Machine Learning on Energy Consumption

Chapter 1: Introduction

1.1 Study objective

The aim of this study is to develop a machine learning model to forecast Greece's energy needs and analyze the current state of renewable energy adoption, electricity demand, and carbon emissions. By integrating data from multiple sources, this research provides a comprehensive overview of global and local energy patterns. The focus areas include the progress of renewable energy use, global electricity consumption, and carbon dioxide emissions from energy production. In addition, a comparative analysis of energy sources for the G10 countries, a survey of energy potential in the Balkan region, and a model for forecasting Greece's energy needs for the next five years are carried out. The findings of the analyses highlight the critical role of policy, technological developments and economic factors in shaping energy developments, underlining the importance of predictive modeling in energy planning and sustainability actions.

1.2 Dataset description

International" dataset was utilized . Energy Statistics ", available on Kaggle , is a comprehensive repository compiled by the United Nations and includes global energy data from 1990 to 2014. This dataset serves as a key information vehicle for understanding energy production, energy trade and consumption across a variety of energy products, including solid, liquid and gaseous fuels, as well as electricity and heat. It comes from the United Nations Statistics Division (UNSD) annual questionnaire on energy statistics, which is an integral part of the broader energy data collection program.

The dataset information includes extensive details on the production, conversion and use of energy products, offering a holistic picture of energy potential at a global level. The core elements of the dataset include production data, which covers the collection, extraction or manufacturing of energy products; trade data, which details imports and exports; and consumption data, which highlights the final use of energy in various sectors. Finally, the dataset includes conversion factors for the various energy products to facilitate consistent reporting and analysis.

UNSD guidelines , data are carefully categorized and recorded to ensure accuracy and consistency. Energy products are classified into solid products, such as coal and peat, liquid products, such as crude oil and refined petroleum products, gaseous products, including natural gas, and renewable energy sources and wastes. Each category includes specific definitions and categories to standardize data collection across countries. For example, solid products include various types of coal and derived products, while liquid products cover crude oil, natural gas liquids, and a range of refined petroleum products, such as motor gasoline and jet fuel.

International " dataset Energy Statistics " comprehensively records transformation processes, energy industries and final consumption flows. Transformation processes refer to the conversion of energy from one form to another, such as refining crude oil into other petroleum products or generating electricity from heat. The data records the different stages and technologies involved in the process of these transformations, ensuring a detailed understanding of the energy transformation. Energy industries record consumption within the energy sector itself, such as energy use by coal mines, oil refineries and power plants. These data help to assess the efficiency and energy requirements of the energy sector. Flows related to final consumption include non-energy uses, such as chemical raw materials, as well as actual energy use such as households, transport. Finally, this detailed categorization ensures that the dataset is meticulously structured to include comprehensive definitions of flow data and reporting guidelines, maintaining data integrity and comparability across different reporting entities.

1.3 What is the usefulness of forecasting?

Accurate energy forecasts are essential for multiple factors. According to a study published in the scientific journal Energy by Maaouane , Zouggar Krajačić and Zahboune [5], reliable forecasts allow governments and policymakers to make informed decisions on energy policy and infrastructure investments. These decisions are crucial for anticipating future energy needs, avoiding shortages and ensuring a stable energy supply. Furthermore, accurate forecasts enhance environmental sustainability by providing information on the potential impacts of different energy sources on carbon dioxide emissions. Understanding this helps shape strategies to mitigate climate change and achieve carbon dioxide emission reduction targets.

From the perspective of businesses and investors, accurate energy forecasts play a crucial role in strategic planning and investment decisions. As researchers Chen and Tan [7] analyze in their article on electricity quantity forecasting using matrix factorization, reliable forecasts allow for the optimization of energy production and consumption, reducing operating costs and improving efficiency. This is particularly important for industries that are highly dependent on energy, where small changes in energy availability or cost can significantly affect profitability and operational stability. Finally, valid forecasts facilitate the integration of renewable energy sources into the grid, enhancing energy security and supporting the transition to a low-energy, low-carbon economy.

In addition to the economic and environmental benefits, accurate energy forecasts play a crucial role in the operational management of energy systems. As highlighted by Rahman , Srikumar and Smith [4], forecasting models are extremely important for load forecasting, which is a fundamental element of effective network management. Energy load forecasting contributes to balancing supply and demand, optimizing the operation of power plants and reducing the risks of power outages or energy surpluses. This operational efficiency is essential for maintaining the reliability and resilience of the energy infrastructure.

Integrating predictive models into energy planning also addresses the variability and uncertainty associated with renewable energy sources. Renewable energy sources, such as wind and solar, are inherently volatile, and their integration into the energy mix requires accurate forecasts to effectively manage their variability. Predictive models can predict renewable energy generation based on weather conditions and historical data, allowing grid operators to plan backup power sources and storage solutions accordingly. This ensures consistent and reliable energy supply, even as the share of renewable energy sources in the energy mix increases.

