Forecasting Electricity Prices using Artificial Neural Networks

by

Shivani Goyal

This thesis has been submitted in partial fulfillment for the degree of Master of Science in Data Science and Analytics

in the
Faculty of Engineering and Science
Department of Mathematics

August 2020

Declaration of Authorship

This report, Forecasting Electricity Prices using Artificial Neural Network, is submitted in partial fulfillment of the requirements of Master of Science in Data Science and Analytics at Cork Institute of Technology. I, Shivani Goyal, declare that this thesis titled, Forecasting Electricity Prices using Artificial Neural Networkand the work represents substantially the result of my own work except where explicitly indicated in the text. This report may be freely copied and distributed provided the source is explicitly acknowledged. I confirm that:

- This work was done wholly or mainly while in candidature Master of Science in Data Science and Analytics at Cork Institute of Technology.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at Cork Institute of Technology or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project report is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Shivani Goyal

Date: 23rd August, 2020

CORK INSTITUTE OF TECHNOLOGY

Abstract

Faculty of Engineering and Science
Department of Mathematics

Master of Science

by Shivani Goyal

The introduction of Ireland's new wholesale electricity market called I-SEM (Integrated Single Electricity Market) has reformed the way electricity market functions. Forecasting day-ahead electricity prices accurately have become an essential task to facilitate the bidding strategies of traders and suppliers of electricity. This thesis aims to use mutual information and correlation feature selection methods to find the best explanatory variables of interest. The models to predict electricity prices are build using machine learning algorithms such as random forest (RF), extreme-gradient boosting (XGBoost), support vector machine (SVM), linear regression (LR), and artificial neural network (ANN) with and without the selected features achieved by proposed feature selection techniques. Furthermore, the performance of the models is compared based on evaluation metrics, mean absolute error (MAE).

Acknowledgements

I would like to thank and acknowledge people who have guided me and supported me throughout this research.

My first and foremost gratitude to my supervisor **Dr. David Hawe** for constant support, patience, guidance and his timely feedback that helped me to write this thesis. Secondly, I thank **CapSpire** to provide me this opportunity to work on such an interesting project and my industry mentor **Mr. Joseph Collins** for helping me to understand the problem statement and providing me the industry domain knowledge. During this tough time of epidemic and lockdown, **CIT** was of immense support.

I'm thankful to my family and friends for the constant support and motivation. Also, my special thanks to my friend, **Phalguni Rathod**, **Neha Bendre**, **Sunita P.** without them passing through my masters would be difficult. Thank you **Rajat Upadhayay**, **Prerna Thakur**, **Abhishek Munjal** & **Rahul Tripathi** for being my best support system in life.

Contents

List of Figures

List of Tables

Dedicated to my Parents

Mr. Lakhmi Chand Goyal and Mrs. Poonam Goyal

Siblings

Pallavi Goyal and Deepak Goyal

Thanks for the Unconditional Love, Support and Believing in me

——Shri Ganeshay Namah——

Chapter 1

Introduction

1.1 Motivation

Since the liberalization of the electricity markets in 1996, the dynamics of trading electricity has been evolved over the time [1]. Forecasting electricity prices has become one of the crucial tasks for the participants in the electricity market [2]. This has raised competition among the energy sector industries and has introduced several opportunities to generators and suppliers of electricity. Moreover, this has reshaped the way market participants (suppliers, generators, and traders) trade the electricity [3].

A generator, utility, or large industrial customer who can forecast unpredictable market prices with a fair degree of precision may change their bidding strategy and their output or consumption schedule to reduce the risk of loss or increase the profits in day-to-day trading. The ability to forecast electricity prices has also hugely benefited the energy supply companies. The suppliers can optimize their strategies and can schedule their operations in accordance with the low price zones depending on the price forecast [2].

Electricity is a unique commodity as storing it in large amount is difficult. The generation of electricity must be balanced with its demand[4]. The electricity demand varies on an hourly basis. Hence, this market is highly volatile in nature. In recent years, the production of electricity from renewable energies that strongly depend on weather such as wind has been increased significantly [1]. Therefore, it is an influential factor in forecasting electricity prices. The distinct characteristics of electricity make the electricity market different from other markets. There are several important characteristics of electricity prices such as sharp price spikes, high volatility, and seasonality that make forecasting electricity prices a challenging task [2].

The machine learning and deep learning, the sub-fields of artificial intelligence are the techniques that enable the machine to improve the tasks with experience. The models built using these techniques are widely used for classification and regression problems. The behavior of machine learning and deep learning models are sensitive to the data passed to them. Hence, feature selection is considered as one of the important techniques while preparing the data .

Suitable selection of input variables is a key element to the success of any load or price forecast method [5]. Therefore, the motivation of this thesis is to determine the best input candidates by applying recognized feature selection techniques that are aimed to remove the redundant information from the data and to build models using machine learning algorithms and artificial neural network (ANN) using with and without the selected features to predict the electricity prices.

1.2 CapSpire

CapSpire is a global consulting and solutions company founded in 2009. They provide the best expert advice and deep business solutions to the complex problems for commodity-focused organizations. Furthermore, provides services in various markets such as commodities/agribusiness, crude, natural gas, natural gas liquids, power, and refined products. Recently, the company has actively engaged in developing and implementing technology solutions for the integrated single energy market (I-SEM) [6].

1.3 Contribution

In the literature on electricity price forecasting, selecting the appropriate features is an important task [1]. This thesis involves the studying of two different feature selection techniques and the implementation of various models based on machine learning and artificial neural network algorithms. The best model to predict the day-ahead electricity prices is evaluated by comparing the models based on the performance metric.

1.3.1 Research Questions

- Which are the features affecting the electricity prices and how to select them using feature selection techniques?
- Which approach: Machine Learning or Deep Learning gives better results?

- Does feature selection techniques give improved results?
- What metric should be used for evaluating the models?

1.4 Structure of This Document

This document has been structured chapter wise as visually described in figure 1.1.

