INTERNATIONAL JOURNAL of SMART GRID
M. Beken et al., Vol.12, No.3, September, 2022

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

Murat Beken*‡, Onder Eyecioglu*, Nursac Kurt **

* Department of Computer Engineering, Bolu Abant Izzet Baysal University, Bolu, Turkey

** Department of Software Engineering, Istinye University, Istanbul, Turkey

(murat.beken@ibu.edu.tr, onder.eyecioglu@ibu.edu.tr, nursac.kurt@istinye.edu.tr)

‡ Computer Engineering Department, Bolu Abant Izzet Baysal University, Bolu, Turkey, Tel: +90 374 254 10 00, murat.beken@ibu.edu.tr

Received: 03.09.2022 Accepted: 30.09.2022

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 (SVM), K nearest neighbour (k-NN), 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, prediction of oil consumption

1. Introduction

European Union countries are calculated according to their share in the final consumption of energy obtained from renewable sources 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 many studies explaining the relationship between energy consumption and macroeconomic indicators (EU-28 countries [3], and new EU Member States [4]). Sadorsky explained two empirical models to examine the relationship between renewable energy consumption and income in emerging markets [5]. Acknowledging a cointegration relationship, Sadorsky showed that real growth in per capita income is positively and significantly related to per capita renewable energy consumption. He verified that in the long run, per capita renewable energy consumption in emerging markets will increase by 3.5% as a result of 1% increase in real income per capita. Many studies focus on the relationship between renewable energy consumption and macroeconomic variables from different perspectives and methods [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.

2. Method

2.1. Machine Learning

We can describe machine learning as computer programs that optimize a performance criterion using instance data and experiences [8]. We can consider artificial intelligence, which consists of many sub-titles such as deep learning, computer vision, and neural networks, as the main subject. Machine learning is also part of this main topic.

Machine learning, which started to develop again as a separate field in the 1990s, generally uses the same methods as data mining without much difference [9]. The difference between machine learning and data mining is that data mining tries to find unknown features in the data. Machine learning relies on predictions made from learned data based on certain typical properties [10].

These two fields are similar in many ways. Data mining has different purposes while using many machine learning methods. However, machine learning aims to increase learner accuracy by using data mining methodologies such as unsupervised learning or pre-processing steps [8-11].

2.2. Support Vector Machine (SVM)

Support vector machines (k-NN) developed by Vapnik are one of the supervised machine learning methodologies used for regression and classification [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.

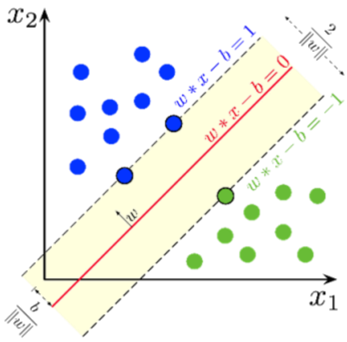

Fig. 1. Graphical representation of Support vector machine algorithm.

2.3. Naive Bayes

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

2.4. 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]. For this reason, it is one of the most 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.

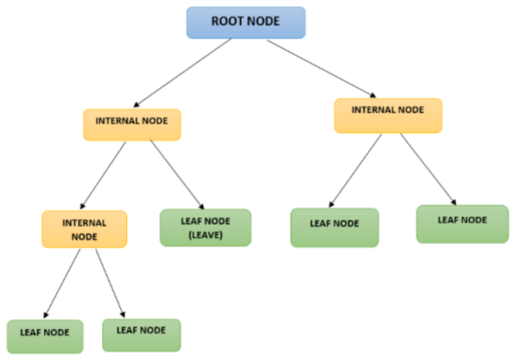

Fig. 2. Graphical representation of Decision tree algorithm.

2.5. k-Nearest Neighbour (k-NN)

k-nearest neighbour (k-NN) is a methodology 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 neighbour method, objects in n-dimensional feature space use nearest neighbour samples to classify or predict. In order to make a classification, the nearest neighbour 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.

