UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Short Term Electricity Load Forecasting using Machine Learning Techniques

MSc Thesis

Zaridis Vasileios

Supervisor: Bargiotas Dimitrios

February 2023

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Short Term Electricity Load Forecasting using Machine Learning Techniques

MSc Thesis

Zaridis Vasileios

Supervisor: Bargiotas Dimitrios

February 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου με χρήση τεχνικών Μηχανικής Μάθησης

Μεταπτυχιακή Διπλωματική Εργασία

Ζαρίδης Βασίλειος

Επιβλέπων: Μπαργιώτας Δημήτριος

Φεβρουάριος 2023

Approved by the Examination Committee:

Supervisor	**Bargiotas Dimitrios** Professor, Department of Electrical and Computer Engineering, University of Thessaly
Member	**Tsoukalas Eleftherios** Professor, Department of Electrical and Computer Engineering, University of Thessaly
Member	**Fevgas Athanasios** Specialized and Laboratory Teaching Staff, Department of Electrical and Computer Engineering, University of Thessaly

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

Being fully aware of the implications of copyright laws, I expressly state that this MSc thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work/contributions of third parties for which the permission of the authors/beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and/or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

The Declarant

Zaridis Vasileios

Acknowledgements

First of all, I would like to thank Dr. Dimitrios Bargiotas, who inspired me and contributed to my research process, from figuring out my thesis topic to my final proofread. I could not have undertaken this journey without him. His expertise and thorough guidelines were invaluable in completing this research. Dr. Dimitrios Bargiotas provided insightful feedback to help my polishing efforts.

Special thanks to the examination committee, Dr. Eleftherios Tsoukalas for his precious knowledge, as well as Dr. Athanasios Fevgas for his continued professionalism and support throughout my studies.

Finally, I would like to express my gratitude to my parents, my brother and my girlfriend for their continuous moral support and faith in me. Their love and support have sustained me throughout this academic journey.

MSc Thesis

Short Term Electricity Load Forecasting using Machine Learning Techniques

Zaridis Vasileios

Abstract

The core of load forecasting is the factor of accuracy so as to make sure the scientific nature of power system optimization. By analyzing historical and present data on electricity consumption it is quite possible to predict future load trends and patterns. For that reason, short-term load forecasting is particularly crucial for making decisions related to safety evaluations and economic dispatch.The scope of the present thesis is to propose an innovative data processing strategy that highlights the significance of data preparation and analysis for the optimum input data to be used in simple neural networks to forecast electricity load consumption. The aforementionedmethod is used with the assistance of data related to the usage of electricity within the Greek power system which can be improved by a satisfactory forecasting operation.In a nutshell, the initial chapters present a theoretical analysis and retrospective of the matter of predicting electricity demand in the near future whereas the last chapter analyzes the implementation of the short-term forecast of the load of the Greek energy network. Specifically, a minimal architecture Multilayer Percepton (MLP) was used and the outcomes regarding the factor of accuracy were placed around 98%.

Keywords: Load Forecasting, Smart Grids, Machine Learning, Newral Networks, Multilayer Percepton.

Μεταπτυχιακή Διπλωματική Εργασία

Βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου με χρήση τεχνικών Μηχανικής Μάθησης

Ζαρίδης Βασίλειος

Περίληψη

Ο πυρήνας της πρόβλεψης φορτίου είναι ο παράγοντας της ακρίβειας, ώστε να διασφαλιστεί ο επιστημονικός χαρακτήρας της βελτιστοποίησης του συστήματος ηλεκτρικής ενέργειας. Με την ανάλυση ιστορικών και σημερινών δεδομένων σχετικά με την κατανάλωση ηλεκτρικής ενέργειας είναι αρκετά πιθανό να προβλεφθούν μελλοντικές τάσεις και πρότυπα φορτίου. Για το λόγο αυτό, η βραχυπρόθεσμη πρόβλεψη φορτίου είναι ιδιαίτερα κρίσιμη για τη εξαγωγή αποφάσεων που σχετίζονται με τον έλεγχο της ασφάλειας και την οικονομική αποστολή. Σκοπός της παρούσας διατριβής είναι να προταθεί μία καινοτόμα στρατηγική επεξεργασίας δεδομένων που τονίζει τη σημασία της προετοιμασίας και ανάλυσης δεδομένων προκειμένου να χρησιμοποιηθούν τα βέλτιστα δεδομένα εισόδου σε απλά νευρωνικά δίκτυα για την πρόβλεψη της κατανάλωσης ηλεκτρικού φορτίου. Η ανωτέρω αναφερόμενη μέθοδος χρησιμοποιείταιμε τη βοήθεια δεδομένων κατανάλωσης από το ελληνικό σύστημα ηλεκτρικής ενέργειας, το οποίο μπορεί να βελτιωθεί με την χρήση μιας ικανοποιητικής λειτουργίας πρόβλεψης των τιμών του. Συνοπτικά, στα πρώτα κεφάλαια λαμβάνει χώρα μια θεωρητική ανάλυση και αναδρομή του προβλήματος της βραχυπρόθεσμης πρόβλεψης φορτίου ενώ στο τελευταίο κεφάλαιο αναλύεται η υλοποίηση της βραχυπρόθεσμης πρόβλεψης του φορτίου του Ελληνικού ενεργειακού δικτύου. Συγκεκριμένα, χρησιμοποιήθηκε minimal αρχιτεκτονική Multilayer Percepton (MLP) και τα αποτελέσματα της ακρίβειας του μοντέλου τοποθετούνται γύρω στο 98%

Λέξεις-κλειδιά: Πρόβλεψη ηλεκτρικού φορτίου, Εξυπνα δίκτυα, Μηχανική Μάθηση, Νευρωνικά δίκτυα, Multilayer Percepton

Table of Contents

Abbreviations

AIS	Artificial Immune System
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
APS	Autonomic Power System
BP	Back Propagation
BPF	Band Pass Filter
DG	Generation Distributed Generation
DOE	Department of Energy
DSM	Demand Side Management
e.g.	exempli gratia (for example)
EPRI	Electric Power Research Institute
GB	Gigabyte
GIS	Geographic Information System
G2V	Grid-to-vehicle
ITs	Information Technologies
k-NN	k-Nearest Neighbor
MB	Megabyte
MAPE	Mean Average Percentage Error
MDP	Markov decision process
ML	Machine Learning
MLP	Multilayer Perceptron
NIST	National Institute of Standards and Technologies

PB	Petabyte
PMUs	Phasor Measurement Units
PNNL	Pacific Northwest National Laboratory
PPMCC	Pearson product-moment correlation coefficient
RES	Renewable Energy Sources
RF	Random Forest
RL	Reinforcement Learning
SG	Smart Grid
STLF	Short-Term Load Forecasting
SVM	Support Vector Machine
SVR	Support Vector Regression
TB	Terabyte
V2G	Vehicle-to-grid

Chapter 1 - Introduction

Forecasting load demand is an essential aspect of power system management and includes predicting energy usage for a future period. When this prediction is made for a period ranging from an hour to a week, it is referred to as **short-term load forecasting (STLF).** The precision of short-term load forecasting (STLF) has a significant impact on the economic performance and reliability of a power system. If STLF is inadequate, it can result in insufficient reserve capacity and costly allocation of peaking units, or even lead to unnecessary reserve capacity. Both the first and the second outcome as described above increase operating costs, highlighting the critical role of accurate load forecasting in energy market analysis and economic dispatch within the power industry. In future smart grids, reliable and precise STLF will be even more critical for operators to manage grids cost-effectively.

Many factors affect the load demanddue to the fact that it is a non-stationary process. Indicatively, weather, socioeconomic factors and fortuitous events have an impact on the load demand, which makes it really difficult to predict. Numerous methods have been suggested for short-term load forecasting (STLF), the majority of which employ either artificial intelligence algorithms or statistical techniques. [1]

According to a research performed ataregion of Abu Dhabi, the main affecting factors of electricity peak demand are the following [2]:

- **Weather**

  Weather is the primary independent factor that affects electricity demand, with temperature and humidity being among the key factors within it. A study has shown that a 3°C increase in Toronto's daily maximum temperature during summer would cause the average and standard deviation of peak load to rise by 7% and 22%, respectively. However, weather is often referred to as the breaking point that can result in system unreliability by reducing the effective power supply.

  (a) Temperature

Temperature can be defined as an indication of the degree of hotness or coldness. Ambient temperature plays a significant role in hourly energy consumption, with hotter days leading to a higher consumption load curve than colder days, whereas colder days result to lower consumption load curve. As it is commonly known, temperature rises rapidly in the summer and hence, it disturbs people's comfort. As a result, individuals will most likely pursue the use of electricity for cooling purposes (e.g.air conditioning as well as fans). The number one reason for extremely high consumption load seems to be the cooling system. As regards Al Ain city, during the summer, the demand on the cooling systems is highest because it's hot outside. This means the peak load for this type of equipment is usually traced in the summer. The recorded temperature is presented in detail in Table 1.1.

Table 1.1: Comparison table between
temperature and peak demand in the city of Al Ain

Year	Peak Demand (MW)	Temperature °C	Peak Demand Day
2014	2090	48	11 – June
2015	2150	45	30 – August
2016	2166	46	11 – August
2017	2220	46	3 - August

Table 1.1 indicates that peak demand occurs at temperatures ranging from 45-48°C, which is considered a high temperature that increases the demand for cooling systems. As expected, the load decreases at night when temperatures are lower. Additionally, higher ambient temperatures result in increased energy losses and decreased efficiency for cooling systems. Consequently, the load and temperature have a positively proportional relationship due to these factors.

(b) Humidity

Another factor included in weather is humidity. It is a term that can be used for the amount of water vapors in air. People in their everyday lives call it*"relative humidity"*which is expressed in percentage (%). It is a fact that humidity can affect the apparent temperature and cause its increase, whereas it has absolutely no effect on the real temperature. Humans'sensitivity to humidity can be explained due to the fact that it affects how well their bodies regulate their temperature.

When there is a lot of humidity in the air, it takes longer for the sweat on their skin to evaporate. This means that someone's body may not cool down as quickly as it would under normal conditions. Table 1.2 shows the humidity as it was recorded at a period of 4 years in the city of Al Ain.

Table 1.2: Comparison between humidity and peak demand in city of Al Ain

Year	Peak Demand (MW)	Humidity (%)	Peak Demand Day
2014	2090	6	11 – June
2015	2150	16	30 – August
2016	2166	18	11 – August
2017	2220	14	3 - August

- **Population effect**

This statement suggests that as the population of Al Ain grows, the demand for electricity will also increase. This is because, as the city's population expands, so does the demand for energy in different areas of life. In 2030, Al Ain is projected to have a population of over 13,000 people, and this growth is expected to result in an increase in energy demand. However, it's also important to note that both population and energy grow together - meaning that as the population of Al Ain grows, the demand for energy will also increase.

As already highlighted in the abstract of this master's thesis, the purpose of our research is about the forecast of electricity load consumption. Using simple neural networks to forecast electricity load consumption, the present thesis proposes an innovative data processing strategy that demonstrates how valuable data preparation and analysis are. In more detail, the research analysis structure is as follows:

- **Chapter 2** is about Smart Grids. To be more precise, it refers to the fact that SGs are created to address various issues and presents both the advantages/characteristics as well as the challenges of them.

- Next, there is **Chapter 3** which presents some necessary definitions - a brief theoretical analysis - of some essential terms regarding machine learning applications in Smart Grids.
- **Chapter 4** is a review of literature in the problem of short-term load forecasting methods.
- Finally, **Chapter 5** highlights the proposal that constitutes the core of this thesis, which is the creation of a minimal solution of a Multilayer Neural Network for the short-term forecast of Greece’s electricity hourly load consumption.

Chapter 2 - Smart Electrical Grids

2.1. Definitions

Before delving into the research, it is important to establish some definitions. The term **"Smart Grid" (SG)** does not have a single definition that accurately describes the phenomenon. The US Energy Independence and Security Act of 2007 similarly defines the smart grid as a modernization of the electrical network that monitors and improves grid resiliency to disruptions, automatically optimizes grid operation, and facilitates the bi-directional flow of energy. [3]

The smart grid is characterized by the US National Institute of Standards and Technologies (NIST) as a contemporary power distribution system that is designed to accommodate energy flows in both directions, leverages bidirectional communication and control capabilities, and enables a diverse range of innovative features and applications. According to the Electric Power Research Institute (EPRI), the smart grid represents a shift from the existing power grid to a system that fosters consumer interaction on a peer-to-peer basis, distributed power generation, and advanced control centers.

The UK's Department of Energy and Climate Change emphasizes the pivotal role of intelligent electricity grids in facilitating more effective management of supply-demand balance data and optimizing demand patterns to decrease peak load periods. In Australia, the "Smart Grid-Smart City" initiative promoted by the government espouses the smart grid as a groundbreaking and exceedingly astute approach to electricity distribution that incorporates sophisticated communication infrastructure, state-of-the-art sensing capabilities, and advanced metering technologies.

Incorporation of smart sensing technologies can play a crucial role in minimizing disruptions and power outages on the electricity grid. Additionally, smart metering can assist end-users in effectively managing their energy consumption. The smart grid represents a revamped and advanced network architecture that integrates digital computing capabilities and highly automated services into the existing power infrastructure. It leverages distributed intelligence and bidirectional communication and power flow systems, thereby significantly improving the efficiency, reliability, and sustainability of the overall system. The study highlights the disparities between the traditional power grid and the smart grid and stresses the criticality of the judicious deployment of information and communication technology for the successful implementation of the smart grid paradigm. The adoption of two-way communication capabilities in the traditional power grid network is a crucial hurdle for the realization of smart grids, and standardization protocols must be in place to ensure a smooth transition to incorporate smart grid technologies. [4]

Table 2.1: Conventional grid versus smart grid

Conventional Grid	**Smart Grid**
Operated by a mechanical system on one side only	Digitized Bi-directional
Centralized Power generation	Distributed Generation
Radially connected	Dispersed
Limited sensors	Many sensors
Small monitoring capabilities	Highly monitored
Manual control	Automated control
Not many security issues	Prone to security issues

2.2 Structure and Characteristics

The National Institute of Standards and Technology (NIST) has undertaken the task of categorizing smart grids into seven distinct classes that encompass a broad range of actors and applications. These actors encompass various devices that facilitate the exchange of data. The tasks or applications that can be accomplished by actors within each class include energy management, site automation, and energy storage. Furthermore, actors within the same class can interact with those belonging to different classes, and a specific class may feature components from other classes. For instance, the distribution utility class may comprise actors from both the operation and customer categories. A comprehensive breakdown of the categorization of smart grids is presented in Table 2.2. [4]

Table 2.2: Classification and elements of SGs

Category	Description
Customer	End-users are the consumers of electricity, which can be categorized as domestic and large consumers like commercial and industrial loads. Various actors have the ability to generate, control, and store electricity
Market	Electricity markets involve the exchange of assets, with the operator and participants acting as the key actors.
Service Provider	A company that offers services related to the secure establishment and operation of smart grids, tailored to the needs of both consumers and utilities.
Operations	The managers responsible for overseeing the flow of electricity monitor the power system to ensure it is functioning properly.
Bulk Generation	Bulk generation is where the delivery of large quantities of electricity to consumers begins, and the actors involved are the generators who produce electricity in large amounts. Energy can also be stored for later distribution.
Transmission	Transmission refers to the transfer of bulk power from generation centers to distribution points. The actors involved are the carriers of electricity who may also have the ability to generate and store electricity.
Distribution	Distribution is the process where Distributed Generation, Distributed Storage, transmission, and consumers interconnect. The actors involved are the entities responsible for distributing electricity to and from customers.