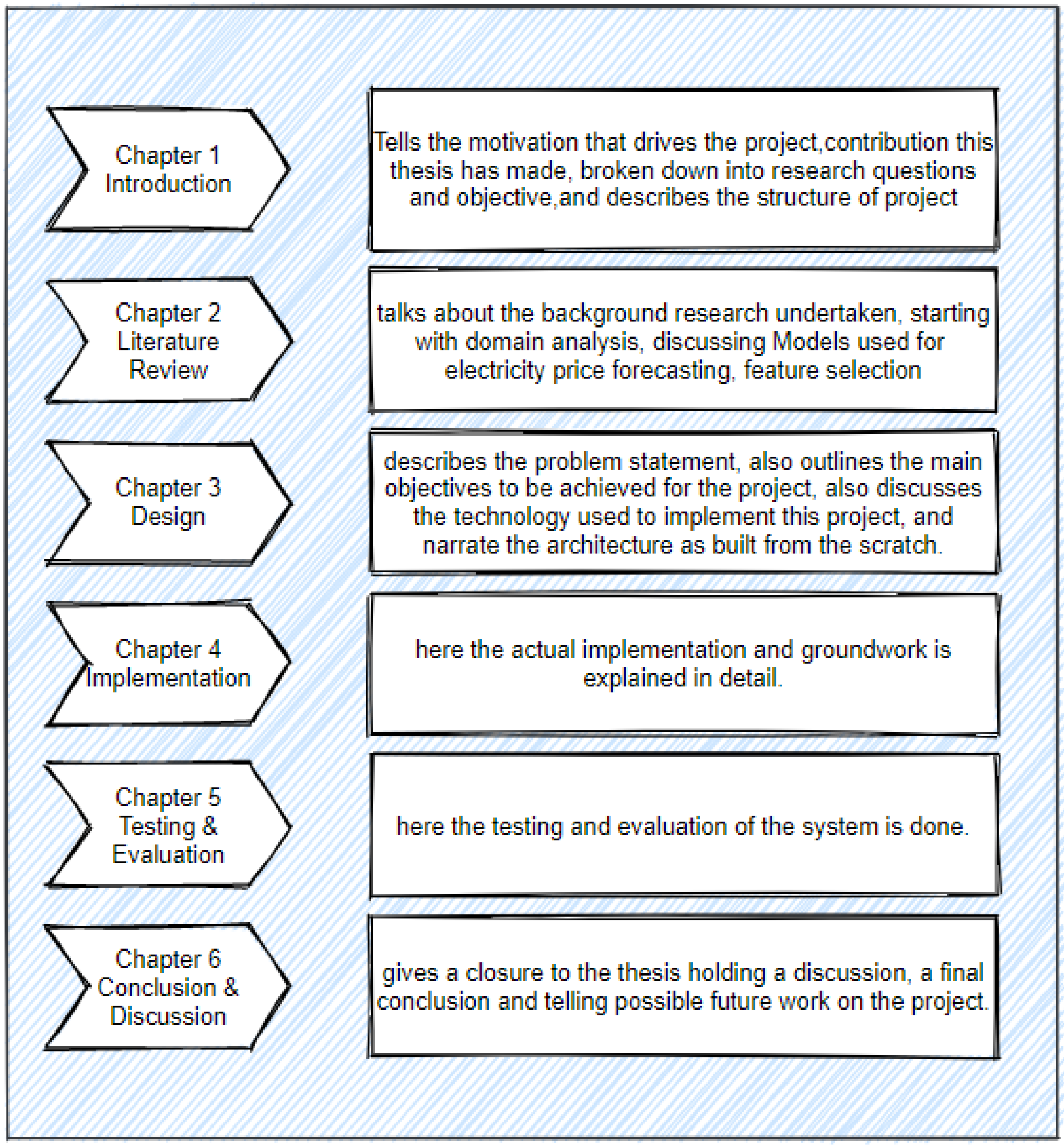

Figure 1.1: Structure of the Document

Chapter 2

Literature Review

2.1 Domain Analysis

2.1.1 European Electricity Market

Traditionally, the infrastructure of the gas and electricity market of Europe was built on distinct markets depending on a country-by-country basis and each market has its own set of laws and behaviors to regulate the electricity to homes and businesses. Over the years Europe has passed several laws that seeks to achieve a single and integrated European electricity market [7]. The single electricity market (SEM) was first introduced on 1st November, 2007 for the island of Ireland and is regulated jointly by Commission for Energy Regulation (CER), and its counterpart in Belfast, the Utility Regulator. Furthermore, in 2018 the Integrated Single Electricity Market (I-SEM) was launched to bring the Irish electricity market in line with the rest of Europe [8].

2.1.1.1 Single Electricity Market (SEM)

The SEM was designed to accommodate the minimum-cost source of electricity generation to meet customer demand across the island at any time, and maximizing the sustainability and reliability over the long term. The main objective for launching the SEM in Ireland's electricity market were as follows [4]:

- To deliver transparent and efficient wholesale electricity prices.
- It bring new investments in modern power production, such as gas-fired plants and wind farms.

- It ensures the security of electricity supply and provides environmental benefits.
- To dispatch the cheapest generators across the island to meet the electricity demand.

The features of the electricity market are mostly dependent on the characteristics of electricity, as electricity generation cannot be easily stored in large amounts. Therefore, bidding of costs in advance of real-time must be balanced with demand and generators. Figure 2.1 shows the high level design of SEM market. SEM is designed on four layers: Generators, Pool layer, Suppliers, and Customers. Generators are the companies that produces the electricity and the suppliers are responsible for supplying the power to households and businesses (customers). The backbone of this structure is the pool layer as it is the centralized and mandatory all-island wholesale market layer, where the generators and suppliers bid their short-run costs for each half-hour of the following day.

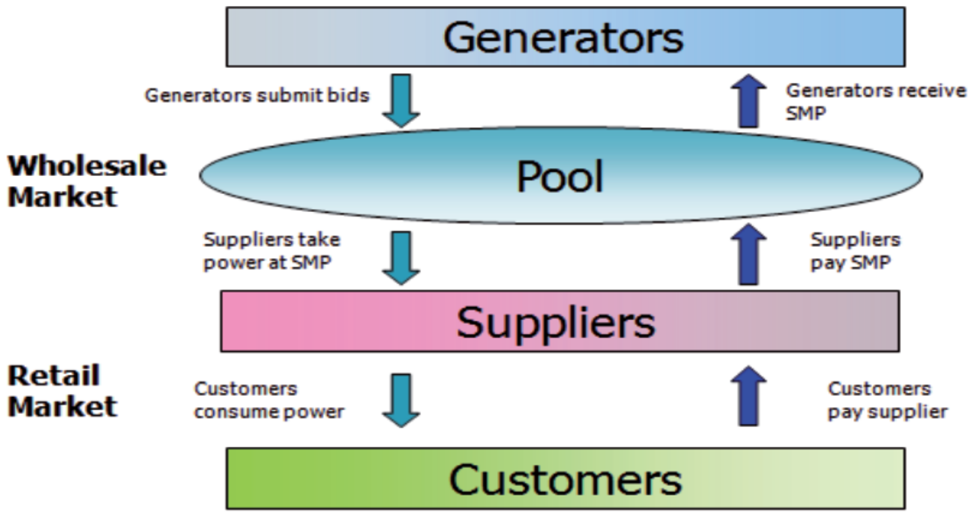

FIGURE 2.1: Design of Single Energy Market(SEM)
[4]

The system marginal price (SMP) for each half-hour trading period, is the final market price at which the electricity is traded and is dependent on the set of generators costs and electricity demand. This price is determined by the single electricity market operator (SEMO - a joint partnership between the transmission system operators EirGrid and SONI in Republic of Ireland and Northern Ireland, respectively) that uses the stack of generation cost bids, to meet customer demand across the island [4]. Figure 2.2 shows the process of estimating the SMP. The information at x-axis and y-axis is the amount of electricity produced by generators and the generator cost bids respectively. These bids are stacked by the cheapest generator costs. The generators that are able to meet the demand are paid marginal price and known as efficient generators. The "out of merit" consists the expensive or insufficient generators which are not able to to meet demand in the half hour and hence, they are not paid SMP.

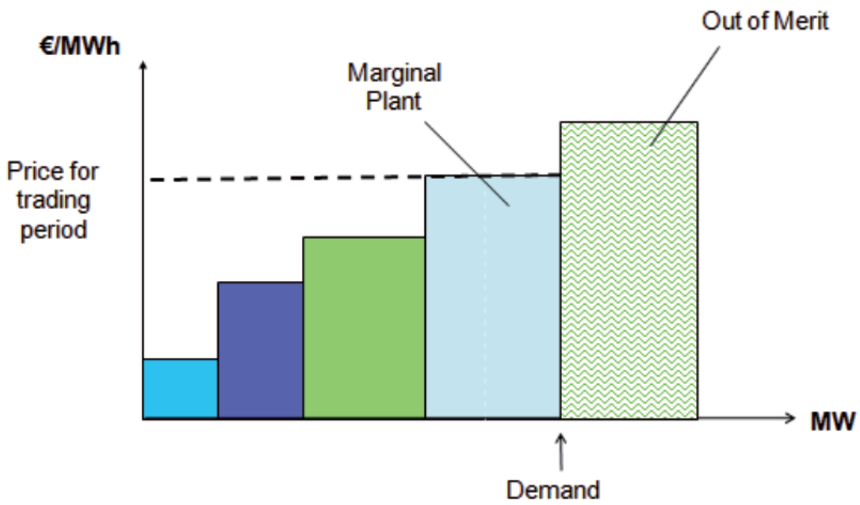

FIGURE 2.2: Estimating System Marginal Price (SMP) [4]

2.1.1.2 Integrated Single Electricity Market (I-SEM)

The main objective of I-SEM is to enable the free flow of energy throughout the Europe. It is a new wholesale electricity market arrangement for the Republic of Ireland and Northern Ireland that integrates the electricity market with the European electricity market by the transmission of the electricity using the cross-borders inter-connectors. It is the major reformation of the market in the history of the electricity market and opened several new opportunities for the participants. The key considerations while designing the I-SEM were as follows [9]:

- How energy will be purchased and sold.
- How generators are adequately compensated for availability.
- Arrangements for market liquidity and forward trading.
- Policies to operate the market.

