UNIVERSITY OF THESSALY
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
Programme of MSc Studies

INPUT SELECTIONS METHODS FOR NEURAL NETWORKS FOR LOAD AND GENERATION

Master's Thesis

Theodoros Zerlentis Kafarakis

Supervisor: Loutridis Spyridon

Volos 2023

UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Programme of MSc Studies

INPUT SELECTIONS METHODS FOR NEURAL NETWORKS FOR LOAD AND GENERATION

Master's Thesis

Theodoros Zerlentis Kafarakis

Supervisor: Loutridis Spyridon

Volos 2023

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
Πρόγραμμα Μεταπτυχιακών Σπουδών

ΜΕΘΟΔΟΙ ΕΠΙΛΟΓΗΣ ΕΙΣΟΔΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΚΑΤΑΝΑΛΩΣΗΣ

Μεταπτυχιακή Διπλωματική Εργασία

Θεόδωρος Ζερλέντης Καφαράκης

Επιβλέπων: Λουτρίδης Σπυρίδων

Βόλος 2023

Approved by the Examination Committee:

Supervisor **Loutridis Spyridon**

Associate Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Bargiotas Dimitrios**

Professor, Department of Electrical and Computer Engineering, University of Thessaly

Member **Daskalopulu Aspassia**

Associate Professor, Department of Electrical and Computer Engineering, University of Thessaly

Acknowledgements

I would like to dedicate this thesis to my parents and my brother for their inspiration, motivation and encouragement through my entire life.

I would also like to express my gratitude to my supervisor Dr. Loutridis Spyridon for his guidance and support during the course of this thesis. I would also like to express my gratitude to the late professor Ioannis Panapakidis for his help during my studies and authoring this thesis.

DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS

«Being fully aware of the implications of copyright laws, I expressly state that this diploma 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

Theodoros Zerlentis Kafarakis

Master's Thesis

INPUT SELECTIONS METHODS FOR NEURAL NETWORKS FOR LOAD AND GENERATION

Theodoros Zerlentis Kafarakis

Abstract

In this thesis multiple prediction simulations were performed, including load, wind, and solar generation forecasts, in order to assess the efficiency of various feature selection techniques while comparing their significance. The prediction model was developed using a multi-layered perceptron, which had one hidden layer made up of 24 neurons. Additionally, the prediction model was trained and tested using three years' worth of historical load, generation, and weather data from the years 2019, 2020, and 2021 based on Greece's energy grid. Three different locations were chosen, as the source of information for the weather parameters, based on the load and generation spatial distribution of Greece.

The impact of a feature selection method was thoroughly examined, as multiple evaluation error metrics and variables were calculated and showed that the accuracy and the overall prediction time of the load and generation forecast was greatly enhanced. Additionally, for all three different prediction models, a comparison study was perform in order to identify the optimal amount of feature for each feature selection method. Furthermore, the results showed that the RReliefF algorithm was able to achieve the best performance over the majority of the other methods, mainly for the load and solar generation. The algorithm was able to better predict the variations of the load curve and the high volatility of solar generation, with less available features. Additionally, the MRMR was also very effective as it was able to properly forecast the wind generation, with the smallest computational time while minimizing the forecast errors. The RReliefF algorithm also showed similar performance for the wind generation forecast, but it was outperformed by the MRMR.

Μεταπτυχιακή Διπλωματική Εργασία

ΜΕΘΟΔΟΙ ΕΠΙΛΟΓΗΣ ΕΙΣΟΔΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΚΑΤΑΝΑΛΩΣΗΣ

Θεόδωρος Ζερλέντης Καφαράκης

Περίληψη

Στην παρούσα διπλωματική πραγματοποιήθηκαν προσομοιώσεις πρόβλεψης φορτίου, αιολικής και ηλιακής παραγωγής, προκειμένου να αξιολογηθεί η αποτελεσματικότητα των διαφόρων τεχνικών επιλογής χαρακτηριστικών. Το μοντέλο πρόβλεψης αναπτύχθηκε με τη χρήση ενός πολυεπίπεδου perceptron, το οποίο είχε ένα κρυφό στρώμα αποτελούμενο από 24 νευρώνες. Επιπλέον, το μοντέλο πρόβλεψης εκπαιδεύτηκε και δοκιμάστηκε χρησιμοποιώντας ιστορικά δεδομένα φορτίου, παραγωγής και καιρικών παραμέτρων τριών ετών από τα έτη 2019, 2020 και 2021 με βάση το ενεργειακό δίκτυο της Ελλάδας. Επιλέχθηκαν τρεις διαφορετικές τοποθεσίες, ως πηγή πληροφοριών για τις καιρικές παραμέτρους, με βάση τη χωρική κατανομή του φορτίου και της παραγωγής στην Ελλάδα.

Ο αντίκτυπος των μεθόδων επιλογής χαρακτηριστικών εξετάστηκε διεξοδικά, καθώς υπολογίστηκαν πολλαπλά στατιστικά σφάλματος, τα οποία έδειξαν ότι η ακρίβεια και ο συνολικός χρόνος πρόβλεψης του φορτίου και της παραγωγής βελτιώθηκε σημαντικά. Επιπλέον, και για τα τρία διαφορετικά μοντέλα πρόβλεψης, διεξήχθη συγκριτική μελέτη προκειμένου να προσδιοριστεί η βέλτιστη ποσότητα χαρακτηριστικών για κάθε μέθοδο. Επιπλέον, τα αποτελέσματα έδειξαν ότι ο αλγόριθμος RReliefF ήταν σε θέση να επιτύχει την καλύτερη απόδοση έναντι της πλειοψηφίας των άλλων μεθόδων, κυρίως για την πρόβλεψη φορτίου και την ηλιακή παραγωγή. Ο αλγόριθμος ήταν σε θέση να προβλέψει καλύτερα τις διακυμάνσεις της καμπύλης φορτίου και την υψηλή μεταβλητότητα της ηλιακής παραγωγής, με λιγότερα διαθέσιμα χαρακτηριστικά. Επιπλέον, ο MRMR αλγόριθμος ήταν επίσης πολύ αποτελεσματικός, καθώς ήταν σε θέση να προβλέψει σωστά την αιολική παραγωγή, με τον μικρότερο υπολογιστικό χρόνο και σφάλμα πρόβλεψης. Ο αλγόριθμος RReliefF παρουσίασε επίσης παρόμοια απόδοση για την πρόβλεψη της αιολικής παραγωγής, ωστόσο ο MRMR ήταν πιο αποτελεσματικός.

Table of contents

List of figures

List of tables

Abbreviations

ANN	Artificial Neural Networks
FSM	Feature Selection Method
LSTM	Long Short Term Memory
MI	Mutual Information
MIC	Maximum Information Coefficient
MRMR	Maximum Relevancy, Minimum Redundancy
MRMRMS	Maximum Relevancy, Minimum Redundancy, Maximum Synergy
RES	Renewable Energy Source
RF	Random Forest
STLF	Short Term Load Forecast

Chapter 1

Introduction

The escalating growth of energy consumption combined with the ever lasting climate change threat that the modern society is faced with, has pushed countries and big corporate organizations to optimize the energy grid distribution and consumption. With the introduction of the Renewable Energy Sources (RES), an extra factor for contribution but at the same time of instability has been introduced to the energy grid. This is the main reason that energy companies strives to understand and estimate the grid demands but also the energy generation from RES.

The prediction of the energy for load and generation is a crucial aspect of the energy industry, affecting not only the energy grid but also the consumers. Multiple forecast algorithms have been developed in the years by using advance mathematics methods and machine learning algorithms. The later one has been growing exponential the past few year with the fast growth of computer capacity and computer algorithms. The accuracy of these predictions depends on the quality, quantity and availability of the related data. With the advancement of technology, vast amounts of data and parameters are available to be utilized for energy load and generation prediction, allowing the scientist to create even more advanced and accurate forecast models. However, some of these parameters might not be relevant to the desired prediction, leading to more complex and inaccurate models.

Feature selection methods have been proposed as a viable solution to this problem. The purpose of these a algorithms is to identify the most relevant and significant features, that have the most impact on the prediction accuracy while in the same time outlining the redundant features that are not needed or lead to inaccuracies.

1.1 Thesis Subject

The objective of this thesis is to review and evaluate the performance of various feature selection methods for energy load, wind and solar generation forecasting by using an artificial neural network. The feature selection methods chosen, after an extensive literature review, are the following:

- RReliefF
- Random Forest
- Mutual Information
- MRMR
- Maximal Information Coefficient

The performance for each method will be measured using a variety of error evaluation metrics such as the Absolute Error (AE), the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). Each method is compared with the original model, which does not includes any feature selection method, in matters of accuracy and time performance. Furthermore, a comparison between each method is performed for each prediction model in order to locate the best feature selection algorithm. Finally, an evaluation based on the amount of features provided by the feature selection method is presented. The results of this study will provide insights into the effectiveness of different feature selection methods and their impact on the prediction accuracy of energy load and generation.

1.2 Thesis Structure

The Diploma Thesis is structured as follows:

In Chapter 2 the literature review for various Feature Selection Methods is presented. During this chapter the state of the art Feature Selection algorithms are analyzed based on their advantages and disadvantages, which are presented by the authors. In Chapter 3 the main aspects of this thesis are presented, starting with describing the algorithms that were created for forecasting the load, wind and solar generation models. Furthermore, in the same Chapter a brief presentation for each Feature Selection Method parameters are discuss. In

Chapter 4 the results from the simulations are presented and compared. Multiple evaluation error metrics are considered in order to compare the performance of each Feature Selection methods. Finally, in Chapter 5 the conclusion of this thesis is presented.

Chapter 2

Literature review

2.1 Introduction

In this chapter a detailed overview of the Artificial Neural Network (ANN) principles and the state of the art Feature Selection Methods (FSM) are presented. Concerning the Artificial Neural Networks, their general structure is briefly discussed combined with the purpose of the activation function, the necessity for data normalization and the different evaluation metrics that are used for comparing the performance of an ANN. Furthermore, the importance of Feature Selection Methods is presented, accompanied with a variety of methods used from the scientific community for energy forecasting.

2.2 Short Term Load and Generation Forecast

Every year the energy needs of society are increasing due to technological developments. In order to cope with these demands, humanity has to improve and develop the existing power grid by introducing alternative sources of energy. Until now, this problem has been solved by using fossil fuels such as coal, natural gas and oil. However, global warming and the environmental pollution that has occurred in recent years, combined with the deterministic amount of fossil fuels available around the globe, has prompted most countries to adopt a more sustainable policy. Some countries are turning to other alternative sources, such as nuclear energy [1], which, despite its advantages, is facing problems of widespread adoption. Renewable energy sources (RES) seems to have provided the answer to this issue, as in recent years there has been a huge interest from countries and major investors in these technologies.

With the increasing penetration of Renewable energy sources in the energy market, many analysts are turning their attention to the implications and difficulties that these technologies present to the grid's stability. In particular, wind power is highly volatile in terms of the amount of electricity produced over a relatively long period of time, as it is entirely dependent on the prevailing weather conditions, since a day without high wind speeds, significantly reduces energy production. These volatile patterns in electricity generation can have a huge impact on the electricity grid if not managed properly. Furthermore, the generation of energy from fossil fuels faces similar issue concerning the amount of energy production. Depending on the grids demands and performance the energy companies are either producing more energy than the grid requires or less which can lead to black outs. This issue mostly arises due to the slow time to adjust of the energy generation facilities.

2.2.1 Benefits of Energy Forecast

As it was discussed previously, the energy load and generation technologies are facing different difficulties concerning the amount of energy that is provided to the grid. A relatively simple approach was considered by the scientific community, by introducing the idea of the load and generation forecast [2]. By analyzing historical and forecasted data, based on various analytical and statistical methods, the scientist were able to identify patterns in the energy load of the grid while at the same time estimate within a good accuracy level the amount of energy generated by renewable source of energies.

A big interest of the scientific community has moved towards technologies for forecasting energy related topics. It can be seen in Figure 2.1 that in the recent year the amount of publication in the field of energy forecasting has almost doubled in the last five years. This can be surely traced due to the increased amount of data that are available as more and more smart devices with sensors and networking capabilities are introduced to the energy grid. This trend can also be linked to the economic and environmental factors that the energy forecasting technologies provide. By using such technologies the energy grid's demand and supply can be optimized, by utilizing at the out most the renewable sources of energy, while at the same time producing the minimum required amount of energy for big generation facilities, which leads to less energy waste, less environmental foot print and more profits [3].

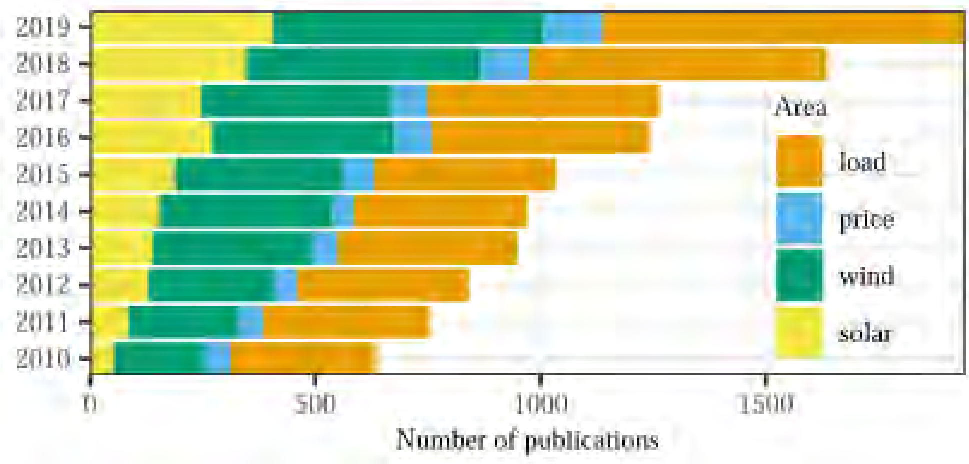


Figure 2.1: Amount of publications for different types of energy forecasts [2].

2.2.2 Forecast Methods Comparison

The utilization and favor of different load forecast methodologies varies in the literature, as various method have their own pros and cons. In the recent years, a big emphasis and research has been given to Artificial Neural Networks, which are increasingly used for machine learning and load forecast applications. In addition, another well known method is the Support Vectors Machines, which have been used for many years in a variety of applications. On the other hand, Random Forests and k-Nearest Neighbors are showing less growth in the load forecast and generation sector, but in general application are still popular. Figure 2.2 shows the use graph of these methods over the years.

One of the main factors that affect the performance and robustness of machine learning algorithms, in the field of energy load and generation forecast, is the number and type of features characteristics derived from meteorological data. More specifically, the algorithms Random Forests, k-Nearest Neighbors and Support Vector Machines show fairly good accuracy for a big amount of features [4, 5, 6], thus enabling the reduction of noise from any erroneous or irrelevant data that might be preset in the data set and cause performance degradation. On the other hand, Artificial Neural Networks face major problems of over-adaptation and instability when the data set is quite extensive. However, this problem can be addressed by pre-processing the data [7], by data normalization and feature selection methods, thus reducing the discrepancies. Moreover, the advantage of the ANN, after proper tuning, is located in the superior accuracy of the results for non-linear values, which makes them one of the most attractive and desirable methods of energy forecast applications.

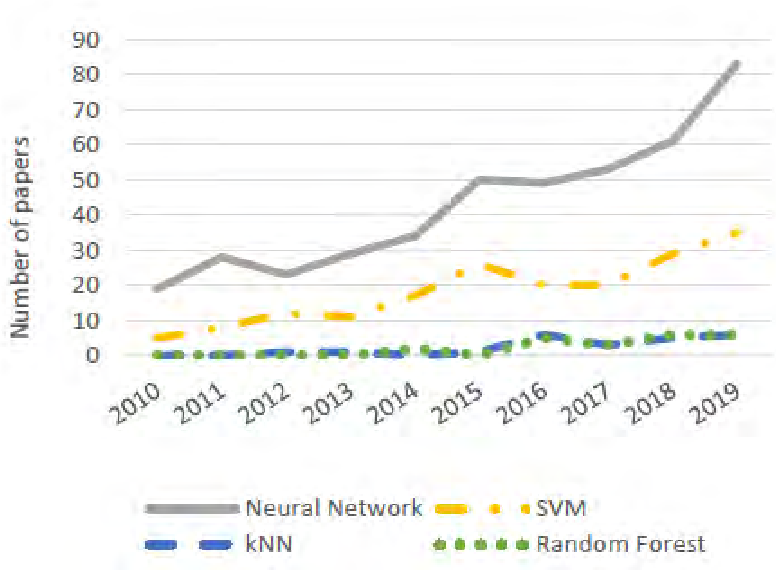


Figure 2.2: Different load forecast methods [8].

2.3 Artificial Neural Networks

A computational model known as an artificial neural network (ANN) is modeled after the structure and operation of biological neural networks, such as the human brain. It is a machine learning algorithm made to spot patterns in data, learn from it, and then predict or decide. An artificial neural network's basic building blocks are interconnected nodes, also known as artificial neurons or units, arranged in layers. Each neuron receives input, processes the information, and then generates an output. The connections between neurons are weighted, which means that each connection's strength or significance is denoted by a weight, a numerical value. In Figure 2.3 an artificial's neural network structure is presented.

Depending on the architecture used and the problem being solved, an artificial neural network's structure may change. The feedforward neural network, which has an input layer, one or more hidden layers, and an output layer, is a popular architecture. Information moves in a straight line, without loops or cycles, from the input layer via the hidden layers to the output layer in a feedforward network.

An artificial neural network learns to modify the weights of its connections based on a supplied dataset throughout the training process. The difference between the network's projected output and the desired output is utilized to update the weights in a process known as backpropagation to make this adjustment. The network continuously increases its capacity to generate precise predictions or classifications by iteratively updating the weights. The

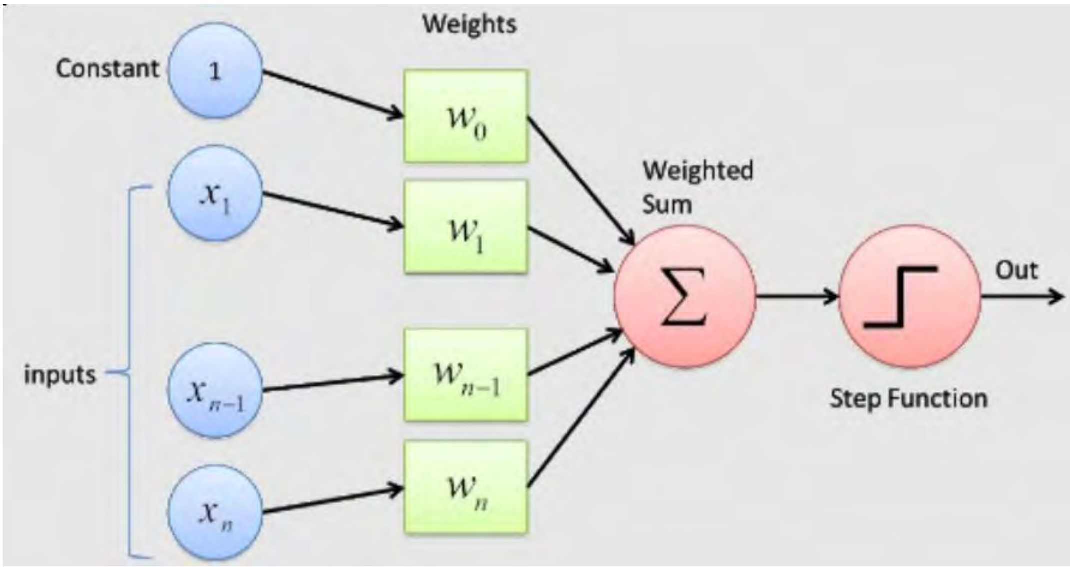


Figure 2.3: Artificial Neural Network [9].

artificial neural networks have excelled in a number of fields, including audio and picture recognition, recommendation systems, natural language processing, and many other challenging tasks. They have the capacity to learn from massive data sets, identify important features, generalize patterns, and make predictions on fresh, unforeseen data.