The following seven points, which are briefly stated, summarize **the main characteristics of SGs** that have been reported according to U.S. Department of Energy (DOE) [6]:

1. Optimize the utilization of assets and increase operational efficiency.
2. Accommodate diverse generation and storage options.

3. Ensure power quality to support the diverse requirements of a digital economy.
4. Proactively anticipate and react to system disturbances in a self-healing manner.
5. Operate with resilience against both physical and cyber threats, as well as natural disasters.
6. Facilitate active consumer participation.
7. Create new opportunities for products, services, and markets.

Likewise, E.U. perceives the SG as an active network [5]:

1. Overcoming limitations hindering the expansion of distributed generation and storage.
2. Ensuring the interoperability of systems and maintaining a secure supply.
3. Providing open access to a liberalized market for all users.
4. Minimizing the environmental impact of electricity production and delivery.
5. Facilitating the active participation of consumers in the demand-side management.
6. Generating consumer interest and engagement.

2.3 Advantages

In a nutshell, the advantages of Smart Grids can be described briefly as follows [6]:

- **Self healing**: The Smart Grid (SG) can use real-time sensors such as PMU to monitor the condition of the power system. Thanks to its self-healing capability, the SG can predict, identify, and respond to issues or power outages by utilizing PMU and automatic control center.
- **Interactive**: A key characteristic of the Smart Grid (SG) is the ability to allow for bidirectional power flow, enabling users to both consume and supply power. To reduce peak demand during high-use periods, dynamic pricing is used to incentivize users to decrease their power consumption. This is done in order to achieve the

desired Load Factor, which is the optimal balance between power demand and supply.

- **Security**: The Smart Grid's network and control system are specifically designed to function as a safeguard against cyber attacks. To take preemptive measures, operators proactively utilize real-time monitoring technology such as PMU to predict and prevent potential issues. Additionally, the use of distributed generation and micro-grids ensures that there is a secure and reliable power supply.
- **Asset Optimization**: Geographic Information System (GIS) can be utilized to create a streamlined network design that minimizes the need for transmission equipment. The utilization of Geographic Information Systems (GIS) presents an opportunity to adopt condition and performance-based maintenance practices that can significantly enhance operational efficiency. Moreover, the integration of Smart Grids (SGs) can help minimize both technical and non-technical losses associated with transmission and distribution, leading to an overall improvement in efficiency.
- **Distributed**: Distributed Generation (DG) refers to the use of decentralized generating units instead of a centralized network. The benefits of DG can be categorized into two main groups: economic and operational. Economically, DG can provide power support during peak demand periods, reducing interruptions that can cause system failures. Additionally, DG offers flexibility in terms of capacity and installation placement, thereby reducing investment risks. Operationally, DG can reduce costs by avoiding the need for upgrading or setting up new transmission and distribution networks when installed close to the customer load. Furthermore, by using local renewable energy sources (RES), DG can decrease dependence on fossil fuel imports and reduce international energy prices.
- **Market Empowerment**: Consumers are engaged in the electricity market by participating in a smart grid. A smart grid not only offers more transparency and options in energy procurement, but it also empowers utilities to adapt to evolving consumer demands. Consequently, it will lead to a demand for energy- and cost-saving products. With smart grids, consumers will be educated, new energy

management services will be developed and the cost and environmental impact of electricity delivery will be reduced.

- **Environment Friendly**: The smart grid has the potential to reduce carbon dioxide emissions by promoting energy conservation and improving end-use efficiency. According to a report by the Pacific Northwest National Laboratory, the introduction of smart grid technologies could result in significant carbon savings by 2030. This includes direct carbon savings of 12 percent from equipment like smart meters, as well as indirect savings of 6 percent from stronger grid support for renewable electricity generation. The smart grid also offers environmental benefits by managing peak load through demand response and reducing transmission and distribution loss. While reducing these losses would require an investment of $20,000 to $75,000 per MW, the smart grid's ability to provide continuous feedback on electricity consumption enables consumers to make informed decisions and adjust their usage patterns in response to pricing and consumption, thereby leading to a reduction in annual CO_2 emissions. Furthermore, the optimized utilization of existing infrastructure enabled by the smart grid can minimize the requirement for new infrastructure construction.

2.4 Challenges

Apart from the advantages, there are also some challenges of smart grids that have to be faced (as reported in the literature)[6]:

- **Financial Resources**: The most important challenge is that a tremendous amount of capital is required for the journey of Smart Grid.
- **Government Support:** Further to the above, the lack of financial resources is not the sole obstacle to implementing a Smart Grid. A government that is willing to support such an initiative and effective energy policies are also crucial factors for the success of Smart Grid implementation.

- **Compatible Equipment:** SG technologies are not compatible with the existing equipment and hence, the older equipment must be replaced. The aforementioned situation could pose challenges not only to utilities and regulators but also to consumers. Additionally, a large number of equipment that is compatible with the system is required to cover the controlling program of the entire grid system.
- **Consumer Education:** In order for the implementation of the SG to be concluded successfully, active and educated consumers are a perquisite. A vital portion of the Smart Meter advantages are based on consumer engagement. As a result, since the goal is to achieve the maximum benefit, consumersneed to be educated and well-informed to make intelligent decisions.
- **Cost Assessment:** Due to the fact that the standards and protocols required to design and operate an advanced metering infrastructure are still in a state of flux, costs can be higher than expected. Therefore, investments made prior to the settlement of the standards, are at a greater risk of becoming obsolete.
- **Cyber Security and Data Privacy:** The implementation of Smart Grid is accompanied by concerns regarding cybersecurity and data privacy, as there is potential for misuse of private information through digital communication networks and more detailed data on consumption patterns.
- **Capacity to Absorb Advanced Technology**: Smart Grid relies on advanced technology that is constantly evolving, which requires the system to have the capacity to absorb such modern technology.
- **Strengthening the Grid:** It is also important to strengthen the grid against unpredictable hazards such as cyber-attacks, weak infrastructure, inefficient control systems, corrosion, smart meter authentication, and blackouts. The aforementioned list is not exhaustive and serves as an indication of potential issues that may arise.

2.5 Smart Grid in the future

According to a really interesting paper [7], Autonomic Power System (APS) is envisaged to be "self" (self-configuring, self-healing, self-optimizing and self-protecting). The concept of an Advanced Power System (APS) involves a comprehensive approach to power management, where various parts of the network work in coordination with each other based on the priorities of the system. For instance, during a storm, certain parts of the network can be disconnected to ensure that the essential parts remain connected and operational. APS can also incorporate new parts of the network, such as power generators, into the system as they become available.

To achieve this, no manual intervention is required, and the power system of the future, by itself, will decide what is best. The future of electricity networks will have to adapt to technological advancements and market rules that address challenges such as population growth, fluctuating energy prices, and an increasing number of electric vehicles and devices.

Customers who previously only consumed electricity may become sellers, and the development of new markets for electric energy is expected due to technological progress and free information accessibility.

Electric vehicles are an excellent example of the transformation in the energy sector. Tesla's introduction of electric cars to the mass market and the "grid-to-vehicle" (G2V) and "vehicle-to-grid" (V2G) schemes allow anyone to purchase or sell energy by merely plugging their vehicle into the grid. This implies that electric vehicle owners can become independent components of the electricity market, and their choices and behaviors can impact the demand, supply, and prices of electric energy.[7]

While the idea of a self-functioning power system by 2050 may seem like science fiction, it is crucial to consider all possible scenarios without ruling out any possibilities. We could not have imagined the significance of smartphones in our lives 20-25 years ago, and we could not have predicted the widespread use of personal computers in every home 35-40 years ago. Therefore, we must not limit ourselves and exclude any futuristic scenario, no matter how improbable it may seem.

Chapter 3 - Machine Learning Applications in Smart grids

3.1 Definitions

In order to study the present chapter, some definitions will be required due to the fact that the old fashioned Electrical engineering is not too familiar as regards new computer science technologies.

Machine learning (ML)

Machine Learning is an area of research that aims to develop methods that improve performance on certain tasks by leveraging data. It is a subset of artificial intelligence. By analyzing training data, which is a sample data, machine learning algorithms build a model that allows predictions or decisions to be made without explicit programming. Machine learning algorithms are applied in a wide range of fields, including medicine, email filtering, speech recognition, agriculture, and computer vision, where it may be difficult or impractical to create traditional algorithms to carry out the necessary tasks. Although not all machine learning is statistical learning, a subset of machine learning is closely related to computational statistics, which focuses on using computers to make predictions. Optimization theory provides methods, theory, and application domains for the field of machine learning. Unsupervised learning is a field of study related to data mining, which focuses on exploratory data analysis. Some machine learning implementations use data and neural networks that mimic the functioning of a biological brain. Machine learning is also referred to as predictive analytics in business problem applications. [3]

Deep learning

Deep learning, also referred to as deep structured learning, is a subcategory of machine learning that utilizes artificial neural networks with representation learning. It is a part of a larger family of machine learning methods. Deep learning can involve three types of learning: supervised, semi-supervised, and unsupervised.[5]

Reinforcement Learning (RL)

Reinforcement Learning is a type of machine learning that focuses on helping intelligent agents take actions in an environment in order to maximize their cumulative reward [8]. There are three basic types of machine learning: Reinforcement Learning, supervised learning, and unsupervised learning. The main differences between RL and supervised learning are that RL does not require labeled input/output pairs and does not require sub-optimal actions to be corrected explicitly. Reinforcement learning algorithms often use dynamic programming techniques, with the environment represented as a Markov decision process. The key difference between traditional dynamic programming techniques and RL algorithms is that the latter do not require a precise mathematical representation of the MDP and are better suited to handle large MDPs.

Deep reinforcement learning (deep RL) is a specialized area within machine learning that combines reinforcement learning (RL) with deep learning techniques. RL focuses on how computational agents can learn to make decisions by trial and error, while deep learning enables agents to make decisions based on unstructured input data without requiring manual engineering of the state space. This combination of RL and deep learning is what sets deep RL apart from traditional reinforcement learning methods. [8].

Artificial neural networks (ANNs)

One possible way to define neural networks (NNs) or artificial neural networks is as computing systems that take inspiration from the biological neural networks found in animal brains. In ANNs, a network of interconnected nodes or artificial neurons model the behavior of neurons in the brain. These neurons receive signals through connections, which can then be processed and sent to other neurons. Nonlinear functions of the input signal determine the output of each neuron. Connections in ANNs are referred to as edges and have weights that allow them to adjust the strength of the signal they transmit. Additionally, neurons may have a threshold that must be crossed before a signal is sent.

During the learning process, both neurons and edges adjust their weights to improve performance.

Neurons in an artificial neural network are organized into layers, where each layer may apply a unique transformation to its input. The flow of signals in the network typically begins at the input layer, passes through the intermediate layers, and terminates at the output layer, potentially passing through the layers multiple times. [9]

The fields of deep learning, machine learning, and data science are intertwined. To comprehend the correlation among the three, a Venn diagram illustrated in Figure 3.1 can be used. This diagram portrays the relationship mentioned above.

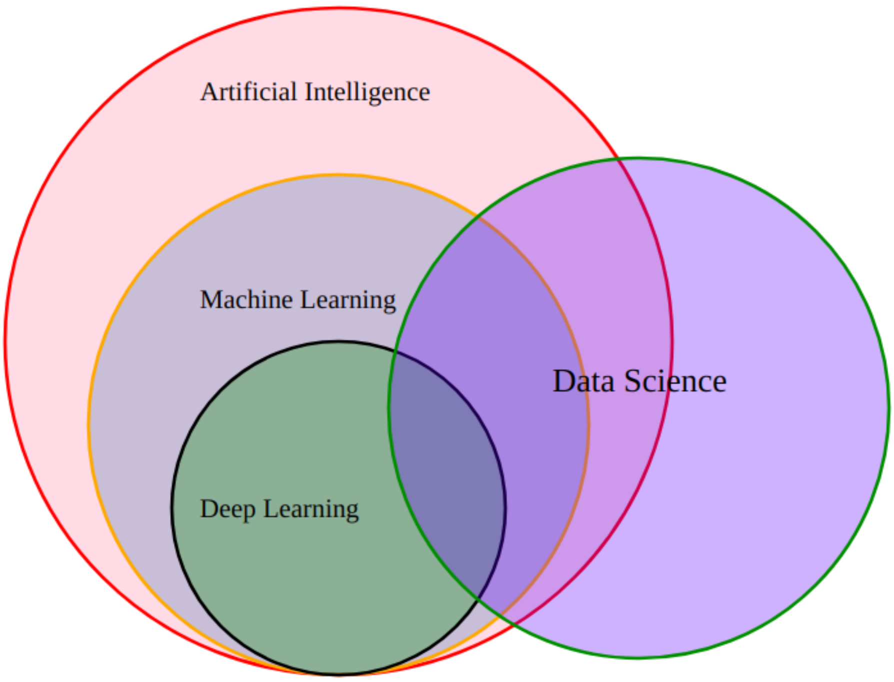

Figure 3.1: AI Venn diagram [9]

3.2 Topics applied to Smart Grids

According to literature research, there have been several common machine learning applications in the context of smart grids over the past few years [10]. These include:

1. **Energy, demand forecasting or management**, which deals with electricity planning from both a generation and load perspective, including distributed generation, demand management, cost reduction, and energy pricing.
2. **Security and reliability** of the electric grid, which involves researching cyber-attacks on smart grids, network anomalies, power failure detection, network operation, and prevention of electricity theft.
3. **Intelligent measurements** refer to the use of advanced sensors and monitoring technologies to collect high-quality data and enable accurate measurement and analysis of various parameters in the power system.
4. **Quality and efficiency of the network**, which covers topics such as improving energy quality, reducing electric losses, and increasing energy efficiency.
5. **Demand side management**, which involves detecting electrical equipment and controlling energy consumption by users.
6. **Electric vehicles:** The sixth most common machine learning application to smart grid is about trying to manage the electric network on an automated basis in order to lower the cost of charging electric vehicles.
7. **Equipment sizing:** Last but not least, a remarkable number of articles refer to the size of electrical network equipment (e.g. cables).