The implementation of an Internal Energy Market (IEM) for electricity and gas, enables the cross-border trade of electricity across Europe. In Republic of Ireland and Northern Ireland, the transmission network is connected to European markets through Scotland

& Wales via the Moyle interconnector (Northern Ireland to Scotland) and the East-West interconnector (Republic of Ireland to Wales) [9].

Figure 2.3 shows the IEM containing 38 cross-borders interconnectors connecting the 20 coupled markets (including Republic of Ireland and Northern Ireland) and having around 3,000 terawatts (TW) of generating capacity across Europe [3].

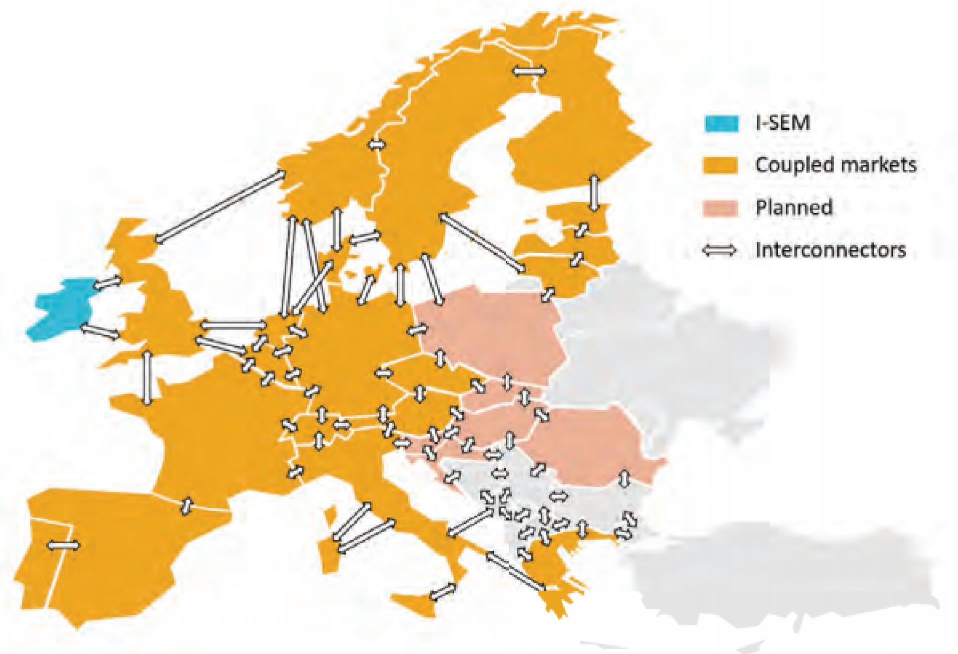

FIGURE 2.3: Internal Energy Market (IEM) of Europe showing coupled markets and cross-border interconnectors

[9]

"Coupled Markets" term is used for the countries of Europe whose energy markets are integrated by ISEM. The coupled markets follows the European target model that consists of a common set of rules and standardised wholesale trading arrangements. The features of target model as follows [3]:

- Establish movement of electricity through geographic regions and use an intermediate price coupling algorithm for organizing all the estimated markets (day-ahead and intra-day market).
- A near real-time buying and selling of energy within regions and across boundaries.
- To handle the price difference that can occur because of congestion in transmission, a hedging facility is setup.
- To promote balancing among nearby regions, there are integrated balancing arrangements to facilitate trading with them.

I-SEM consists of three distinct markets: Day-Ahead, the Intraday, and the Balancing Market. Day-Ahead market and the intraday market are also reffered to as ex-ante market as this occurs before the market closes [8]. However, the balancing market takes place after the trading is ceased. The balance market is more volatile than the other two markets. Figure 2.4 shows the timelines for submission of orders, market clearing, publishing market schedules, and settlement for each market.

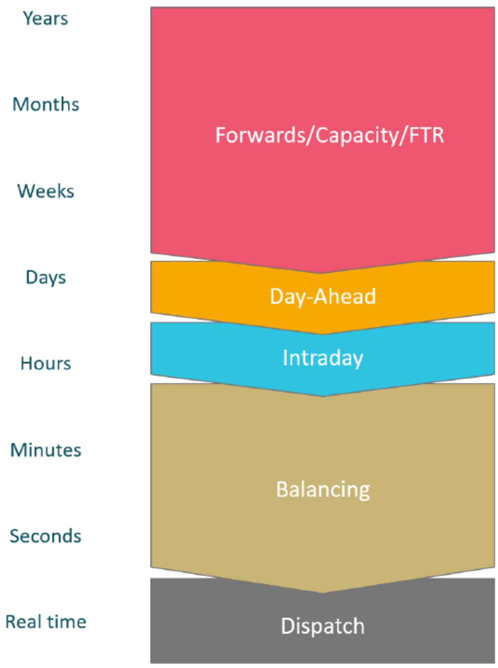

FIGURE 2.4: I-SEM's each market timeframe
[9]

Generators are the companies that produce the electricity and trade it to the suppliers and further sell this electricity to customers. Both generators and suppliers submit orders to buy or sell electricity within these three markets (day-ahead, intraday & balance market) time framework.

1. Day-Ahead Market (DM): Day-ahead market opens a day before the electricity is supplied to the customers. For the day 'D', delivering hourly electricity in the delivery period [11 pm D to 10 pm D+1], the participants submit orders at 11 am on the day (D) [8].

2. Intraday Market (ID): As time passes by, generators get more certainty on how much electricity can be delivered by them. So, even after clearance of day-ahead market auction, generators and suppliers have the opportunity to continue the bid depending on the current availabilities. These auctions are held closer to delivery

time in the Intraday market (ID). Hence, the participants can submit orders for delivery from 1 hour up to 24 hours into the future [9].

3. Balance Market (BM): After the day-ahead and intraday market ceases, the balance market opens one hour before delivery. The transmission system operator (TSO) compares the demand forecast with the generation forecast and ensures that demand meets supply.

The price coupling market algorithm, Euphemia adjusts the demand curve (suppliers bid to buy electricity) and supply curve (generators bid to sell electricity) by managing trades between bidding zones. When these both curves cross, the intersection point is known as the clearance point. Hence, the price at clearance point is the market price at which the electricity is traded for that particular hour (shown in Figure 2.5, where red line is the supply curve and blue line is the demand curve).

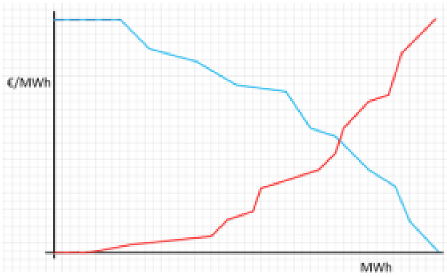

FIGURE 2.5: Euphemia Algorithm: Demand-Supply Curve [10]

The market price changes with the increase or decrease in the demand, or generation of electricity. Consider there are two zones A & B such that zone A has relatively more demand compared to B and zone B has more supply than zone A. The market price is evaluated depending on the market type, I-SEM independent market, coupled market without congestion, or coupled market with congestion.

For I-SEM independent markets, the markets are not coupled. In figure 2.6, Zone A with high demand has a higher market price while high-supply zone B has a low market price. Therefore, each zone's market price is different and have a large difference in the prices [9]. For coupled markets without congestion shown in figure 2.7, euphemia considers these coupled markets as single market and sets a single market price for all bidding zones. Some of the generation in cheaper market serves the load in the most expensive market. Hence, it leads to the efficient energy flows between the two markets and equalizes the market prices for both the regions. The coupled markets with congestion shown in figure 2.8, the market prices for each zone diverges as the capacity of inter-connectors reaches its limit.