It's vital to remember that there are various kinds of artificial neural networks, including the Generative Adversarial Networks (GANs) for creating fresh data samples, the Recurrent Neural Networks (RNNs) for processing sequence data, the Convolutional Neural Networks (CNNs) for processing images and many more. The artificial neural networks are a versatile and effective tool in the field of machine learning and artificial intelligence since each form of network is designed to tackle particular problems and tasks.

2.3.1 Structure

Depending on the particular architecture and problem being solved, an ANN's structure might change, but there are some basic elements and ideas that remain constant:

- The Input Layer is the neural network's first layer which represents the characteristics or properties of each neuron. The dimensionality of the input data affects how many neurons are present in the input layer.

- The Hidden Layers are the layers between the input and output. The network can learn and represent complicated patterns and relationships in the data thanks to the hidden layers. Depending on the complexity of the problem and the network's design, the number of hidden layers and the number of neurons in each hidden layer may change.

- The Output Layer is the neural network's final layer and is responsible for producing the output or prediction. The type of problem being solved determines how many neurons are present in the output layer. For instance, in a binary classification task, a single neuron may represent the likelihood of falling into one class, whereas in a multi-class classification problem, numerous neurons may represent the likelihoods of falling into each class.

- Weights and Connections: Through connections or edges, neurons in one layer are linked to neurons in the next layer. Each connection has a weight attached to it that controls how strong the connection is. The network can adapt and modify its behavior based on the input data since the weights are learned throughout the training process.

- The Activation function represent the generated based on weighted sums of it's input. The network can learn intricate associations thanks to the activation function's introduction of non-linearity. The activation processes sigmoid, tanh, ReLU, and Softmax are frequently used.

- The Bias represents an extra variable for each neuron that can be changed during training. The bias enables the network to change the activation function, which has an impact on the network's general behavior and adaptability.

- The Feedforward and Backpropagation are two different type of process or flow of infomation inside the neural network. In the feedforward secnarion the information feeds forward from the input layer via the hidden layers to the output layer. During the backpropagation process the network updates the weights depending on the discrepancy between the projected output and the desired output during training. The error is reduced and the network's performance is enhanced by this iterative procedure.

Depending on the issue and the desired outcome, an artificial neural network's structure and architecture might differ greatly. To handle particular types of data and tasks, many architectures, including feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), have been developed.

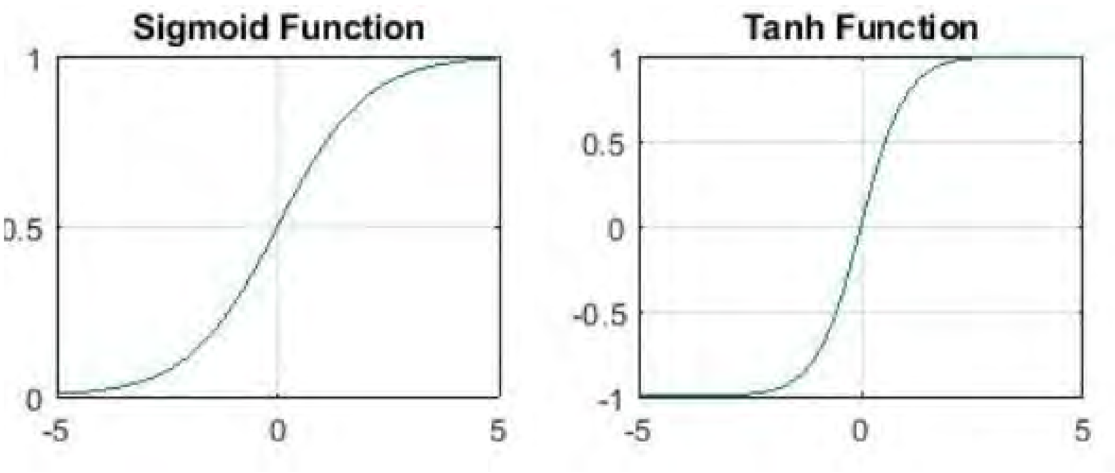
2.3.2 Activation Function

An activation function in a neural network is a mathematical function that establishes a neuron's or node's output depending on the weighted sum of its inputs. The network gains non-linearity from the activation function, which enables it to learn and simulate intricate interactions between inputs and outputs. The activation function generates an output value or activation value based on the weighted sum of the inputs to a neuron, also known as the activation. The neural network's following layer receives this output as input [10].

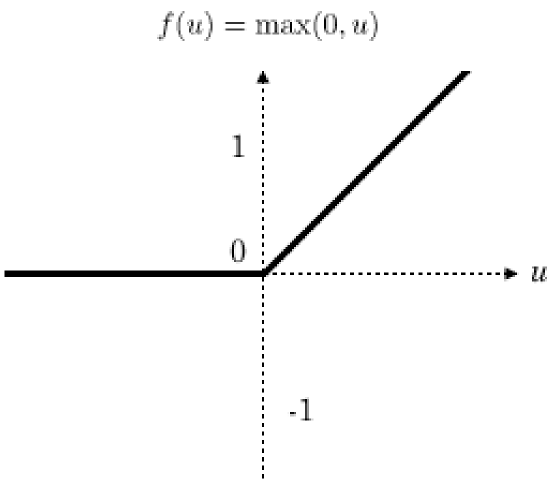
An activation function's primary objective is to create non-linearities into the neural network. A neural network would only be able to convert the input data linearly if no activation function is presented. Furthermore, it can learn and represent more complex interactions between the inputs and outputs by employing non-linear activation functions. Given enough neurons and the right parameters, it enables the network to approximate any continuous function. Some commonly used activation functions include:

- The Sigmoid function converts the input into a probability represented by a number between 0 and 1. It improves output quality and helps with binary classification issues.
- Similar to the sigmoid function, the hyperbolic tangent (tanh) function transforms the input into a number between -1 and 1. Problems with classification can also benefit from it.
- The Rectified Linear Unit (ReLU) maintains the positive values while setting all negative values to zero. It is frequently used in deep learning because it makes big neural networks' training more effective.
- Leaky ReLU is a variant of ReLU that permits a negligible, non-zero gradient for the input. This lessens the ”dying ReLU” issue, in which neurons may become imprisoned in a state of dormancy.
- The Softmax function is frequently employed to solve multi-class classification issues. The outputs are transformed into a probability distribution, where each output represents the likelihood that the input belongs to a particular class.

The type of the issue and the network architecture determine the activation function to use. In Figure 2.4 the three most common activation functions are presented. Different activation functions have various characteristics that can impact the network's generalization,



(a) Sigmoid and Tanh activation functions [11].



(b) ReLu activation function [12].

Figure 2.4: Activation function of a neural network.

convergence rate, and training dynamics. To ensure that the neural network performs at its peak, it is crucial to choose an activation function that is appropriate for the task at hand.

2.3.3 Data Normalization

Machine learning algorithms are data driven processes, meaning that the performance and the outcome of the end product is solely data dependent. Although nowadays there is an abundant number of data to use in such application, sometimes missing or extreme values of theses data can cause huge performance and accuracy degradation. This issue is even more serious for short term load forecast, as noise in data or outliners or even missing values are quite common in the data acquisition processes which then feeds these values to the forecast model. Another common issue to the forecast models is the existence of non-numeric features, such as the day of the week or special days of the year etc, that can only be translated to certain values, most of the time either 0 or 1. This can lead to worse performance of the forecast model as the weights of each neuron cannot predict and produce a good outcome to values that are very high, like the load of the energy grid, and the same time very low values, like the non-numeric features. In order to solve this issue researchers developed several ways to normalize the data, so that all features have a specific range of values that is proportional to their original values. The most common data normalization method is the min-max, which takes into account the maximum and minimum value of each feature and then translate each value to a proportion that is equivalent to the difference of the maximum and minimum value [13]. The following equation describes the min-max method:

$$x^*_{ij} = \frac{x_{ij} - min_i}{max_i - min_i} \tag{2.1}$$

where i is the feature, j is the data line, x^* is the normalized data, x is the data to be normalized, max_i and min_i is the maximum and minimum value of the feature i, respectively.

2.3.4 Evaluation Metrics

One of the most important steps for creating a good machine learning algorithm, is the ability to evaluate the outcome of the model, meaning that whether or not the end result is worthy. In order to evaluate these algorithms multiple metric have been developed, with almost everyone of them being based around the performance error. For the neural networks and the load and generation forecast the error can be easily defined, by simply subtracting the actual value that we want to predict with the one that the neural network calculated. Although this is the main idea of the error metric, multiple variations of this measurement have been introduced, like the Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE) and the Mean Absolute Error (MAE) [14, 15]. In the following, the three main evaluation metrics equations that will be used in this thesis are presented.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{2.2}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{x_i} \tag{2.3}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{2.4}$$

where n is the number of predicted loads, x_i and y_i is the actual load and the predicted load, respectively.

2.4 Feature Selection Methods

As more and more data and features are available during load and generation forecasting, it becomes more urgent to filter those redundant or unworthy features. This is issue can be solved by utilizing a big variety of feature selection methods. More importantly, FSM can locate and isolate the best features for the forecast algorithm in order to improve both overall performance and accuracy of the model, reduce overfitting and reduce computational time and power [16]. Multiple methods have been developed and proposed for short-term load forecasting problems, with each method having advantages and disadvantages over the other ones. In this section multiple Feature Selection Methods are discuss by comparing the finding from the literature.

2.4.1 Mutual Information

Mutual Information is based on the theory of information between two instances. When two instances are considered independent then Mutual information is equal to zero. If the two instances are dependent with each other then Mutual Information has a positive value. As a feature selection method, Mutual Information evaluates how much information each individual feature provide to the target value. Mutual Information has been used extensively as a feature selection method as it is able to capture both linear and non-linear correlation between instances [17]. In Figure 2.5 a generic representation of Mutual Information is shown.

In the bibliography a thorough investigation have been done around the performance of Mutual Information as a feature selection for short term load forecasts. In [18] the researchers used 1 year of load data to evaluate the performance of Mutual Information, by splitting the data set into three different time periods, one for the summer season, one for winter and one for the rain season. For each season they used the Mutual Information algorithm to find out the best features out of a pool of twenty features in total. Those twenty features were consisting out of four different types, as energy load data, load difference, temperature and

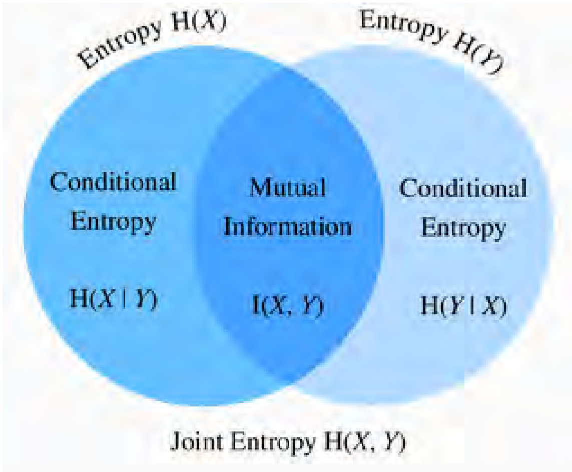


Figure 2.5: Mutual Information [19].

irradiance data, of the previous five days from STLF targeted day. Afterwards, they created six different scenarios based on the number and type of features. The first scenario included all twenty data, while the second and third one had only fifteen features, with the difference being between those two that the third scenario did not include temperature data. The fourth scenario had ten feature while the fifth had fifteen feature without load difference data. Finally, the sixth scenario did not include load data variables. After evaluation of the final results the researcher find out that by using MI for feature selection improves significantly the accuracy of the forecast. In more details, the second scenario showed the best results minimizing the error almost up to half from the forecast without the use of MI. Furthermore, out of three seasons the winter season data set showed slightly better results, while in general the overall forecasts showed small deviations between seasons. This resulted in a more stable and robust forecast that is capable of anticipating season's changes.

In [20] the authors used a two stage feature selection algorithm based on the Mutual Information theory. The first stage calculates the relevancy of the feature by using a modified version of the Relief algorithm, that offers fast and robust performance, and assigns a relevance weight to each feature. Furthermore, based on that weight, the best more relevant feature are selected in order to create a new subset of features. Afterwards, the second stage calculates the redundancy of the new subset of features by utilizing the Interaction Gain criterion, which compares the mutual information of two variables and a target value, with the mutual information of each individual variable and the target. Based on that criterion, multiple redundant features can be identified and eliminated from the final prediction. For the case studies the researchers compared the proposed feature selection method with multiple algorithm like, the principal component analysis, correlation analysis, mutual information, two stage correlation analysis, Numerical Sensitivity Analysis etc. The results showed that the proposed two stage algorithm can achieve better results and accuracy than the rest of the feature selection methods.

In another research, [21] the authors used a Conditional Mutual Information algorithm as a feature selection method for a short term load forecast model using hybrid neural networks. During their approach they used the measurement of entropy, which is based on the information theory, in order to measure the uncertainty of a feature V and a variable C. In the beginning, they calculated the Mutual Information of all the features based on the target values and ranked theses features. The top best performing features then are selected and moved to a new subset V_{new}, with the one feature to maximize the mutual information to be moved to another subset new S. Afterwards, the conditional mutual information for all pairs of the subset V_new and S is calculated in correspondence to the target values of C. By comparing the conditional mutual information values of all the pairs, the authors were able to select the next best feature that is actually providing accuracy to the model based on the already existing top features. By performing the above algorithm multiple times, a new more refined subset of the original features is identified. In order to evaluate the performance of the algorithm the researchers compared the proposed method with algorithms based on Correlation analysis, Mutual information and the RReliefF. The results showed that for the short term load forecasting the Conditional Mutual information was superior from the other methods, by almost having the half Mean Absolute Percentage Error (MAPE).

2.4.2 Maximum Relevancy Minimum Redundancy

Maximum Relevancy and Minimum Redundancy, also know as (MRMR), is a well known feature selection method in machine learning, similar to Mutual Information. The idea behind this method is to calculate and find the most relevant features while at the same time trying to minimize the redundancy of features. The main concept of this method is based on the information theory, by comparing the mutual information of each features. Specifically, the relevant feature achieves high values of mutual information based on the target value, while the redundant features have high value of mutual information based on other features [22]. The MRMR method is well used in short term load forecast applications as a feature selection method, as it is able to find the best feature for the forecast while minimizing the

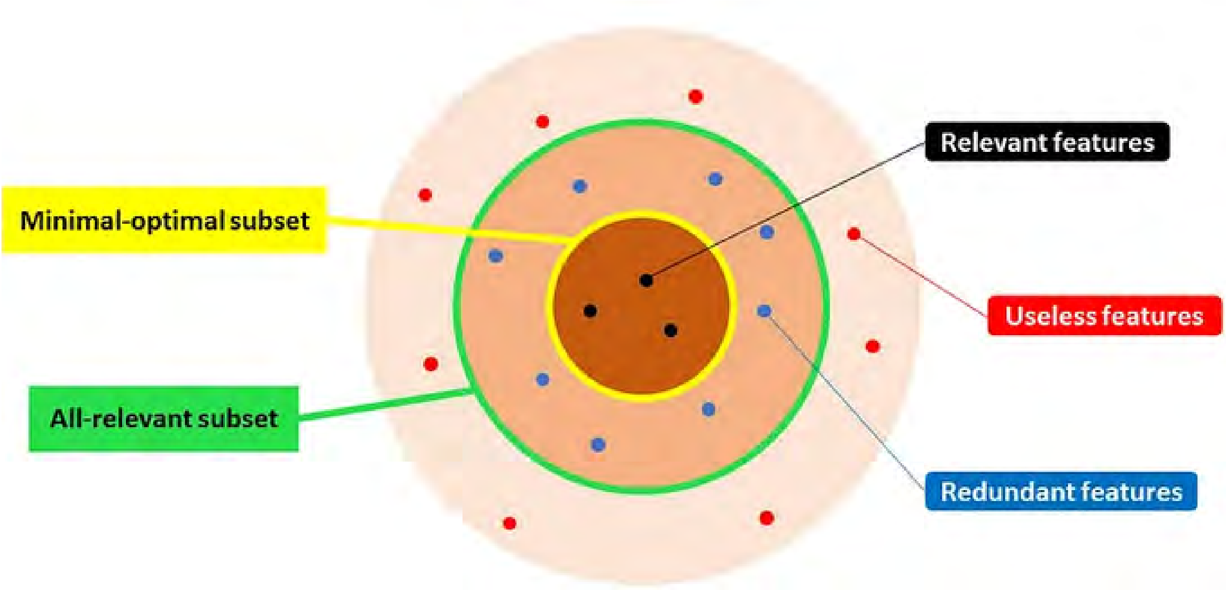


Figure 2.6: MRMR representation.

redundant feature, leading to better accuracy and more robust forecast algorithms. Figure 2.6 shows a representation of the MRMR.

An improvement to the classic MRMR algorithm was utilized in [23], which not only takes into account the relevancy and the redundancy of each feature, but also involves the synergy between two or more features. Specifically, the synergy criterion helps to identify features that may improve the accuracy of the forecast when they are combined with other features. While the redundancy or relevance filter might show that these feature are not the best our of the feature pool, when combine with others the overall outcome might be better. This method gives an extra step of selecting the best features that in most cases of other algorithms those features will be thrown out of the top feature set. The algorithm is called Maximum Relevancy, Minimum Redundancy, Maximum Synergy (MRMRMS). In this paper the authors introduced a three stage filter algorithm as the feature selection method. In the first filter the relevancy of each feature is calculated based on the target values of the data set. Furthermore, the synergy of each feature with another is calculated in the first stage, in order to combine both values of synergy and relevancy, of each feature, in order to rank them. In the second filter, the output from the first filter is used, which involves the best ranked features. During the second stage the redundancy of each feature is calculated in order to identify which feature affects the forecast similar to another. Furthermore, for each feature all three criteria, relevancy, redundancy and synergy, are combined and compared with a threshold value, which decides whether the feature is good or not. All the worthy features are transferred to a separate subset S, while the unworthy ones are added to the subset $-S$. The threshold value is defined by the user, and in the case of the paper it was fine tuned after multiple simulations. Finally, in the third filter, the two worthy and unworthy subsets S and $-S$ are compared once more based on the synergy between the two subset's features. During this stage, any remaining worth synergy between two features is considered and decided whether or not those will be transferred to the final set of the selected features. In order to evaluate the performance of this algorithm, the authors compared multiple evaluation error metrics with other algorithms like correlation analysis, mutual information, numerical sensitivity analysis, maximum relevancy etc. The results showed that the proposed algorithm achieved better accuracy with an average computational time than the rest of the methods.