In alignment with the machine learning technique, Figure 3.2 shows the most researched themes in the past three years using machine learning techniques are related to the reliability and safety of electrical networks and energy management and forecasting during the years of 2017 and 2019. [10]

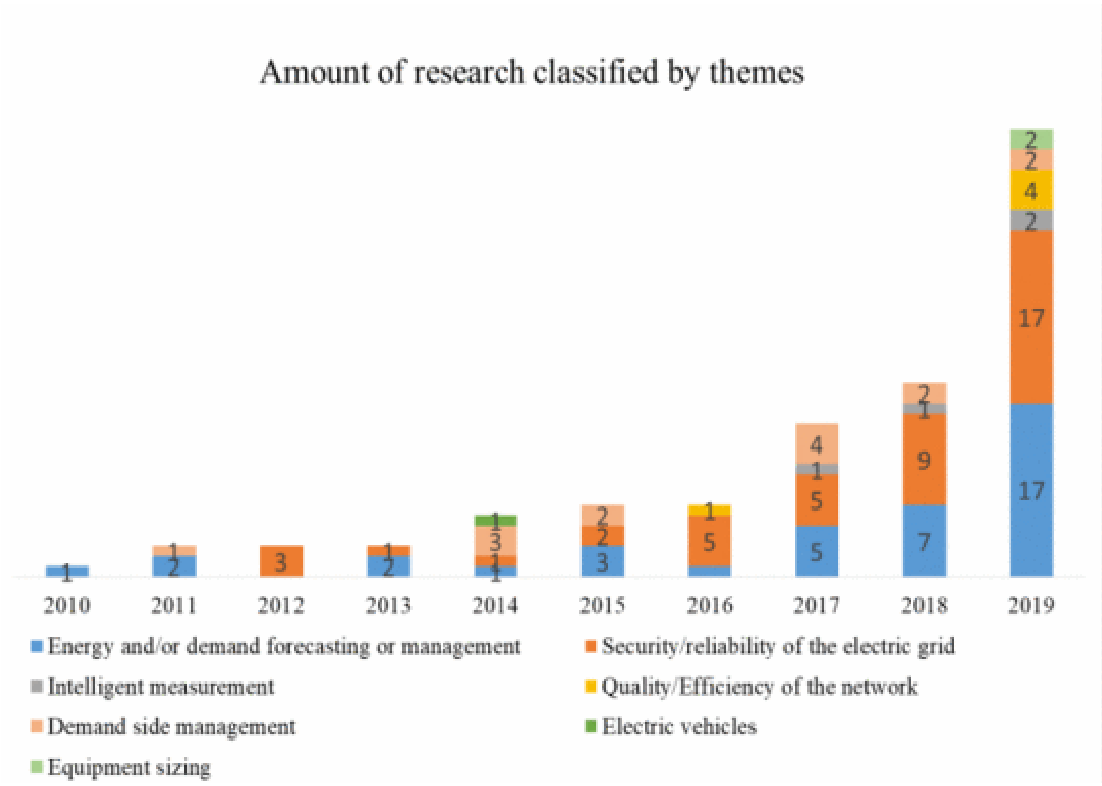

Figure 3.2: Papers amount regarding ML applications in smart grids [10]

3.3 Machine Learning Algorithms Applied in Smart Grids Researches

Similar to the previous section, based on literature research [10], the top 10 techniques used for improving power system operations are illustrated in Figure 3.3. These techniques are indeed widely applied in the field of smart grids for various purposes such as load forecasting, fault diagnosis, anomaly detection, and energy management. It's interesting to see how different machine learning algorithms can be used to solve specific problems in the smart grid domain. Further details on these techniques are discussed below.

1. ***Artificial neural networks (ANN):***

   Artificial neural networks (ANNs) are a type of machine learning algorithm inspired by the structure and function of the human brain. ANNs consist of interconnected nodes, or neurons, organized in layers. Each neuron receives input from the neurons in the previous layer, performs a computation, and passes its output to the neurons in the next layer. By adjusting the strength of the connections between neurons, ANNs can learn to recognize patterns and make predictions. They are particularly useful for tasks such as image recognition, natural language processing, and speech recognition. They have been used in a wide range of applications, including self-driving cars, recommender systems, and medical diagnosis.

2. ***Support Vector Machine (SVM):***

   Support Vector Machine (SVM) is a popular supervised machine learning algorithm that is widely used for pattern recognition by classifying input data into specific categories using a set of training examples. SVM has been extensively used in the field of smart grids, particularly for identifying cyber-attack risks on the network, managing demand on the user side by detecting electrical devices, and predicting energy demand.

3. ***Boosting algorithms:***

   Boosting is a machine learning algorithm that involves the use of basic classifiers and the combination of the boosted models to obtain more accurate results. In this approach, weak models are iteratively trained to improve their performance, and their predictions are combined to form a stronger model. The boosting techniques identified in the literature review include adaptive boosting, extreme gradient boosting, and gradient boosting. Some applications of boosting in smart grids include demand response and network fault diagnosis.

4. ***Decision tree:***

   The Decision trees are a popular method used for both classification and regression tasks. They are a type of supervised learning algorithm that makes a series of binary decisions to classify or predict the outcome of a given input. Decision trees are created by recursively splitting the data into smaller subsets based on the features that best separate the data into groups that have similar outcomes. This is done by choosing the feature that best separates the data at each node of the tree. The decision tree continues to split the data until a stopping criterion is met, such as when all of the data in a subset belongs to the same class or when the tree has reached a certain depth. They are simple to understand and interpret, which makes them very famous in applications such as medicine, finance, and marketing. However, they can be prone to overfitting and require careful tuning to achieve good performance.

5. ***Random Forest:***

Random Forest is a machine learning technique that uses multiple decision trees to classify large datasets. Each tree generates a value for a specific subset of randomly selected variables, resulting in more accurate predictions.

6. ***Bayesian networks:***

They are based on Bayes' theorem, which allows for the calculation of conditional probabilities. In such a network, nodes represent random variables and edges represent the conditional dependencies between them. By specifying the conditional probabilities between the nodes, a Bayesian network can be used to make predictions about the likelihood of different events or outcomes. One of their advantages is that they can handle incomplete or missing data, which makes them particularly useful in situations where data may be noisy or uncertain. Moreover, they have been applied to a variety of problems, including medical diagnosis, natural language processing, and robotics. However, they can be computationally expensive to train and require careful selection of prior probabilities to avoid bias.

7. ***K Means:***

K Means is an unsupervised machine learning method that uses clustering techniques to classify data with similar behaviors. This technique is commonly applied in load forecasting.

8. ***K-Nearestneighbor (KNN):***

K-nearest neighbors (KNN) is a popular machine learning algorithm used for classification and regression tasks. The algorithm works by finding the K closest data points in the training set to a given input and using the majority class or average value of those data points to predict the outcome for the input. KNN is a non-parametric method, meaning that it does not make assumptions about the underlying distribution of the data. KNN is simple to implement and can be effective for datasets with low noise or outliers. However, KNN can be sensitive to the choice of distance metric and the value of K, and may perform poorly on high-dimensional datasets. KNN has been applied to a wide range of applications, including image recognition, recommender systems, and gene expression analysis.

9. ***Hidden Markov Model (HMM):***

Hidden Markov Models (HMMs) are a type of statistical model used in machine learning and artificial intelligence. HMMs are used to model systems that involve a sequence of observations that are not directly observable but are related to underlying hidden states. The model consists of two main components: a transition model and an observation model. The transition model defines the probability of moving from one state to another, while the observation model defines the probability of observing a given output given the current state. HMMs are used in a variety of applications, including speech recognition, natural language processing, and bioinformatics. One of the key advantages of HMMs is their ability to handle sequential data and model complex systems with a large number of states. Despite the above, HMMs can be difficult to train and require careful tuning of model parameters to achieve good performance.

10. ***Logistic Regression:***

Logistic Regression is a type of machine learning algorithm used for binary classification tasks. The algorithm models the relationship between a set of input variables and a binary output variable using a logistic function. The logistic function transforms the output of a linear regression into a probability value between 0 and 1, which can be interpreted as the likelihood of the binary outcome. Furthermore, it is a parametric method, meaning that it assumes a specific form for the relationship between the input variables and the output variable. It is widely used in many applications, including medical diagnosis, finance, and marketing. Among its advantages, simplicity and interpretability are included, as the coefficients of the input variables can be easily interpreted as the effect of each variable on the probability of the binary outcome.

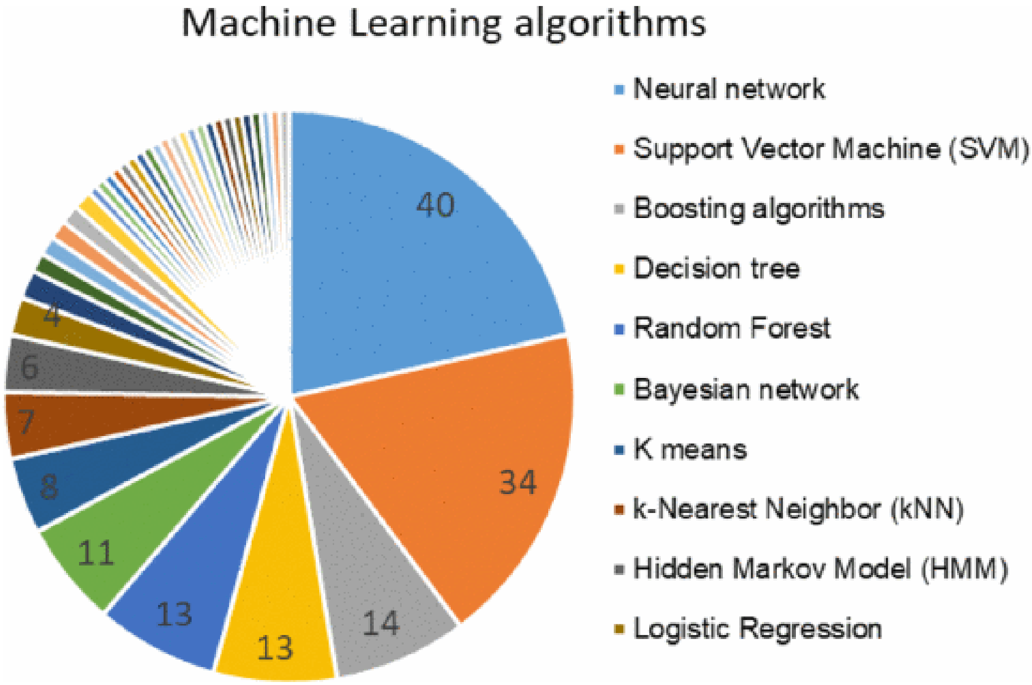

Figure 3.3: Quantity of machine learning algorithms in smart grids applications. [10]

3.4 Big Data and Smart Grids

Smart Grids are the future of power systems and are distinguished from traditional systems by their strong integration with communication and information infrastructures to enhance monitoring, control, and management of the grid and facilitate the bidirectional flow of electricity and information. However, managing the enormous amount of data, or "big data," generated by numerous data sources in SGs is a challenge that needs to be addressed in order to support and enhance the capabilities of smart infrastructures in new power systems. The management of large-scale data has become a primary area of focus for research and development in the field of Smart Grids (SGs). This involves addressing challenges posed by multiple sources of data, including data gathered from advanced metering infrastructure (AMI) through smart meters, power consumption patterns, phasor measurement units (PMUs), power market pricing and bidding data, as well as data associated with the monitoring, control, maintenance, automation, and management of power system equipment. [11]

In this section, we aim to provide a brief explanation of the historical background, architectures, technologies, and tools related to big data. Recent database technologies have emerged due to the challenges of managing vast amounts of data and a lack of effective storage capacity. The different stages of data evolution, including the megabyte (MB), gigabyte (GB), terabyte (TB), petabyte (PB), and exabyte, are well-known and have occurred over the years from the 1970s to 2011.

Big Data can be described using six features known as the "6 Vs" refer: volume, velocity, variety, veracity, value, and visualization. These characteristics describe the unique challenges associated with processing and analyzing large and complex datasets. Smart grids, which are modern electricity networks that use digital technology to optimize energy distribution, generate a vast amount of data from sensors and smart meters. The analysis of this data can provide insights into energy consumption patterns, grid performance, and renewable energy integration. However, the data generated by smart grids exhibits the same characteristics as big data, which requires specialized tools and techniques for processing and analyzing. The application of big data analytics to smart grids can enable more efficient and reliable energy distribution, reduce costs, and support the transition to renewable energy sources.

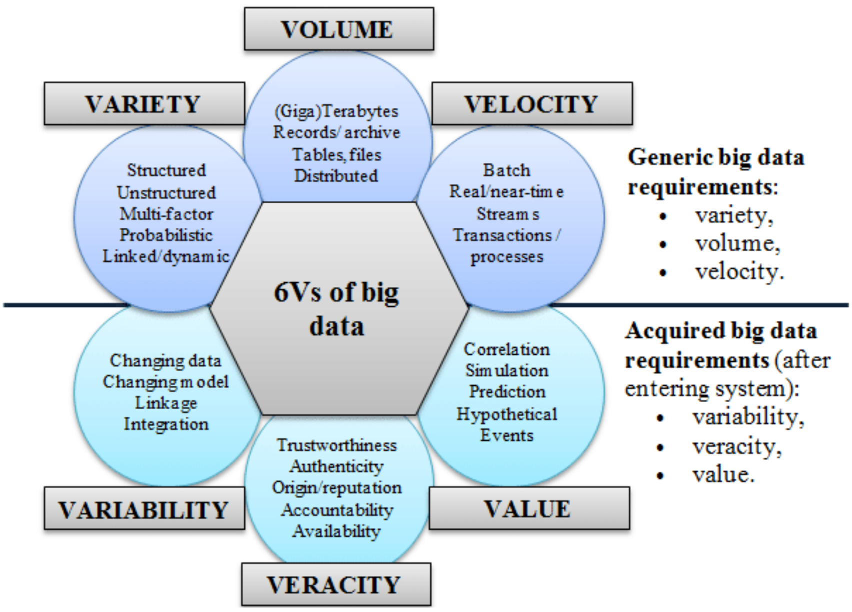

Figure 3.4: 6V's of Big Data [11]

Indicatively, among the main Big Data applications in SM the following are included [11]:

1. ***Power Generation Management***

In Power generation management refers to the process of optimizing the production and distribution of energy from various sources. Big data analytics has become increasingly important in power generation management as it enables the analysis of large and complex datasets generated by energy systems. Big data can provide insights into energy consumption patterns, renewable energy integration, grid performance, and equipment health. By analyzing this data, power generation managers can optimize energy production and distribution, reduce costs, and improve system reliability. Big data analytics tools, such as machine learning algorithms and predictive analytics, can be used to develop models that can forecast energy demand, predict equipment failures, and identify opportunities for energy savings. The use of big data in power generation management is crucial for ensuring a reliable and sustainable energy supply in the face of changing energy demands and increasing adoption of renewable energy sources.

2. ***Renewables and Microgrid Management***

Renewables and microgrid management refer to the management of decentralized energy systems that incorporate renewable energy sources. Big data applications in renewables and microgrid management involve the use of data analytics tools to optimize energy production and distribution from these systems. Big data analytics can provide insights into energy consumption patterns, renewable energy integration, and grid performance, enabling managers to make informed decisions about energy production and distribution. By analyzing data from sensors and other sources, machine learning algorithms can be used to develop models that can forecast energy demand, predict equipment failures, and identify opportunities for energy savings. The application of big data analytics in renewables and microgrid management can lead to more efficient and reliable energy distribution, reduce costs, and support the transition to renewable energy resources. This is particularly important given the increasing adoption of distributed energy systems and the need to ensure a reliable and sustainable energy supply.