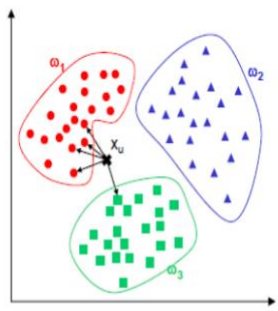

Fig. 3. Graphical representation of k- Nearest Neighbour (k-NN) algorithm

2.6. Ensemble Boosted Trees

Ensemble boosted tree is a classification method in which too many classifiers are used in the same classification duty. In this method, the results of classifiers with different accuracy rates are compounded with different methodologies. Thus, we can achieve better conclusions from a single classifier.[18]. The success of these methods is the learning success of the basic learners and their differences from each other.

3. Experimental Results

In this study, various machine learning methodologies were used according to the annual oil consumption, per capita oil consumption, and crude oil purchasing power of the European Union countries. Estimated values of the crude oil purchasing power of the countries depend on oil consumption using machine learning regression models are presented in Fig.4., Fig.5., Fig.6., Fig.7., and Fig.8. As a result, countries are estimated from oil consumption changes. The accuracy results of Support Vector Machines, Naive Bayes, Decision Trees, K Nearest Neighbour, and Ensemble Boosted Trees models are listed in Table 1. Results show that among the all methods we used, the support vector machine approach is the most successful one while the k-Nearest Neighbour is the least successful one. It is also important to note that we didn't optimize any methods we used.

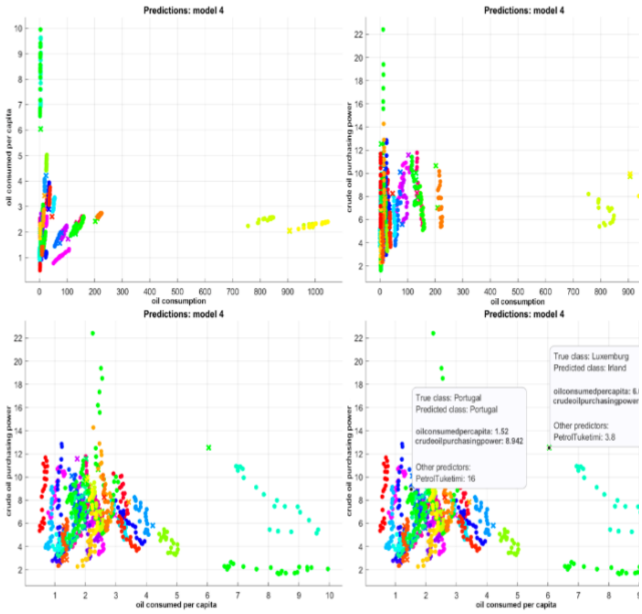

Fig. 4. Estimated values of the crude oil purchasing power by SVM.

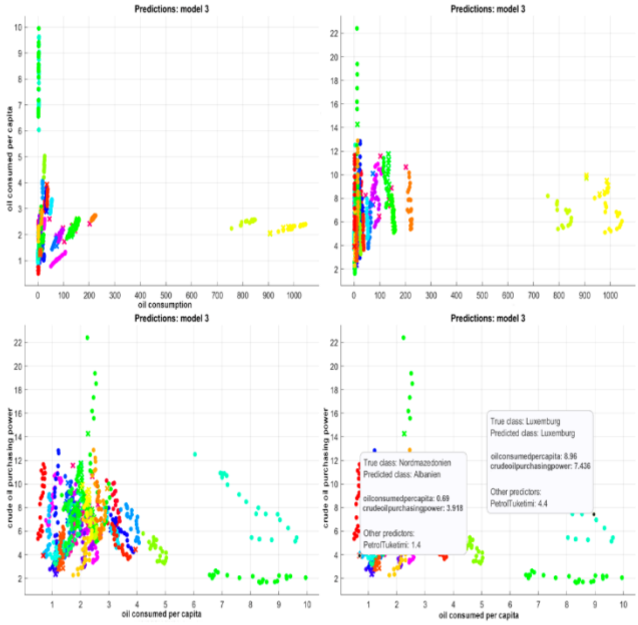

Fig. 5. Estimated values of the crude oil purchasing power by Naive Bayes.