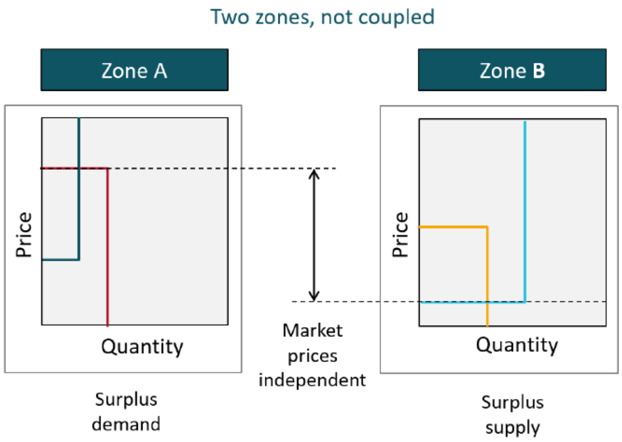

FIGURE 2.6: I-SEM independent market

[9]

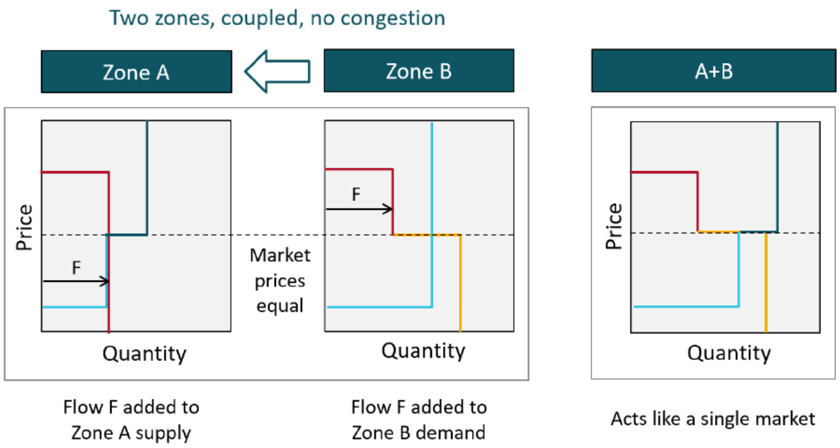

FIGURE 2.7: I-SEM coupled market without congestion

[9]

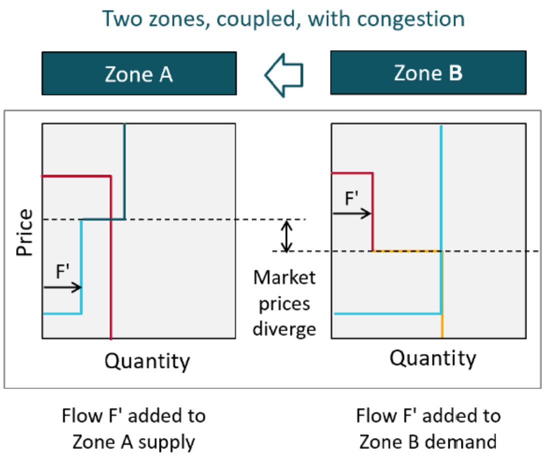

FIGURE 2.8: I-SEM coupled markets with congestion

[9]

2.2 Modelling Approaches for Electricity Price Forecasting

Electricity price forecasting (EPF) literature began to evolve in the early 2000s. D.W. Bunn outlines the main methodological problems and techniques that have been effective in the estimation of daily loads and prices in the competitive power markets in his paper titled "Forecasting loads and prices in competitive power markets" published in February 2000 [11]. He concludes that the load and price forecasting are mutually interconnected activities and further used methods like variable segmentation, neural techniques, and the combination of forecasts to predict the prices.

In an article published in IEEE Power & Energy Magazine (2006), Amjady and Hemmati explains the need to forecast the electricity prices for the short-term horizon and proposes the models for predictions after reviewing the problems related to electricity price forecasting [12]. Short-term electricity price forecasting shows the crucial significance in day-to-day market operations. In general, the short-term electricity price forecasting includes predictions from a few minutes to a few days ahead [13]. In [12], mentioned that the statistical time series techniques like auto-regressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), autoregressive (AR) are generally successful for the less periodic data. Hence, they furthered the use of hybrid (neural networks) and artificial intelligence-based methods for forecasting the load and price as such approaches are capable of observing the nonlinear behavior of electricity prices and related fundamental variables[12].

Weron (2014) reviewed several articles and papers, each proposing their classification of various approaches for analyzing and predicting electricity prices. Based on the studies the methods of forecasting the electricity prices are sub categorized into five main groups: multi-agent, fundamental, reduced-form, statistical, and computational intelligence models [13]. Fig 2.9 shows the proposed taxonomy of electricity price modeling approaches.

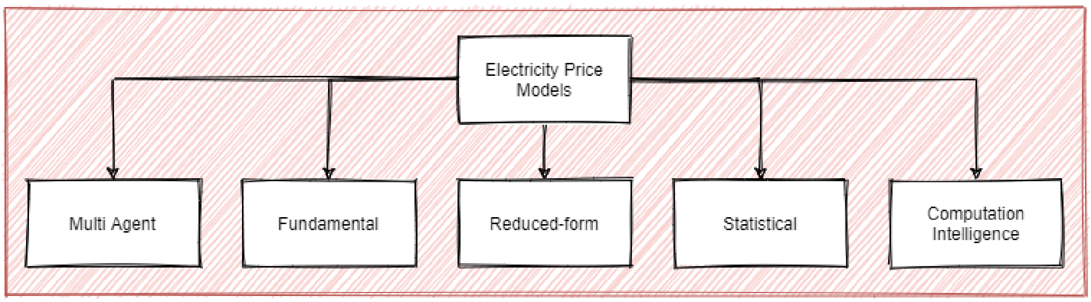

FIGURE 2.9: A taxonomy of electricity price modeling approaches [13]

1. **Multi-agent models:** These models consider system operation and predict electricity prices by matching the demand and the supply. In an article, Ventosa et al. (2005) identifies three main multi-agent modeling trends: optimization, equilibrium, and simulation models [14].

   Optimization models focus on the maximization of profit for competitors in the market and considered as production-cost models (PCM). Hence, these models work better for regulated markets having a stable structure compared to competitive electricity markets. The equilibrium models determine the overall market behavior. They take the strategic bidding practices into account and follow the game-theoretic approach. These models are useful for markets having no price history, but known supply costs and demand in the market. Simulation models are the extended version of the equilibrium models. When the problem is too complex to address within a formal equilibrium framework, then the simulation models are considered. Moreover, agent-based simulation techniques address the characteristics of electricity markets [13].

   Some common models based on the equilibrium approach are the Nash-Cournot system, supply function balancing, and strategic production cost models (Figure 2.10).

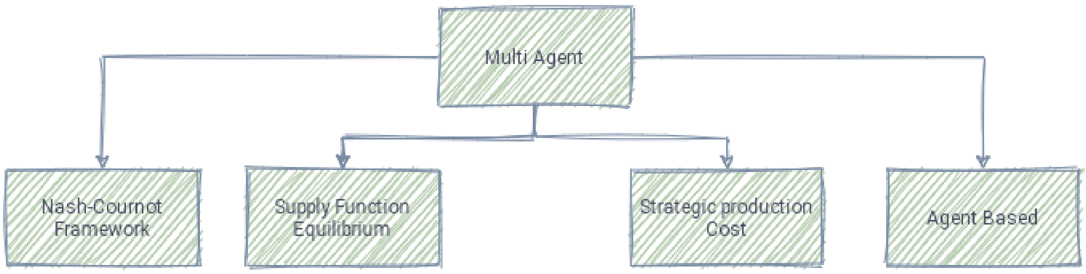

FIGURE 2.10: Tyes of multi-agent model

2. **Fundamental or structural models:** These models capture the impact of physical and economic factors on the electricity prices while trading and producing electricity. In this approach, the several fundamental drivers of forecasting electricity prices like loads, weather conditions, system parameters are modeled and predicted independently by applying statistical (like auto-regressive type), reduced-form (like jump-diffusions), or machine learning techniques (support vector machines).