3. ***Microgrids***

Big data is also useful in microgrids, which are new distributed power generation models that incorporate RERs. Microgrid investment planning and optimal load distribution are the two primary applications of big data in this context.

4. ***Demand Side Management***

Demand Side Management (DSM) is the practice of managing and optimizing the consumption of energy on the consumer side of the power grid. In recent years, DSM has been enabled by big data analytics, which provide the ability to collect, process and analyze large amounts of data from various sources, including smart meters, weather forecasts, energy prices and consumer behavior. With this data, DSM programs can predict and adjust energy consumption patterns to balance the supply and demand of electricity on the power grid. Big data analytics can also help to identify areas of inefficiency and waste, allowing for targeted improvements to be made to energy systems. Overall, the use of big data in DSM programs has the potential to significantly reduce energy consumption, lower costs and increase energy efficiency, leading to a more sustainable and resilient energy future.

5. ***Electric Vehicles***

Electric vehicles (EVs) are becoming increasingly popular as a more environmentally friendly alternative to traditional gas-powered vehicles. With the growth of EVs, big data analytics are playing an important role in managing the charging infrastructure and improving the overall efficiency of the EV ecosystem. By collecting and analyzing data from EV charging stations, energy providers and grid operators can better predict and manage demand, ensuring that the charging infrastructure is available when and where it is needed. Big data analytics can also help to optimize charging schedules, taking into account factors such as energy prices, time of day and travel patterns. In addition, data analytics can provide insights into vehicle performance, allowing manufacturers to improve their products and extend the lifespan of EV batteries. As the use of EVs continues to grow, big data analytics will play an increasingly important role in managing the charging infrastructure, reducing costs, and improving the overall efficiency of the EV ecosystem.

6. *Power System Monitoring, Protection and Management*

Power system monitoring, protection, and management are critical components of ensuring the reliability and stability of the power grid. With the growth of renewable energy sources, distributed energy resources, and smart grids, the amount of data generated by power systems has grown exponentially. Big data analytics are essential to processing and analyzing this data, providing insights into the performance of the power system and enabling more effective management and protection. For example, big data analytics can help identify potential faults or disturbances in the power system and trigger protective measures to prevent blackouts or other disruptions. Big data analytics can also help optimize the use of distributed energy resources and manage the integration of renewable energy into the power grid.

7. *Load Forecasting and Classification*:

Load forecasting and classification are essential elements of managing the demand and supply of electricity on the power grid. With the advent of big data analytics, power system operators can collect and analyze massive amounts of data from various sources, including weather forecasts, historical energy consumption patterns, and customer behavior. By using machine learning algorithms, load forecasting models can predict future electricity demand more accurately, allowing power system operators to adjust the supply accordingly, and avoid overloading the grid or wasting energy. Furthermore, big data analytics can be used for load classification, which can help to identify specific patterns or trends in energy consumption. By analyzing data from individual consumers or groups of consumers, power system operators can better understand and predict energy demand, as well as identify areas of inefficiency and waste. In conclusion, the use of big data analytics in load forecasting and classification has the potential to significantly improve the efficiency and reliability of the power grid, leading to cost savings and a more sustainable energy future.

Chapter 4 - Literature Review of Short Term Load Forecasting Methods

4.1 Forecasting using ANNs

4.1.1 ANN with Back Propagation (BP) Algorithm

To improve load forecasting, various techniques have been proposed by researchers. One such method involves using a neural network trained with the back propagation momentum training algorithm. Another approach involves using an Artificial Neural Network (ANN) trained with the Artificial Immune System (AIS) learning algorithm, specifically for short-term load forecasting [12]. In load forecasting, neural networks have traditionally been trained using the backpropagation algorithm. However, a new algorithm has been developed that utilizes a similarity degree parameter to select relevant historical load data for training the neural network [13]. This approach provides several advantages over backpropagation, including greater accuracy, faster convergence, more efficient use of historical data, and improved mean average percentage error (MAPE).

4.1.2 ANN with Fuzzy Logic

Artificial neural networks (ANNs) have shown great promise in solving complex problems due to their ability to learn from data. However, in some cases, the use of fuzzy logic can further enhance the performance of ANNs. Fuzzy logic allows for the representation of imprecise or uncertain data, which is common in real-world applications. By combining ANNs with fuzzy logic, it is possible to create hybrid models that can handle complex and uncertain data more effectively. This can lead to more accurate predictions and better decision-making in areas such as finance, medicine, and engineering. The use of ANNs with fuzzy logic represents a powerful approach for tackling real-world problems, and it is an active area of research in the field of artificial intelligence.

A method has been proposed to simplify system structure and improve load forecasting precision [14], while constructing membership functions based on short-term load characteristics and modifying load heft has been shown to enhance load forecasting results [15]. Another forecasting method considers the effect of temperature and humidity by selecting similar days with weight factors [16]. Additionally, the use of Interval Type-2 Fuzzy Logic Systems (IT2 FLSs) has been proposed to handle uncertainties and improve prediction accuracy in short-term load forecasting. [17]

4.1.3 Hybrid models

A really interesting research has been performed related to hybrid models by Ioannis P. Panapakidis [14], which is also well-analyzed by Mr. Arvanitidis Athanasios Ioannis in his MSc thesis [18]. IoannisPanapakidis developed a sturdy hybrid model to predict day-ahead and hour-ahead load using hourly values from 10 buses in Thessaloniki, Northern Greece.The proposed method is highly targeted and centers on the implementation of the minCEntropy clustering algorithm on the training data to form k clusters. To create the hybrid model, the clustering and integration of historical load and temperature data in a Multilayer Perceptron (MLP) neural network is considered the most effective approach. Each subset is trained with a separate Artificial Neural Network (ANN), using the corresponding cluster data. To determine the best ANN for each test set pattern, the system computes the Euclidean distance between each pattern and k centroids, and feeds the output into the appropriate ANN.

4.2 Forecasting using Machine Learning Methods

4.2.1 Random Forest

The random forest (RF) ensemble method is a technique that combines individual models to improve the accuracy of the overall forecasting model. In a study referenced as [15], the authors applied this principle to predicthourly consumption of office buildings for the next day. They employed several ensemble algorithms, including RF, and also took into account environmental factors such as temperature and humidity, as well as previous load data, to improve the results. Ultimately, their use of RF resulted in a mean absolute percentage error (MAPE) of 6.11%.

4.2.2 ARIMA model

The ARIMA (Autoregressive Integrated Moving Average) model is a widely used time-series analysis methodology for forecasting. It is a statistical model that combines autoregression, which analyzes the relationship between an observation and a number of lagged observations, and moving averages, which analyzes the error term between actual and predicted values. ARIMA models are particularly useful in situations where the data has a clear trend or seasonal pattern. The model can be adjusted to account for trends, seasonality, and other factors that may impact the data being analyzed. The ARIMA model has been successfully applied in various fields, including finance, economics, and energy forecasting.

Although the ARIMA model is commonly used, it does not always result in a significant improvement in forecast accuracy and may be computationally expensive. This highlights the need to supplement these models with external inputs to achieve better results. Overall, time-series analysis is a powerful tool for load forecasting, but its success heavily relies on the quality and quantity of data available and the selection of appropriate models and techniques.

4.2.3 Support Vector Machines

The SVR model, which is commonly used for short-term load forecasting, is particularly effective when employing a linear kernel that reflects the linear relationship between input and forecasted data. In Reference [19], the authors compared the SVR model with other models like MLR and multivariate adaptive regression splines and concluded that the SVR model performed better with a MAPE of less than 2.6% for the day-ahead prediction. Similarly, in Reference [20], the authors suggested using the SVR model to forecast the Portuguese electricity consumption for 48 hours instead of an Artificial Neural Network (ANN) after efficiently tuning the hyper-parameters of the SVR model. The results showed a MAPE between 1.9% and 4% for the first and second-day forecast, respectively. In Reference [21], the authors proposed a variant of SVR called nu-SVR, which improved the optimization problem and automatically adjusted the epsilon tube width to adapt to data. A comparison was made between the nu-SVR and ANN models for forecasting the day-ahead hourly load in south-Iran, and the nu-SVR had a lower average MAPE of 2.95% compared to 3.24% for ANN, suggesting that the SVR model could be more effective than ANN in certain situations.

4.3 Forecasting using Reinforcement Learning

A paper of particular interest involves the integration of reinforcement learning to enhance a novel approach, making it adaptable to the environment. The proposed approach utilizes optimal forecasting models to improve prediction accuracy based on the unique characteristics of individual substations [22]. Overall, this paper presents a promising approach for STLF that takes advantage of advanced machine learning techniques and individualized analysis.

This paper proposes a solution that combines forecasting models with reinforcement learning (RL) and integrates them into a database framework. The solution is also parallelized to increase forecasting throughput. In the electric power industry, each substation has its unique load characteristics, and utility companies may have hundreds or thousands of substations that supply power to a diverse range of customer demographics. A forecasting model's primary objective is to make future predictions as accurately as possible, but there is no single model that can accurately predict the load for all substations globally. Therefore, the predictive models must be tailored to the specific characteristics of each substation's load profile, which may exhibit various time series characteristics.

Additionally, the models must adhere to certain guidelines, such as the reduction of stationary and seasonality, to improve the accuracy of the predictions. The integration of RL with forecasting models allows for the automatic selection of the best model for each substation's load profile, resulting in more accurate predictions. To address the challenge of creating the best forecasting model for every substation, an alternative approach is proposed. The first approach is to create a unique model for each substation, which is time-consuming since each substation must be tested individually. The proposed solution suggests utilizing multiple models for each substation to enable cross-referencing and different interpretations of predictions, rather than relying on a single model.The implementation of such an approach is not feasible considering the vast number of substations requiring daily management.Therefore, the proposed solution is to use multiple models and parameters to perform load forecasting on the loads at each substation, compare the results, and choose the model that produces the best results to use.The system is designed to interact with the environment by continuously receiving new load data and comparing it with its forecasts. If the error rate exceeds the acceptable threshold, the system uses reinforcement learning to make adjustments. It runs the same routine on the substation's cumulative time series and selects the best models with the proposed method. RL allows for iterative retuning of the models to make the best predictions. The ultimate objective is to select the optimal model for each substation during various time periods to ensure the prediction process is as accurate as possible.

The objective of RL is to gain knowledge from its environment and increase the likelihood of achieving better forecasting outcomes. The ***environment*** in this case is a substation that supplies power to consumers and is linked to the distribution network. The term "state" in this context describes the difference or deviation between the predicted load values and the actual load values observed during a certain time period, typically denoted as h.The ***new state*** is the updated difference between the actual loads and the new set of forecasted results obtained through the RL agent's optimized model's process. The ***action*** taken involves selecting and utilizing the most recent model picked by the RL agent's algorithm and comparing it to the latest actual load received. The ***scalar reward*** is based on the sum of variances or residuals from the models used by the agent. The objective is to achieve higher precision, resulting in lower total variance and a positive reward.

The **adaptive forecasting process** is shown in Figure 4.1. Firstly, the agent identifies the substation and obtains its load time series data. The data is then divided into a training set and a test set. The agent uses the training set to train the forecasting model and then applies it to forecast for the same period as the test set, which is period t. The residual is obtained from the model, and this is used to determine the model's preferential status with the RL agent. The agent possesses a collection of forecasting models stored in its knowledge base. It iterates through these models to obtain their respective residuals. Based on this, it identifies the most suitable model to utilize for predicting the substation's future load for a given time interval h. [22]

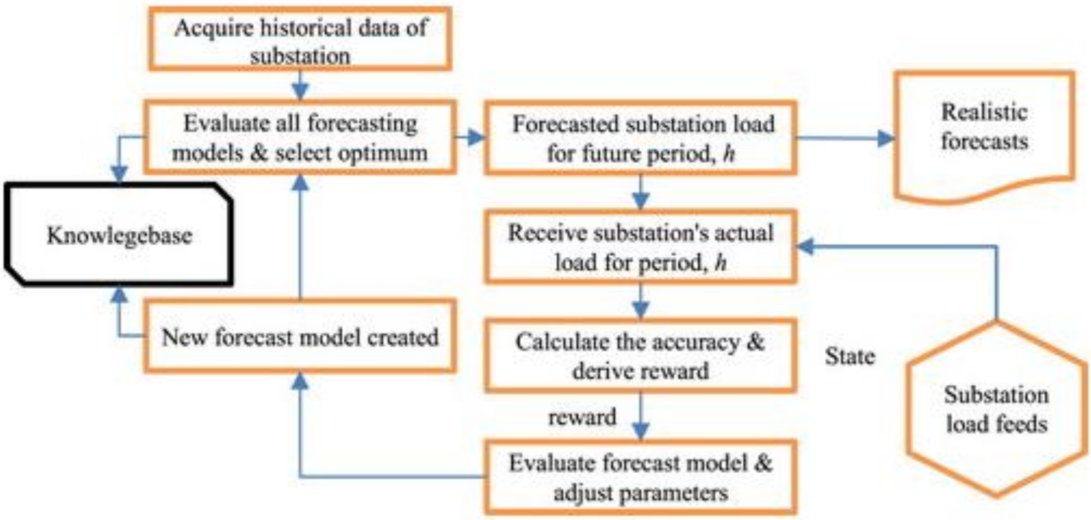

Figure 4.1: Proactive creation of RL agent's forecasting solutions. [22]

4.4 Observations

According to the above literature review it is clear that over the years the use of neural networks dominates the use of classical machine learning methods. It also seems that RL will play a very important role in the successful selection of models when we refer to huge systems that often require different types of models, e.g. load, demand, price, weather forecasting and much more. This fact can be explained by the increasing computing power and of course by the evolving research in the field of machine learning. However, because of their great complexity, neural dictations are usually a time-consuming solution. Consequently, a crucial topic that we will attempt to develop in the next chapter is the creation of minimal and highly efficient networks.

Chapter 5 - Data Analysis and Proposed Approach for Short-Term Load Forecasting

5.1 Minimum prerequisites

As mentioned in the previous chapter, the most critical issue in forecasting is a combination of two (2) factors: **model accuracy** and **complexity of the model**.

First of all, the model accuracy is the most vital aspect in machine learning. When it comes to load forecasting, it is quite important to build a model as accurate as possible. In this domain the acceptable accuracy according to the recent bibliography is 96%-99%.

Secondly, the complexity of the model is an equally important factor, especially as regards the prediction of electricity load. A model with many layers and neurons is kind of time-consuming, especially during training. Therefore, when looking for a scalable solution that will address not only research purposes but also widespread use by power companies, it is really important for the model to be minimal. This allows it to support both the hourly forecast and the hourly training with fresh data which is clearly a much heavier processing procedure.