A hybrid feature selection approach was used in [24] by combining the Maximum Information Coefficient (MIC), as an early stage of non-numeric features selection, and the Maximum Relevancy, Minimum Redundancy algorithm for the rest of the features. Specifically, in the first stage, the MIC is used to calculate the nonlinear dependencies between non numeric features and the target values. Those non numeric features are composed of workdays, Saturday, Sunday and seven different holidays. By utilizing the MIC method, the authors were able to identify whether or not those features improved the overall accuracy of the STLF or introduced fluctuations and inaccuracy to the forecast model. Furthermore, after the first stage is completed, the worthy features were combined with the rest of the numeric one, in order to be evaluated by the MRMR method. During that stage, for each feature the relevancy and the redundancy is calculated in order to be ranked from the most worthy features to the least one. The authors used an Improved Long-Short Term Memory network and a classical Long-Short Term Memory network with and without the hybrid feature selection method, in order to evaluate it's performance. They additionally compared the results with some well know algorithms like the Support Vector Regression and the Gated Recurrent Unit methods. The results showed that for each algorithm that used the hybrid feature selection method, the accuracy of the end results was greatly improved. The only drawback that the researchers noticed, was the additional computational time that the hybrid method introduced, which in some cases was more than eight times.

2.4.3 Random Forest

Random Forest (RF) is an ensemble learning method most commonly used for regression and classification applications. The Random Forest algorithm calculates the importance of each feature based on the model's predictive performance. One of the main reasons that this method is popular as a feature selection method in energy load and generation forecast is the ability to handle multi-dimensional data-sets with relevant easy and the ability to achieve good accuracy even with nonlinear features [25]. In Figure 2.7 the representation of the Random Forest algorithm is shown.

In the [26], the researchers utilized the Random Forest algorithm as a feature selection method combined with a convolutional neural network to perform a short-term load forecast. During their research, they used two different data sets for their short term load forecast, one based on hourly historical data of 1 year from New Zealand and one based on half hourly historical data of 1.5 years from China. The researchers used the Random Forest algorithm for both data sets in order to identify the best features. The results showed similarities in the ranking of features, with the top ones being the temperature, temperature variation of previous days and humidity. In order to evaluate the overall performance of the algorithm the authors compared the results from both data sets with four other algorithms. The results showed that in both cases of different data set the proposed algorithm with the use of Random Forests as a feature selection method performed better than the rest.

The ability of Random Forest (RF) Algorithms to identify and separate the important data, that greatly affect the accuracy of the prediction, combined with the robustness they offer, have prompted researchers to combine multiple AI algorithms in order to improve the accuracy of the results. In the study conducted by Wenting et al. [27], Random Forests were utilized to pre-process the data in order to identify and isolate the important features that affects the prediction. Then, by combining a new type of Convolutional Neural Networks called Temporal Convolutional Networks, the researchers were able to predict the electrical output of a wind farm, for a short period of time, with very good accuracy. Furthermore, the researchers compared the prediction model with and without the Random Forest algorithm for feature selection. They also used the Random Forest algorithm combined with two other prediction models, one based on the Long Short Term Memory (LSTM) model, and the other one based on the variant of the classical LSTM, which is called Gated Recurrent network. Firstly, the comparison between the models with and without the Random Forest algorithm,

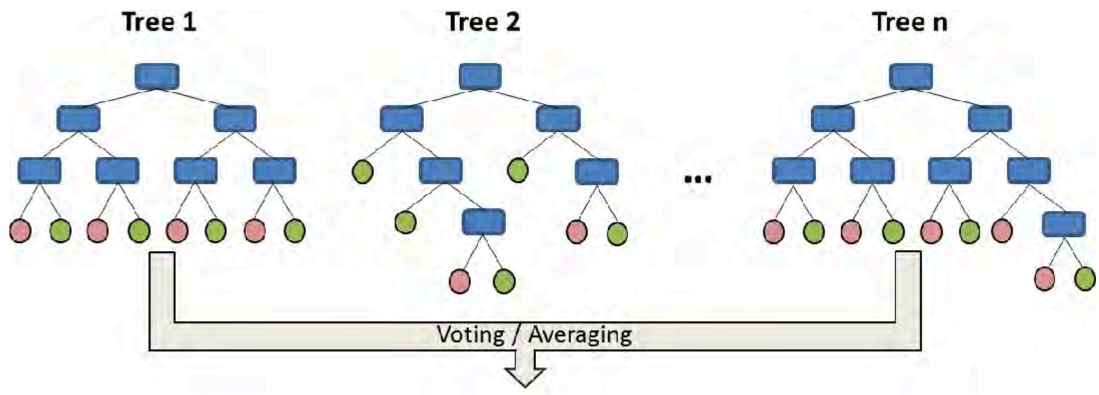


Figure 2.7: Random Forest representation [28].

showed that the overall performance of the model with the RF, was greatly improved almost up to 25% from the one without feature selection. The researchers noticed also a huge improvement in the total computation time of the model, which is an additional benefit of Random Forest algorithm. Finally, the comparison between the different prediction models showed that all three algorithms were benefited by the introduction of the feature selection method of Random Forests, that allowed better overall performance.

2.4.4 Maximal Information Coefficient

In 2011 Reshef et al. [29] introduced the Maximal Information Coefficient (MIC) as a measure of dependence between two different instances. In general, MIC is a statistical measure, based on the concept of information theory, that is capable of calculating the linearity and non-linearity association of two variables. By measuring the amount of information shared between two instances, this method is able to rank all features and decide which is best suitable for the application. Some of the many advantages of this method is the ability to maintain good performance even if outliners exists in the data set, and also handle mixed data types like numeric, continuous or non continuous values. This method although faces some issues with big data sets and high dimensional data sets, which can lead to performance degradation.

Tang et al. [15] utilized the Maximal Information Coefficient algorithm as a feature selection methods combined with a Temporal Convolutional Network as the predictors and the Fuzzy c-means as a clustering method. The authors used six years of data in total from the public grid in order to perform a short term load forecast. For this data set, they used the MIC to identify the relevancy between features and the target load values and also outline redundant features. The feature selection method indicated that there is a strong relevancy between the load and the maximum, minimum and average temperatures from the data sets. In order to evaluate the overall performance of the algorithm, the authors compared the results with other prediction models like the Support Vector Regression, Temporal Convolutional Networks, Least Square Support Vector Machine etc. The results showed that the proposed algorithm showed superior performance from the other ones, by reducing the overall error of the prediction.

In [30], the authors used a Long Short Term Memory network as a predictor for a short term load forecast, combined with the Maximal Information Coefficient method for feature selection. Specifically, the MIC algorithm was able to correlate the importance of temperature, relative humidity and wind speed variables as the most relevant features. Less importance showed other meteorological data, like rainfall, atmospheric pressure and solar radiation intensity. By filtering the numerous features that were present in the data set, the authors were able to improve the overall accuracy and performance of the prediction model.

A hybrid approach was considered by Yao et al. [31], by using the Maximal Information Coefficient method as an early stage of feature selection and ranking. Afterwards, the top selected features were fed to a second algorithm , the so called LightGBM, in order to further refine the features by creating a correlation between the features and the targeted values. Those features were then fed to a regression analysis model, based on LightGBM and XGboost algorithms, to perform predictions. The process was repeated until certain termination thresholds were met. In the end, the feature set with the best accuracy is selected as the end result. The authors in order to correctly evaluate the performance of the proposed algorithm, calculated a variety of evaluation errors based on models that utilized only the Maximal Information Coefficient, models with only XGboost or LightGBM for feature selection and a model having only the predictor, without any further data correlation and analysis. Furthermore, they utilized six different predictors based on the LightGBM, XBboost, Support Vector Machines, Random Forest, the ARIMA model and a Back Propagation Neural network. The results from the simulations showed that the standalone model with the Maximal Information Coefficient, for all the different predictors, achieved the second best results. The proposed algorithm achieved the best results for all the prediction models, minimizing the error and at the same time achieving small computational time.

2.4.5 RReliefF

In 1992 the first Relief algorithm was introduced by utilizing the instance based learning for classification problems with only two classes [32]. The algorithm is able to distinguish values between near instances by estimating the quality of their attributes. When it was introduced the algorithm faced problems with data quality, due to either noise or missing values in the data sets. An extension of the original Relief Algorithm was developed in 1994 by [33] and it was called ReliefF. This extended version of the original algorithm was not limited to only two classes for classification, it was more robust and was able to cope with bad quality data. Finally, the further improved version of the original, the RReliefF algorithm, was developed in 1997 as a regression problem solving method [34]. This version selects randomly instances from the data set and then utilizes the K-nearest neighbors algorithm in order to rank each individual feature and add a weight value to it. This extension further enhanced the robustness and performance of the algorithm.

In [35] the authors proposed a new method for Short Term Load Forecast (STLF) by utilizing a combination of the Innovative Features method and the RReliefF as a feature selection algorithm. Firstly, the Innovative Features gather a pool of features that represent the dynamic and non-linear attributes of the grid in order for the RReliefF algorithm to select the best feature for forecasting. The authors compared this combined method with other STLF algorithms, like the Support Vector Regression method or the Levenberg-Marquardt neural network algorithm etc, and the results showed that the proposed method had the best performance.

An extensive comparison of different feature selection methods was performed in [36]. In this case study, two year of electricity load data were used to compare three different prediction models, a Neural Network, a Linear Regression and a Model Tree Rule algorithm. For all three prediction methods, four different feature selection algorithms were used. More specifically, the authors used the Mutual Information, the RReliefF, the Autocorrelation and the Correlation-Based Selection methods for feature selection. The results showed that for the Neural Networks and the Linear Regression model the RReliefF achieved the best results but with the highest computing time. For the Model Tree Rule algorithm the RReliefF achieved the worst results out of the other three feature selection methods. However, comparing the overall performance for the load forecast with the prediction model, the combination of Neural Networks and the RReliefF as the feature selection method, achieved the best accuracy.

2.5 Summary

In this chapter a brief literature review was presented for the energy forecast and the machine learning algorithms. Starting on, the importance of the short term load and generation forecast was presented, by explaining the benefits but also the current trends of forecast methods. Moreover, a brief explanation of the artificial neural network structure and parameters are presented, with a reference also to the state of the art algorithms that are available in the literature for performing energy load and generation forecasts. Finally, five different feature selection methods were discussed based on the available literature review. The Mutual Information, the Random Forest, the RReliefF, the Maximal Information Coefficient and the MRMR were the five discussed algorithms.

In the next chapter, the three prediction models for load, wind and solar generation are presented. Specifically, the parameters selected as the overall feature pool will be presented and explained. Furthermore, a detailed overview of the artificial neural network will be presented, elaborating with the amount of hidden layers, the amount of neurons and the type of activation function used.

Chapter 3

Load and Generation Forecast Models

3.1 Introduction

This chapter presents the different short term load and generation forecast models that will be used in order to compare all the different feature selection methods. Specifically, a detailed overview is presented for the load forecast model, by analyzing all the available features and the parameters for the artificial neural network. Afterwards, a similar approach will be presented for the wind and solar generation algorithm that will be used for the evaluation of the variety of feature selection methods.

3.2 Load Forecast Model

A correlation between the day of the week can be seen, as the energy demand of a community follows the patterns of the daily life of people. For example during the weekdays, in the morning, the majority of the people are at their jobs, meaning that the factories and companies are consuming more energy than the weekends. This correlation can be exploited to further enhance the performance of the load forecasting. In Figure 3.1 the correlation between the forecasted day and the previous ones is exhibited. By analyzing the diagram, it can be seen that the day before the forecast and one week before, achieves the best results, leading to the conclusion that these days can influence greatly the performance of the forecast model. In Figure 3.1 the correlation between the day is shown. Additionally, this implies and further enforce the perception that each specific day of the week has a different influence and demand on the energy load of the grid. For this reason multiple prediction models utilizes

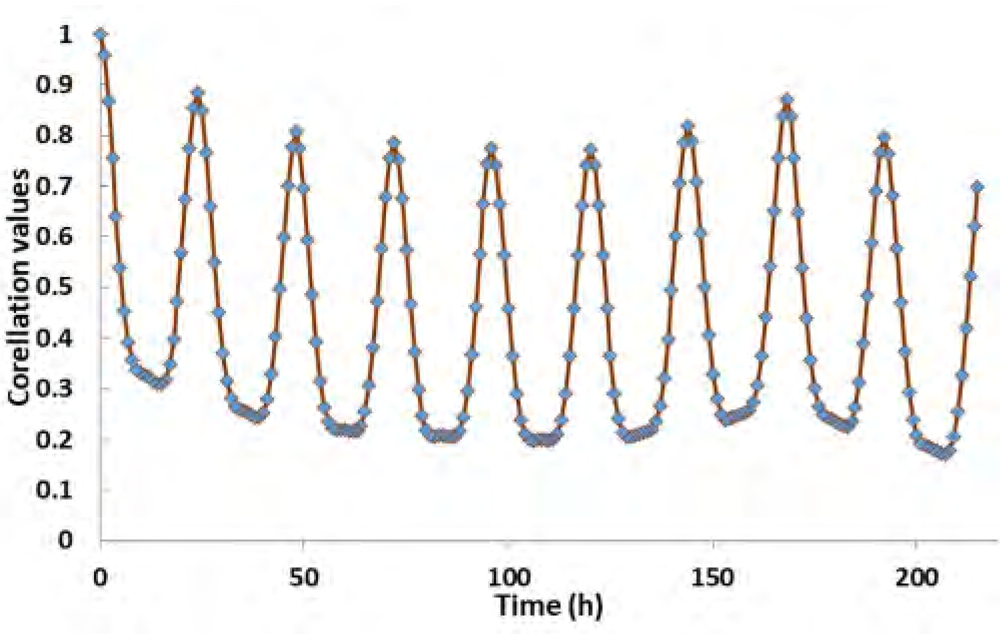


Figure 3.1: Correlation between previous days of the week [37].

this phenomenon to improve the prediction accuracy [37].

In order to predict energy load accurately, temperature information is essential. Since temperature directly influences how much heating and cooling is needed in the residential, commercial, and industrial sectors, there is a strong correlation between temperature and energy demand. By utilizing this kind of data, forecast simulations can find patterns that aid in predicting future energy demand by analyzing historical temperature patterns and correlating them with data on energy consumption [38]. For instance, higher temperatures typically result in increased electricity use for air conditioning during the hot summer months. In a similar situation, colder winter weather raises demand for heating. Energy providers can decide wisely about resource allocation, generation capacity planning, and grid stability by incorporating temperature data into energy load forecasts.

- Inputs 1 to 24: Load per hour for day $d-1$.
- Inputs 25 to 48: Load per hour for day $d-7$.
- Input 49 and 50: Minimum and Maximum Temperature for day d.
- Input 51 and 52: Minimum and Maximum Temperature for day $d-1$.
- Input 53 and 54: Square deviation of maximum temperature for day d and day $d-1$

with the comfort temperatures.

$$CT = \begin{cases} (T_{max} - Tc_{min})^2, & \text{if} \quad T_{max} < Tc_{min} \\ 0, & \text{if} \quad Tc_{min} \leq T_{max} \leq Tc_{max} \\ (T_{max} - Tc_{max})^2, & \text{if} \quad T_{max} > Tc_{max} \end{cases} \tag{3.1}$$

where $Tc_{max} = 25^oC$ is the maximum and $Tc_{min} = 17^oC$ is the minimum comfort temperature.

- Input 55: Difference between maximum temperature of day d and day $d-1$.
- Input 56 to 62: Days of the week, where value 1 corresponds to the day of the week, and the value 0 is set to the rest of the days.
- Input 63: Holidays, where 1 means that this day is a holiday and 0 means it is not a holiday.

where day d is the desired forecasted day, $d-1$ is the previous day of the forecasted day and $d-7$ is one week before.

3.3 Wind Generation Forecast Model

Similarly to the load forecast, the most important historical data are the Wind production for the previous days. As wind production is mainly influenced by the wind, short-term historical production data can help identify and predict patterns for the wind generation forecasts. For this reason the historical hourly wind production data of the two previous days are considered as inputs for the prediction model. Furthermore the minimum and maximum temperature data for the forecasted day and the previous one were chosen as there is a correlation between wind speed and temperature.

Wind speed is a crucial parameter for wind power planning and operations because it is crucial for predicting wind generation. Wind speed directly affects how much electricity is produced by wind turbines. For wind farms to operate as efficiently as possible, accurate wind speed forecasts are essential because they tell operators when and where to deploy resources. Operators of wind farms can use this knowledge to maximize energy output, control grid integration, and guarantee the dependability and stability of the power supply. Additionally, wind speed data enables utilities to decide on backup power sources and grid balancing in an informed manner, reducing costs. For this reason the maximum and minimum wind speed of the forecasted day is considered to the prediction model. Additionally, the average wind speeds for the previous two days is also taken into consideration [39].

- Inputs 1 to 24: Wind production per hour for day $d-1$.
- Inputs 25 to 48: Wind production per hour for day $d-2$.
- Input 49 and 50: Minimum and Maximum Temperature for day d.
- Input 51 and 52: Minimum and Maximum Temperature for day $d-1$.
- Input 53: Maximum forecasted wind speed for day d.
- Input 54: Minimum forecasted wind speed for day d.
- Input 55: Average Wind speed for day $d-1$.
- Input 56: Average Wind speed for day $d-2$.

where day d is the desired forecasted day, $d-1$ and $d-2$ is previous and the second day before the forecasted day, respectively.

3.4 Solar Generation Forecast Model

For the solar generation forecast, the two previous days of historical hourly solar data was chosen as input for the artificial neural network. There is a high correlation between the pattern of the solar energy production and the future output, considering short term predictions. As the solar generation has a high volatility and it is impacted by variables that are hardly predicted and forecasted, the time window for the available and useful information is very short. Furthermore, the minimum and maximum temperature for the forecasted day and the day before is considered closely related to the solar generation, as it is impacting the performance but also the weather conditions that the solar panels are operating [39].

For precise solar forecast simulations, solar irradiation data is crucial. The solar irradiation has a direct impact on the output and effectiveness of solar photovoltaic systems. Forecasters can more accurately predict future solar power generation by analyzing historical solar irradiation data to find patterns and variations in the solar energy availability. Solar power operators can optimize system performance, create maintenance schedules, and guarantee reliable electricity generation with the aid of solar irradiation data, which offers insightful information about the anticipated solar energy resource. By allowing utilities to forecast the solar energy contribution and manage the fluctuating nature of solar generation, it also helps with grid integration and load balancing. For this precise reason the average solar irradiation of the forecasted day and the day before is taken into account.

- Inputs 1 to 24: Solar production per hour for day $d-1$.
- Inputs 25 to 48: Solar production per hour for day $d-2$.
- Input 49 and 50: Minimum and Maximum Temperature for day d.
- Input 51 and 52: Minimum and Maximum Temperature for day $d-1$.
- Input 53: Average forecasted solar irradiation for day d.
- Input 54: Average forecasted solar irradiation for day $d-1$.

where day d is the desired forecasted day, $d-1$ and $d-2$ is previous and the second day before the forecasted day, respectively.

3.5 Artificial Neural Network Approach

In order to evaluate the performance of multiple feature selection methods, a forecasting model is required in order to produce an outcome based on the available data. As it was described in the previous chapter, the artificial neural networks are excellent candidates for performing load and generation forecast. Based on the neural network's structure and settings, different outcomes and performances can be achieved, allowing this kind of methods to be considered very versatile.