In conclusion, the ability to accurately forecast energy quantities enhances decision-making processes in multiple dimensions of the energy sector. It supports the strategic planning of energy infrastructure, optimizes operational efficiency, mitigates environmental impacts and ensures energy security. By leveraging advanced machine learning techniques, this study aims to contribute to the development of a forecasting model that addresses these critical needs, ultimately supporting the transition to a sustainable and resilient energy future.

Chapter 2: Data Analysis

2.1 Trends in the adoption of renewable energy sources

The adoption of renewable energy sources is influenced by a complex interplay of factors, including technological developments, economic incentives, regulatory frameworks and social acceptance. According to the Renewable and Sustainable Energy Reviews (Painuly, 2001) [12], several constraints hinder the widespread adoption of renewable energy sources. These include high initial costs, lack of infrastructure and inadequate regulatory support. Overcoming these barriers requires concerted efforts in many areas, including policy formulation, financing mechanisms and technological innovation.

Economic incentives play a key role in promoting the adoption of renewable energy sources. Subsidies, tax incentives and feed-in tariffs are some of the economic mechanisms that have been successfully implemented in various countries to encourage investment in renewable energy projects. Technological advances, especially in solar and wind energy, have significantly reduced the costs of these technologies, making them more competitive with conventional energy sources. In addition, public awareness and social acceptance are crucial for the successful implementation of renewable energy projects. Community engagement and educational campaigns can help address concerns related to the environmental and aesthetic impacts of renewable energy installations.

Analysis of the dataset revealed significant trends in renewable energy adoption from 1990 to 2015. The graph created depicts trends across various renewable energy sources, including wind, hydroelectric, solar, geothermal, and biomass. The data was grouped by year, and the total amount of each renewable source was summed for each year to illustrate adoption trends.

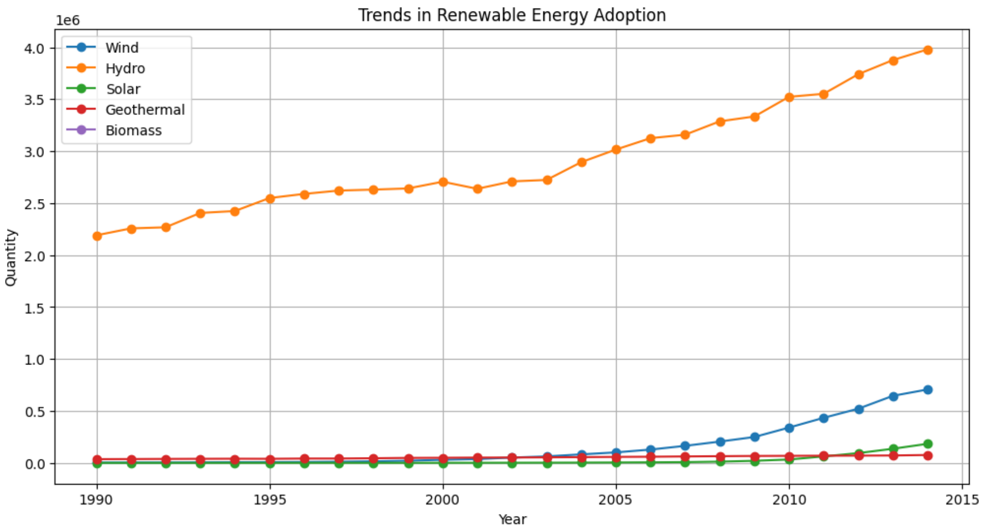


Figure 1: Renewable energy patterns graph

The graph shows that hydropower has consistently been the largest contributor to renewable energy production over the years, with a steady increase in its quantity. Wind and solar power show significant growth, especially in the latter part of the timeline, reflecting technological advances and increasing investment in these sectors. Biomass and geothermal energy show relatively stable trends with moderate growth.

2.2 Electricity demand

Electricity demand is a critical aspect of energy management and planning, as it exhibits significant seasonal fluctuations and is correlated with economic activities. The analysis of global electricity demand from 1990 to 2015, as depicted in the first graph, shows a steady increase in electricity consumption over time. This trend is influenced by several factors, such as population growth, industrialization, and technological advancement.

Economic growth significantly affects electricity consumption patterns. As economies expand, industrial and commercial activities increase, leading to higher electricity consumption. Developing countries, in particular, experience sharp increases in electricity demand as they undergo industrialization and urbanization. The correlation between economic growth and electricity demand highlights the importance of integrating energy planning with economic development strategies. As noted by the World Bank (2020), sustainable economic growth relies heavily on the availability of reliable and affordable electricity.