5.2 Proposed Approach

Following the prerequisite specs mentioned in the previous section, this thesis presents as a proposal the creation of a minimal solution of a Multilayer Neural Network for the short-term (one hour ahead) forecast of Greece's electricity load. The data used refer to the Greek power system for the years 2015, 2016, 2017, 2018, 2019whereas low, medium and high voltage is combined. In addition, hourly temperature data of the Greek territory was used (for the same period as load data) in combination with the load consumption data, since it is proven that high and low temperatures have a considerable effect on the shape of the load consumption curve as they are also the ones that usually determine the "peaks". The data used was retrieved from the Open Power System Data

Project (OPSD) free download dataset containing the average load and temperature values of different EU power systems. [23]

For the process of training the model, data from the period 2015 - 2018 was used, while it was modified and analyzed in such a way that the final model takes as input data (load and temperature) of the previous day, week or even month depending on the results of the autocorrelation analysis which will be thoroughly presented in section 5.3.

Moreover, through mathematical and data mining models, an attempt was made to create new features that are a combination of the original raw data and the aim of increasing the accuracy of MLP at a small computational cost.

Finally, for the evaluation process, data of the same format was used relating to the period 2019. In addition, efforts were made for the most appropriate metrics to be used in order to rigorously and robustly evaluate the final model.

5.3 Data analysis and preparation

The original data structure consisted of the date time and the load consumption in MW at that time (Table 5.1). Subsequently, the temperature information from a different data source but for the same time range was added.

5.3.1 Feature engineering

Besides temperature, there are more factors which affect the fluctuation of the load consumption curve. Specifically, both **time** and the **type of day** shall be considered. As regards the first factor, namely **time**, Figure 5.1shows the load curve of a random day over time. It is obvious that there are two peaks which represent the peak hours of electricity consumption. According to data, the above-mentioned hours are between 9-10am and 8-10 pm. Furthermore, one more factor that affects the shape of the load curve (and not the absolute values) is the **type of day**. For example, it is important whether the day being studied is a working day or a weekend (where most of the population does not work), since the energy consumption is potentially shifted to different hours.

Finally, the season is a quite useful guide for forecasting electric load, taking into consideration thatduring summer and winter months, the higher consumption is a result of increased demand for heating and cooling purposes.

In conclusion, necessary conversions were made in the date time column, in order to study the above features in a form that is understandable and processable by a mathematical model such as neural networks.

Table 5.1: Raw load data

index	utc_timestamp	cet_cest_timestamp	GR_load_actual_entsoe_transparency
2	2015-01-01T01:00:00Z	2015-01-01T02:00:00+0100	5226.83
3	2015-01-01T02:00:00Z	2015-01-01T03:00:00+0100	4987.34
4	2015-01-01T03:00:00Z	2015-01-01T04:00:00+0100	4879.48
5	2015-01-01T04:00:00Z	2015-01-01T05:00:00+0100	4909.53
6	2015-01-01T05:00:00Z	2015-01-01T06:00:00+0100	4950.78
...	...	...	...
43820	2019-12-31T19:00:00Z	2019-12-31T20:00:00+0100	7530.86
43821	2019-12-31T20:00:00Z	2019-12-31T21:00:00+0100	6585.61
43822	2019-12-31T21:00:00Z	2019-12-31T22:00:00+0100	6108.25
43823	2019-12-31T22:00:00Z	2019-12-31T23:00:00+0100	5657.63
43824	2019-12-31T23:00:00Z	2020-01-01T00:00:00+0100	5337.38

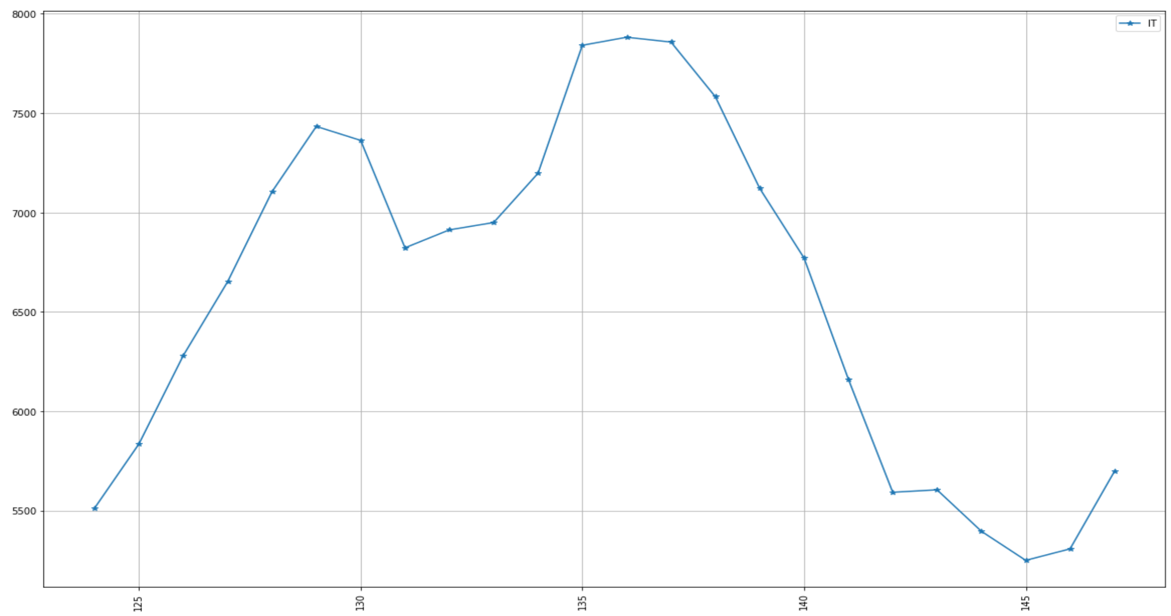

Figure 5.1: Electricity consumption curve of a random day.

(x-axis: Hour, y-axis: Load consumption in MW)

Table 5.2 shows the “utc” date time conversion with more useful information for the upcoming model.

Table 5.2: Date time conversion in dataset form.

utc_timestamp	Day_of_year	Time_num	Day_name	Week_day	Weekend	Cos_seasson	Sin_seasson
2015-01-01 01:00:00+00:00	1	1	Thursday	3	0.0	0.999852	1.721336e-02
2015-01-01 02:00:00+00:00	1	2	Thursday	3	0.0	0.999852	1.721336e-02
2015-01-01 03:00:00+00:00	1	3	Thursday	3	0.0	0.999852	1.721336e-02
2015-01-01 04:00:00+00:00	1	4	Thursday	3	0.0	0.999852	1.721336e-02
2015-01-01 05:00:00+00:00	1	5	Thursday	3	0.0	0.999852	1.721336e-02
...	...	...	...	...	...	...	...
2019-12-31 19:00:00+00:00	365	19	Tuesday	1	0.0	1.000000	6.432491e-16
2019-12-31 20:00:00+00:00	365	20	Tuesday	1	0.0	1.000000	6.432491e-16
2019-12-31 21:00:00+00:00	365	21	Tuesday	1	0.0	1.000000	6.432491e-16
2019-12-31 22:00:00+00:00	365	22	Tuesday	1	0.0	1.000000	6.432491e-16
2019-12-31 23:00:00+00:00	365	23	Tuesday	1	0.0	1.000000	6.432491e-16

Cos (5.3.1) and sin (5.3.2) season new features represent the seasonality of data in a numerical format.

$$cos_{seasson} = cos\frac{2\pi * day_{year}}{365} \tag{5.1}$$

$$sin_{seasson} = sin\frac{2\pi * day_{year}}{365} \tag{5.2}$$

Figure 5.2 shows a graph with cos and sin seasonality values during an entire year.For example, for the days of the winter period, cos-seasonality gives values that tend to the maximum, for the summer days to the minimum whereas for spring and autumn we notice prices that are located around zero. On the other hand, in sin-seasonality average values are observed for the winter and summer periods, maximum for spring and minimum for autumn. Consequently, these two features can concurrently produce continuous values in the range [-1,1] to describe the time each observation is in the dataset.

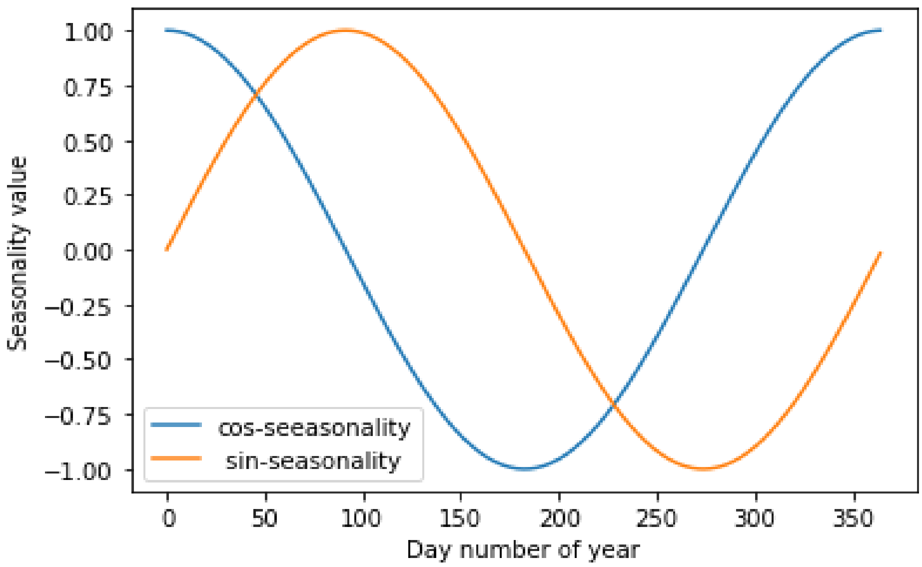

Figure 5.2: Seasonality curves

5.3.2 Features correlation

The prediction of electrical load is translated into a supervised learning problem, in which it is essential to keep their complexity low in terms of the number of columns that we will use during the training process. Therefore, it is important to know which features have a high correlation with the prediction goal, so as not to introduce 'noise' which can confuse the model and lead it to over-fitting or under-fitting.

To assess the correlation between two continuous variables, the **Pearson correlation** is a useful metric. This statistical measurement, also known as Pearson's r or the Pearson product-moment correlation coefficient, is based on the ratio of the covariance of two variables and the product of their standard deviations. This produces a normalized value between -1 and 1, which reflects the strength and direction of a linear relationship between the variables. However, the Pearson correlation can only indicate linear correlation and does not account for other types of relationships or correlations between the variables. For instance, if one were to examine the age and height of a group of high school students, the Pearson correlation coefficient would likely be greater than 0 but less than 1, as a perfect correlation is unlikely in reality.

$$Cor_{var1var2} = \frac{covariance(var1,var2)}{std_{var1} * std_{var2}} \quad (5.3)$$

Figure 5.3 shows the correlation matrix of the current dataset. When two variables have a positive correlation, it means that as one variable increases, the other variable tends to increase as well. In other words, there is a direct relationship between the two variables such that when one variable goes up, the other tends to go up too. This positive correlation can be measured using statistical methods such as Pearson correlation coefficient. It can be seen that temperature (as expected) has the highest correlation with the target, while seasonality has the lowest. Therefore, it is possible to reduce the complexity of the problem by removing the columns with low correlation with load consumption.

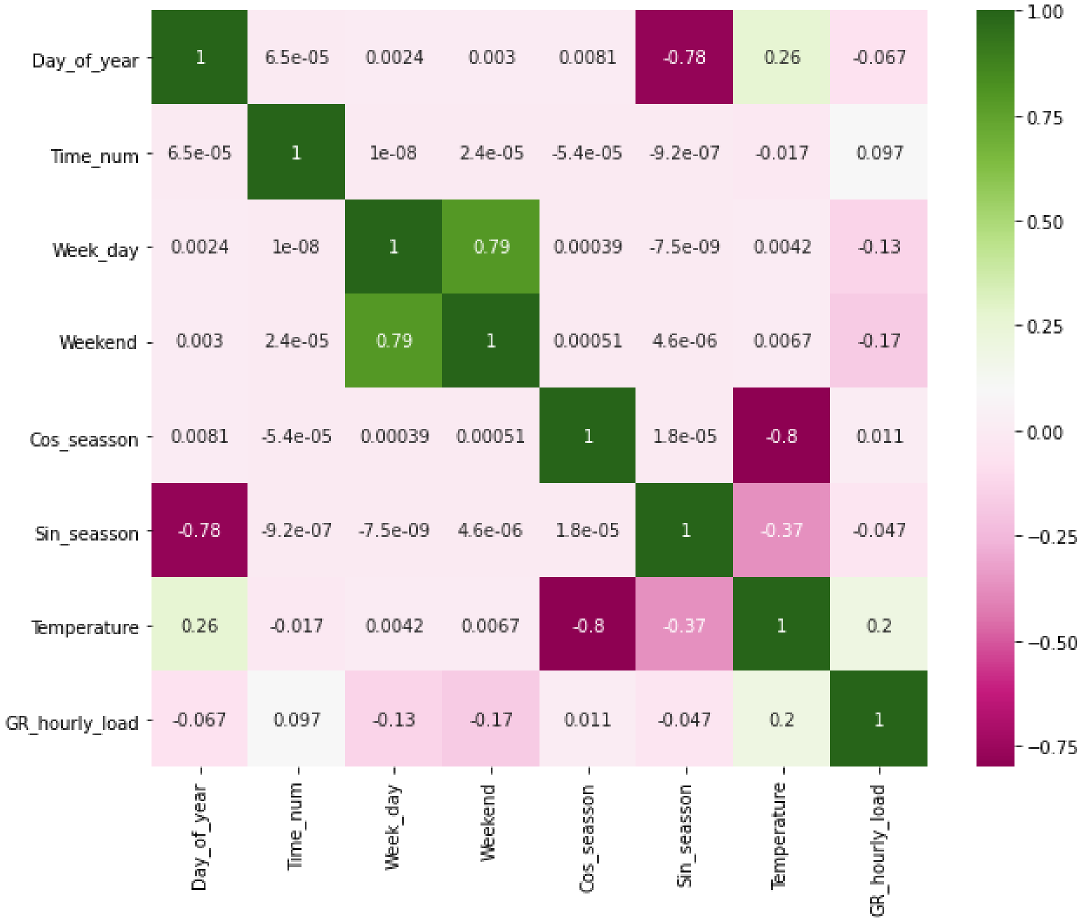

Figure 5.3: Correlation matrix

5.3.3 Autocorrelation analysis

First and foremost, it is crucial to define the concept of autocorrelation followed by the mathematical sequence. Autocorrelation, as defined by Wikipedia, is the correlation of a signal with a delayed version of itself where the amount of delay is the independent variable. Autocorrelation measures the similarity between observations of a random variable over time, and is a mathematical technique used to identify repeating patterns, such as periodic signals that may be obscured by noise. It is also useful in identifying the fundamental frequency of a signal by analyzing its harmonic frequencies. Autocorrelation is commonly employed in signal processing to analyze time domain signals or series of values.