3.5.1 Proposed Structure

In this thesis, a simple Multi Layer Perceptron is chosen in order to compare the performance of the feature selection methods. More specifically, the multi layer perceptron is comprised of three layers in total, one input, one hidden and one output layer. Figure 3.2

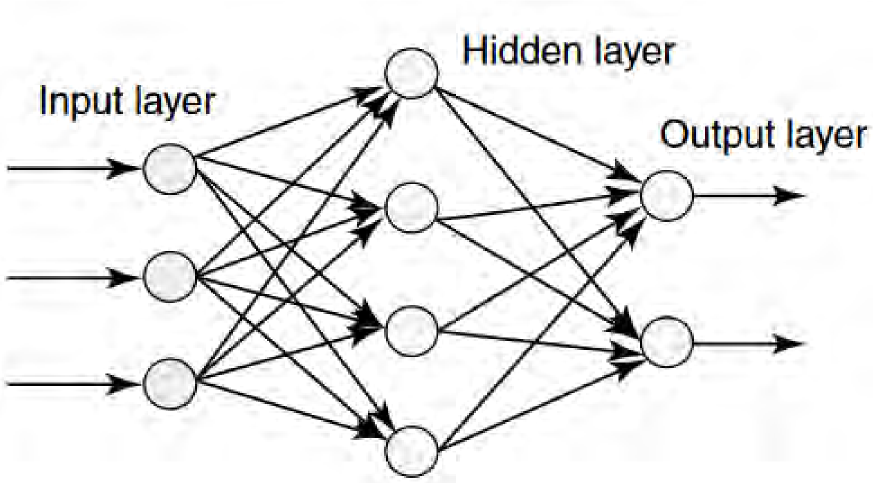


Figure 3.2: Artificial Neural Network [9].

shows a simplified structure of a multi layer perceptron. This simple but robust neural network is the perfect candidate to check and evaluate feature selection methods and how well they can perform. Moreover, this allows for a more generic approach to the topic, without involving special characteristics and combinations of other case sensitive forecast methods combine with feature selection methods, which might lead to wrong conclusions.

Considering the details of the Multi Layer Perceptron's structure, the input layer consist of a determined amount of neurons based on the available features. In this thesis, an investigation for the best amount of features for each method is also completed. In order to achieve this a variable amount of input layers is required.

In order to determine the optimal amount of neurons in the hidden layer, an analysis was performed by comparing the mean absolute percentage error for different amount of neurons. In principal, a load forecast model was used with a total of 63 features was used to make the initial estimation of the neurons. As can be seen in Figure 3.3, the optimal amount of neurons for the hidden layer is 24, as the minimum MAPE is achieved. Similar performance with less accuracy can be seen for the scenario with the 30 neurons. Moreover, some scenarios, like the ones with the 26 and 18 neurons, can offset the overall performance of the results significantly. Finally for the output layer, 24 output neurons are chosen, one for each hour of the forecasted day.

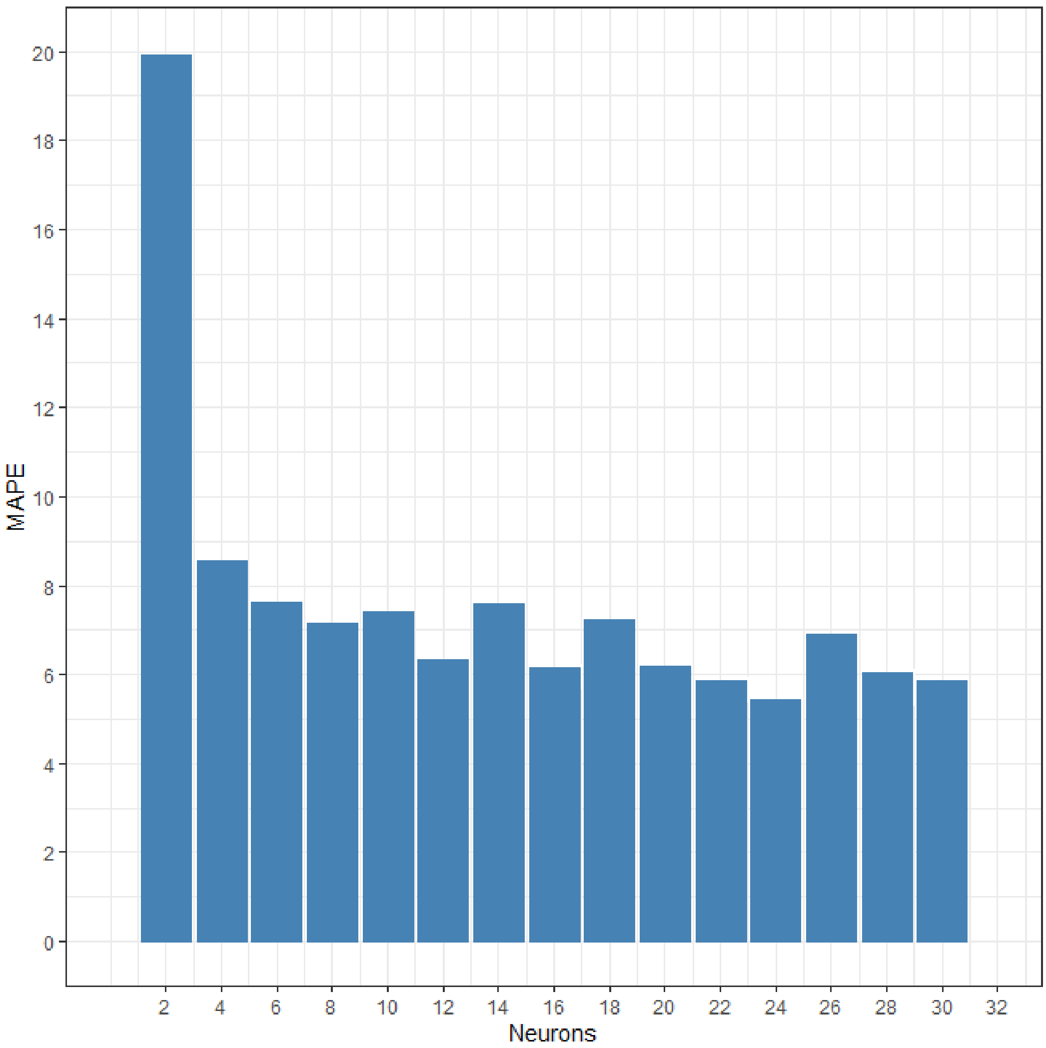


Figure 3.3: Number of neurons in the hidden layer.

3.5.2 Activation Function

Due to its efficiency in deep learning, the Rectified Linear Unit (ReLU) activation function is frequently used for the hidden layers and output layers of a neural network. ReLU is a well known option due to a number of benefits. In comparison to other activation functions like sigmoid or tanh, computing the ReLU activation is computationally efficient. Second, ReLU adds non-linearity to the network, allowing it to learn nuanced patterns in the data and model complicated relationships. Another advantage is that ReLU prevents saturation for positive inputs by allowing them to pass through unchanged, which lessens the vanishing gradient issue that might arise during training. ReLU also gives the network sparsity by mapping negative inputs to zero, which enables the network to concentrate on crucial features and ignore unimportant ones. For these reasons, the ReLU activation function was chosen to be used in the hidden and the output layer, as provides simplicity, computational effectiveness and non-linearity.

3.6 Summary

In this chapter, three different forecast model were presented. Starting with the load forecast model, a detailed overview of the available features that will be used in the prediction models was presented. From a big pool of irrelevant data 63 in total features were selected as the inputs for the neural network. Moving on, for the wind generation model, a similar approach was presented, leading to a total of 56 features that will be used for prediction. Lastly, for the solar generation forecast 54 candidates were selected as the inputs for the neural network. Furthermore, the approach of the artificial neural network that was utilized during these simulations was presented. In details, a the Multi Layer Perceptron with one hidden layer was chosen as the building block for performing the load and generation forecasts. Additionally, a study case was performed to determine the amount of neurons in the hidden layer. The results showed that the optimal amount of neurons in the hidden layer should be 24, as it minimizes the error of the forecast.

In the next chapter, the results from the simulations for the load, wind and solar generation, using the feature selection methods, are presented. More specifically, multiple case scenarios are simulated by using the Mutual Information, the Random Forest, the RReliefF, the Maximal Information Coefficient and the MRMR as the feature selection methods. For each method multiple scenarios with different amount of features are presented, in order to identify the best method for feature selection.

Chapter 4

Feature Selection Simulations Results

4.1 Introduction

In this chapter the final forecast simulation are presented and compared with a variety of feature selection methods. Firstly, a brief demonstration of the used data and prediction horizon is presented, explaining the year and location of data that were used for predicting the load and generation of the energy grid. Furthermore, an early data preparation was perform in order to remove noise and outliners from the data in order to further improve the accuracy and performance of the models. Finally, a detailed explanation of the results for each feature selection method, with different amount of used features is presented.

4.2 Data and Prediction Horizon

The right mix of training and test data for an artificial neural network depends on a number of variables, including the dataset's size, the difficulty of the task, and the data's accessibility. A typical strategy is to divide the data into training and testing, with the remaining 20–30% going to each. This guarantees that the network has enough data to discover the underlying links and patterns during the training phase. On the other hand, test data is used to assess how well the trained network performs and how well it generalizes to new data. It's critical to strike a balance between having enough training data to support efficient learning and sufficient test data to appropriately gauge the network's performance. Additionally, it is important to avoid using the test data of any model for tuning or parameter selection because doing so can result in overfitting and false results of model's performance.

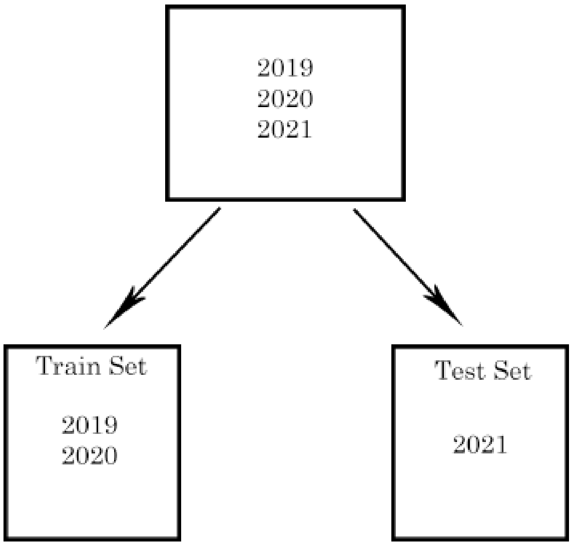


Figure 4.1: Historical data divided into Training and Test Sets.

Three years of hourly historical data from Greece's energy grid, obtain by [40], were used in the forecast models and feature selection method. Particularly, the years of 2019, 2020 and 2021 were chosen, which included hourly load and generation data. From these data sets the actual energy load, the wind generation and solar generation for the entire Greece was extracted in order to be used in the forecast models. As represented in Figure 4.1 the year of 2019 and 2020 were chosen as the training data to train the artificial neural network, while 2021 was chosen as the test set.

The location of historical meteorological data used to in load and generation forecast, using neural networks is crucial. Weather conditions can vary greatly from one region to another, and these differences have a direct impact on the patterns of energy production and consumption. Neural networks can sufficiently capture the complex interactions between weather variables and energy demands by using historical meteorological data that is relevant to the target area. Local meteorological variables like temperature, humidity, wind speed, and solar radiation can give important information on seasonal variations in energy use. Furthermore,artificial neural network models may more accurately comprehend and forecast the energy requirements and fluctuations particular to that area by taking the actual location into account, leading to more accurate and efficient energy predictions.

Regarding the source of the historical weather data used during this thesis simulations, the NASA's project POWER [41] was used to obtain the required data. Two major location were chosen as the starting points based in Greece. The location of Athens, as the capital

Figure 4.2: Spatial Distribution of Greek Wind Capacity [42].

of Greece with the majority of the population located around there, combined with the location of Thessaloniki, which is the second largest city, were chosen as the data points for the weather parameters. For a better evaluation of the weather parameters the average values of the two locations were extracted in order to further enhance the performance of the prediction models. These weather data were utilized for the load forecast models as the entirety of Greece's energy grid was taken into account.

Furthermore, for the wind prediction model the location of the extracted data was different from the previous cases. This has to do with the Wind farms distribution in Greece, which can be found in multiple locations. The majority of the wind farms are located in the central Greece as it can be seen from Figure 4.2. This location is set to be around the town of Lamia. For this reason the same location was chosen as the weather data extraction location, in order to better anticipate the fluctuations of the energy generation.

Regarding the solar generation models a similar approach with the wind generation was

Table 4.1: Spatial Distribution of Greek Solar Capacity 2021 [43].

Location	Solar Capacity (MW)
East Macedonia	361.816
Central Macedonia	675.133
West Macedonia	230.23
Aegean	9.101
Epirus	16.77
Thessaly	637.706
West Greece	326.56
Central Greece	1185.94
Peloponnisos	314.08
Attiki	209.35
Crete	78.291

used. As it can be seen from Table 4.1 the majority of the solar capacity can be detected around the central Greece. This is because the biggest part of central Greece has flat planes used for agriculture and solar farms, with an average yearly solar irradiation, compared with the rest of the country. For this reason the same location as the one for the wind generation forecast model was chosen to extract weather data related to solar generation.

For neural networks to be trained and perform well, it is essential that invalid data, outliers, and extreme values be removed from the data. For neural networks to produce reliable predictions or classifications, the data must contain patterns and relationships. The dataset may, however, produce biased or incorrect model results if it contains invalid or false data points, such as missing values or inconsistent entries. Similar to outliers and extreme values, learning can be disproportionately affected, which affects how well the network can generalize and predict outcomes. The main concept of data preparation is to improve the used data in order to allow the neural network to focus on useful patterns and generalization. This method improves the network's predictions' accuracy, effectiveness, and interpretability, allowing for better decisions and insights based on the model's outputs.

For the above mentioned reason a data preparation was performed in the original data sets in order to purify and improve the overall data quality. Specifically, the linear approximation

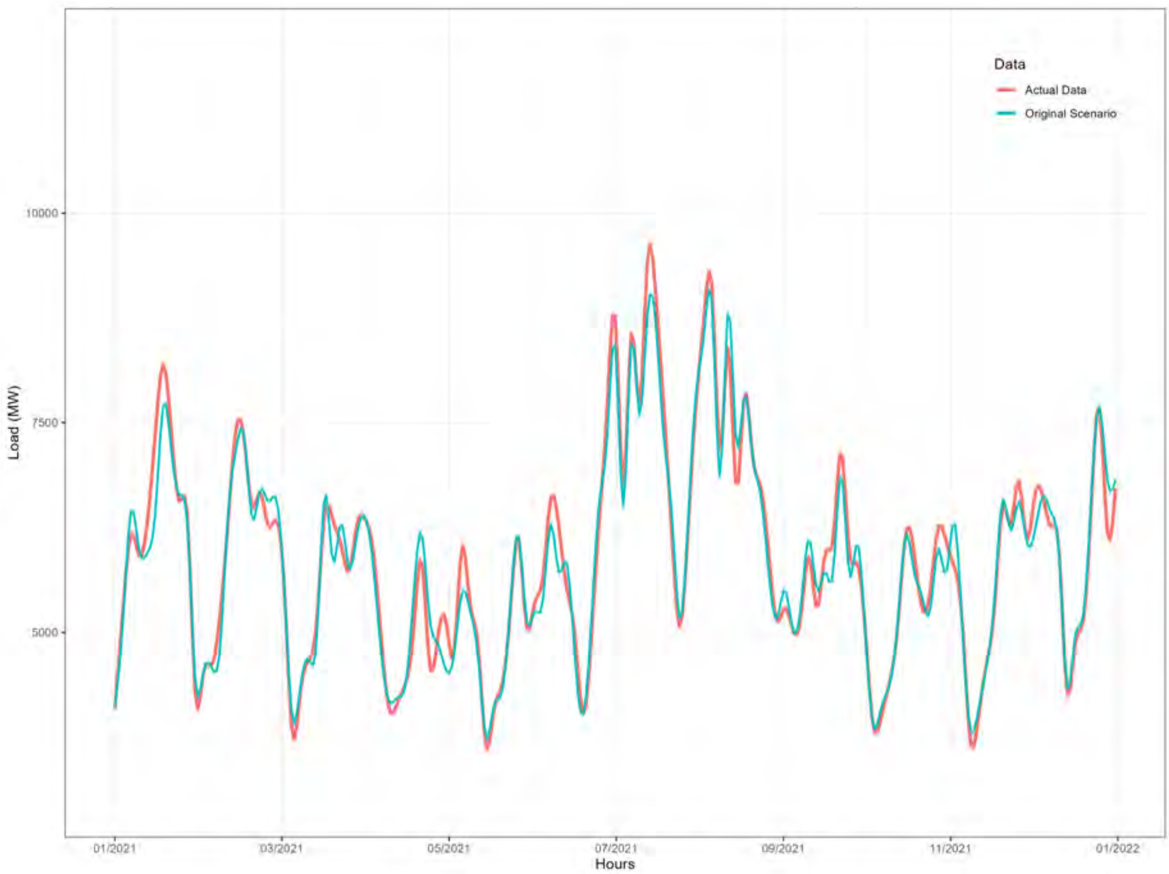


Figure 4.3: Actual and forecasted load based on the model without an FSM.

Table 4.2: Evaluation errors for the original load forecast model without a FSM.

MAPE (%)	MAE (MW)	RMSE (MW)
5.63	286.79	390.13

method based on the previous and the next available values was used in order to fill any missing or invalid data in the data set. Furthermore, values that had unnaturally high values, most probably due to logging errors or noise, were removed in order to improve the prediction model's performance.

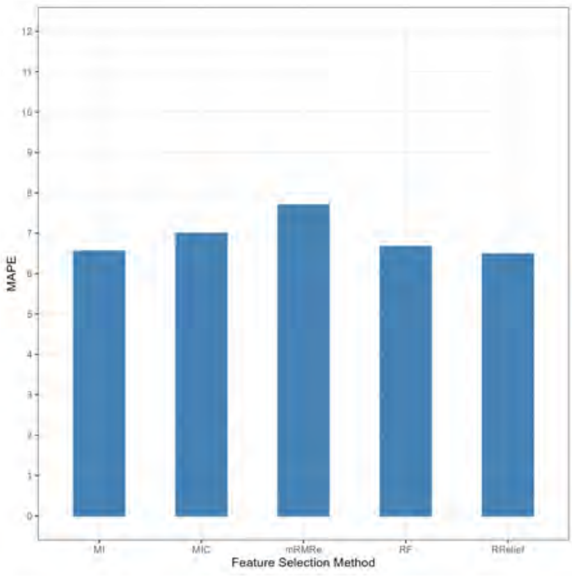
Finally, all the simulations were executed using R 4.3.0, a popular programming language for statistical analysis and data manipulation. The hardware used for these simulations consisted of an Intel Core i7-5500U CPU operating at a frequency of 2.4 GHz, equipped with four cores, and an NVIDIA GeForce 840M graphics card with 2GB of memory.