   In general, parameter rich and parsimonious structural models of supply and demand are the two sub-classes of fundamental models [13], shown in Figure 2.11

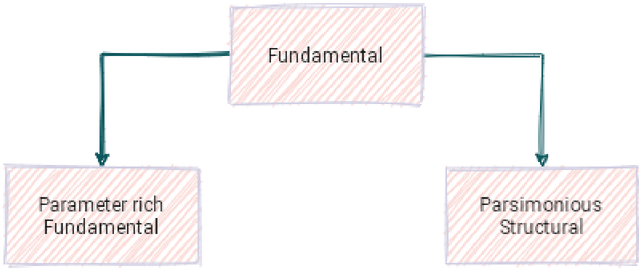

Figure 2.11: Types of Fundamental model

In a study, Vahviläinen and Pyykkönen (2005)[15] developed a parameter-rich fundamental model considering 27 scalar parameters (13 climate, 4 demand, and 10 supply parameters) for the Nordic market and demonstrate that the model is capable of capturing the market price movements on a monthly scale. This parameter-rich fundamental model predicts the electricity spot prices by calculating the weighted average of the supply power of both the hydro-power and the condensing power. Furthermore, Aïd, Campi, and Langrené (2013) [16] introduced a scarcity function which studies the deviations of the day-ahead price from the marginal fuel price. In this part of the literature, Karakatsani and Bunn [17], and Weron and Misiorek [18] considered hybrid solutions for predicting electricity prices. They exercised time series, regression, and neural network models using fundamental factors as input variables.

3. **Reduced-form models:** The ultimate objective of these models is the evaluation of the derivatives and risk management. These models focuses on to highlight the characteristics of the electricity prices such as spikes in the price, or correlation between the commodity prices rather than predicting the prices accurately [13]. Therefore, these models are considered comparatively better in terms of handling spikes than the structural and statistical models. Markov regime-switching and jump-diffusion models are the most popular models of the reduced form approach [2] (Figure 2.12).

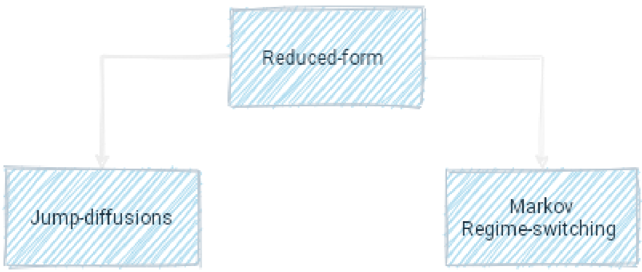

Figure 2.12: Types of Reduced Form Models

Geman and Roncoroni [19] used the mean-reverting jump-diffusion (MRJD) model. Their method incorporates both the electricity price trend and the statistical components. The drawback of jump-diffusion models is that they cannot manifest consecutive spikes for the high-frequency data. However, markov regime-switching (MRS) models work better with spike clustering observed on the daily time as well as the hourly time scale. Furthermore, Christensen, Hurn & Lindsay [20] proposed the poisson autoregressive framework and studies the persistence of spikes in electricity prices as an important factor in developing an effective model.

4. **Statistical Models:** These models are direct applications of statistical techniques for forecasting prices. In this methodology, the previous prices and previous or current values of exogenous factors (such as consumption and production of load, or weather variables) are combined mathematically to forecast the current prices [13].

   The two methods popularly used for analysing the series are additive and multiplicative models. The time series comprises of four components: the seasonal component (S, a pattern or trend that occur within one year), the trend (T, a pattern observed over a long term for years), the cyclical component (C, observed when data recorded for long period of time), and the residual component (R, occurs due to unpredictable events like recession). In multiplicative models the seasonal fluctuations increases with the series increases. However, for additive models the seasonal fluctuations remains constant with the series increases. The mathematical equation 2.1 & the equation 2.2 depicts the formula for additive model (A) and multiplicative model (M) respectively.

$$A = T + S + C + R, \tag{2.1}$$

$$M = T * S * C * R, \tag{2.2}$$

   To obtain the electricity price forecasting models, researchers did a simple substitution of prices for loads. In this field, more and more contemporary statistical, econometric, or signal processing techniques have been implemented over time. Many competitors, suppliers bid the prices and quantities for the next day's 24 hours in the day-ahead electricity price forecasting. Therefore, the first approach is to estimate all of the prices as a 24 -step-ahead forecast from a single price time series in a univariate framework. Moreover, the other option is to forecast the prices from 24 different time series as one-step-ahead forecasts in a multivariate framework [2]. The widely used statistical approach based models are shown in Figure 2.13

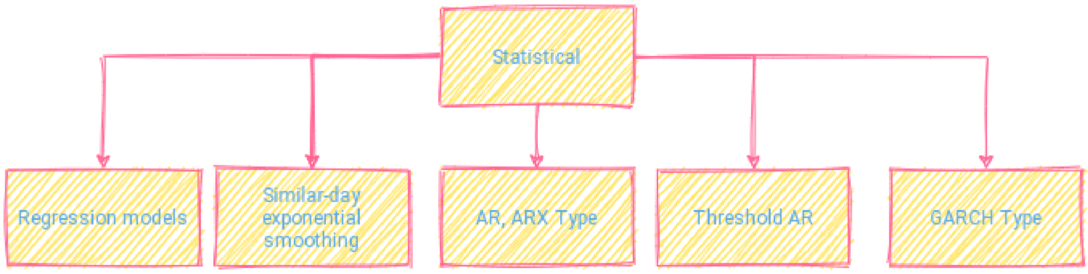

FIGURE 2.13: Types of Statistical Models

The exponential smoothing modeling approach is widely used for load forecasting relatively compared to electricity price forecasting [21], and the forecast is computed on an exponentially weighted average of past observations. Ziel and Weron [22] compared the the univariate and multivariate modeling frameworks for forecasting the day-ahead electricity prices. Nogales (2002) [23], developed a dynamic regression and transfer function models based on time series analysis for predicting prices accurately. [24] studies that adding the exogenous variables while modeling with the ARIMA model does not necessarily raise the prediction accuracy. Cruz, Muñoz, Zamora, and Espinola (2011) [25] show in their study that the double seasonal exponential smoothing models outperform the both naive and ARIMA models for hourly spot prices prediction from the Spanish market. The major drawback of the statistical approach is the limitation of modeling the non-linear behavior of electricity prices and fundamental variables.

5. **Computational Intelligence Models:** These are the techniques that incorporate elements of science, evolution, and fluidity to construct solutions that can respond to complex dynamic systems, and hence regarded as intelligent [13]. In the electricity price forecasting literature, the term artificial intelligence (AI) is widely used interchangeably with computational intelligence (CI). Soft computing, machine learning, data mining, and cybernetics such field of studies can also be incorporated with computational intelligence models [26]. These models are flexible and can manage the complexity and non-linearity. Feed forward neural network, fuzzy neural network, recurrent neural network, and support vector machines are the popular classes of computational intelligence models (Figure 2.14).

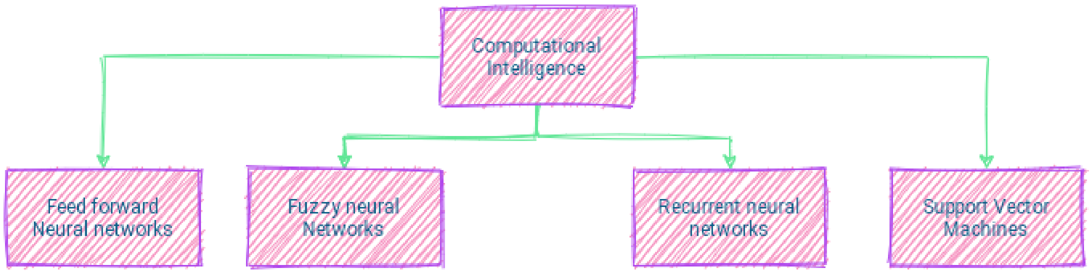

FIGURE 2.14: Types of Computational Intelligence Models

Zhang and Cheng [27], Mandal et al [28], Catalão et al [29] proposed the model using artificial neural network (ANN) to forecast 1-6 hour-ahead load and electricity prices for Australian market. The mean absolute percentage error (MAPE) for one-step ahead forecast was relatively less compared to MAPE value of six-step ahead forecast. Panapakidis and Dagoumas [30] developed the several ANN models with different architectures of layers, neurons, and number of variables by applying clustering approach to predict electricity prices and got 20 percent better results. Hence, the researchers' study shows that the artificial neural network performs better with the clustering approach and predicts electricity prices for short term horizon with minimum error.