The average of a time series y1, …, y*n*is:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (5.4)$$

The auto covariance function at lag *k*, for *k*≥ 0, of the time series is defined by:

$$s_k = \frac{1}{n}\sum_{i=1}^{n-k}(y_i - \bar{y})\,(y_{i+k} - \bar{y}) = \frac{1}{n}\sum_{i=k+1}^{n}(y_i - \bar{y})\,(y_{i-k} - \bar{y}) \qquad (5.5)$$

The autocorrelation function (ACF)at lag*k*, for *k*≥ 0, of the time series is defined by:

$$r_k = \frac{s_k}{s_0} \qquad (5.6)$$

It is clear that in matters of forecasting given time series, the most important factor for forecasting the target is its own values in the past and hence, they shall be used as the most essential inputs in the model. It remains only to be determined which moments of the past best describe the moments of the future. For this reason, we used the autocorrelation metric to leverage past moments as much as possible.

Figure 5.4 shows the autocorrelation plot, where x-axis shows the past hourly value, while y-axis shows the autocorrelation values. It appears that, at the moment t-0 (the moment we are referring to) the autocorrelation value is 1, whereas if we go back in time this value decreases. It is worth noting that the high autocorrelation values are observed one hour before, one day before, one week before and one month before. This pattern can be considered as a very useful guide.

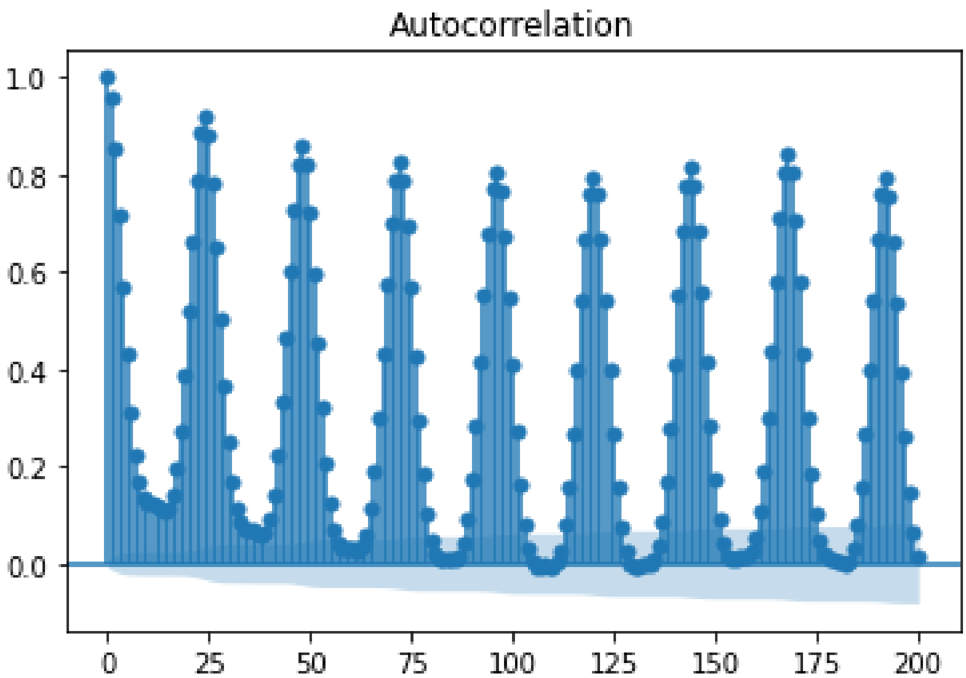

Figure 5.4: Autocorrelation graph

(x-axis: Hours, y-axis: Auto-correlation score)

Therefore, based on the values in the diagram above, the following moments were chosen to be used: t-1, t-24, t-23, t-25, t-48, t-2, t-168, t-72, t-47.

5.3.4 Input preparation

The observations of the previous unit lead to the need to use past data as inputs to the model. In order to achieve something like this, it is necessary to modify the structure of the dataset. Since the oldest point in time (that was decided to be used as an input to the model) is a month ago, the training process as well as the evaluation will be initiated one month later than the oldest time input. For example, the raw data used in this work starts from timestamp 01-01-2015T01-00-00. The oldest time point that will be used as input is one week (168 hours). Consequently, the target feature will be shifted 168 points ahead of time, while the independent variables will be cut off 168 points earlier than the end of the set. With this practice independent and dependent variable have the same length and the model can make the most appropriate use of the past moments to predict an hour ahead from the present. Table 5.3 shows the final structure of the dataset in which the last column consists of the prediction target, while all the others are the independent variables that will be given as inputs to the model.

Table 5.3: Dataset after load consumption input

	GR_hourly_load-168	GR_hourly_load-1	GR_hourly_load-2	GR_hourly_load-24	Week_day	Weekend	Temperature-1	Temperature	Cos_seasson	Sin_seasson	GR_hourly_load
0	5226.83	5869.19	5890.13	5396.53	3.0	0.0	-0.927	-0.990	0.990532	1.372788e-01	5707.95
1	4987.34	5707.95	5869.19	5250.28	3.0	0.0	-0.990	-1.046	0.990532	1.372788e-01	5552.57
2	4879.48	5552.57	5707.95	5308.04	3.0	0.0	-1.046	-1.198	0.990532	1.372788e-01	5562.97
3	4909.53	5562.97	5552.57	5697.59	3.0	0.0	-1.198	-1.348	0.990532	1.372788e-01	6053.18
4	4950.78	6053.18	5562.97	6449.47	3.0	0.0	-1.348	-1.311	0.990532	1.372788e-01	6986.42
...	...	...	...	...	...	...	...	...	...	...	...
43650	6303.19	8401.76	8733.89	7449.33	1.0	0.0	4.673	4.503	1.000000	6.432491e-16	7530.86
43651	5662.10	7530.86	8401.76	6821.04	1.0	0.0	4.503	4.342	1.000000	6.432491e-16	6585.61
43652	5237.44	6585.61	7530.86	6274.54	1.0	0.0	4.342	4.219	1.000000	6.432491e-16	6108.25
43653	4764.68	6108.25	6585.61	5589.98	1.0	0.0	4.219	4.143	1.000000	6.432491e-16	5657.63
43654	4337.98	5657.63	6108.25	5025.98	1.0	0.0	4.143	4.128	1.000000	6.432491e-16	5337.38

One more helpful technique for machine learning is the conversion of discrete value columns into multiple binary columns. Table 5.4 shows the final form of the dataset after converting column Weekday into several binaries, whereas each of them declares the existence or not of the day.

Table 5.4: Final dataset, ready for training

	GR_hourly_load-168	GR_hourly_load-1	GR_hourly_load-2	GR_hourly_load-24	Weekend	Temperature-1	Temperature	Cos_seasson	Weekday_6	Weekday_5	Weekday_4	Weekday_3	Weekday_2	Weekday_1	Weekday_0	Sin_seasson	GR_hourly_load
0	5226.83	5869.19	5890.13	5396.53	0.0	-0.927	-0.990	0.990532	0	0	0	1	0	0	0	0.137279	5707.95
1	4987.34	5707.95	5869.19	5250.28	0.0	-0.990	-1.046	0.990532	0	0	0	1	0	0	0	0.137279	5552.57
2	4879.48	5552.57	5707.95	5308.04	0.0	-1.046	-1.198	0.990532	0	0	0	1	0	0	0	0.137279	5562.97
3	4909.53	5562.97	5552.57	5697.59	0.0	-1.198	-1.348	0.990532	0	0	0	1	0	0	0	0.137279	6053.18
4	4950.78	6053.18	5562.97	6449.47	0.0	-1.348	-1.311	0.990532	0	0	0	1	0	0	0	0.137279	6986.42

The "feature scaling" method is an important technique used in data processing to normalize or standardize the range of independent variables or data features. This process involves rescaling the values of the data features so that they fall within a specified range or distribution. In simpler terms, feature scaling allows data scientists to adjust the scale of different features so that they can be compared on the same scale. This technique is particularly important in machine learning, where algorithms often rely on distance-based metrics to compare data points. If the range of values of different features is not normalized, the feature with a larger scale will dominate and impact the results more significantly. Feature scaling ensures that each feature is given equal importance in the analysis and modeling process.

This leaves a model in the lurch when trying to learn from such data or fit a model to such data. In the case of existing datasets there is indeed a different scale between features. For example, the consumption data is in the order of thousands (MW) while the temperature data is in the order of tens. Consequently, min-max normalization (5.7) was used for this case.

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{5.7}$$

5.4 Implementation

In this section, the architecture of the neural network will be presented along with the choice of hyper parameters that are appropriate for the accuracy of the predictions. In view of the above, two more different approaches will be extensively discussed and compared, in order to draw robust conclusions.

5.4.1 MLP proposed architecture

As mentioned at the beginning of this chapter, the aim of the present thesis is to keep the complexity of the problem at low levels as much as possible. The first step is to utilize only sparingly the independent variables, and the second is to use a MLP whose size is such that it does not lose accuracy, but is also small enough to have live retraining capabilities.

After several tests and in conjunction with publications of various researches on the compared or similar object (time series prediction) an MLP was decided which will consist of 1 input layer of 16 neurons (that is, the independent variables) 3 hidden layers 50, 50 and 16 neurons respectively and finally the output layer which includes a neuron that basically makes the prediction of the electric charge one hour ahead in the future. Figure 5.5shows the graphical representation of this architecture.

To add, it should be noted that the rectified linear unit (ReLu) activation function is used at each level. ReLu is an activation function that is defined as the positive part of its input. In order to get a clearer picture of it, Figure 5.6 shows the graphical representation of ReLu. Additionally, the commonly called "*kernel initializer parameter*" was initialized as normal. The term "kernel" has been borrowed from other traditional methods such as SVM. The concept is to convert the data from a particular input space to another space by utilizing kernel functions. In neural networks, we can perceive layers as non-linear maps that carry out these transformations, and therefore, the terminology "kernels" is applicable. As a result, in the current case when weights of the layers are initialized as normal, it means that a normal distribution will be used for the transformations.

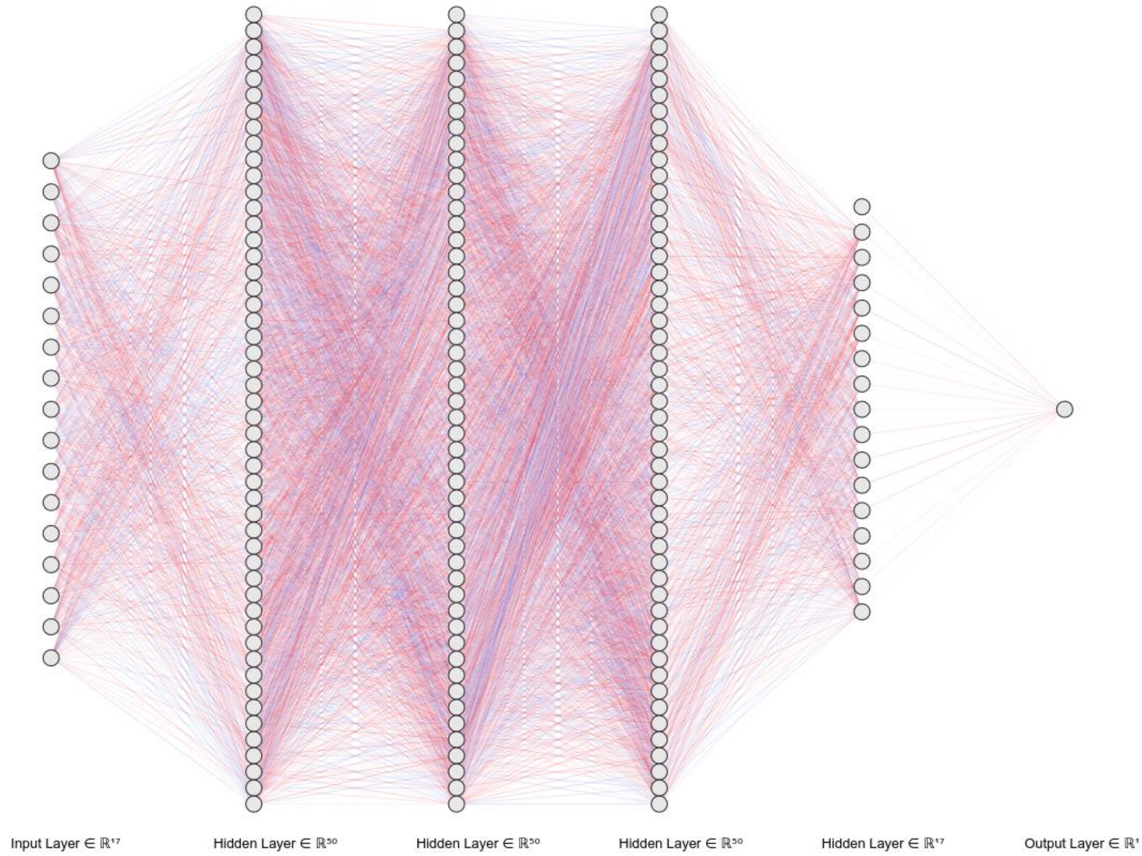

Figure 5.5: MLP architecture

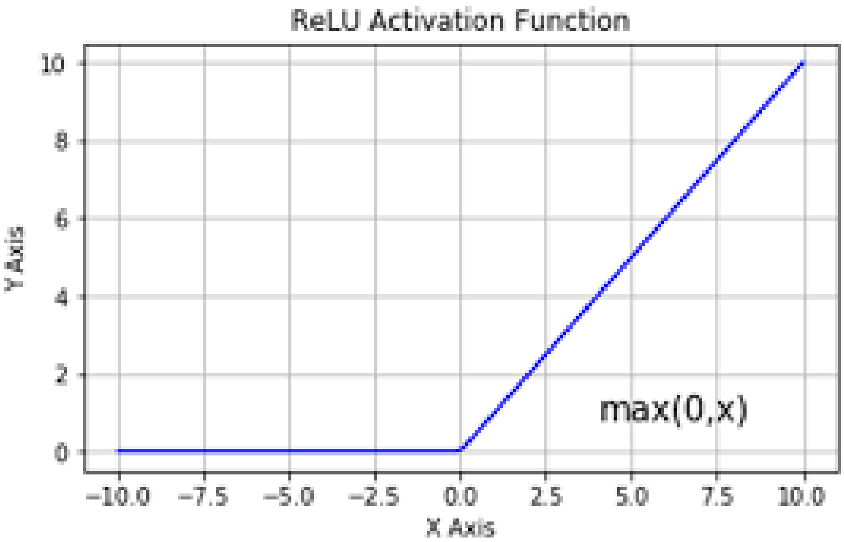

Figure 5.6: ReLu activation function

The maximum duration of the training process was set to 1.000 epochs, since early stopping was used. The term early stopping defines the “patience” during the training of the model. For example, in the current implementation the limit was set to 80 epochs, which means that if the validation error does not decrease during 80 consecutive epochs then the training process stops and the model uses the best weights up to that point. The metric of the validation error was set to Mean Absolute Error (MAE) (5.8) which is suitable for problems related to the prediction of continuous values such as the consumption of the electrical load.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (5.8)$$

5.4.2 Case 1 - Simple Hourly Load Forecasting

For the first case of load forecasting, the training process lasted 1.000 epochs, which means that the model was constantly being improved, in a way that a longer duration for training could be used. However, tests that were performed showed minor differences and due to the fear of over fitting the 1.000 epochs remained as such. Training data was divided into two (2) sets: **training** and **validation,** whose errors during training are shown in Figure 5.7.