Technological developments also play a critical role in shaping electricity demand. The rise of digital technologies and the increased use of electronic devices have contributed to the increase in electricity consumption. The advent of electric vehicles (EVs) is a prime example of how technological innovation is increasing electricity demand. The International Energy Agency (IEA , 2020) highlights that the transition to a digital and electrified economy will continue to lead to significant increases in electricity demand.

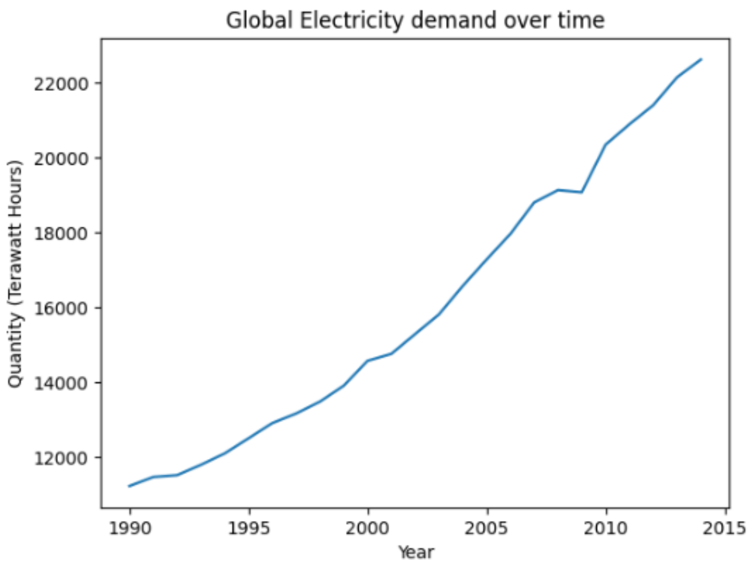


Figure 2: Graph of electricity demand over time

The chart presents an analysis of global electricity demand from 1990 to 2015, measured in terawatt hours (TWh). The data illustrates a steady upward trajectory in electricity consumption over this 25-year period. The increase in global electricity demand can be attributed to a number of factors, including population growth, economic development, and technological advances. The steady increase in demand highlights the need for continued investment in energy infrastructure and the importance of strategic energy planning to ensure that supply can meet growing demand.

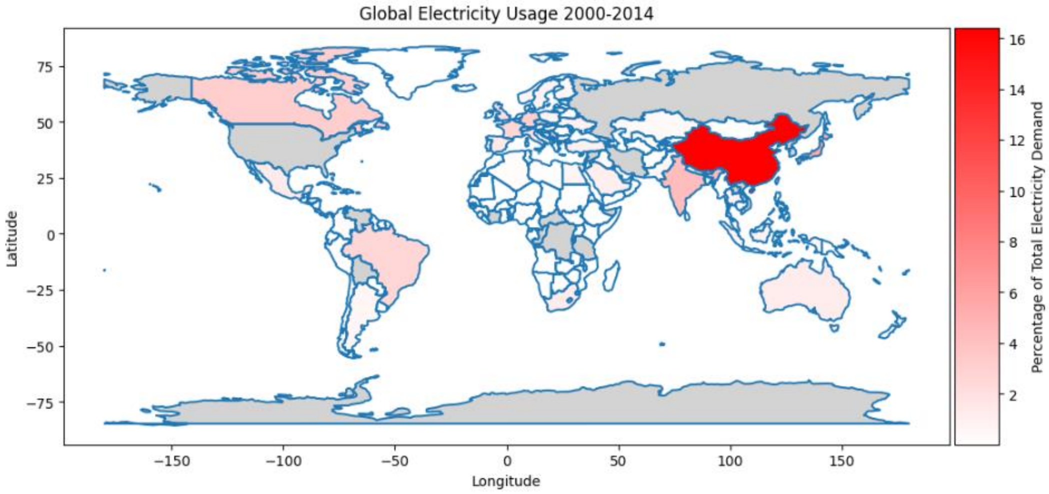


Figure 3: Visualization of electricity use on a map

The graph provides a spatial representation of global electricity use from 2000 to 2014, highlighting the percentage of total global electricity demand attributable to each country. This visual analysis reveals significant regional disparities in electricity consumption. In particular, countries such as China and the United States emerge as major consumers, reflecting their large populations and significant industrial bases. China's rapid industrialization and economic expansion are particularly evident, as it accounts for a significant portion of global electricity use. In contrast, other regions show lower levels of consumption, which can be attributed to varying degrees of economic development, industrial activity, and energy policy implementation. The map highlights the need for tailored energy strategies that respond to specific regional requirements and promote equitable access to energy resources. Furthermore, the data highlight the importance of international cooperation to address global energy challenges, particularly in regions with rapidly growing electricity demands.