4.3 Load Forecast

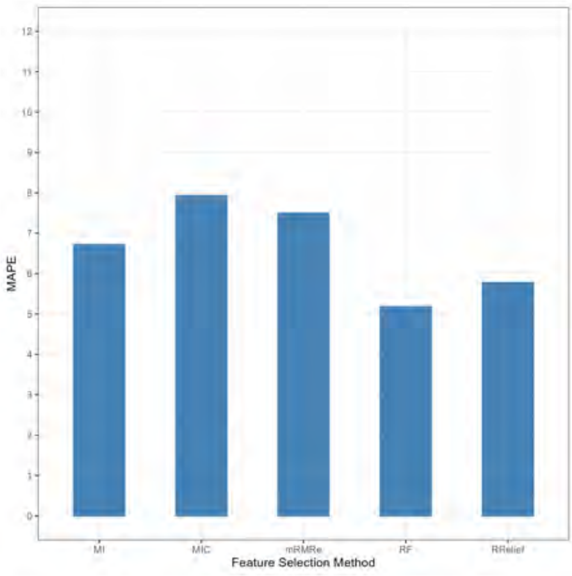
Starting with the first scenario for the load forecast, the original prediction model did not include a feature selection method, meaning that it utilized all the available features. In Figure 4.3 the hourly load data for the time period of 2021 is presented. Specifically, a comparison between the actual and the forecasted data is shown. By analyzing the figure, an overall good performance can be seen for the proposed scenario as it can greatly anticipate the fluctuation of the energy grid load. At some specific points, for example between the months of January and February, the prediction model underestimated the load requirements by some hundred up to thousand of MW. This can be cause due to big fluctuation of temperatures or demands in the energy grid that the neural network could not anticipate. Furthermore, by examining the Table 4.2 it can be seen that an overall of 5.63% of the Mean Absolute Percentage Error is achieved. In general this could be described as a good overall performance for a prediction model, as it leads to up to $286.8MW$ of Mean Absolute Error, which compared with the total amount of the available load of the grid, the error is relative small.

In order to select the most effective features for load forecast, a comprehensive evaluation was conducted using the Mutual Information, the Maximum Relevancy and Minimum Redundancy, the Random Forest, the RReliefF and the Maximal Information Coefficient as the feature selection methods. Multiple scenarios were tested, based on the proposed algorithm in Chapter 3.2, to determine the optimal number of features for accurate predictions. Specifically, eleven different scenarios were examined, each varying in the number of features included, in order to identify the best case. These scenarios encompassed 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, and 60 features, which were compared against the original scenario that included all 63 features. The interval between the scenarios was chosen to be four features, as the gap between the scenarios are not that big and it can capture the variations of performance between them very well. Additionally, the total amount of scenarios remained relatively small.

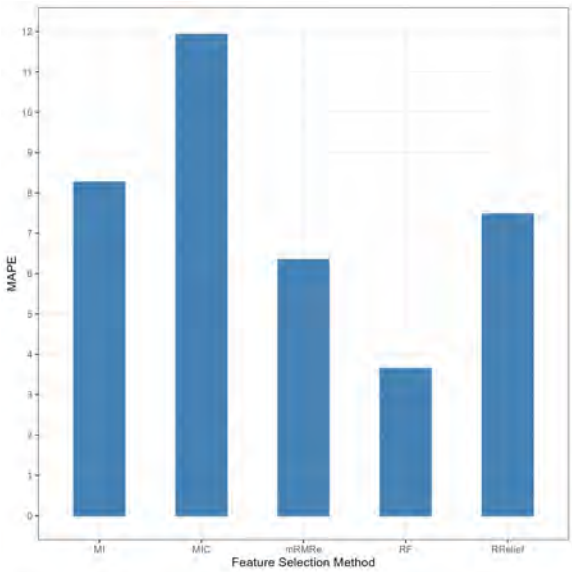
Regarding the first set scenarios of 20, 24, 28 and 32 features, the results for each feature selection method can be shown in Figure 4.4. Specifically, it can be seen that by reducing the available features for the forecast models, their overall performance decreases. For the scenarios with the 20 and 24 feature the evaluation errors based on the Mean Absolute Percentage Error, shows that the error is almost double in most of the cases from the original scenario. This can be explain by the incompetence of the artificial neural network to pre-



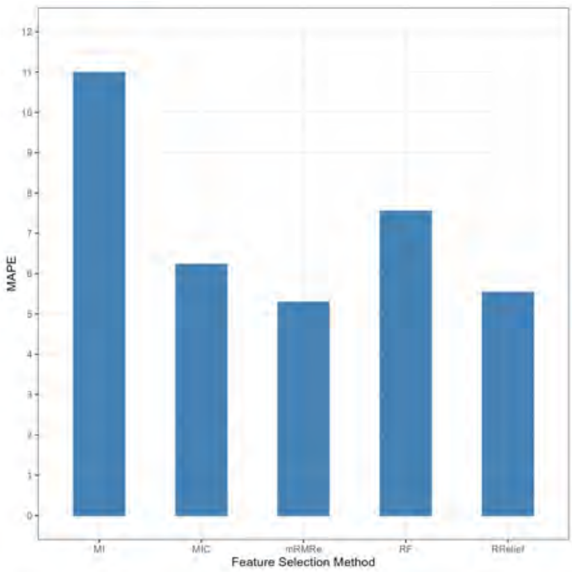
(a) MAPE of the top 20 features.



(b) MAPE of the top 24 features.



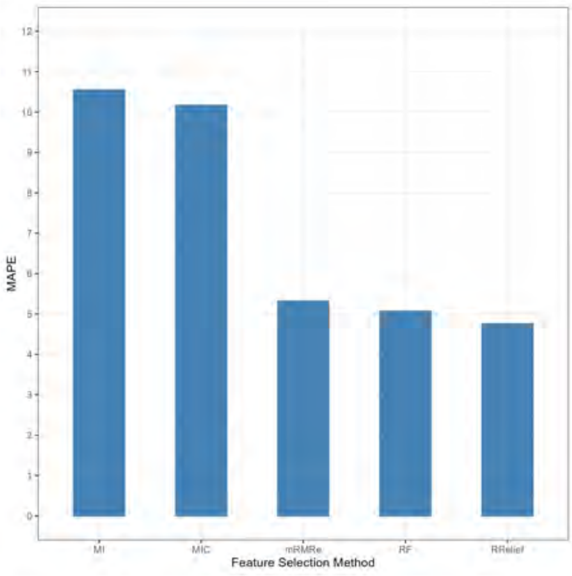
(c) MAPE of the top 28 features.



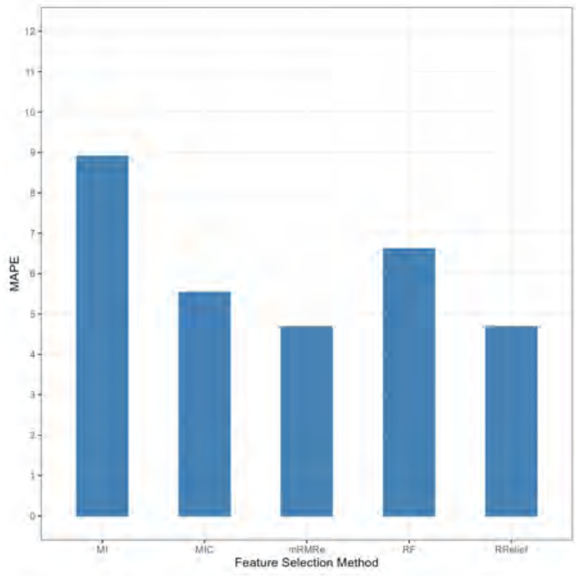
(d) MAPE of the top 32 features.

Figure 4.4: Load forecast performance with feature selection for the 20, 24, 28 and 32 top features.

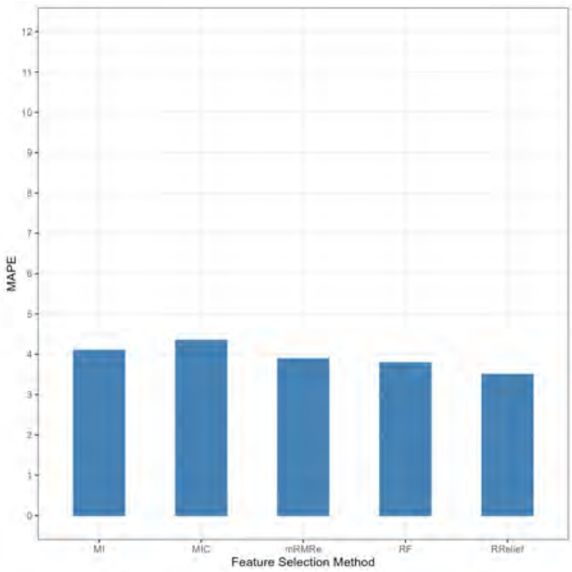
dict the fluctuations of the energy load based on the available features. The information and correlation between each individual feature could not give enough weight to each individual neuron in order to produce the desired output. From the other hand, the Random Forest algorithm is able to better correlate the best features from the available pool and improve the performance of the prediction model compared with the other feature selection methods. Additionally, for the scenario of the 28 total feature the MAPE for this scenario could be identified as 3.71%, which is far superior that the original scenario.



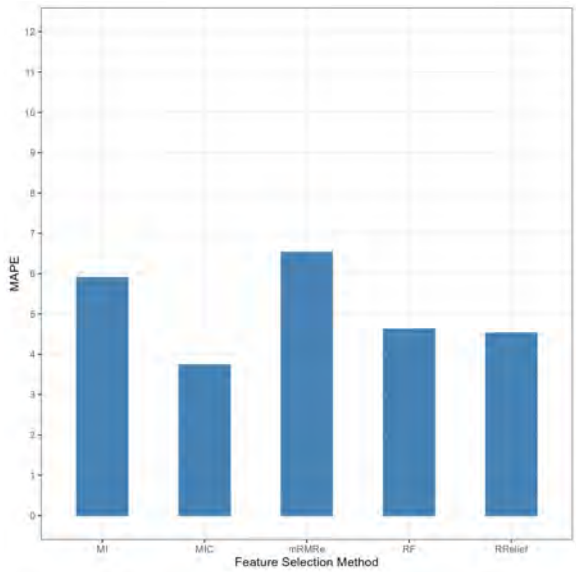
(a) MAPE of the top 36 features.



(b) MAPE of the top 40 features.



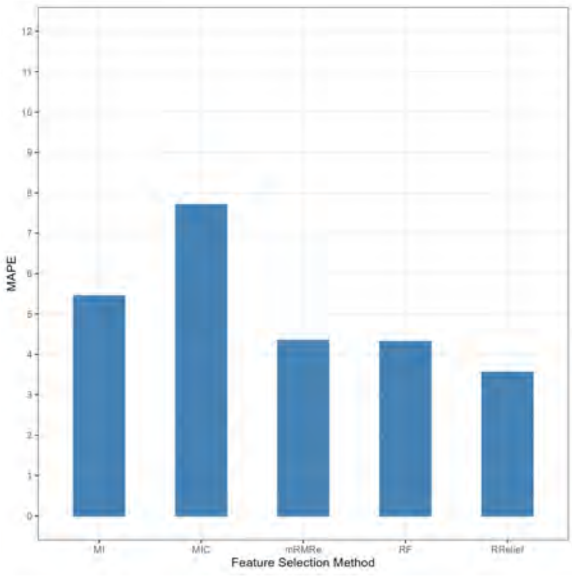
(c) MAPE of the top 44 features.



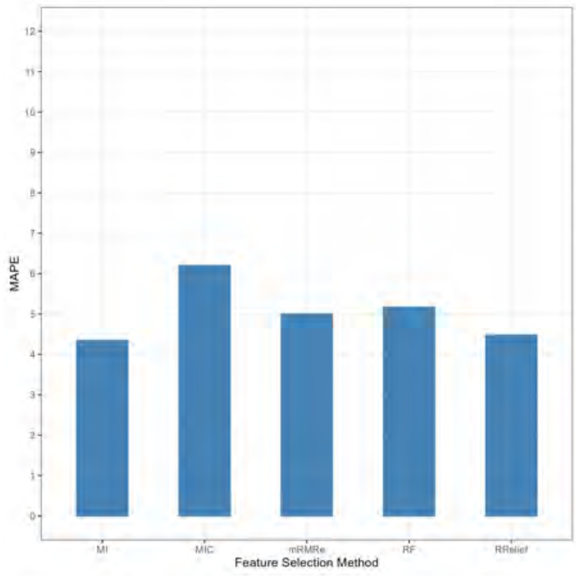
(d) MAPE of the top 48 features.

Figure 4.5: Load forecast performance with feature selection for the 36, 40, 44 and 48 top features.

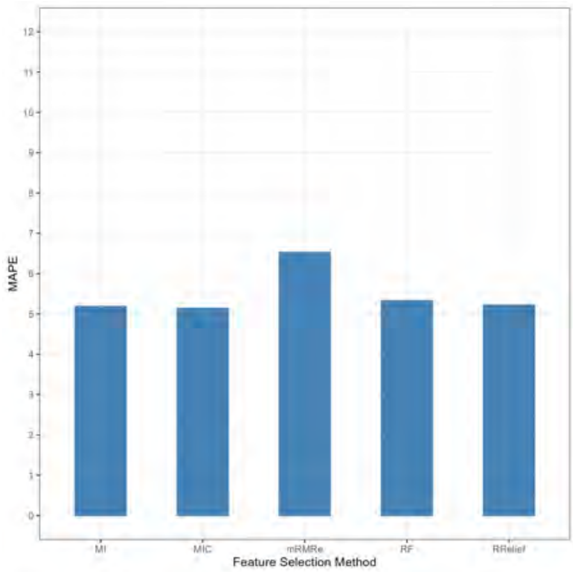
Moving on with the second set of scenarios, with 36, 40, 44 and 48 available feature, the results in Figure 4.5 shows that the performance of the prediction models, greatly increases. Particularly, for the scenarios with the 44 and 48 features, the results are almost all better than the original model. For the scenario with the 44 features in Figure 4.5c, the RReliefF algorithm shows the best results with a 3.45% MAPE, which surpasses any previous performance. For the 48 scenarios the Mutual Information Coefficient also shows great promise as it achieves 3.82% MAPE, while the rest of the algorithm perform worse. Furthermore,



(a) MAPE of the top 52 features.



(b) MAPE of the top 56 features.



(c) MAPE of the top 60 features.

Figure 4.6: Load forecast performance with feature selection for the 52, 56, and 60 top features.

Mutual Information seems to perform worse with fewer available features, as the error is decreasing while the features are getting more and more. For the scenario with the 44 features MI seems to perform very well compared with the other scenarios of features, but compared with the other methods the results are still worse. Regarding the Maximum Relevancy and Minimum Redundancy algorithm, during the scenario with the 40 features, it surpasses the performance of the other results, leading to 4.62% MAPE, while RReliefF follows closely with a 4.69% MAPE.

Table 4.3: Overall performance of the best feature selection methods.

Method	Num. of Features	MAPE (%)	MAE (MW)	RMSE (MW)	Calc. Time (sec)
Original	63	5.63	286.79	390.13	56.2
RReliefF	52	3.62	215.12	303.73	55.9
MIC	48	3.82	226.61	319.74	97.1
RReliefF	44	3.45	212.12	300.49	54.6
MRMR	44	3.89	257.02	358.24	56.1
MI	44	4.12	266.73	454.95	57.1
RF	28	3.71	221.58	316.36	73.1

Lastly, for the scenarios with the 52, 56 and 60 features, the results for all the feature selection methods are presented in Figure 4.6. As was expected Mutual Information performs even better with more available features, compared with the majority of the previous scenarios. In Figure 4.6b with the 56 features, Mutual Information achieves the best results, with a 4.34% MAPE. Furthermore, once more RReliefF achieves the best performance for the 60 and 52 features scenarios, with the last one achieving a 3.62% MAPE.

Based on the above mentioned simulations, multiple feature selection methods with different amount of feature show promising results. In Table 4.3 a summary from the best scenarios is presented. Particularly, the type of method and the amount of used features is presented with all three evaluation metrics. Additionally, in the last column the overall calculation time of the neural network plus the feature selection method is presented.

By comparing the results from Table 4.3 it can be seen that RReliefF shows the best performance for multiple scenarios, while at the same time the lowest calculation time. Specifically, the lowest time is achieved due to the low computational time that the RReliefF needs to calculate the most relevant features. Additionally, due to the removal of the redundant features the artificial neural network is able to converge faster, leading in the end with lower overall computational time. Furthermore, the scenario with the 44 features achieves the best results from the rest of the simulations, by having the lowest evaluation errors. While other feature selection methods achieves better results from others, like the Random Forest scenario with the 28 features, the overall computational time is much bigger. This is due to the performance of each feature selection method and the ability to produce fast results.

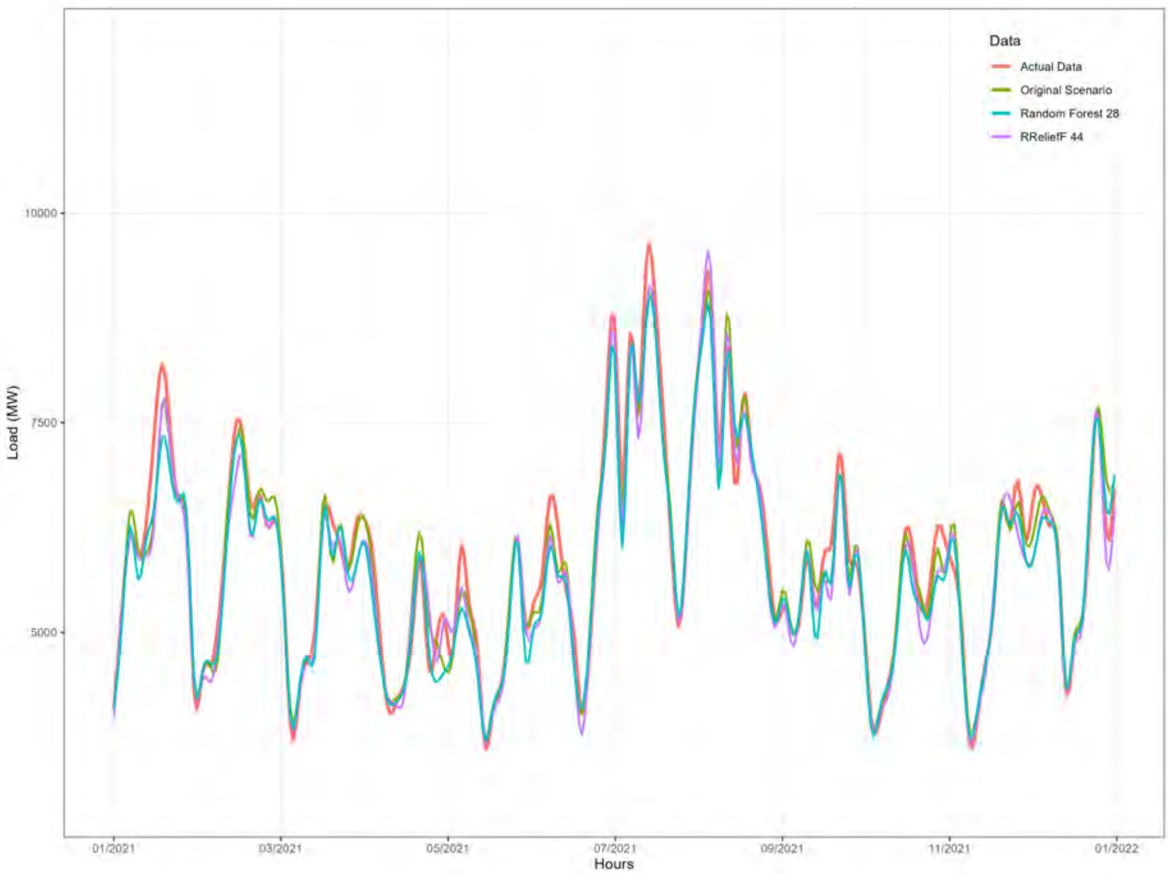


Figure 4.7: Hourly actual data compared with the original scenario, RReliefF and Random Forest with 44 and 28 features, respectively.