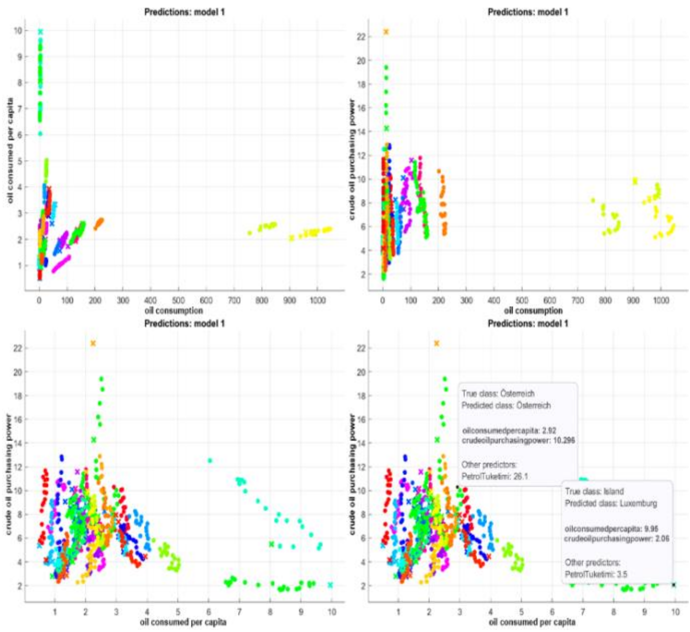

Fig. 6. Estimated values of the crude oil purchasing power by Decision Trees.

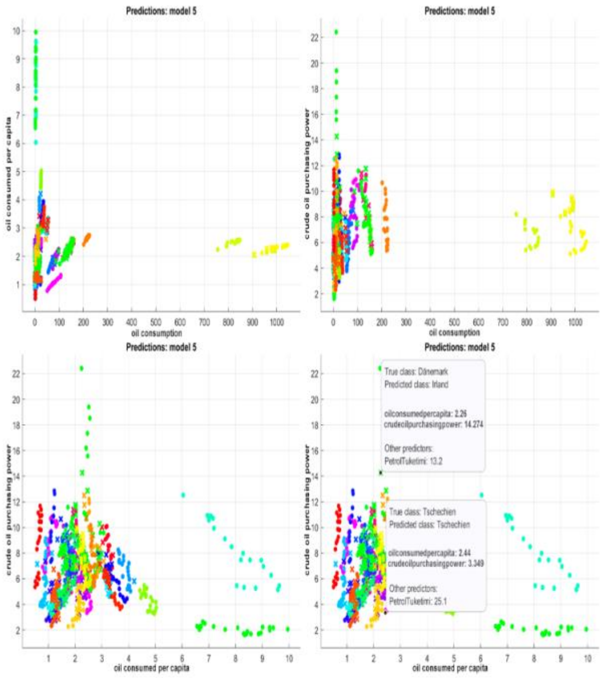

Fig. 7. Estimated values of the crude oil purchasing power by k-NN.

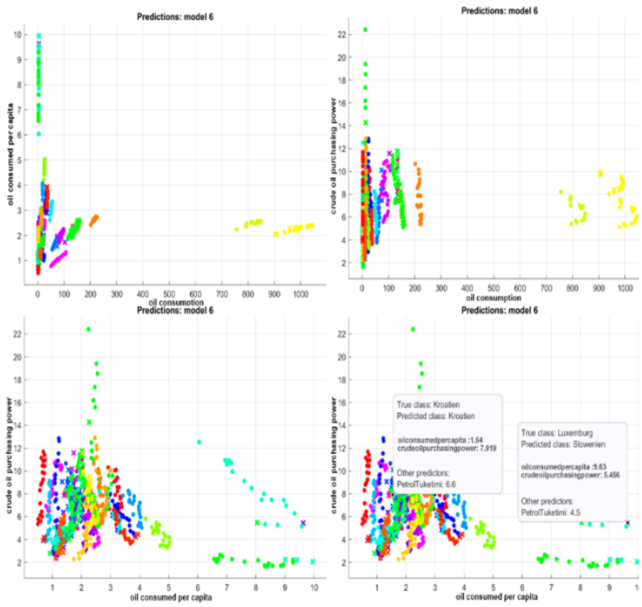

Fig. 8. Estimated values of the crude oil purchasing power by Ensemble Boosted Trees.