In conclusion, the analysis of global electricity demand and its contributing factors provides valuable insights into current and future energy needs. By understanding these trends, policymakers, energy planners and investors can make informed decisions to ensure a sustainable and reliable energy future. Incorporating advanced forecasting models and continuously monitoring energy trends is crucial to addressing the challenges and opportunities in the energy sector. This holistic approach will help balance economic growth with environmental sustainability and energy security.

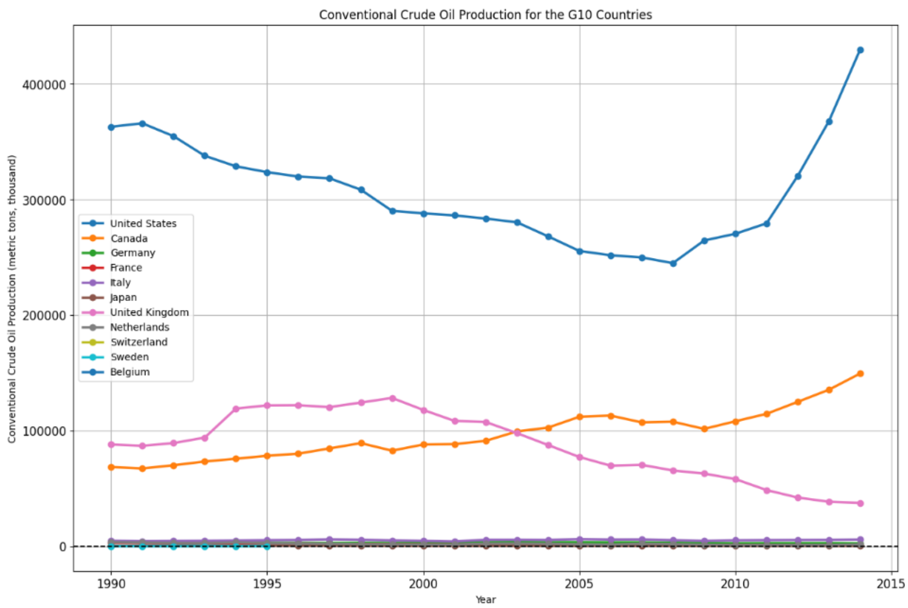
2.3 Comparative analysis of energy production of the G 10

The Group of Ten (G10) comprises countries with the world's most important and advanced economies, including the United States, Canada, Germany, France, Italy, Japan, the United

Kingdom, the Netherlands, Switzerland, Sweden, and Belgium. These nations collectively represent a significant portion of global economic activity, contributing over 46% of world gross domestic product (GDP) in nominal terms and more than 62% of global net wealth, amounting to approximately $300 trillion (World Population Review , 2024).

Crude Oil Production

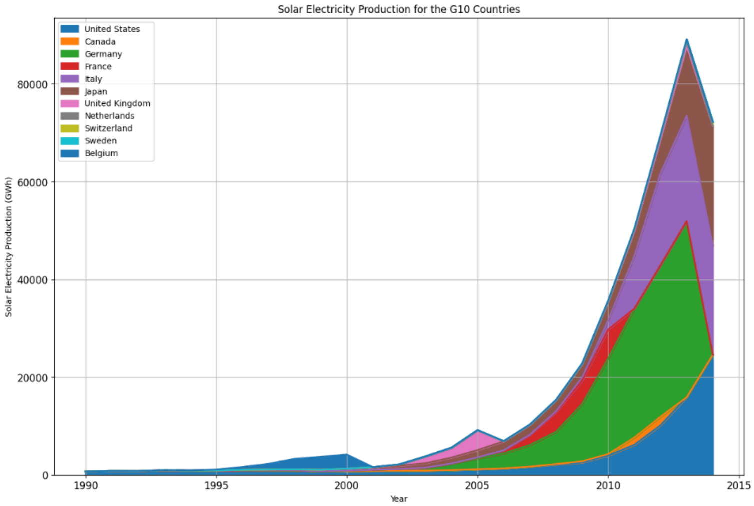
The pie chart depicts trends in conventional crude oil production among the G- 10 countries from 1990 to 2015. The data reveal distinct production patterns, reflecting each country's resource availability and technological advancements. The United States, for example, experienced a decline in crude oil production from about 1990 to 2008. This decline was followed by a significant increase, attributed to advances in extraction technologies, particularly hydraulic fracturing and horizontal drilling. This technological innovation has reestablished the United States as a leading oil producer in recent years. Similarly, Canada has shown a steady increase in crude oil production, capitalizing on its significant oil resources and investments in extraction technology. In contrast, traditional European producers such as the United Kingdom and Italy are experiencing a gradual decline in production, indicative of mature oil fields and the challenges associated with extracting the remaining reserves.