Finally, in Figure 4.7 the hourly actual data is presented and compared with the original scenario without any feature selection method, the RReliefF with 44 features and the Random Forest algorithm with 28 features. It can be clearly seen that the RReliefF method can better predict the fluctuation of the energy grid, compared with the other methods. Additionally, the Random Forest with the impressive 28 features shows also very good results and better performance than the original scenario. In Table 4.4 the actual features for the best scenario of the RReliefF with the 44 features is presented. The majority of the most relevant features are the historical load data of the previous day and the day before one week. Furthermore, a big importance seems to have the maximum temperature of the forecasted day, combined with two features of the squared temperature variables. Finally, the type of day for whether it is or not Monday, Saturday or Sunday seems to highly influence the results. The holiday type also seems to influence the end results.

Table 4.4: Most relevant features based on the RReliefF method.

Input	Feature	Input	Feature
1	H1,d-1	23	H23,d-1
2	H2,d-1	24	H24,d-1
3	H3,d-1	25	H1,d-7
4	H4,d-1	26	H2,d-7
5	H5,d-1	27	H3,d-7
6	H6,d-1	28	H4,d-7
7	H7,d-1	29	H5,d-7
8	H8,d-1	30	H6,d-7
9	H9,d-1	31	H7,d-7
10	H10,d-1	32	H8,d-7
11	H11,d-1	33	H9,d-7
12	H12,d-1	34	H10,d-7
13	H13,d-1	38	H14,d-7
14	H14,d-1	44	H20,d-7
15	H15,d-1	46	H22,d-7
16	H16,d-1	50	Max Temp,d
17	H17,d-1	53	CT,d
18	H18,d-1	54	CT,d-1
19	H19,d-1	56	Day type
20	H20,d-1	61	Day type
21	H21,d-1	62	Day type
22	H22,d-1	63	Holidays

4.4 Wind Generation Forecast

Regarding the wind generation forecast, the original prediction model with the total of 56 features did not include a feature selection method. In Figure 4.8 the hourly wind generation data for the time period of 2021 is presented. Specifically, in the figure it can be seen a comparison the actual and the forecasted data. The results shows that the overall performance of the wind forecast model is less accurate compared to the one for the load forecast. This has to do with the nature of the forecast, as multiple factors, like weather conditions, highly contribute to big fluctuation of the energy generation. Furthermore, by examining the figure it seems that the accuracy of the model is better when the fluctuation of wind generation are not that rapid, meaning that for bigger period of time when the generation transit more smoothly, the prediction model is able to achieve better results. This can be explained by the dependency of the feature data based on the previous two days for wind generation data, but also for weather conditions for the forecasted days. So if rapid changes in weather happens is a short time of period, the model will not be able to predict very well the expected wind generation. Additionally, by examining the Table 4.5 it can be seen that an overall of $345.23MW$ of the Mean Absolute Error is achieved. In general this could be described as a good overall performance for a wind prediction model, compared with other models in the literature.

In order to select the most effective features for wind generation forecast, a comprehensive evaluation was conducted using the five proposed feature selection methods. Multiple scenarios were tested, based on the proposed algorithms, in order to determine the optimal number of features for an accurate and fast predictions. In total nine different scenarios were examined, each varying in the number of features included. Similarly to the load forecast scenarios, the interval of four features was chosen. These scenarios encompassed 20, 24, 28, 32, 36, 40, 44, 48, 52 which were compared against the original scenario that included all 56 features.

For the first set scenarios of 20, 24 and 28 features, the results for each feature selection method can be shown in Figure 4.9. Specifically, by analyzing the diagram, the RReliefF algorithm shows the best results for all three scenarios. Particularly, for the scenario with the 28 features the RReliefF algorithm achieved a $275.63MW$ MAE, which is far better from the original scenario. For the same figure the Maximal Information Coefficient achieved a $289.01MW$ MAE which is the second best result out of the three diagrams. MRMR algorithm showed the least accuracy out of all the other feature selection method, implying that this method is sensitive to scarce available features.

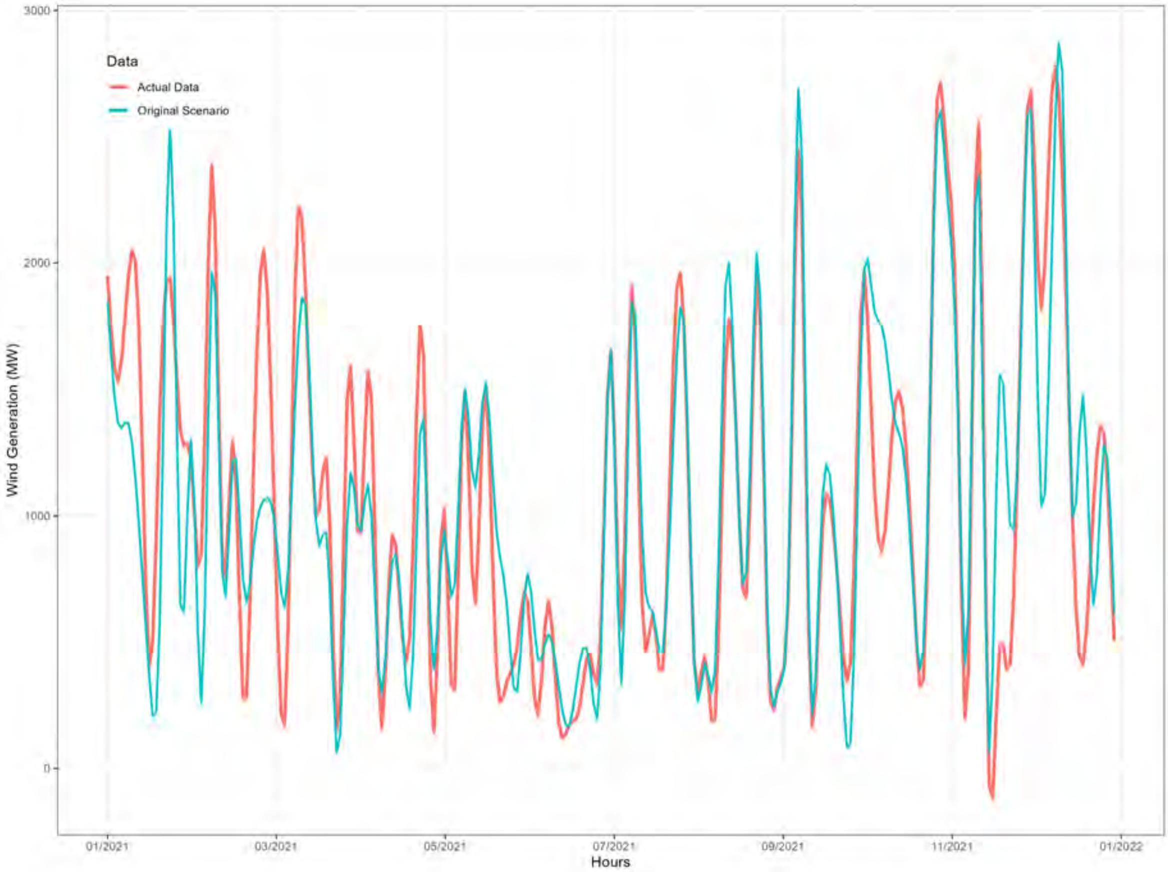
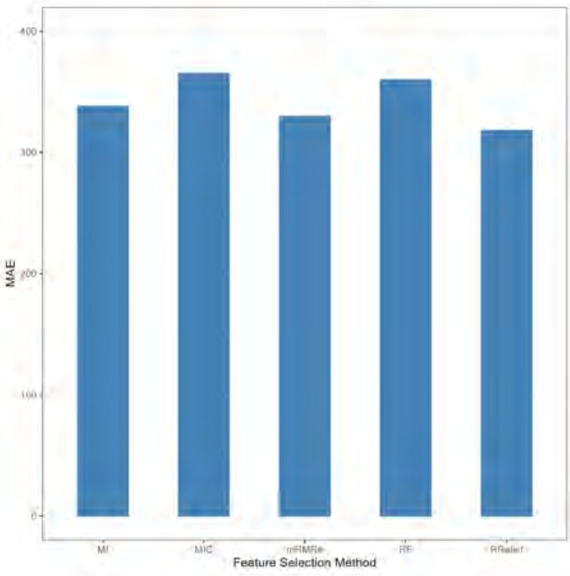


Figure 4.8: Actual and forecasted wind generation based on the model without an FSM.

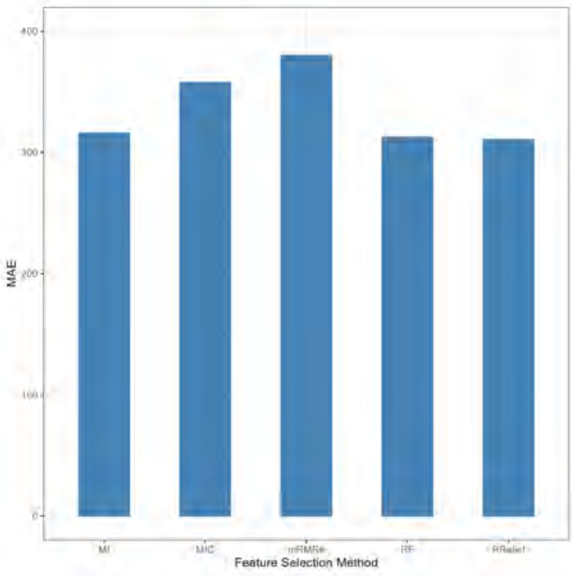
Table 4.5: Evaluation errors for the original wind generation model without a FSM.

MAE (MW)	RMSE (MW)
345.23	404.58

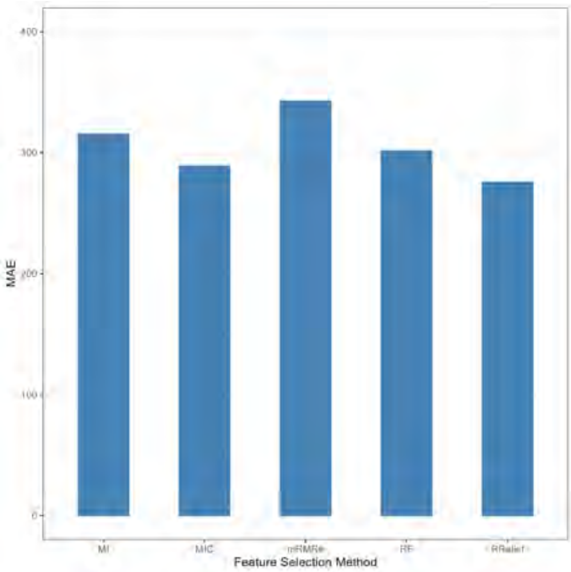
Moving on with the scenarios of 32, 36 and 40 features, RReliefF is again in the lead with the best results, as it can be seen from Figure 4.10. For example in the scenario with the 36 features RReliefF achieved $262.12MW$ MAE, which is even lower that the rest of the figures. In contrast with the previous scenarios MRMR algorithm show a better and promising performance, as the more available features allowed MRMR to make better estimation regarding the most relevant features. Specifically, for the scenario 36 and 40 MRMR achieved a $264.21MW$ and $285.81MW$ MAE, respectively. On the other hand for the scenario of 32 features Maximal Information Coefficient showed promising results with an $269.12MW$ MAE, achieving the third best results for the mentioned scenarios.



(a) MAE of the top 20 features.



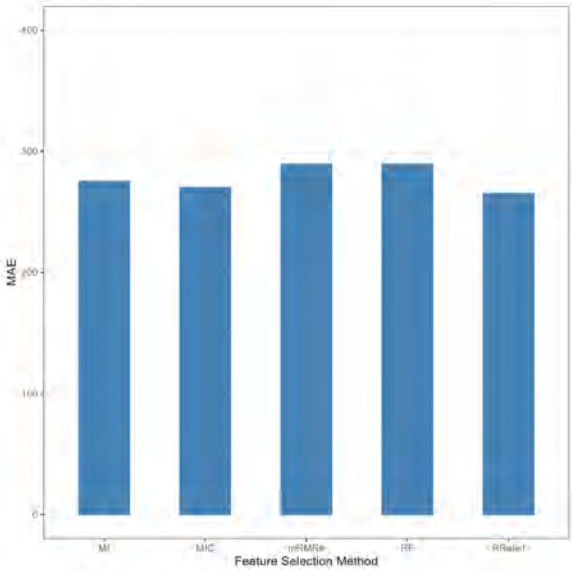
(b) MAE of the top 24 features.



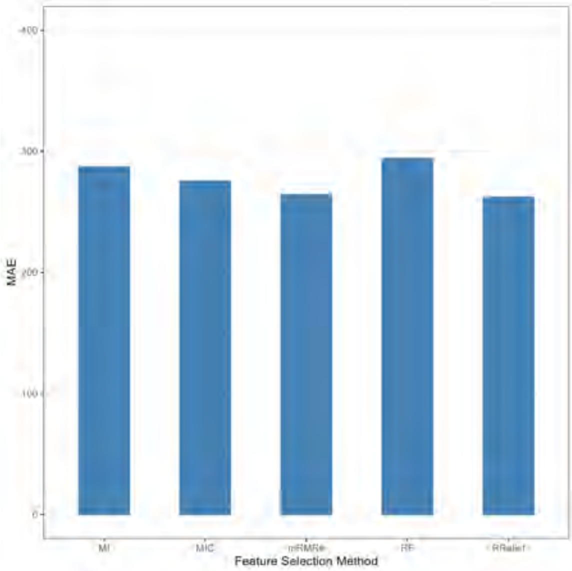
(c) MAE of the top 28 features.

Figure 4.9: Wind generation forecast performance with feature selection for the 20, 24 and 28 top features.

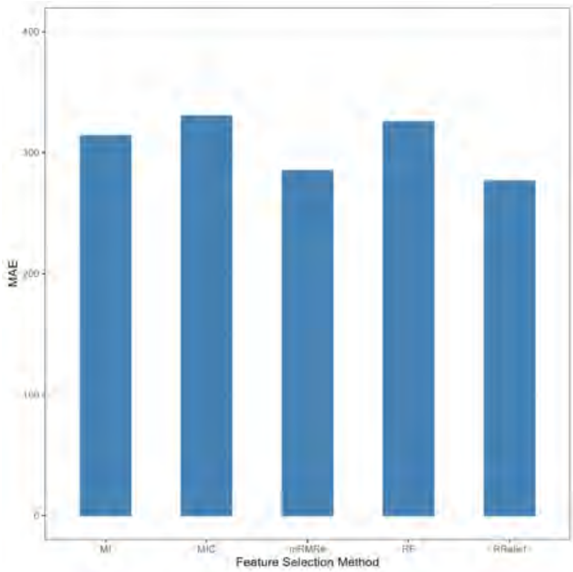
Finally, for the last three scenarios of 44, 48 and 52 features, the results can be seen in Figure 4.11. For those three scenarios RReliefF acomplished one of the worst results compared with the other methods. The same pattern seems to be followed by the MIC algorithm which had the worst overall performance. In contrast, MRMR showed better overall results compared with the other scenarios with fewer features, allowing it to achieve a $273.32MW$ MAE. Additionally, Mutual Information showed an excellent performance for the scenario of 48 features with an $278.92MW$ MAE. By comparing all the results it seems that the RReliefF algorithm can better estimate the relevant features based on an average amount of features, while having more variables can lead to overfitting and less accuracy.



(a) MAE of the top 32 features.



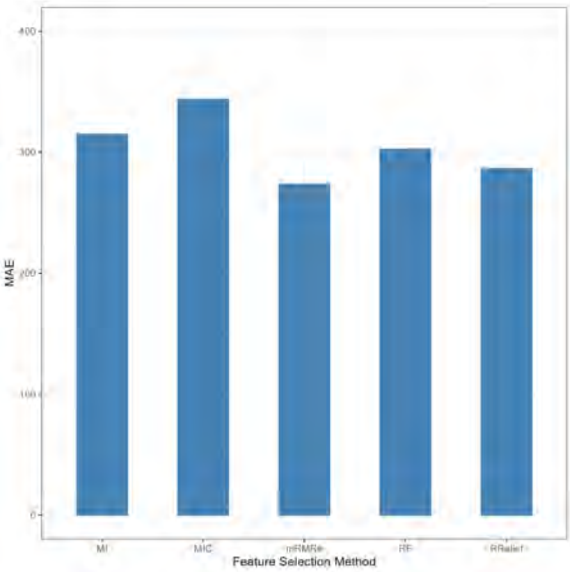
(b) MAE of the top 36 features.



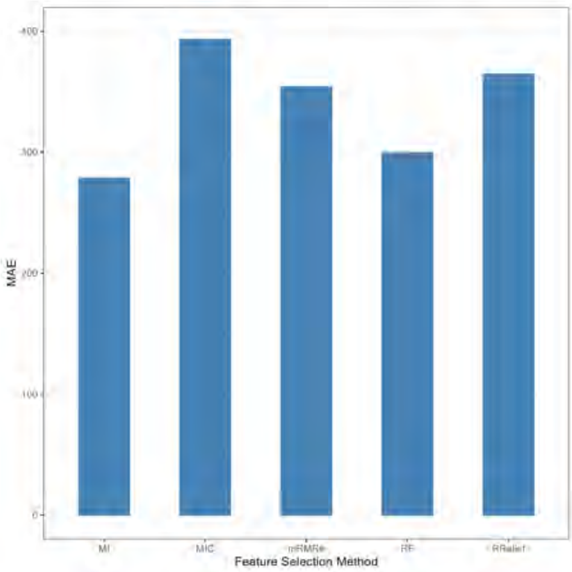
(c) MAE of the top 40 features.

Figure 4.10: Wind generation forecast performance with feature selection for the 32, 36 and 40 top features.

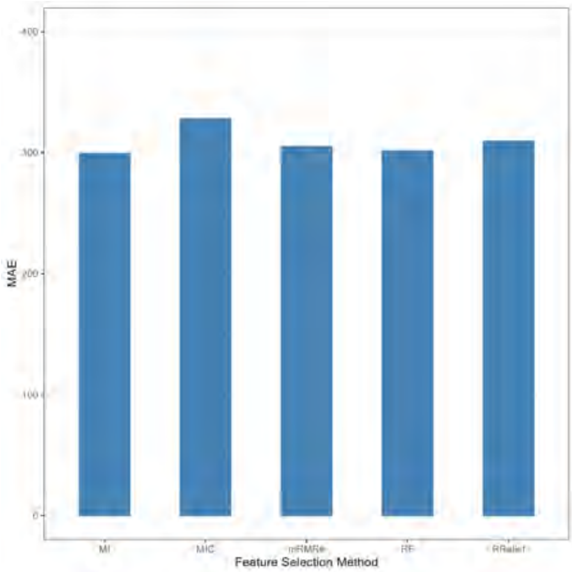
Based on the above mentioned simulations, multiple feature selection methods with different amount of feature show promising results. In Table 4.6 a summary from the best scenarios is presented. A comparison between this features is completed based on the two evaluation error MAE and RMSE. Furthermore, in the last column the overall calculation time of the neural network plus the feature selection method is presented.



(a) MAE of the top 44 features.



(b) MAE of the top 48 features.



(c) MAE of the top 52 features.

Figure 4.11: Wind generation forecast performance with feature selection for the 44, 48 and 52 top features.