Table 1. 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

4. 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 neighbour, 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 neighbour method gave the lowest success, while the support vector machines gave the highest success. Even though support vector machine gave best results in terms of accuracy, it should also be noted that run-time speed, number of data points required are also important factors to select a classification algorithm. Besides, we didn't use artificial neural networks in this study. For our future work, we plan to investigate the performance of the classification methods in a more detailed manner.

5. References

[1] F. Cucchiella, I. D'Adamo, M. Gastaldi, and M. Miliacca, "Efficiency and allocation of emission allowances and energy consumption over more sustainable European economies," Journal of Cleaner Production, vol. 182, pp. 805-817, 2018.

[2] E. Commission, "A policy framework for climate and energy in the period from 2020 to 2030", Official Journal of the European Union Brussels, Belgium, p. 15, 2014.

[3] M. T. García-Álvarez, B. Moreno, and I. Soares, "Analyzing the environmental and resource pressures from European energy activity: a comparative study of EU member states," Energy, vol. 115, pp. 1375-1384, 2016.

[4] F. Cucchiella, I. D'Adamo, M. Gastaldi, S. L. Koh, and P. Rosa, "A comparison of environmental and energetic performance of European countries: A sustainability index," Renewable and Sustainable Energy Reviews, vol. 78, pp. 401-413, 2017.

[5] P. Sadorsky, "Renewable energy consumption and income in emerging economies," Energy policy, vol. 37, no. 10, pp. 4021-4028, 2009.

[6] J. A. Duro and E. Padilla, "International inequalities in per capita CO2 emissions: a decomposition methodology by Kaya factors," Energy Economics, vol. 28, no. 2, pp. 170-187, 2006.

[7] J. Sun, "The decrease in the difference of energy intensities between OECD countries from 1971 to 1998," Energy Policy, vol. 30, no. 8, pp. 631-635, 2002.

[8] N. Kurt, O. Ozturk, and M. Beken, "Estimation of gas emission values on highways in Turkey with machine learning", 2021 10th International Conference on Renewable Energy Research and Application (ICRERA), IEEE, pp. 443-446, 2021.

[9] E. Aladağ. "Makine Öğrenmesi Nedir." https://www.emrealadag.com/makine-ogrenmesi-nedir.html (accessed 17 June, 2022).

[10] K. Süer, "Yapay zeka ile meme kanseri lenf nodu analizi," Master of Science, Fen Bilimleri Enstitüsü, Beykent Üniversitesi, Turkey, 547247, 2019. [Online].

[11] M. A. Alkan. "Makine Öğrenimi Nedir?" https://www.endustri40.com/makine-ogrenimi-nedir/ (accessed 20 August, 2022).

[12] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273-297, 1995.

[13] G. Harman, "Destek Vektör Makineleri ve Naive Bayes Sınıflandırma Algoritmalarını Kullanarak Diabetes Mellitus Tahmini," Avrupa Bilim ve Teknoloji Dergisi, no. 32, pp. 7-13, 2021, doi: 10.31590/ejosat.1041186.

[14] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition," IEEE transactions on electronic computers, no. 3, pp. 326-334, 1965.

[15] F. Köktürk, "K-En Yakın Komşuluk, yapay sinir ağları ve karar ağaçları yöntemlerinin sınıflandırma başarılarının karşılaştırılması," Doctor of Philosophy, Sağlık Bilimleri Enstitüsü, Biyoistatistik Ana Bilim Dalı, Bülent Ecevit Üniversitesi, Turkey, 2012. [Online].

[16] J. Han, J. Pei, and H. Tong, Data mining: concepts and techniques. Morgan Kaufmann, 2022.

[17] P. Cunningham and S. J. Delany, "k-Nearest Neighbour Classifiers," ed: University College Dublin. School of Computer Science and Informatics, 2007.

[18] E. Şirin. "Ensemble Yöntemler (Topluluk Öğrenmesi): Basit Teorik Anlatım ve Python Uygulama." https://www.veribilimiokulu.com/ensemble-yontemler-topluluk-ogrenmesi-basit-teorik-anlatim-ve-python-uygulama/ (accessed 14 August 2022).