These trends illustrate the varying stages of oil production maturity and highlight the critical impact of technological advances and resource availability on national production capabilities. Understanding these dynamics is crucial for shaping energy policy, economic planning, and strategic investments within the G10 and globally.

Solar Electricity Generation

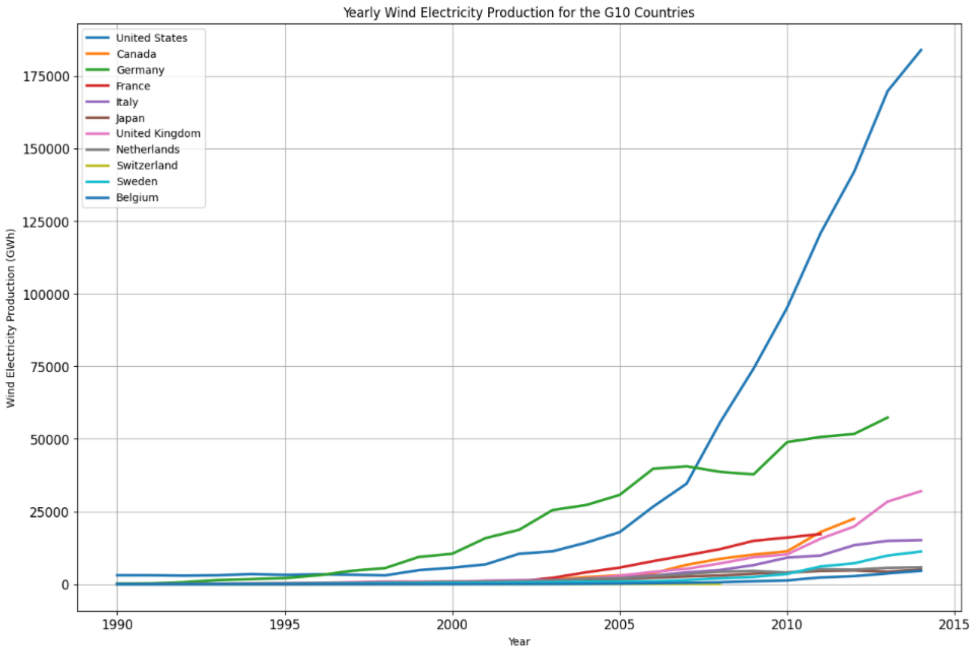
The chart highlights the trends in solar power generation in the G10 countries from 1990 to 2015. A dramatic increase in solar power generation is evident, especially since 2005. Germany is emerging as a leader in solar power generation, driven by strong government policies, economic incentives and significant technological developments. The country's commitment to renewable energy, as evidenced by its energy transition policy, has facilitated significant investment and innovation in solar technology.



Other G10 countries, such as the United States, Italy and Japan, are also showing significant growth in solar electricity generation. These trends reflect concerted efforts to diversify energy sources and reduce carbon emissions. The rapid growth in solar power generation underscores the global shift towards renewable energy sources and highlights the importance of supportive policies, technological advances and economic incentives for the adoption of clean energy sources.

<u>Wind Power Generation</u>

The chart shows wind power generation trends among G10 countries from 1990 to 2015. The United States leads in wind power generation, showing sharp growth, especially after 2005. This growth can be attributed to favorable regulatory frameworks, significant investments and advances in wind turbine technology. Germany and the United Kingdom also show significant increases in wind power generation, supported by strong government policies and investments in wind farms.



The increase in wind power generation across the G10 countries highlights the growing importance of wind power as a sustainable and sustainable energy source. This trend reflects global efforts to reduce dependence on fossil fuels, mitigate the impacts of climate change and transition to a low-carbon energy system.

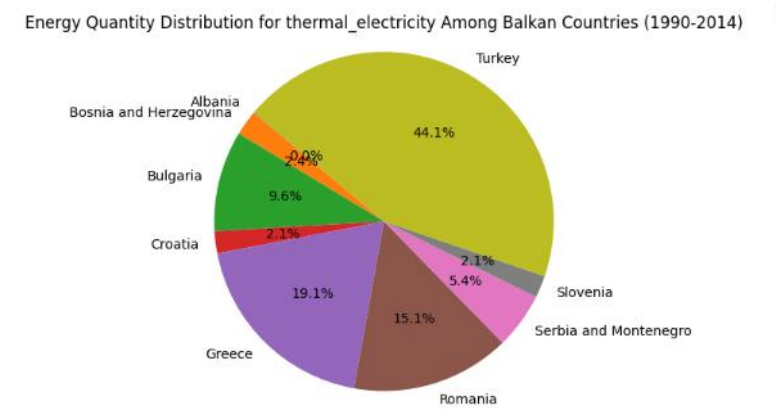
In conclusion, the comparative analysis of energy production trends in the G10 countries highlights the different approaches and achievements in the conventional and renewable energy sectors. This knowledge is crucial for shaping future energy policies, promoting technological innovations, and promoting sustainable energy practices worldwide. By leveraging the lessons learned from the G10 countries, other nations can improve their energy strategies, contributing to a more sustainable and resilient global energy system.

