region and candidate countries, the per capita oil consumption in this region, and the oil purchasing power of the people in this region are obtained from EUROSTAT and the countries in the EU region are determined by Machine Learning methods in the future depending on the consumption size. In addition, the success levels of the different machine learning methods we use were tested and compared with each other.

II. METHOD

A. Machine Learning

We can define machine learning, a sub-branch of artificial intelligence, as programming computers that optimizes a performance criterion using sample data and experiences [8]. We can think of artificial intelligence as the main topic where image processing, cognitive science, neural networks, and much more are combined. Machine learning is also a component under this primary heading.

Machine learning began again in the 1990s as a separate field. Machine learning and data mining often use the same methods, which overlap importantly [9]. Machine learning focuses on predictions from learned data based on known typicals. Data mining focuses on the discovery of unknown features in data [10].

These two areas overlap in many ways. Data mining uses many machine learning methods, but it has different objectives. Machine learning, on the other hand, uses data mining methods such as unsupervised learning or preprocessing steps to improve learner accuracy [8,11].

B. Support Vector Machines

Support vector machines developed by Vapnik are one of the supervised machine learning methods used for classification and regression [12]. It is built on statistical learning theory techniques. Therefore, its theoretical background is strong.

Support vector machines aim to find a function in a multidimensional space that can separate training data with class labels [13]. Class labels are usually divided into positive and negative. The reason for this is to find the optimal hyperplane. The working logic of the support vector machine is shown in Figure 1.

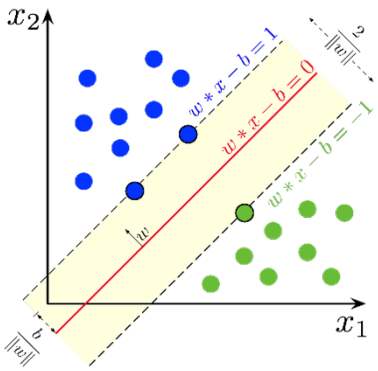

Fig.1. Support Vector Machine

C. *Naive Bayes*

Gaussian Naive Bayes classification algorithm, which is based on the Bayes theorem, is a simple probability-based classification method. The Naive Bayes method is a simple yet powerful algorithm for predictive modeling. Therefore, it is one of the most used algorithms especially, in signal and image processing fields [14].

D. *Decision Trees*

Decision trees are tree-based algorithms used in classification and regression problems. Decision trees are easy to interpret and have good reliability [15]. Therefore, it is one of the most widely used methods among classification algorithms. The decision tree is made up of three main parts: node, branch, and leaf. It is a method that is very easy to understand [16]. In this tree structure, each variable is represented by a node. The branches and leaves are the other part of the tree structure are the elements. The leaves under the decision tree give us the result.

Decision trees are based on historical data, to which class of new data belongs. It decides what it is by making rules [15]. Decision tree, questions asked, and acts by the received answers. It creates the rules according to the answers it receives from the questions. This method makes classification based on rules. The structure of the decision tree is shown in figure 2.

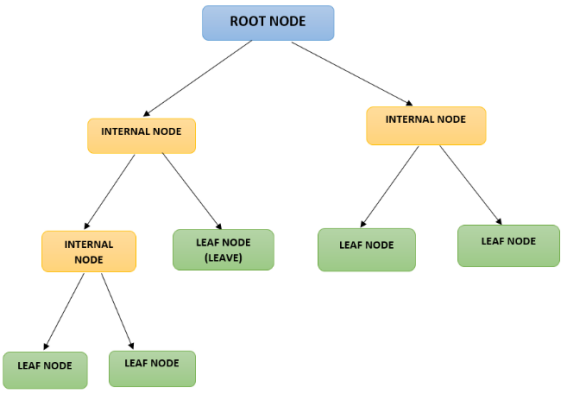

Fig. 2. Decision Trees

E. *K Nearest Neighbor (KNN)*

K-nearest neighbor is a method that makes classification based on distance. This method, which is easy and simple to interpret, is one of the supervised machine learning methods [17]. K-nearest neighbor method, objects in n-dimensional feature space use nearest neighbor samples to classify or predict. In order to make a classification, the nearest neighbor is determined using the Euclidean distance. According to the k number, the data are called positive and negative [17]. Samples are grouped by proximity or distance. The structure and groups of this method are given in figure 3.