By analyzing the results from Table 4.6 it can be seen that MRMR shows the most promising performance for multiple scenarios, while at the same time it achieves the most rapid calculation time. MRMR shows very good stability and robustness in highly volatile scenarios like the wind generation forecast, being able to capture very well the non linearities of the forecast. Particularly, for the scenario with the 36 features, MRMR achieves the second best MAE but the first best RMSE compared with the RReliefF. By comparing the two scenarios, it is very hard to decide which feature selection method is better. Overall MRMR achieves

Table 4.6: Overall performance of the best feature selection methods.

Method	Num. of Features	MAE (MW)	RMSE (MW)	Calc. Time (sec)
Original	56	345.23	404.58	70.2
MI	48	278.92	372.26	65.9
MRMR	44	273.32	385.93	49.1
MRMR	40	285.81	382.31	50.6
MRMR	36	264.21	372.21	50.3
RReliefF	36	262.12	378.71	51.7
MIC	32	289.01	398.62	83.4
RReliefF	28	275.63	385.12	51.3

smaller calculation time which might be considered as an additional advantage. Furthermore, Mutual Information with the scenario of the 48 features achieves very good results, although requiring more computational time from the rest of the scenarios. Similarly, Maximal Information Coefficient also is capable to achieve a good result with only 32 features, but the impact in the calculation time can be considered as significant as it requires much more time.

Finally, in Figure 4.12 the hourly actual data is presented and compared with the original scenario without any feature selection method, the RReliefF with 36 features and the MRMR algorithm with 36 features. The difference between the two feature selection method is very small, as both outperform the original scenario. MRMR seems to perform much better with higher values of wind generation over the bigger spans of the time, while RReliefF is slightly better in predicting the changes of wind energy in shorter time periods. Overall both methods perform very well with the MRMR having the advantage of smaller computational time compared with the RReliefF.

In Table 4.7 the actual features for the best scenario of the MRMR with the 36 features is presented. The majority of the most relevant features are the historical wind generation data of the previous day and the day before. Furthermore, a big importance seems to have the maximum and minimum wind speed of the forecasted day, as the main drive for wind energy production is the air speed. Furthermore the average wind speed of the previous two days have an important impact to the prediction model. Finally, the minimum temperature of the forecasted day was considered as a relevant feature.

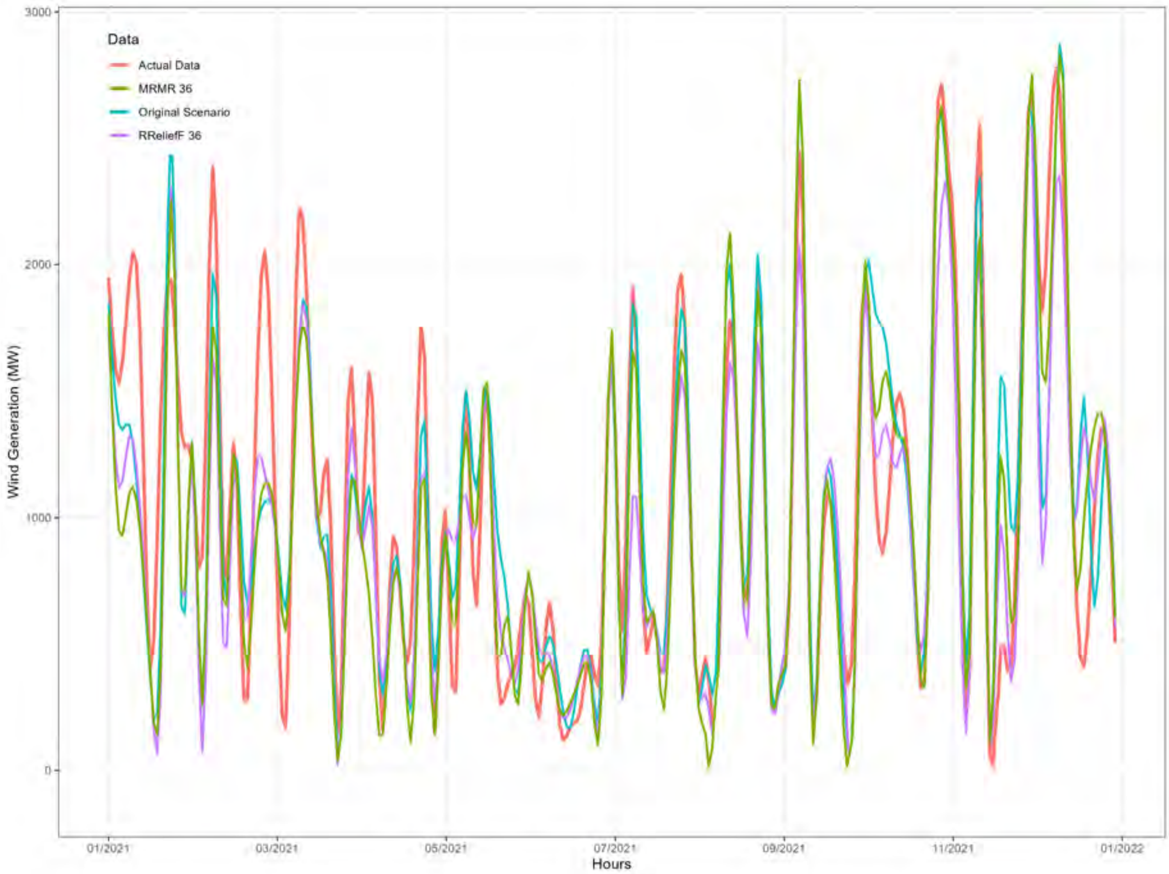


Figure 4.12: Hourly actual data compared with the original scenario, RReliefF and MRMR with 36 features.

Table 4.7: Most relevant features based on the RReliefF method.

Input	Feature	Input	Feature
4	H4,d-1	22	H22,d-1
5	H5,d-1	23	H23,d-1
6	H6,d-1	24	H24,d-1
7	H7,d-1	37	H13,d-2
8	H8,d-1	38	H14,d-2
9	H9,d-1	39	H15,d-2
10	H10,d-1	40	H16,d-2
11	H11,d-1	41	H17,d-2
12	H12,d-1	42	H18,d-2
13	H13,d-1	43	H19,d-2
14	H14,d-1	45	H21,d-2
15	H15,d-1	46	H22,d-2
16	H16,d-1	47	H23,d-2
17	H17,d-1	49	Min Temperature,d
18	H18,d-1	53	Max Wind Speed,d
19	H19,d-1	54	Min Wind Speed,d
20	H20,d-1	55	Average Wind speed,d-1
21	H21,d-1	56	Average Wind speed,d-2

4.5 Solar Generation Forecast

About the solar generation forecast simulations, the first scenario considers the original model, with the total of 54 features which did not include a feature selection method. Particularly, in Figure 4.13 the hourly solar generation data for the time period of 2021 is presented. Specifically, in the figure a comparison between the actual and the forecasted data can be seen. The findings indicate that the solar forecast model performs less accurately overall than the load forecast model, and comparatively similar to the wind generation forecasts. This has to do , similarly with the wind generation, about the fluctuation of variables that can be very hardly be predicted like weather conditions and solar irradiance. Additionally, it appears from the figure that the model's accuracy is higher when solar generation fluctuates less dramatically, i.e., the prediction model performs better over a longer time span when the generation transitions more smoothly. This can be explained by the dependence of the featured data based on the solar generation data from the previous two days as well as the forecasted days' weather. Additionally, by looking at the table, it is clear that a mean absolute error of $175.32 MW$ was achieved overall. In comparison to other models in the literature, this could be characterized as a good overall performance for a wind prediction model.

The five suggested feature selection methods were thoroughly assessed in order to choose the most useful features for solar generation forecast. The best number of features for precise and quick predictions was tested in a variety of scenarios based on the suggested algorithms. Eight distinct scenarios were looked at in total, with each having a different number of features. These scenarios, which included features 20, 24, 28, 32, 36, 40, 44, and 48, were contrasted with the initial scenario, which had all 54 features.

Table 4.8: Evaluation errors for the original solar generation model without a FSM.

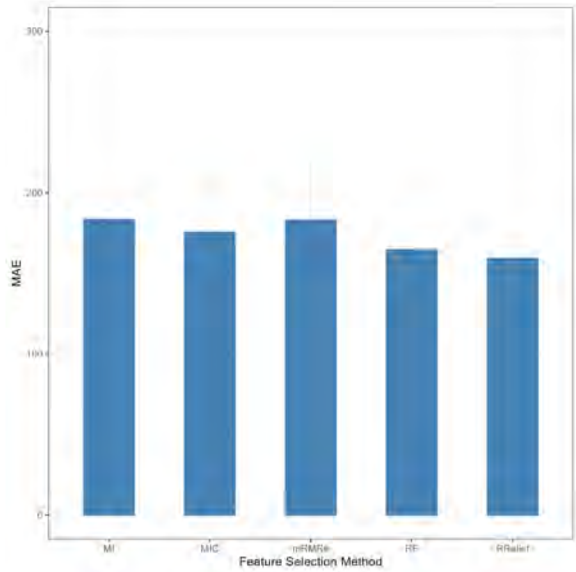
MAE (MW)	RMSE (MW)
175.32	418.68

For the first set scenarios of 20, 24 and 28 features, the results for each feature selection method can be shown in Figure 4.14. Specifically, by analyzing each figure, the Random Forest algorithm shows the best performance for the majority of the scenarios. Particularly, for the scenario with the 28 features the RF algorithm achieved a $156.12 MW$ MAE, which is far better from the original scenario. For the same amount of features the RReliefF achieved

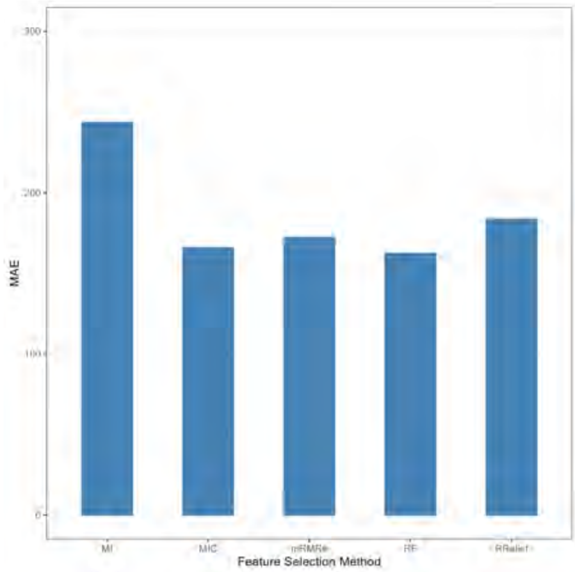
Figure 4.13: Actual and forecasted solar generation based on the model without an FSM.

a $257.14MW$ MAE which is one of the worst results out of the three diagrams. Although, RReliefF showed bad performance for the 28 features, for the least amounts, it was able to capture better the solar generation forecast by achieving an $159.21MW$ MAE. This shows a fluctuation of error for the RReliefF algorithm, which means that this algorithm is sensitive to the correct amount of features. Lastly, Maximal Information Coefficient had a similar performance with the Random Forest algorithm, with the best results of $162.45MW$ MAE for the 28 features.

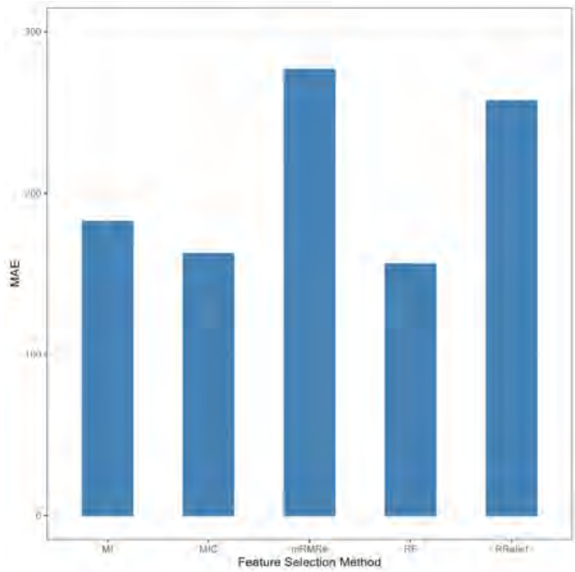
For the next set of features of 32, 36 and 40 the results are depicted in Figure 4.15. For sets of the scenarios, RReliefF shows far better performance compared with the previous simulations. For the scenario with 36 features RReliefF achieved $102.18MW$ MAE, which is the lowest value achieved for theses simulations. Again, RReliefF showed a sensitivity with the amount of features as for the scenario with the 40 variables the error from this algorithm was much higher than the rest. In contrast to the earlier scenarios, the MRMR algorithm performs better and shows promise because it was able to better estimate the solar generation due to the increased number of features that were available. Specifically, for the scenario 36



(a) MAE of the top 20 features.



(b) MAE of the top 24 features.

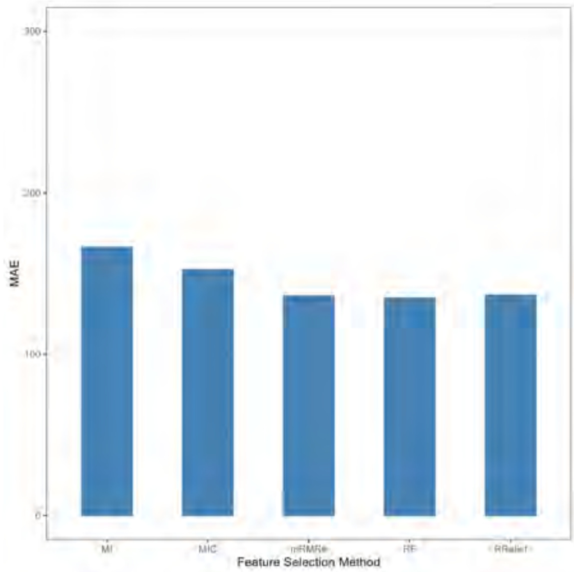


(c) MAE of the top 28 features.

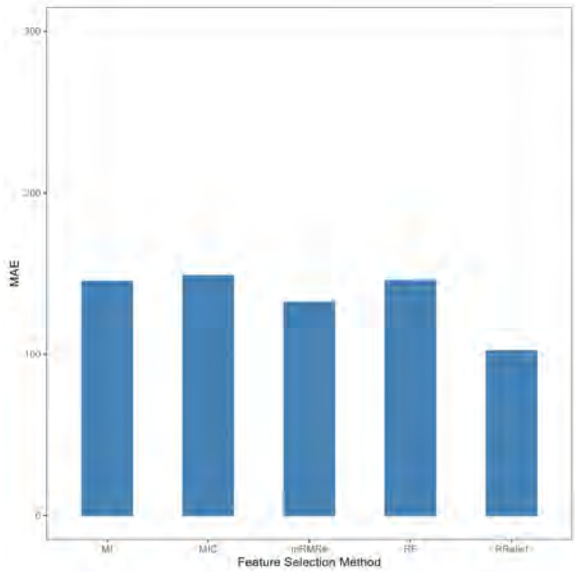
Figure 4.14: Solar generation forecast performance with feature selection for the 20, 24 and 28 top features.

and 40 MRMR achieved a $132.15MW$ and $139.29MW$ MAE, respectively. The Random Forest algorithm also showed robustness and accuracy as it achieved the second best results of the scenarios with the 32 and 40 features. Specifically, for the 32 features it achieved $134.75MW$ MAE.

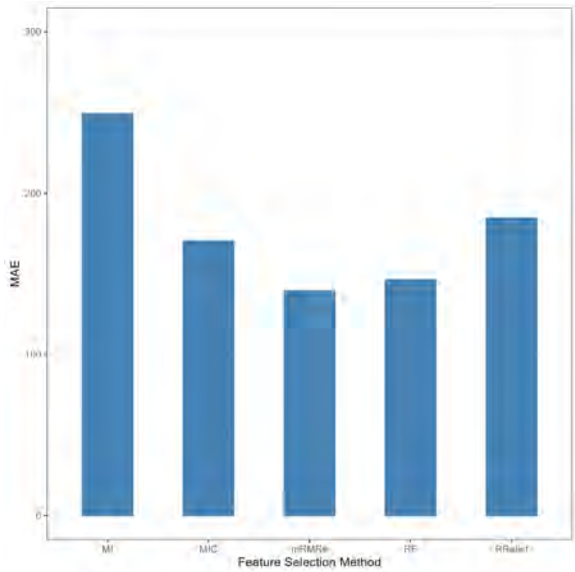
Finally, for the last two scenarios of 44 and 48 features, the results can be seen in Figure 4.16. Comparing the two scenarios, RReliefF outperformed all the other four feature selection methods by achieving $105.77MW$ and $106.33MW$ MAE for the 44 and 48 scenarios, respectively. For the 48 features scenario, Mutual Information achieved the second best results with a $142.33MW$ MAE. Finally, MRMR showed promising results for the 44 scenario as it achieved $141.76MW$ MAE.



(a) MAE of the top 32 features.



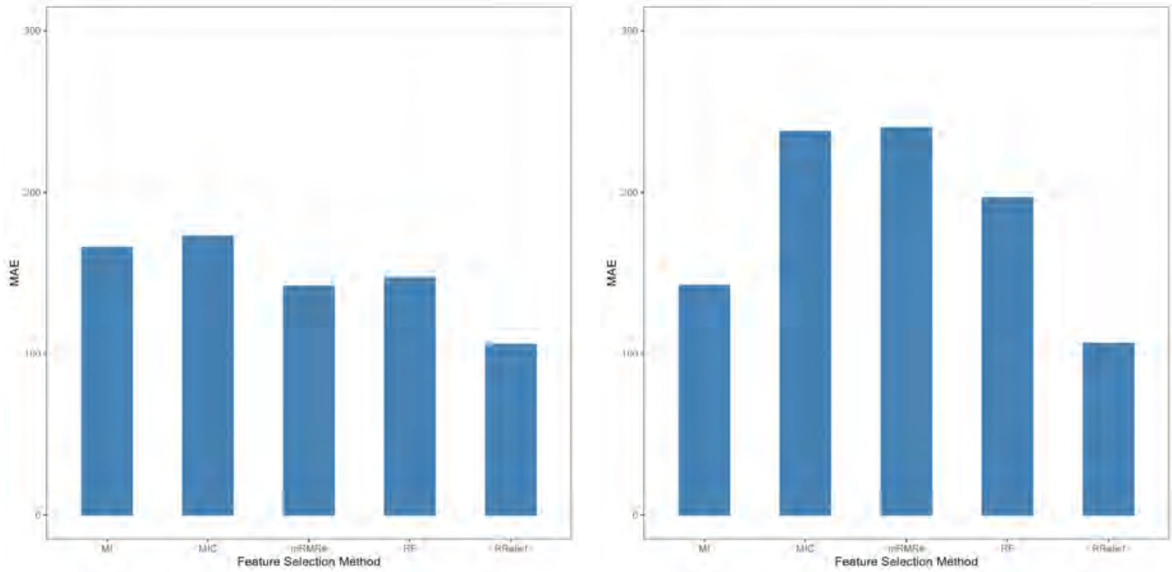
(b) MAE of the top 36 features.



(c) MAE of the top 40 features.

Figure 4.15: Solar generation forecast performance with feature selection for the 32, 36 and 40 top features.

Based on the aforementioned simulations, results from multiple feature selection methods with various features we compared. A summary of the top scenarios is shown in Table 4.9. Based on the two evaluation errors, MAE and RMSE, a comparison of these features is completed. The total amount of time the neural network and the feature selection method took to calculate is also shown in the last column.



(a) MAE of the top 44 features.

(b) MAE of the top 48 features.