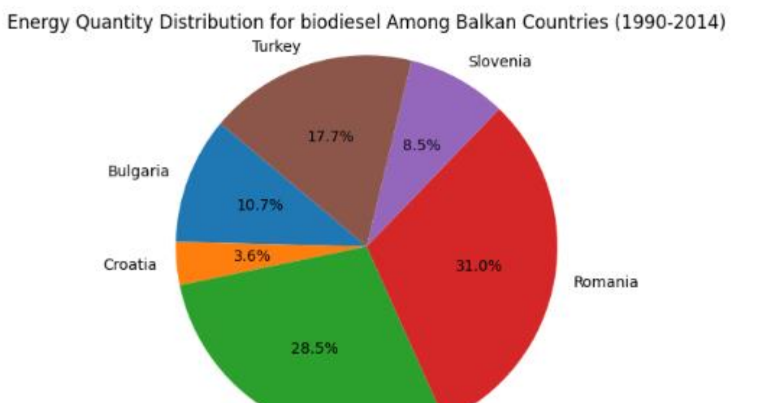
2. 4 Energy patterns in the Balkan Peninsula

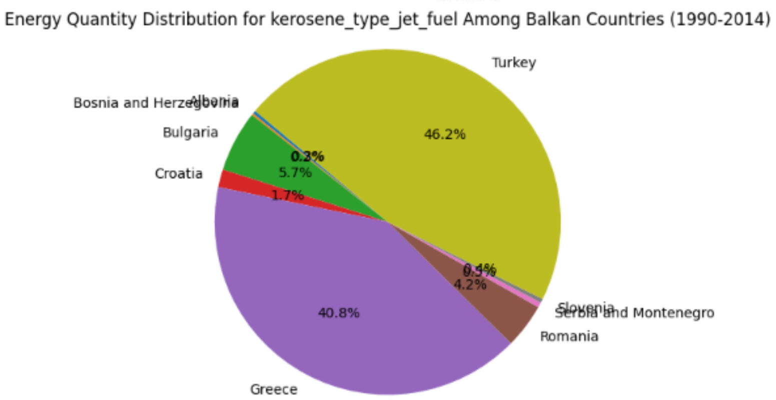
This analysis examines energy trends in the Balkan countries using the International Energy Statistics dataset. The Balkan region, which consists of Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Greece, Turkey, FYROM, Romania, Serbia and Montenegro, and Slovenia, is characterized by a diverse energy landscape and different levels of development and resource availability. Understanding energy trends in this region is crucial for several reasons, including policymaking, regional cooperation, and sustainable development.

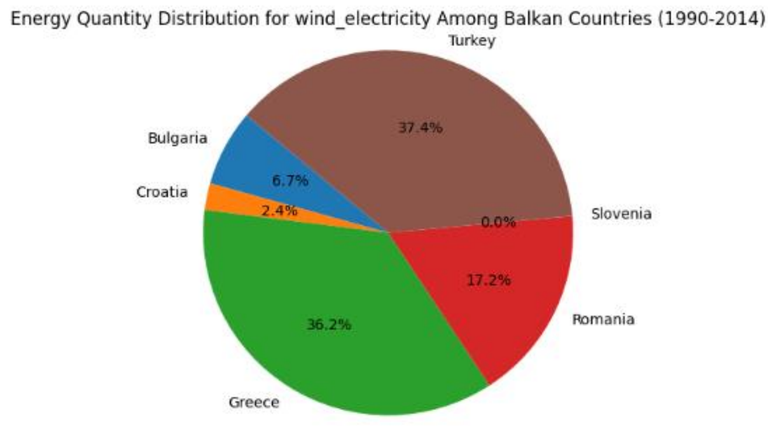
To analyze energy trends in the Balkan countries, we compiled the total energy quantities by country and category for the period from 1990 to 2014. The data were processed to extract relevant trends and illustrate the distribution of energy quantities across different categories and countries within the region.