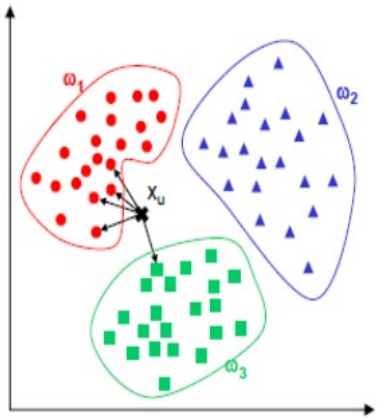

Fig. 3. KNN

F. *Ensemble Boosted Trees*

Ensemble boosted tree is a classification method in which more than one classifier is used in the same classification task. In this method, the results of classifiers with different accuracy scores are combined with different methods. Thus, it is possible to obtain better results from a single classifier [18]. The success of these methods is the learning success of the basic learners and their differences from each other.

III. Experimental Results

In this study, various machine learning methods were used according to the annual oil consumption, per capita oil consumption, and crude oil purchasing power of the European Union countries. As a result, countries are estimated from oil consumption changes. The accuracy of country estimates using Support Vector Machines, Naive Bayes, Decision Trees, K Nearest Neighbor, and Ensemble Boosted Trees are presented in Fig.4., Fig.5., Fig.6., Fig.7., Fig.8.