Figure 4.16: Solar generation forecast performance with feature selection for the 44 and 48 top features.

Table 4.9: Overall performance of the best feature selection methods.

Method	Num. of Features	MAE (MW)	RMSE (MW)	Calc. Time (sec)
Original	54	175.32	418.68	51.2
MI	48	142.33	223.79	55.9
RReleifF	48	106.33	211.28	53.1
RReleifF	44	105.77	219.73	49.6
MRMR	40	139.29	279.59	55.3
RReliefF	36	102.18	208.08	51.7
MRMR	36	132.15	239.36	59.3
RF	28	156.12	293.34	56.4
RReliefF	20	159.21	269.88	46.3

By analyzing the results from Table 4.9 it can be seen that RReliefF shows the most promising performance for multiple scenarios, while at the same time it achieves the most rapid calculation time. RReleifF is able to minimize the prediction error, by better capturing

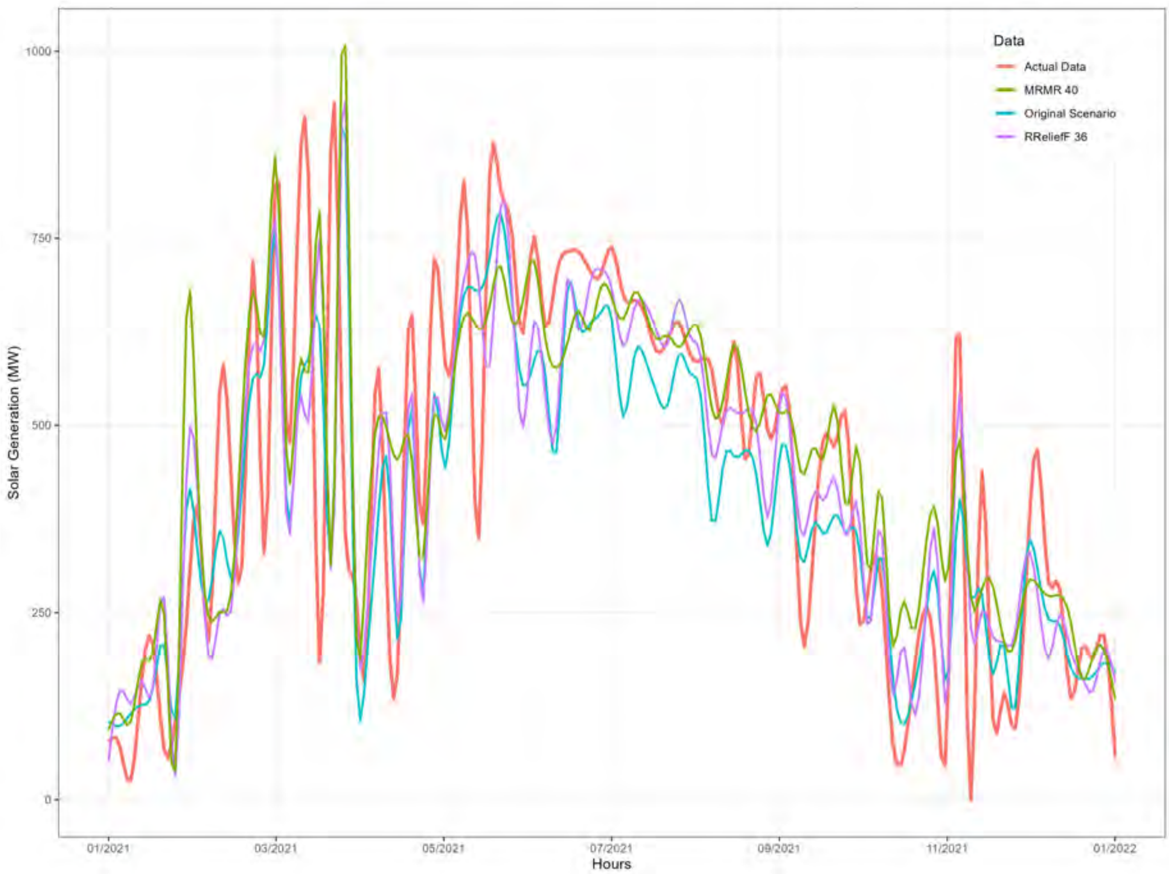


Figure 4.17: Hourly actual data compared with the original scenario, RReliefF and MRMR with 36 and 40 features.

the highly volatility of multiple variables in the data set. Particularly, for the scenario with the 36 features RReleifF achieves the best MAE and RMSE from all the other methods, with a very small penalty of time. In general RReleifF has one of the smallest calculation times compared with other feature selection methods, like Mutual Information or Random Forest. Furthermore, MRMR was the second best algorithm, with the scenario of 40 features. Finally, an amazing result was achieved by RReliefF for a bare minimum of 20 features, which also achieved the lowest calculation time than the rest of simulations.

In Figure 4.17 the hourly actual data is presented and compared with the original scenario without any feature selection method, the RReliefF with 36 features and the MRMR algorithm with 40 features. Since both methods outperform the initial scenario, there is very little difference between them. RReliefF performs marginally better in predicting changes in solar energy over shorter time spans, while MRMR appears to perform significantly better with higher values of solar generation over longer time spans. Overall, the RReliefF shows the best robustness and accuracy from all the other feature selection methods.

In Table 4.10 the actual features for the best scenario of the RReliefF with the 36 features is presented. The majority of the most relevant features are the historical soalr generation data of the previous day and the day before, mostly on the day hours. Furthermore, a big importance seems to have the maximum and minimum temperature of the forecasted day and the day before, as it plays a major role in the solar energy production. Finally, as expected the average solar irradiation was considered as a relevant feature as it is the most important factor for solar energy.

Table 4.10: Most relevant features based on the RReliefF method.

Input	Feature	Input	Feature
7	H6,d-1	35	H11,d-2
8	H7,d-1	36	H12,d-2
9	H9,d-1	37	H13,d-2
10	H10,d-1	38	H14,d-2
11	H11,d-1	39	H15,d-2
12	H12,d-1	40	H16,d-2
13	H13,d-1	41	H17,d-2
14	H14,d-1	42	H18,d-2
15	H15,d-1	43	H19,d-2
16	H16,d-1	44	H20,d-2
17	H17,d-1	45	H21,d-2
18	H18,d-1	46	H22,d-2
19	H19,d-1	49	Min. Temperature,d
20	H20,d-1	50	Max. Temperature,d
31	H7,d-2	51	Min. Temperature,d-1
32	H8,d-2	52	Max. Temperature,d-1
33	H9,d-2	53	Average Solar Irradiation,d
34	H10,d-2	54	Average Solar Irradiation,d-1

4.6 Summary

In this chapter the simulation results were presented. Several cases were performed for load, wind and solar generation, comparing simulations between the original scenario without any feature selection method and models with different amount of features. The results showed that for the load generation the RReliefF and the Random Forest algorithm achieved the best performance with 44 and 28 features, respectively. For the wind generation forecast the MRMR and the RReliefF algorithm outperformed the rest of the methods with 36 available features. Finally for the solar generation model, the RReliefF and the MRMR algorithm achieved the least amount of errors for 36 and 40 features, respectively.

Chapter 5

Conclusion

5.1 Summary

The main objective for this thesis was to perform an analysis for different kind of predictions, like the load, wind and solar forecasts, and evaluate the performance of multiple feature selection methods and their importance. The use of a multi layered perceptron was used in order to create the prediction model, with one hidden layer which was comprised of 24 neurons. Furthermore, three year of historical load, generation and weather data, from the year of 2019, 2020 and 2021 was used to train and test the prediction model.

Multiple feature selection methods were tested during this thesis. For load, wind, and solar generation, several cases were completed, comparing simulations of the original scenario without any feature selection method and models with various numbers of features. The outcome, demonstrated that the RReliefF and Random Forest algorithms, with 44 and 28 features, respectively, achieved the best performance for the load generation models. The MRMR and RReliefF algorithms, which each had 36 features, performed better than the other methods for predicting the wind generation. For the solar generation, the RReliefF and the MRMR algorithms, produced the least amount of errors for 36 and 40 features respectively. Overall, the RReliefF algorithm showed the best performance cross prediction models. Almost for all three type RReliefF achieved the best outcome while at the same time minimazing the computational time required.

5.2 Future extensions

In this thesis the topic of the best performing feature selection method was discussed. Although, feature selection is a very important task concerning the energy prediction and the neural networks, further investigation is needed to identify the weaknesses and the advantages of each method for multiple types of prediction models. A case study about the impact of the performance of feature selection method based on the type or structure of a Neural Network should be further investigated and in order to identify possible weaknesses for each method.

Bibliography

[1] R.B. Duffey et al. "Canadian Solutions to Global Energy and Environment Challenges: Green Atoms". In: *2006 IEEE EIC Climate Change Conference*. 2006, pp. 1–7. DOI: `10.1109/EICCCC.2006.277186`.

[2] Tao Hong et al. "Energy Forecasting: A Review and Outlook". In: *IEEE Open Access Journal of Power and Energy* 7 (2020), pp. 376–388. DOI: `10.1109/OAJPE.2020.3029979`.

[3] Cameron W. Potter, Eric Grimit, and Bart Nijssen. "Potential benefits of a dedicated probabilistic rapid ramp event forecast tool". In: *2009 IEEE/PES Power Systems Conference and Exposition*. 2009, pp. 1–5. DOI: `10.1109/PSCE.2009.4840109`.

[4] A. Lahouar and J. Ben Hadj Slama. "Hour-ahead wind power forecast based on random forests". In: *Renewable Energy* 109 (2017), pp. 529–541. ISSN: 0960-1481. DOI: `https://doi.org/10.1016/j.renene.2017.03.064`.

[5] Tiago Pinto et al. "Short-term wind speed forecasting using Support Vector Machines". In: *2014 IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments (CIDUE)*. 2014, pp. 40–46. DOI: `10.1109/CIDUE.2014.7007865`.

[6] Mehmet Yesilbudak, Seref Sagiroglu, and Ilhami Colak. "A novel implementation of kNN classifier based on multi-tupled meteorological input data for wind power prediction". In: *Energy Conversion and Management* 135 (2017), pp. 434–444. ISSN: 0196-8904. DOI: `https://doi.org/10.1016/j.enconman.2016.12.094`.

[7] Danting Dong, Zhihao Sheng, and Tiancheng Yang. "Wind Power Prediction Based on Recurrent Neural Network with Long Short-Term Memory Units". In: *2018 International Conference on Renewable Energy and Power Engineering (REPE)*. 2018, pp. 34–38. DOI: `10.1109/REPE.2018.8657666`.

[8] Kathrine Lau Jørgensen and Hamid Reza Shaker. "Wind Power Forecasting Using Machine Learning: State of the Art, Trends and Challenges". In: *2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE)*. 2020, pp. 44–50. DOI: 10.1109/SEGE49949.2020.9181870.

[9] Ajith Abraham. "Artificial Neural Networks". In: *Handbook of Measuring System Design*. John Wiley & Sons, Ltd, 2005. Chap. 129. ISBN: 9780471497394. DOI: https://doi.org/10.1002/0471497398.mm421. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471497398.mm421. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471497398.mm421.

[10] Shiv Ram Dubey, Satish Kumar Singh, and Bidyut Baran Chaudhuri. *Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark*. 2022. arXiv: 2109.14545 [cs.LG].

[11] Chigozie Nwankpa et al. *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. 2018. arXiv: 1811.03378 [cs.LG].

[12] Leo Pauly et al. "Deeper Networks for Pavement Crack Detection". In: *Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC)*. Ed. by Min-Yuan (National Taiwan University of Science Cheng et al. Taipei, Taiwan: Tribun EU, s.r.o., Brno, July 2017, pp. 479–485. ISBN: 978-80-263-1371-7. DOI: 10.22260/ISARC2017/0066.

[13] Eduardo Ogasawara et al. "Adaptive Normalization: A novel data normalization approach for non-stationary time series". In: *The 2010 International Joint Conference on Neural Networks (IJCNN)*. 2010, pp. 1–8. DOI: 10.1109/IJCNN.2010.5596746.

[14] J.M. Twomey and A.E. Smith. "Performance measures, consistency, and power for artificial neural network models". In: *Mathematical and Computer Modelling* 21.1 (1995), pp. 243–258. ISSN: 0895-7177. DOI: https://doi.org/10.1016/0895-7177(94)00207-5. URL: https://www.sciencedirect.com/science/article/pii/0895717794002075.

[15] Xianlun Tang et al. "Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network". In: *Electric Power Systems Research* 205 (2022), p. 107761. ISSN: 0378-7796. DOI: `https://doi.org/10.1016/j.epsr.2021.107761`. URL: `https://www.sciencedirect.com/science/article/pii/S0378779621007422`.

[16] Isabelle Guyon and André Elisseeff. "An Introduction to Variable and Feature Selection". In: *J. Mach. Learn. Res.* 3.null (Mar. 2003), pp. 1157–1182. ISSN: 1532-4435.

[17] Mohamed Bennasar, Yulia Hicks, and Rossitza Setchi. "Feature selection using Joint Mutual Information Maximisation". In: *Expert Systems with Applications* 42.22 (2015), pp. 8520–8532. ISSN: 0957-4174. DOI: `https://doi.org/10.1016/j.eswa.2015.07.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0957417415004674`.

[18] Tanwalai Panapongpakorn and David Banjerdpongchai. "Short-Term Load Forecast for Energy Management System Using Neural Networks with Mutual Information Method of Input Selection". In: *2019 SICE International Symposium on Control Systems (SICE ISCS)*. 2019, pp. 9–15. DOI: `10.23919/SICEISCS.2019.8758714`.

[19] Aston Zhang et al. *Dive into Deep Learning*. 2023. arXiv: `2106.11342 [cs.LG]`.

[20] Nima Amjady and Farshid Keynia. "A new prediction strategy for price spike forecasting of day-ahead electricity markets". In: *Applied Soft Computing* 11.6 (2011), pp. 4246–4256. ISSN: 1568-4946. DOI: `https://doi.org/10.1016/j.asoc.2011.03.024`. URL: `https://www.sciencedirect.com/science/article/pii/S1568494611001207`.

[21] Song Li, Peng Wang, and Lalit Goel. "A Novel Wavelet-Based Ensemble Method for Short-Term Load Forecasting with Hybrid Neural Networks and Feature Selection". In: *IEEE Transactions on Power Systems* 31.3 (2016), pp. 1788–1798. DOI: `10.1109/TPWRS.2015.2438322`.

[22] Hanchuan Peng, Fuhui Long, and C. Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238. DOI: `10.1109/TPAMI.2005.159`.

[23] Oveis Abedinia, Nima Amjady, and Hamidreza Zareipour. "A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems". In: *IEEE Transactions on Power Systems* 32.1 (2017), pp. 62–74. DOI: `10.1109/TPWRS.2016.2556620`.

[24] Shaoqian Pei et al. "Multi-Step Ahead Short-Term Load Forecasting Using Hybrid Feature Selection and Improved Long Short-Term Memory Network". In: *Energies* 13.16 (2020). ISSN: 1996-1073. DOI: `10.3390/en13164121`. URL: `https://www.mdpi.com/1996-1073/13/16/4121`.

[25] Jicheng Liu and Yinghuan Li. "Study on environment-concerned short-term load forecasting model for wind power based on feature extraction and tree regression". In: *Journal of Cleaner Production* 264 (Apr. 2020), p. 121505. DOI: `10.1016/j.jclepro.2020.121505`.

[26] Yi Xuan et al. "Multi-Model Fusion Short-Term Load Forecasting Based on Random Forest Feature Selection and Hybrid Neural Network". In: *IEEE Access* 9 (2021), pp. 69002–69009. DOI: `10.1109/ACCESS.2021.3051337`.

[27] Zha Wenting et al. "Short-term wind power forecasting model based on random forest algorithm and TCN". In: *2021 40th Chinese Control Conference (CCC)*. 2021, pp. 5776–5781. DOI: `10.23919/CCC52363.2021.9550052`.

[28] Peter Roßbach. "Neural Networks vs. Random Forests – Does it always have to be Deep Learning?" In: 2018.

[29] David N. Reshef et al. "Detecting Novel Associations in Large Data Sets". In: *Science* 334.6062 (2011), pp. 1518–1524. DOI: `10.1126/science.1205438`. eprint: `https://www.science.org/doi/pdf/10.1126/science.1205438`. URL: `https://www.science.org/doi/abs/10.1126/science.1205438`.

[30] Ziyue Wang and Xianchao Huang. "Short-Term Load Forecasting Based on Maximal Information Coefficient Correlation Analysis and LSTM Optimization". In: *2022 4th International Conference on Power and Energy Technology (ICPET)*. 2022, pp. 602–607. DOI: `10.1109/ICPET55165.2022.9918352`.

[31] Xiaotong Yao, Xiaoli Fu, and Chaofei Zong. "Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost". In: *IEEE Access* 10 (2022), pp. 75257–75268. DOI: 10.1109/ACCESS.2022.3192011.

[32] Kenji Kira and Larry A. Rendell. "A Practical Approach to Feature Selection". In: *ML Workshop*. 1992.

[33] Igor Kononenko. "Estimating attributes: Analysis and extensions of RELIEF". In: *Machine Learning: ECML-94*. Ed. by Francesco Bergadano and Luc De Raedt. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 171–182.

[34] Igor Kononenko, Edvard Šimec, and Marko Robnik-Šikonja. "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF". In: *Applied Intelligence* 7.1 (Jan. 1997), pp. 39–55. ISSN: 0924-669X. DOI: 10.1023/A:1008280620621. URL: https://doi.org/10.1023/A:1008280620621.

[35] Amir Rafati, Mahmood Joorabian, and Elaheh Mashhour. "An efficient hour-ahead electrical load forecasting method based on innovative features". In: *Energy* 201 (2020), p. 117511. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2020.117511. URL: https://www.sciencedirect.com/science/article/pii/S0360544220306186.

[36] Irena Koprinska, Mashud Rana, and Vassilios G. Agelidis. "Correlation and instance based feature selection for electricity load forecasting". In: *Knowledge-Based Systems* 82 (2015), pp. 29–40. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2015.02.017. URL: https://www.sciencedirect.com/science/article/pii/S0950705115000714.

[37] Ioannis P. Panapakidis, Aggelos S. Bouhouras, and Georgios C. Christoforidis. "A Comparison of Feature Selection Techniques for Neural Network Based Load Forecasting". In: *2019 54th International Universities Power Engineering Conference (UPEC)*. 2019, pp. 1–6. DOI: 10.1109/UPEC.2019.8893447.

[38] Alfredo Esclapez et al. "Automatic Selection of Temperature Variables for Short-Term Load Forecasting". In: *Sustainability* 14 (Oct. 2022), p. 13339. DOI: 10.3390/su142013339.

[39] Ioannis Panapakidis et al. “Renewable energy sources generation forecasting in aggregated energy system level”. In: *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*. 2021, pp. 1–6. DOI: 10.1109/EEEIC/ICPSEurope51590.2021.9584682.

[40] *ENTSOE, Transparency Platform*. https://transparency.entsoe.eu/.

[41] *Prediction Of Worldwide Energy Resources (POWER)*. https://power.larc.nasa.gov/.

[42] *ELETAEN: Greece’s Wind Capacity Statistics*. https://eletaen.gr/hwea-wind-statistics/. (accessed: 25.05.2022).

[43] *DAPEEP SA: Bulletin of APE*. https://www.dapeep.gr/dimosieuseis/sinoptiko-pliroforiako-deltio-ape. June 2022.