2.2.1 Machine Learning

Machine learning is a way of solving problems that are difficult to be solved with conventional algorithmic programming methods. For instance, finding out spam or ham mails or predicting stock prices or finding weather forecast are some of the problems which are not impossible yet very tedious and difficult to be solved with the algorithmic coding approach. However, these problems can be solved with ease using the ML approach. ML leverages the huge datasets which have features and target attributes, features can be factors affecting the output values and targets are the output values for each set of feature (not all the dataset have targets). For instance, for predicting house prices, the number of rooms, locality, floor area could be the features and price of the house is the target, as the number of rooms increases price might increase, or with the decrease in floor area price might decrease or many such scenarios where each factor is affecting the target. Hence, it becomes difficult to solve the problem in traditional programming method. Machine learning can be thought of as a black box where input features(X) are fed and the black box maps it to its targets Y, this black box is nothing but the learning algorithm (Figure 2.15).

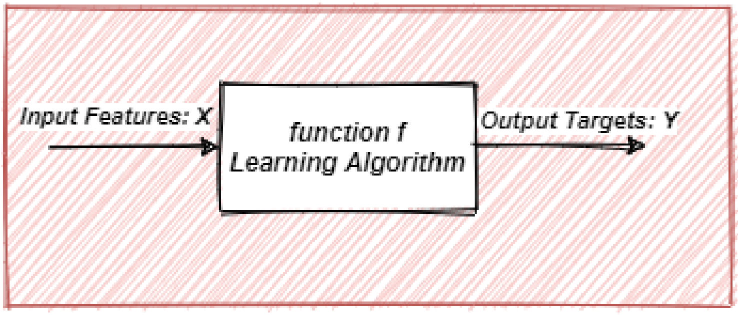

FIGURE 2.15: Machine Learning

Machine Learning problems can be categorized into 3 parts (Figure 2.16):

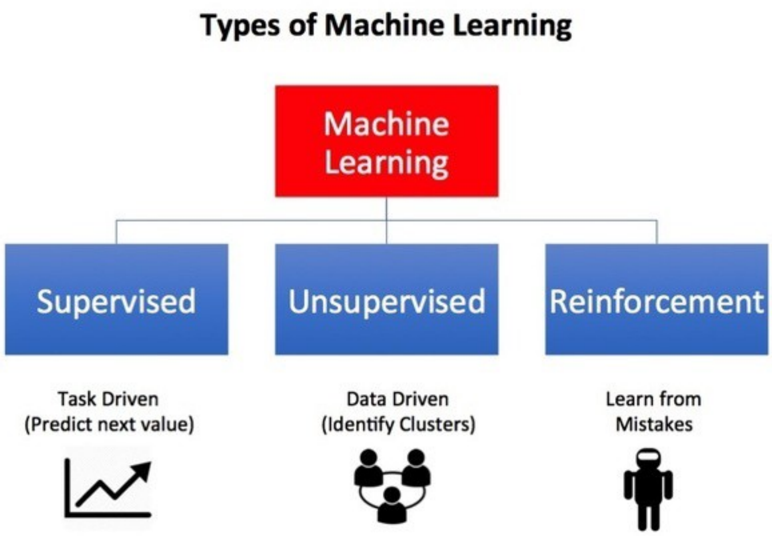

FIGURE 2.16: ML Types
[31]

- Supervised ML: Data passed to these models have both features and targets, can be used to classify the output (classification), or predict the output(regression).
- Unsupervised ML: Data passed only has the features and no associated targets, the learning algorithm tries to find patterns in the dataset.
- Reinforcement Learning: It's a very broad area of AI. Here, there is a learning agent that learns from its environment, looks for the final result, gets a penalty if the result is wrong, and reward if it's right. Main goal of this agent is to maximize the rewards.

2.2.2 Deep Learning

Deep Learning is a branch of AI which mimics the structure of human brain. It tries to learn the same way human brain learns. Any deep learning solution has 3 main components (Figure 2.17):

- Neural Network: It is the core component of the solution. There can be number of layers in a network and the architecture changes with the nature of problem statement and has scope of experimentation with it, more on this in section 2.2.3
- Loss Function: It is function which validates the predicted output Y' with actual output Y and calculates the error and the loss score. There are number functions of calculating the loss score like cross-entropy-loss, negative log loss, etc
- Optimiser: It is a component which updates the weights of the neural network based on the loss score calculated.

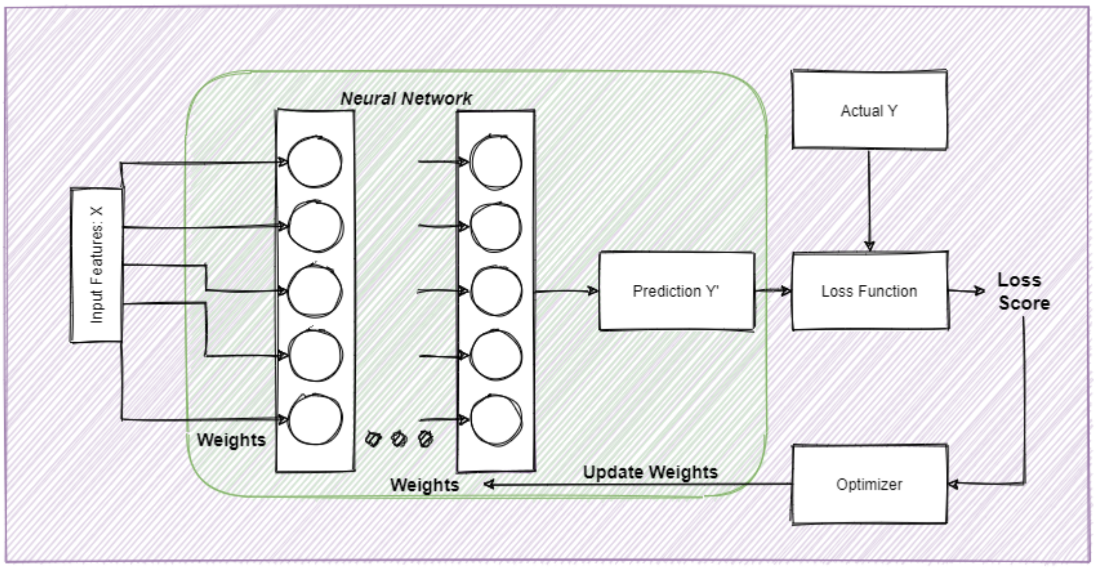

FIGURE 2.17: Deep Learning Solution Flow

Applications of Deep Learning isn't limited here, it can be applied to sounds, images, videos, text, etc (see Figure-2.18).

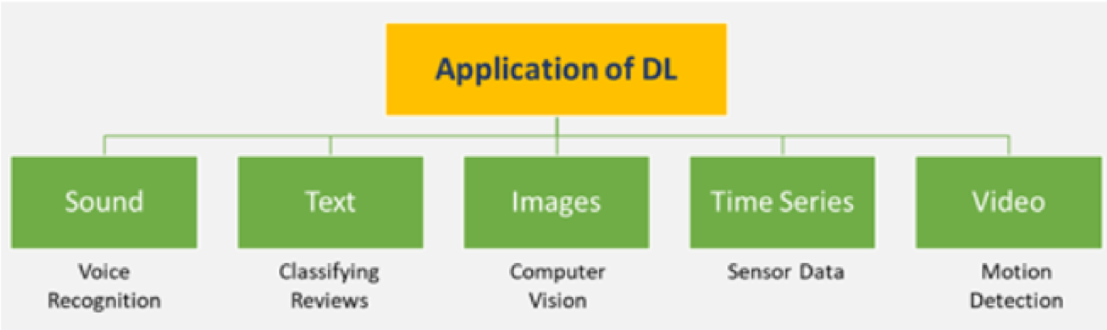

Figure 2.18: Applications of DL
[32]

2.2.3 Artificial Neural Networks

Artificial neural network (ANN) is the core basis of DL (Section 2.2.2). ANN has 3 main parts - input layer, hidden layer(s) and output layer, while ANN is uniquely defined by combination of its architecture and learning algorithm[13].