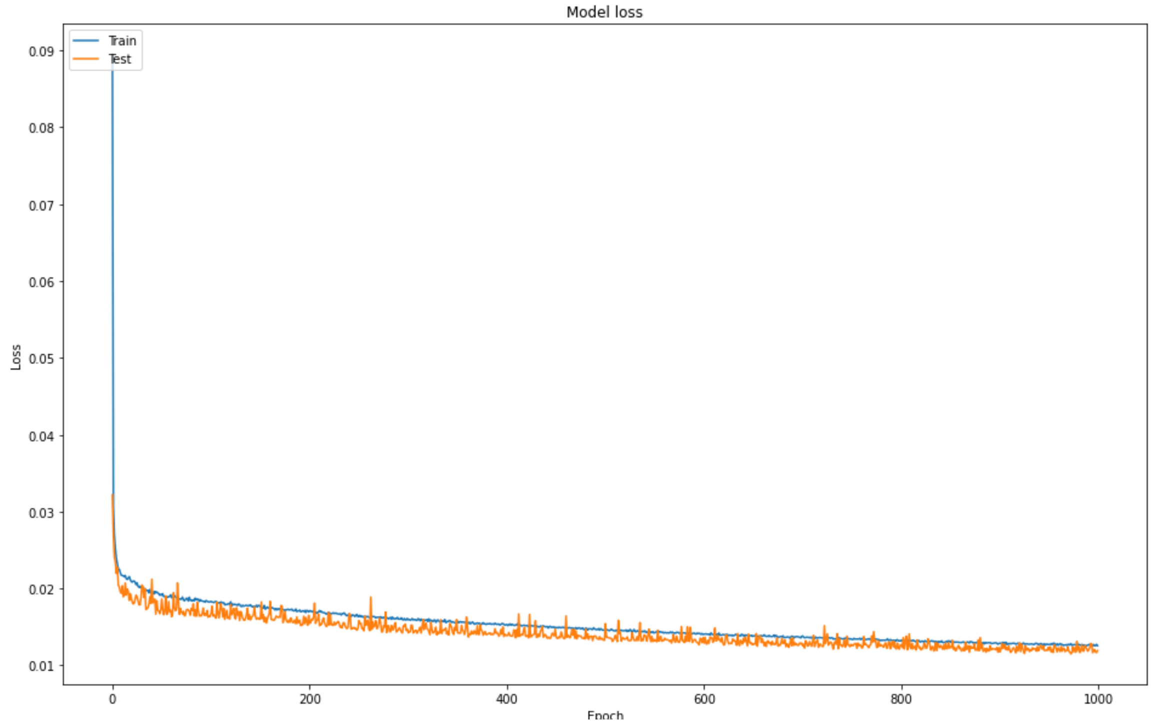

Figure 5.7: Train and test loss during the training

Looking ahead, Figure 5.8 illustrates the graph of forecast results using the R2 score for the test dataset. The R-squared (R2) is a statistical metric that shows the proportion of variance in a dependent variable that is accounted for by an independent variable or a set of independent variables in a regression model. While correlation measures the strength of the relationship between two variables, R-squared quantifies the extent to which the variation in one variable can be explained by the variation in another variable. Therefore, if a model's R2 is 0.50, it implies that half of the observed variation can be explained by the inputs of the model.

.

$$R^2 = 1 - \frac{\sum squaredregression(SSR)}{total \sum of squares(SST)} = 1 - \frac{\sum y_i - \hat{y}_i}{\sum y_i - \bar{y}_i} \tag{5.9}$$

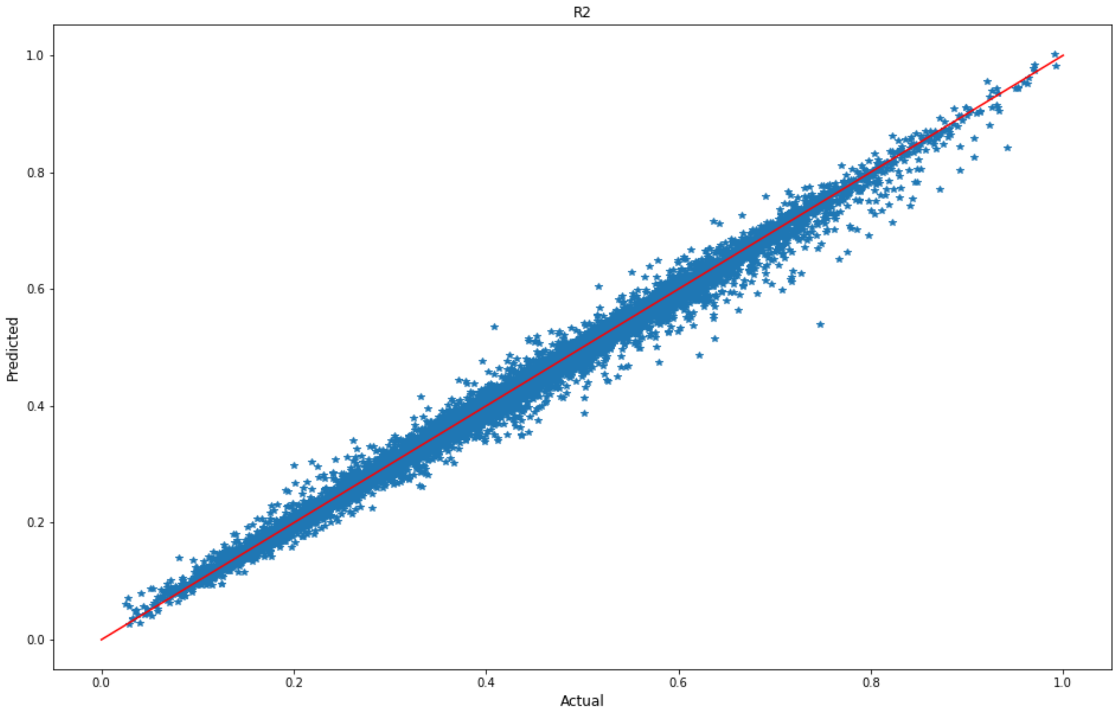

Figure 5.8: R2 score graph for test set

The x-axis represents the actual values of the test set, while the y-axis shows the values predicted by the model. As one can notice, the equation y=x has been plotted, which is used as an aid to better understand the spread of the forecast. Essentially, the points that are above or very close to the equation y=x have been predicted with great accuracy, while those that are further away have a greater error.

In addition, the model's accuracy was assessed using Mean Absolute Percentage Error (MAPE) as shown in section 5.10. MAPE is a metric commonly used to measure the accuracy of a forecasting method, and it calculates the average of the absolute percentage errors of each entry in a dataset. This helps to determine how accurate the forecasted quantities were compared to the actual quantities. MAPE is often useful when analyzing large datasets and requires non-zero values in the dataset.

$$MAPE = \frac{1}{n}\sum\left|\frac{A_t - F_t}{A_t}\right| \tag{5.10}$$

- n = number of times the summation iteration happens

- A = actual values
- F = forecasted values

Therefore, the accuracy of the MLP can be summarized as follows:

- Mean absolute percentage error = 1.33 %
- Accuracy = 98.67 %
- Mean absolute error (MW) = 78.284 MW
- r2 = 99.1 %

Figure 5.9 shows the load consumption sequence of the first week of the test set along with the model predictions. As it seems, the variation is small and mainly concerns the peak load during rush hours. The correct prediction of the peak load is considered one of the most challenging issues in the field of short-term electric load forecasting.

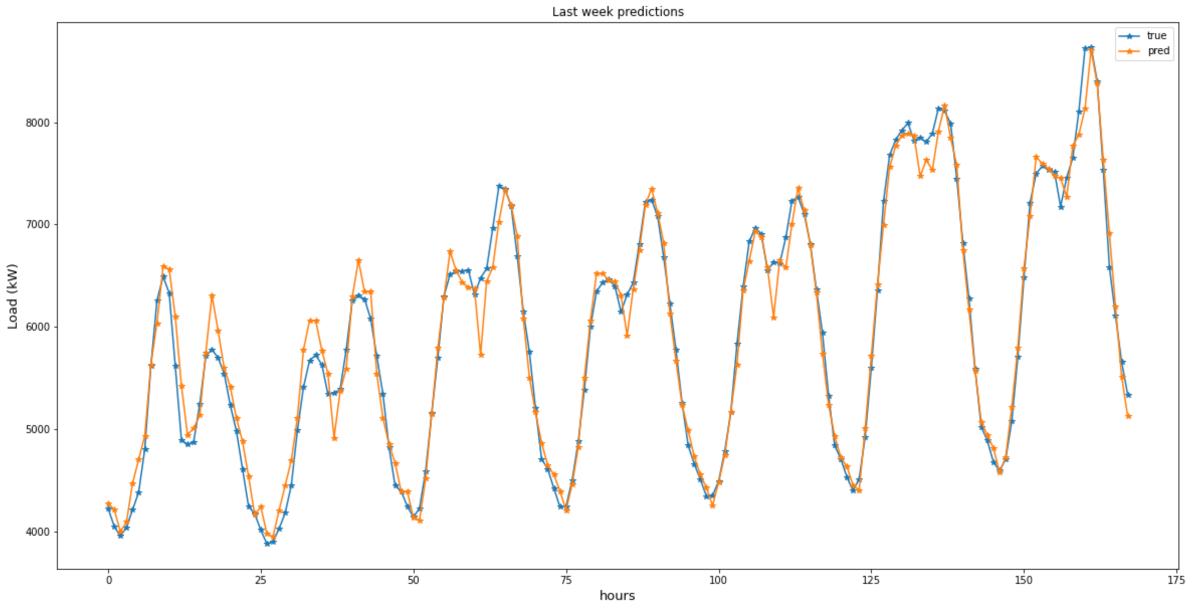

Figure 5.9: Actual and forecasted load consumption of first test set week

Finally, in order to better understand the problem, a categorization was made into the percentage errors of the model's forecast by season of the year. As image 5.10depicts, the season with the smallest error is summer, whereas the one with the largest one is winter. This fact can be explained by assuming that temperature changes during the summer are milder in relation to winter's where the most intense weather events take place such as snowfalls, strong storms and winds, which can cause sharp changes in load consumption and damage the power distribution network.

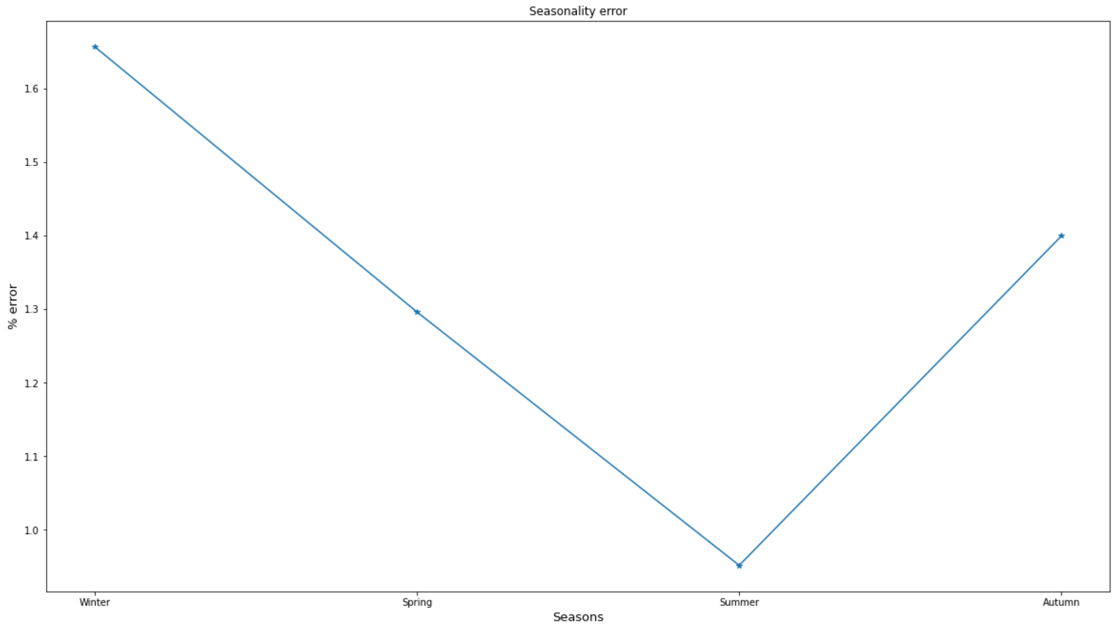

Figure 5.10: MAPE across the seasons

5.4.3 Case 2 - Simple Hourly Load Forecasting using profiling

On the occasion of the implementation [20] a test was carried out for the creation of consumer profiles in order to investigate whether they can contribute to increasing the accuracy of the model. Figure 5.11 shows the raw load consumption. It is clear that the amount covered by this consumption could be divided into categories. For this reason, with the assistance of the machine learning algorithm 11, 10 possible clusters are created and evaluated with the silhouette score metric. To evaluate the clustering performance of the model, the Silhouette Coefficient was used. This coefficient is calculated based on the mean distance within a cluster (a) and the mean distance to the nearest cluster (b) for each sample. The Silhouette Coefficient for a sample is then calculated using the formula (b - a) / max (a, b). The value of b represents the distance between a sample and the nearest cluster that it does not belong to. It is important to note that the Silhouette Coefficient can only be calculated if the number of labels is between 2 and n_samples - 1. The resulting Silhouette scores for the 10 possible clusters are shown in Figure 5.12, and the profile distribution for the clustering with the highest score is shown in Figure 5.13.