Indicative diagrams:









In conclusion, the Balkan region has significant potential for the development of renewable energy sources, particularly in hydropower, solar and wind power, due to its geographical diversity. Mountainous areas are ideal for hydroelectric power, with countries such as Albania and Bosnia and Herzegovina already investing in hydroelectric infrastructure. The region's favourable climatic conditions, characterised by high levels of solar radiation, make it suitable for solar power projects, with notable development in Greece and Turkey. Coastal and elevated areas offer excellent wind resources, providing opportunities to diversify the energy mix and reduce carbon emissions. Exploiting these renewable resources can reduce dependence on imported fossil fuels, enhance energy security and promote environmental sustainability.

Chapter 3: Applying Machine Learning

Selecting and implementing a machine learning model are critical steps in predictive modeling. This study uses Scikit - learn , a flexible and accessible machine learning library in Python , to develop a predictive model to forecast Greece's energy quantities for the next five years.

Scikit-learn was chosen for its extensive array of supervised and unsupervised learning algorithms, making it well-suited for a variety of predictive modeling tasks. As an open source library, it supports a wide range of algorithms, including linear regression, decision trees, random forests, k-nearest neighbors, and support vector machines. This variety allows researchers to experiment with different models and choose the one that best suits their specific data and prediction requirements.

The library's user-friendly interface and comprehensive documentation simplify the implementation of complex machine learning models. Scikit-learn's consistent and easy-to-use API enables rapid model development and deployment. Extensive examples further help users understand and implement various algorithms, which is especially beneficial for researchers and practitioners who need to quickly develop models without years of programming experience.

Another advantage of Scikit-learn is its seamless integration with other Python libraries, such as NumPy, SciPy, and pandas, which are essential for scientific computing and data processing. NumPy supports large, multidimensional arrays and tables, along with a collection of mathematical functions for operating on these arrays. SciPy builds on NumPy by adding a collection of high-level algorithms and commands for data manipulation and analysis, while pandas offers data structures and analysis tools, making it easier to manipulate structured data. This integration improves the overall workflow, allowing researchers to handle large data sets and perform complex analyses with ease.

The library also includes a comprehensive collection of tools for model evaluation, cross-validation, and hyperparameter tuning, which are essential for developing robust and accurate predictive models. Cross-validation techniques, such as k-fold, help evaluate model performance on different subsets of the data, ensuring that the model generalizes even to unseen data. Hyperparameter tuning methods, such as GridSearchCV and RandomizedSearchCV, allow for a systematic search for the best hyperparameters, optimizing model performance. These tools allow researchers to thoroughly evaluate the performance of their models and fine-tune them to achieve the best possible results.

Scikit - learn users benefit from a large and active community that provides ongoing updates, improvements, and support, ensuring that the library remains up-to-date with the latest developments in machine learning. The community contributes by developing new features, fixing bugs, and sharing knowledge through forums, mailing lists, and online platforms like Stack Overflow . Users can benefit from a wealth of shared knowledge and resources, ensuring they can find solutions to their problems and stay informed about best practices.

To forecast Greece's energy quantities for the next five years, from the dataset (2015-2020), a machine learning model was developed. The model incorporated historical energy consumption and production data, economic growth forecasts, and planned renewable energy projects. The development process included data preprocessing, feature selection, model training, model evaluation, and forecasting.

Data preprocessing included cleaning and transforming the data to ensure its quality and compatibility with machine learning algorithms. For example, the " year " column was converted to datetime format and then exported as a numeric year for analysis. Missing values were treated by imputation, and categorical variables were coded using techniques such as single-point coding. In addition, the data was normalized to ensure that features with different scales did not negatively affect the model's performance.

Relevant features that influence energy consumption and production were identified. In this study, the main feature used was " year ", with " quantity " as the target variable . Additional features considered included economic indicators such as GDP growth rate, population growth, and industrial production. Feature selection techniques such as correlation analysis and feature importance ranking were used to identify and retain the most important features, ensuring that the model was both efficient and effective.

Historical data was used to train the machine learning model. Linear regression was chosen as the primary algorithm due to its simplicity and effectiveness in capturing linear trends in the data. The training data included the years up to 2010, while the testing data included the following years. The linear regression model was fitted to the training data using the method of least squares. Other models, such as decision trees and random forests, were also explored to compare performance and robustness.

The model's performance was assessed using metrics such as the mean squared error (MSE) and R-squared (R^2). MSE measures the mean squared difference between the observed actual outcomes and the outcomes predicted by the model, providing a sense of the model's accuracy. R^2, or coefficient of determination, indicates the percentage of the dependent variable's variance that is predictable from the independent variables. These metrics helped assess the accuracy and reliability of the predictions.

Cross-validation was used to ensure that the model generalizes even to unseen data, providing a robust estimate of its performance.