2.2.3.1 Architecture

The architecture of ANN is the structure in which the nodes are arranged to form the network. Generally the architectute is designed as follows:

- Input Layer: The number of nodes of input layer are kept same as the number of features in the dataset
- Hidden Layer(s): The number of hidden layers as well as nodes in each hidden layer are arbitrary which is to be decided with experiments or assigned randomly.
- Output Layer: Number of nodes in the last layer depends on our expectation from problem statement, e.g multi-class classification problem can have multiple nodes telling the probabilities of each class, regression problem generally have single node yet it varies with problem statement.

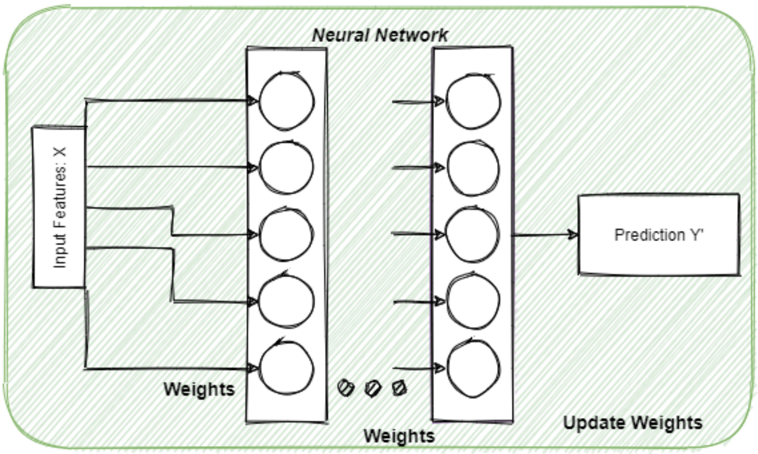

FIGURE 2.19: Artificial Neural Network

Neural networks uses the concept of feed-forward and back propagation to adjust the weights, an optimiser along with loss function is used to find the difference in the predicted and actual target values and further updates the weights accordingly (Figure 2.19).

2.2.3.2 Activation Functions

The architecture of ANN is the structure through which features passes back and forth while adjusting weights appropriately according to the optimiser. When the data is passed through each node, we expect the output of the nodes to be in certain range e.g 0 to 1, 0 or 1, -1 to 1 or only positive values, etc. To achieve this we have to feed the data to an activation function which will be defined for each node. There are variety of activation functions to achieve certain range eg, 0 to 1 is achieved with Sigmoid Function, -1 to 1 with tanh function, positive values with Rectified Linear Unit aka ReLU (which is most widely used activation function and used for implementing this project too) See Figure 2.20.

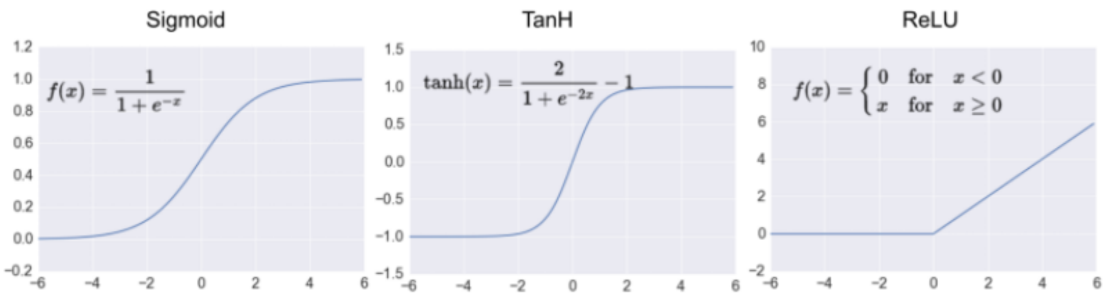

FIGURE 2.20: Commonly used activation functions
[33]

2.3 Feature Selection

Feature selection is the process of selecting features that can be useful to build a constructive model. It is a method widely used in machine learning, where the implementation of a learning algorithm selects a subset of features available from the data. It is considered as an important part of the pipeline of data pre-processing. The main objective of this process is to remove the redundant information from the data that may enhance the model performance [34]. Modeling with a subset of best-selected features can simplify the forecasting process and enhance its ability to generalize the unseen data [35]. The selection of suitable features is a crucial issue for the performance of load or price forecast model [35]. H.T. Pao [36] applies an artificial neural network (ANN) for predicting electricity prices and have yielded mixed results that may largely be associated with problems in data selection and sampling variation. It overcomes the problem with high dimensional data as this method reduces the complexity of the model. Furthermore, the benefit of performing feature selection is that it reduces the training time, over-fitting of the model, and it might improves the accuracy of the model [34]. The feature selection technique is sub-classified into three classes: filter methods, wrapper methods, and embedded methods [34] (Figure 2.21).

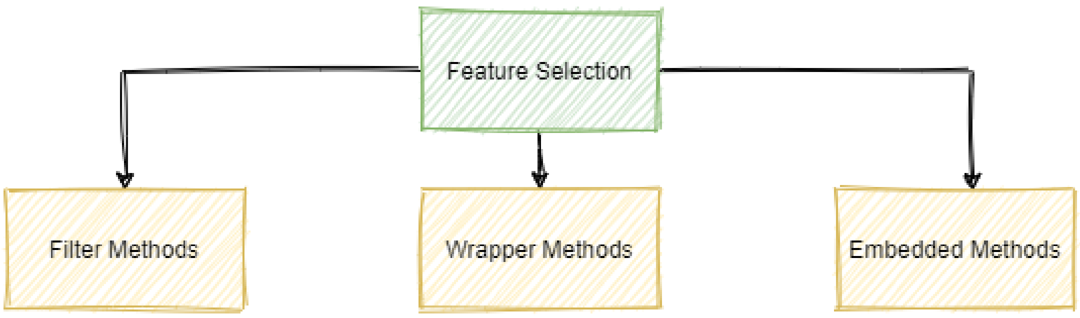

Figure 2.21: Types of Feature Selection Methods

1. **Filter Methods**: This is the type of feature selection method that uses the statistical approach to assign a score to each attribute. The features are ranked on these scores and select the highly ranked scored features accordingly. Chi-squared tests, information gain, and correlation coefficient scores are the few examples of filter methods [34]. The main advantage of the filter method is that it is computationally fast to execute as it does not consider the model. The major limitation of this method is that it may select the redundant information and avoid selecting important features as it does not evaluate a specific model performance [7].

2. **Wrapper Methods**: This is the type of feature selection method that considers finding the best subset of features as a search problem. The different combinations

of features are prepared, evaluated, and compared to other combinations by using a predictive model such as kNN (k-nearest neighbour), logistic regression algorithm. Further, the model assigns a score to each combination and evaluates the different combinations of features based on model accuracy [34]. The search process uses various approaches to add and remove features like methodical approach (best-first search), or heuristics approach (forward and backward passes). The main advantage of this feature selection technique is that it is based on a practical approach as it takes the performance of the model and interrelation of features into account [7]. However, the major drawback of this method is that it requires a long computational time.

3. **Embedded Methods**: This is another type of feature selection method that finds the subset of appropriate features at the same time the model is selected [7, 34]. Regularization methods are the most common form of this feature selection method. It imposes the additional constraints while optimizing a predictive algorithm that reduces the complexity of the model [34]. The main advantage of this method is that it is relatively less computational expensive compared to wrapper methods though it considers the underlying model. However, the limitation of embedded methods is that they can not be applied always as they are specific to learning algorithm [7].

In the electricity price forecasting literature, determining the appropriate feature selection technique varies depending on the prediction model used. For uni-variate time series models such as ARIMA having electricity prices as only variable, auto-correlation plots or akaike information criterion (AIC) are commonly used for feature selection. J.A. Carta, P. Cabrera, J.M. Matía s, and F. Castellano [37] analyses the importance of feature selection for use with artificial neural network (ANN) with a multilayer perceptron (MLP) structure to forecast the mean hourly wind speeds. In the study, they applied correlation feature selection (CFS), which is based on a filter approach, and an multi layer perceptron (MLP) -based wrapper approach (WA) feature selection methods and highlights that the wrapper approach (WA) generated relatively lower mean errors compared to the filter approach (FA) method, but was more computationally intensive [37]. In case of building models based on neural network or any other machine learning algorithms with explanatory variables, the most widely used feature selection technique used by researchers are trial and error or filter methods based on linear analysis techniques such as statistical sensitivity analysis [38], correlation analysis [39], or principal component analysis [40]. The correlation feature selection (CFS) and mutual information (MI) filter based methods are used in this thesis for finding the appropriate variables, discussed in detail in Section 4.3.1

Chapter 3

Design

3.1 Problem Definition

This thesis is based on the project provided by a CapSpire, a software company. They provided the dataset having hourly recorded day-ahead markets electricity prices for period 19-12-2018 to 13-05-2020, based on which the expectation was to apply feature selection techniques and to build ANN and different machine learning models.