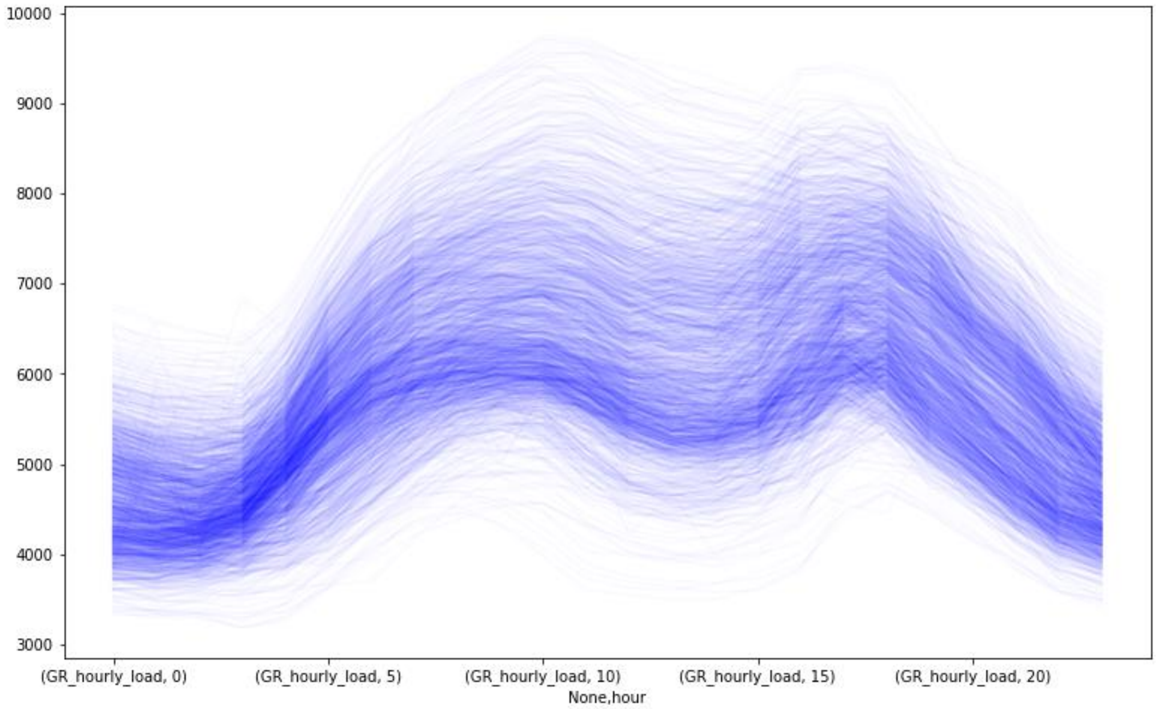

Figure 5.11: Raw load consumption

(x-axis: Hours, y-axis: Load (MW))

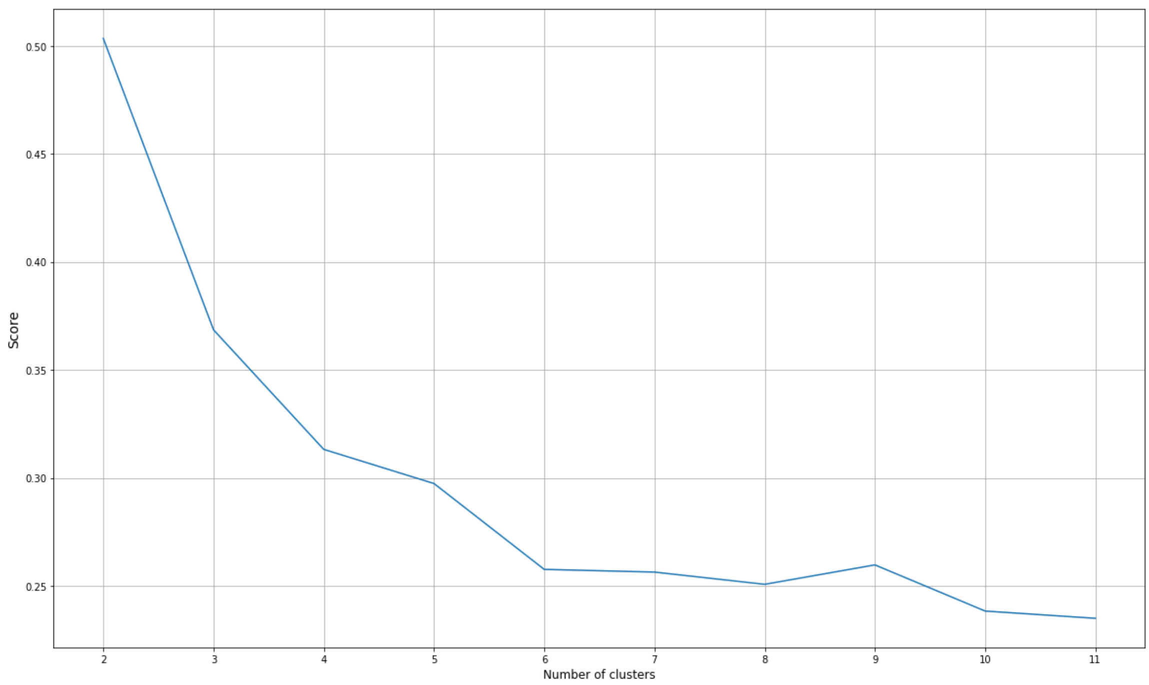

Figure 5.12: Silhouette score for 10 possible clusters

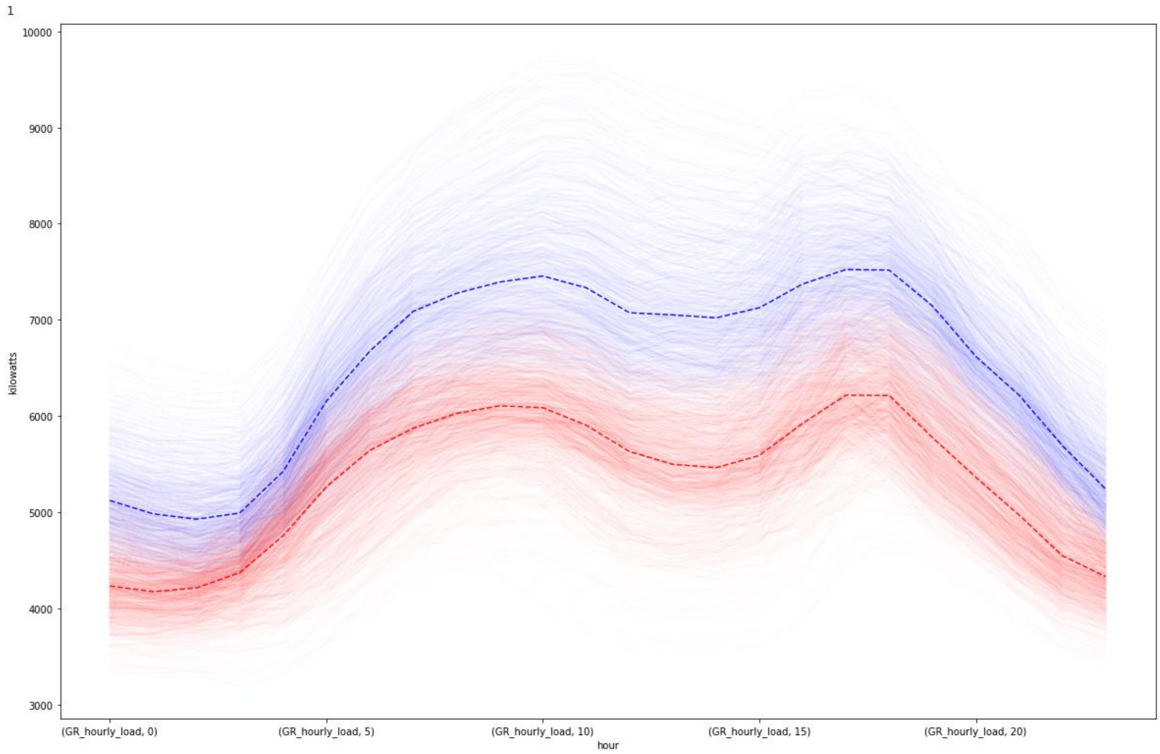

Figure 5.13: Selected clusters

Consequently, a new feature was created which includes the information to which of the two consumer profiles each record of the dataset belongs to. The training of the model, which will also include consumer profiles, was carried out in almost the same way as the previous case with the only difference being that the input layer of the model consists of 17 neurons instead of 16 due to the new column of profiles.

The results obtained from the evaluation of the model in this case are quite similar to the ones from the previous case and can be compared in Table 5.5:

Table 5.5: Load forecasting with and without consumers profile comparison

	Accuracy	Percentage Error	Absolute Error(MW)	R2 score
Case 1	99.7%	1.3%	79.3	99.1
Case 2	99.65 %	1.35%	80	99

5.4.3 Outcome analysis and final proposal

As can be noticed in the Table 5.5from the previous section, there's not much difference between the two implementations. In absolute numbers, the first (and simpler) implementation has an average 0.7MW less error than the second case. This difference is not adequate to determine which of the two models is more suitable for the existing problem. In different data, perhaps not aggregated (of the entire country), consumer profiles could play a more important role in differentiating low, medium and high voltage consumers and, by extension, increase the accuracy of the forecast. Despite all this, in the present case it seems that not only did they not have a positive effect, but perhaps also introduced 'noise' which had a negative impact on the final results. Moreover, as mentioned several times in previous chapters, in the field of machine learning and specifically in load forecasting, the most crucial element is the accuracy of the model. However, in case we need to seek for a commercial solution the factors of time and complexity are equally important.

In view of the above, it is easy to understand that the first implementation is much simpler and cheaper in terms of computational complexity and fast prediction. It should be taken under consideration that with the second implementation, the clustering process for finding the profile is potentially added to a live application, as well as its retraining in order the model is always up-to-date. The same happens with the process of load forecasting in which we have added an extra feature increasing the architecture and computations of the MLP. Consequently, since the first implementation provides at least better results and is more economical and minimal as a solution, it is proposed as the most suitable one for the problem of the short-term load forecasting of the Hellenic energy network.

Chapter 6 - Conclusions

In the year of 2023, in the wake of a war and under an energy and economic crisis, it is critical to reflect on the past, assess the present and plan for the future. The multiple crises have led us to review many constants of our lives prior to all the above mentioned unfortunate developments. Life seems to be more important than ever. Geostrategic and economic games lead humanity to impoverishment, driving it back many decades ago, whereas at the same time planet earth steps faster and faster into an overheating that triggers the ongoing crisis. As a result, it is every citizen’s as well as society’s duty to act directly.

Technology is once again a factor that plays a very important role in the crisis we are currently experiencing. From the negative point of view, technology can be used as a weapon and give the possibility to destroy life and planet with great ease. On the other hand, it seems that it is humanity's only hope to escape from its problems and fears, such as the overheating of the planet, social inequalities, the encroachment of human rights and much more.

As it turns out from the previous chapters, the present thesis refers to a major technological issue which has the potential to contribute positively directly and indirectly to the improvement of many of the problems we face today. The prediction of energy consumption is a really important “tool” that will contribute to the reduction of the production of unused energy, which apparently creates multiple benefits for the society. Saving energy from reducing production will lead to a significant reduction in pollutant emissions and will assist in slowing down global warming. It will also affect the prices which will eventually drop and hence, ease the economic crisis by creating greater energy sufficiency especially in countries facing shortages. Finally, it will provide the technological political community with time in order to develop new technologies and designs with the ultimate aim to make a completely green planet possible without depending on on hydrocarbon mining and human exploitation.

Bibliography

[1] *Independence Energy and Security act of 2007.*p. 153, 110–140. 110th Congr. Public law110-140, 2007.

[2] Maha Al Dahmi,*A Review of Factors Affecting Electricity Peak Demand in Al Ain Area,* IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2018, https://ieeexplore.ieee.org/document/8566401/authors#authors

[3] *MachineLearning,*Wikipedia,https://en.wikipedia.org/wiki/Machine_learning

[4] Md. Golam Rahman, M. Fahad Bin Ramim Chowdhury, Md. Abdulla Al Mamun, Md. Rakib Hasan, Sayeed Mahfuz, *Summary of Smart Grid: Benefits and Issues,* International Journal of Scientific and Engineering Research4(3):7, 2013

[5] *Deep Learning,* Wikipedia, https://en.wikipedia.org/wiki/Deep_learning

[6] *Artificial Neural Network*, Wikipedia, https://en.wikipedia.org/wiki/Artificial_neural_network

[7] W. Strielkowski, *Economics and Sociology,* Social and Economic Implications for the Smart Gridsof the Future, Vol. 10, No. 1, pp. 310-318. DOI:10.14254/2071-789X.2017/10-1/22, https://economics-sociology.eu/files/ES_10_1_Strielkowski.pdf

[8] *Reinforcement Learning,* Wikipedia, https://en.wikipedia.org/wiki/Reinforcement_learning

[9] Keith D. Foote,*Artificial Neural Networks: An Overview,* 2021. (https://www.dataversity.net/artificial-neural-networks-overview/)

[10] Tacio Souza Bomfim, *Evolution of Machine Learning in Smart Grids*, IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), 2020.

[11] Maedeh Ghorbanian, Sarineh Hacopian Dolatabadi,Pierluigi Siano, *Big Data Issues in Smart Grids: A Survey,* IEEE Systems Journal, Volume: 13, Issue: 4, 2019, https://ieeexplore.ieee.org/document/8809368

[12] Khosravi, A, Nahavandi, S.Creighton, *Short term load forecasting using Interval Type-2 Fuzzy Logic Systems*, Fuzzy Systems (FUZZ), IEEE International Conference on, p. 502-508, 2011.

[13] Kuihe Yang &Lingling Zhao, *Load forecasting based on amendment of Mamdani Fuzzy System*, Conference: Wireless communications, networking & mobile computing. p. 1-4, 2009.

[14] Ioannis P. Panapakidis, *Clustering based day-ahead and hour-ahead bus load forecasting models,* International Journal of Electrical Power and Energy Systems, 80:171–178, 2016.

[15] S. Hadri, NaitMalek, Y. Najib, M. Bakhouya, M. Fakhri, Y. El Aroussi, *AComparative Study of Predictive Approaches for Load Forecasting in Smart Buildings,* Procedia Comput. Sci.,p. 160 & p. 173–180, 2019.

[16] Barakat, E.H. Al-Qasem, J.M., *Methodology for weekly load forecasting*, IEEE Trans. Power Syst., p. 1548–1555,1998.

[17] N. Amjady, *Short-term hourly load forecasting using time-series modeling with peak load estimation capability*, IEEE Trans. Power Syst., p. 798–805, 2001.

[18] Arvanitidis Athanasios Ioannis, *Short-Term Load Forecasting Using Artificial Neural Netwroks,* MSc Thesis, p. 29-31, 2021.

[19] K. Chapagain, S. Kittipiyakul, *Short-Term Electricity Demand Forecasting with Seasonal and Interactions of Variables for Thailand*, In Proceedings of the 2018 International Electrical Engineering Congress (iEECON), Krabi, Thailand, 2018.

[20] P.M. Ferreira, I.D. Cuambe, A. Ruano, R. Pestana, *Forecasting the Portuguese Electricity Consumption using Least-Squares Support Vector Machines*. IFAC Proc., p. 411–416, 2013.

[21] T. Pinto, I. Praça, Z. Vale, J. Silva, *Ensemble learning for electricity consumption forecasting in office buildings*, Neurocomputing, p. 423, 747–755, 2021.

[22] Chee Keong Wee, Richi Nayak, *Adaptive load forecasting using reinforcement learning with database technology,* Journal of Information and Telecommunication, Volume 3, 2019, https://www.tandfonline.com/doi/full/10.1080/24751839.2019.1596470?scroll=top&needAccess=true&role=tab

[23] Open Power System Data Project (OPSD), Data Platform (https://data.open-power-system-data.org/time_series/2020-10-06?fbclid=IwAR1Cv6PLvjlPoTf0oRVqHvfSnsDlq6dnoTXGAQfYDGw0c3K0ifUJWbuE_HI).