In the present study, the overall mean square error (MSE) was found to be 1857531075938.05 and the overall R -squared (R^2) was -11.11. A high MSE value indicates that the model predictions deviate significantly from the actual values, suggesting that the model may not be well-fitted to the data or that there are underlying complexities in the data that the model failed to capture. A negative R^2 score further indicates that the model performs worse than a horizontal line (average of the data), highlighting potential problems with model fit or the need for a more sophisticated modeling approach.

The trained model was applied to forecast the energy quantities of Greece for the next five years (2015-2020). Future years were entered into the model and the predicted quantities were calculated. The forecasts were compared with actual data (where available) to validate the model's performance. In addition, confidence intervals were created to quantify the uncertainty of the forecasts, providing a range within which the actual values are expected to vary.

The use of Scikit-learn in this study demonstrates its effectiveness in developing a machine learning model for energy forecasting, although the results highlight the challenges and complexity of accurate forecasting in this area. Its flexibility, ease of use, integration into the Python ecosystem, comprehensive model evaluation tools, and strong community support make it an ideal choice for researchers and practitioners. The forecasts produced by the model provide valuable information on the future energy needs of Greece, aiding in planning and decision-making. Future work may include improving the model by incorporating additional features, exploring more advanced algorithms, and extending the forecast horizon to improve accuracy.

Bibliography

[1] https :// www . Kaggle . com / datasets / unitednations / international - energy - statistics

[2] Zhenxiang Dong, Jiangyan Liu, Bin Liu, Kuining Li, Xin Li, Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification, Energy and Buildings, Volume 241,2021,110929, ISSN 0378-7788, https://doi.org/10.1016/j.enbuild.2021.110929(https://www.sciencedirect.com/science/article/pii/S0378778821002139)

[3] Aowabin Rahman, Vivek Srikumar, Amanda D. Smith, Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, Applied Energy, Volume 212, 2018, Pages 372-385, ISSN 0306-2619, https:// doi.org/10.1016/j.apenergy.2017.12.051 .

[4] Mohamed Maaouane, Smail Zouggar, Goran Krajačić, Hassan Zahboune, Modeling industry energy demand using multiple linear regression analysis based on consumed quantity of goods, Energy, Volume 225,2021, 120270, ISSN 0360-5442, https://doi .org/10.1016/j.energy.2021.120270 .

[5] Vennila, C. , Titus, Anita , Sudha, T. Sri , Sreenivasulu, U. , Reddy, N. Pandu Ranga , Jamal, K. , Lakshmaiah, Dayadi , Jagadeesh, P. , Belay, Assefa , Forecasting Solar Energy Production Using Machine Learning , *International Journal of Photoenergy* , 2022 , 7797488, 7 pages, 2022 . https://doi.org/10.1155/2022/7797488

[6] Yibo Chen, Hongwei Tan, Short-term prediction of electric demand in building sector via hybrid support vector regression, Applied Energy, Volume 204, 2017, Pages 1363-1374, ISSN 0306-2619, https://doi.org /10.1016/j.apenergy.2017.03.070 .

[7] Aurélie Foucquier, Sylvain Robert, Frédéric Suard, Louis Stéphan, Arnaud Jay, State of the art in building modeling and energy performances prediction: A review, Renewable and Sustainable Energy Reviews, Volume 23, 2013, Pages 272-288, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2013.03.004 .

[8] Yabin Guo, Jiangyu Wang, Huanxin Chen, Guannan Li, Jiangyan Liu, Chengliang Xu, Ronggeng Huang, Yao Huang, Machine learning-based thermal response time ahead energy demand prediction for building heating systems, Applied Energy, Volume 221, 2018 , Pages 16-27, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2018.03.125 .

[9] MA Rafe Biswas, Melvin D. Robinson, Nelson Fumo, Prediction of residential building energy consumption: A neural network approach, Energy, Volume 117, Part 1, 2016, Pages 84-92, ISSN 0360-5442, https:/ /doi.org/10.1016/j.energy.2016.10.066 .

[10] Cipriano, J., Aznar, F., Granell, R., & Martínez, I. (2017). Using matrix factorisation for the prediction of electrical quantities. ResearchGate. https://www.researchgate.net/publication/321205832_Using_matrix_factorisation_for_the_prediction_of_electrical_quantities/links/5a1bf8760f7e9be37f9c16aa/Using-matrix-factorisation-for-the-prediction-of-electrical-quantities.pdf

[11] JP Painuly, Barriers to renewable energy penetration? a framework for analysis, Renewable Energy, Volume 24, Issue 1, 2001, Pages 73-89, ISSN 0960-1481, https://doi.org/10.1016/S0960-1481(00)00186-5.

[12] Saleh, Hyatt. (2018) 2018. Machine Learning Fundamentals. 1st ed. Packt Publishing. https://www.perlego.com/book/859002.