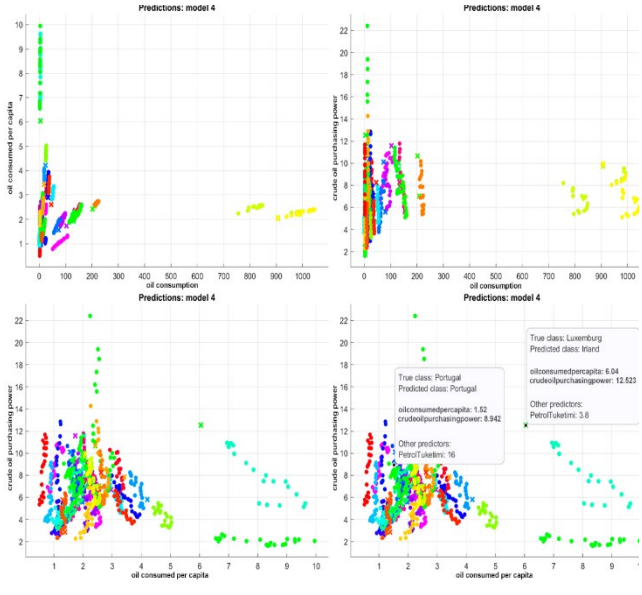

Fig. 4. SVM

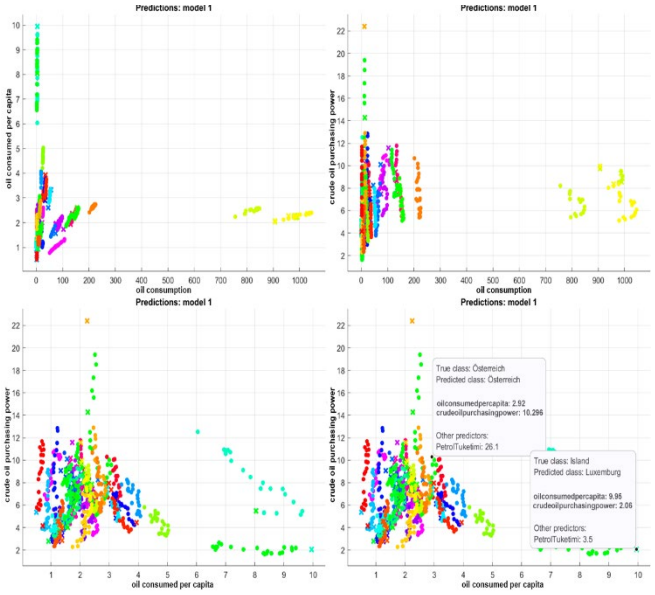

Fig.6. Decision Trees

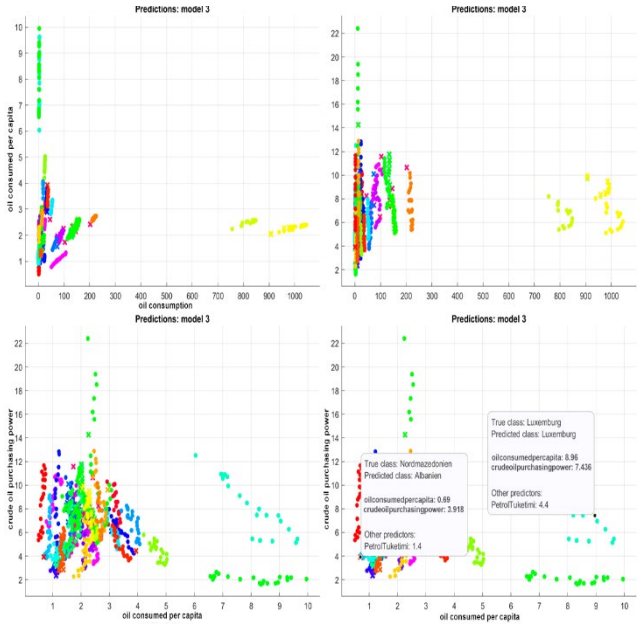

Fig. 5. Naive Bayes

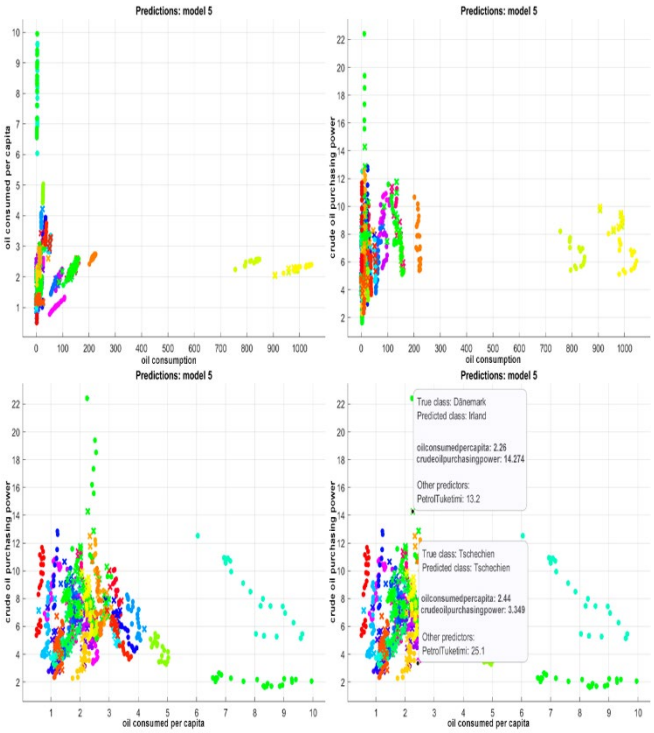

Fig.7. KNN

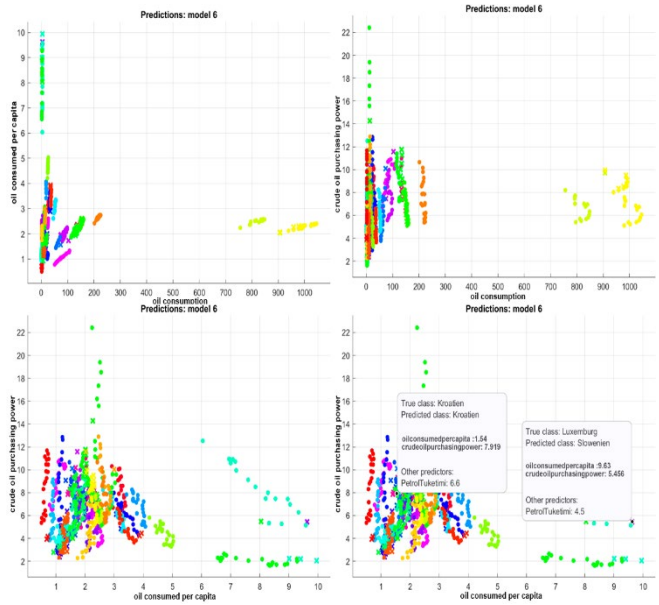

Fig.8. Ensemble Boosted Trees

TABLE II. Accuracy rates of the methods

Method	Accuracy
Decision Trees	%90
Naive Bayes	%87.7
Support Vector Machines	%95.2
K Nearest Neighbor	%67.1
Ensemble Boosted Trees	%82.8

IV. CONCLUSION

In this study, it has been tried to predict which country it is based on the oil consumption, per capita oil consumption, and crude oil purchasing power data of the countries in the European Union region between 2000 and 2020. This study was implemented on the Matlab platform. Support vector machines, Naive Bayes, Decision trees, K nearest neighbor, and Ensemble boosted trees have been used as methods. The accuracy of the success of each method has been tested. When we look at the success accuracy of the methods, the nearest neighbor method gave the lowest success, while the support vector machines gave the highest success.

REFERENCES

1. Cucchiella, F.; D'Adamo, I.; Gastaldi, M.; Miliacca, M. Efficiency, and allocation of emission allowances and energy consumption over more sustainable European economies. J. Clean. Prod. 2018, 182, 805–817.
2. European Commission. A Policy Framework for Climate and Energy in the Period from 2020 to 2030; European Commission: Juba, South Sudan, 2014; Eurostat. Statistics Database; Available online: Http://ec.Europa.Eu/eurostat/data/database (accessed on 22 October 2018).
3. García-Álvarez, M.T.; Moreno, B.; Soares, I. Analyzing the environmental and resource pressures from European energy activity: A comparative study of EU member states. Energy 2016, 115, 1375–1384.
4. Cucchiella, F.; D'Adamo, I.; Gastaldi, M.; Koh, S.C.L.; Rosa, P. A comparison of environmental and energetic performance of European countries: A sustainability index. Renew. Sustain. Energy Rev. 2017, 78, 401–413.
5. Sadorsky, P. Renewable energy consumption and income in emerging economies. Energy Policy 2009, 37,4021–4028.
6. Duro, J.A., Padilla, E., 2006. International inequalities in per capita CO2 emissions: a decomposition methodology by Kaya factors. Energy Economics 28, 170–187.
7. Sun, J.W., 2002. The decrease in the difference of energy intensities between OECD countries from 1971 to 1998. Energy Policy 30, 631–635.
8. Kurt, N., Ozturk, O, & Beken, M. Estimation of Gas Emission Values on Highways in Turkey by Machine Learning, 2021 10th International Conference on Renewable Energy Research and Application (ICRERA). DOI: 10.1109/ICRERA52334.2021.9598769
9. Endüstri4.0. "Makine Öğrenimi Nedir?". Erişim 17 Mayıs 2020. https://www.endustri40.com/makine-ogrenimi-nedir/
10. Kulilik SÜER,Yapay Zeka İle Meme Kanseri Lenf Nodu Analizi (Yüksek Lisans Tezi, Beykent Üniversitesi,2019),13
11. AEA. "Makine Öğrenmesi Nedir?". Erişim 17 Mayıs 2020. https://www.emrealadag.com/makine-ogrenmesi-nedir.html
12. Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20, 273-297. http://dx.doi.org/10.1007/BF00994018
13. Harman, G. (2021) Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini, European Journal of Science and Technology Special Issue 32, pp. 7-13, DOI: 10.31590/ejosat.1041186)
14. Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. Electron. Comput. 14 (3), 326– 334
15. Kokturk, F., K-en yakin komşuluk, yapay sinir ağlari ve karar ağaçlari yöntemlerinin sınıflandırma başarılarının karşılaştırılması (Doktora Tezi, Bülent Ecevit Üniversitesi,2012).
16. Han J, Kamber M. Data mining: Concepts and Techniques (Morgan Kaufmann Publishers), 1st Edition., pp. 332-336, San Francisco, USA, 2000
17. Cunningham P, Delany SJ. K-Neighbor classifiers. Technical Report UCD-CSI2007-4, School of Computer Science and Informatics, University College Dublin, Ireland, 2007.
18. Sirin, E. "Ensemble boosted Trees" (Erişim Tarihi:25.06.2022) https://www.veribilimiokulu.com/ensemble-yontemler-topluluk-ogrenmesi-basit-teorik-anlatim-ve-python-uygulama/
19. Noman, M., Ahsan, M., Experimental Investigation of Diesel Engine Emissions Fueled with Binary and Ternary Blends of Diesel, Biodiesel, and Ethanol, 2019, International Journal of Renewable Energy Research 9(4):2064-2072

20. Aktas, A.Z., A Review and comparison of renewable energy strategies or policies of some countries, 2015 International Conference on Renewable Energy Research and Application (ICRERA). DOI:10.1109/ICRERA.2015.7418490
21. Miguel A. Sanz-Bobi, Cuadra, F., Battle, C., A review of key points of an industrial biogas plant. A European perspective, 2012 International Conference on Renewable Energy Research and Applications (ICRERA)
DOI: 10.1109/ICRERA.2012.6477413