Variable selection is considered as an important topic for electricity price forecasting [2]. Several factors such as demand, balance market prices, production of electricity in the electricity market influences the electricity prices. As per the expectations of capSpire, the goal of this project is to determine the explanatory variables of interest and to build a model for predicting the day-ahead electricity prices.

3.2 Objectives

There are 4 main objective of this thesis. First objective is to use feature selection techniques to determine the best subset of features for predicting the electricity prices. Then the second objective is to implement the machine learning algorithms for forecasting day-ahead market electricity prices. The third objective is to use artificial neural network (ANN) to forecast the electricity prices. Fourth and the last objective is to compare the artificial neural network model with other machine learning models with and without selected features to find the best performing model for forecasting the day-ahead electricity prices.

3.3 Design

3.3.1 Technologies Used

1. Python [41]: Version 3.7 is used for this thesis. All the coding for basic logic as well as modeling is done in python.

2. NumPy [42]: Version 1.18.4 is used. It is used to play with data in the form of nd-arrays and its manipulation as and when required.

3. Pandas [43]: Version 1.0.3 is used. It is used for reading the data from file as dataframe and manipulation of the same.

4. MatplotLib [44]: Version 3.2.1 is used. It is prominently used for data visualization.

5. ScikitLearn [45]: Version 0.23.1 is used. Scikit-Learn popularly known as sklearn is the librrary that holds numerous ML models, data-sets, etc and in this project the models used were imported from this.

6. Keras [46]: Version 2.3.0 is used. It is deep learning based libarary, an abstraction on the popular tensorflow library. In this project it used to build and implement ANN.

3.3.2 Solution Architecture

The solution architecture for this project is proposed and built from scratch and is divided into 4 stages, as shown in figure 3.1 .

- **Stage 1: Data selection and availability**

  Read and load the data using pandas library. The loaded data consists of fundamental variables, lagged version of these fundamental variables (-24 hour, -48 hour -168 hour) and proxy available variables, more on this in section 4.1.1.

- **Stage 2: Data Preprocessing**

  Performed exploratory data analysis on fundamental variables to gain insights on seasonality, trend, and other patterns of electricity consumption and production. Furthermore, normalized the data using MinMaxScaler() function from the SciKitLearn library. Detailed explanation in section 4.2.1.

- **Stage 3: Training**

  The data is split into train and test set with a share of 80 percent and 20 percent respectively. Feature selection techniques such as mutual information method and correlation feature selection method were applied to the training set. Build the models using machine learning algorithms and artificial neural networks on the training data with and without selected features. Detailed explanation in section 4.3.1.

- **Stage 4: Testing**

  Fit the models on test data (unseen) to check the performance of the model. Moreover, compared the ANN model with other learning algorithm models and checked the difference in the performance of the model before and after feature selection. Hence, the best model with the lowest mean absolute error is used to predict the electricity prices. Detailed explanation in section 4.4.

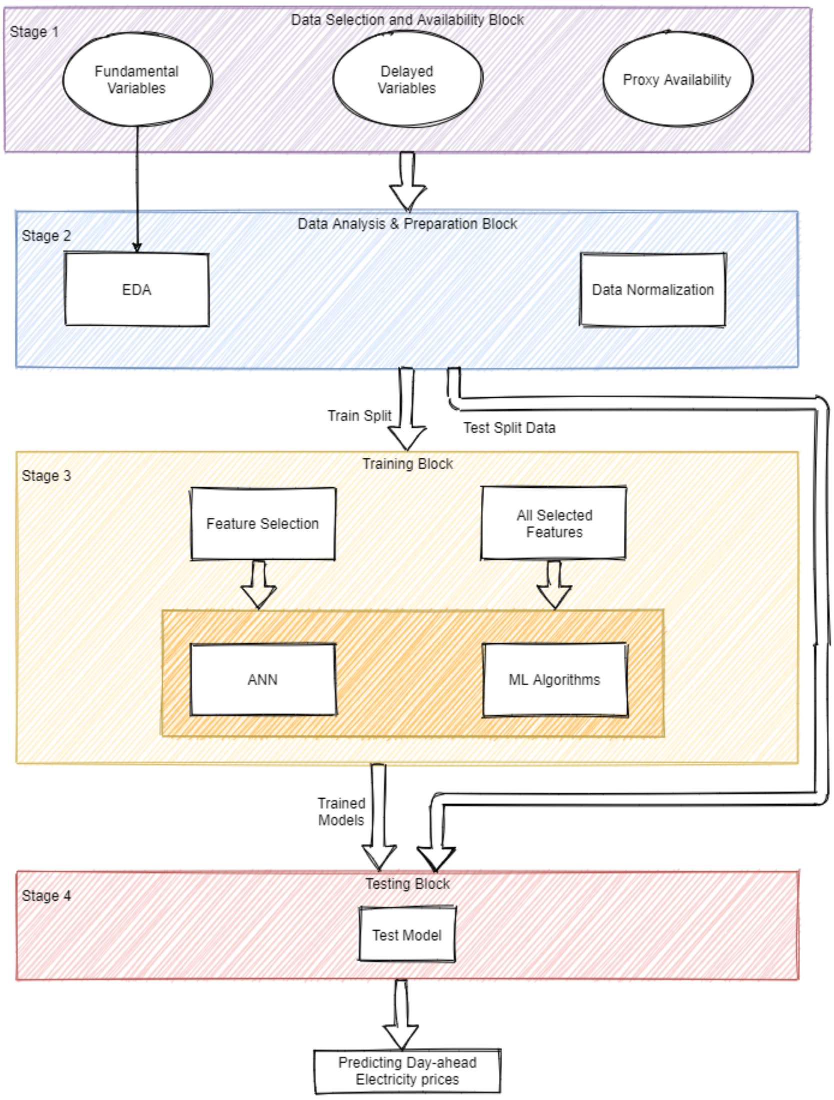

FIGURE 3.1: Solution Architecture

Chapter 4

Implementation

4.1 Stage 1: Data selection and avaibility

4.1.1 Data Description

When more than one variables of the data are time dependent, than the recorded data is referred as Multivariate Time Series Data. The dataset provided by the CapSpire for price forecasting contains the 4 fundamental variables, 15 delayed variables of the fundamental variables, and 13 available proxy variables recorded hourly for time period 19-12-2018 to 13-05-2020.

Table 4.1, shows the considered fundamental variables that consists of day-ahead market prices (DAMPrices), balance market price information (BMInfo), electricity production from wind forecast (WindForecast), and demand forecast(DemandForecast). These are the influential factors that significantly affects in forecasting electricity prices.

Fundamental Variables	Temporal Resolution	Available Since
DAMPrices	Hourly	19/12/2018
BMInfo	Hourly	19/12/2018
WindForecast	Hourly	19/12/2018
DemandForecast	Hourly	19/12/2018

Table 4.1: Overview of Fundamental Factors

Delayed variables of the fundamental variables consists 24-hour, 48-hour, 72-hour and 168-hour lagged variables of each fundamental factors. The pandas shift() function enables to transform the time series data into a supervised learning problem [47]. Table 4.2, shows the brief description of delayed variables.

AvailabilityProxy variables contains the information of the other electricity gererator participants in the market (coal, gas, hydro, fuel, oil) before submitting an order to day-ahead market. It shows the sell order volume by fuel type recorded for each hour.

Delayed Variables	Desciption
-24	24-hour lagged DAMPrices
-48	48-hour lagged DAMPrices
-72	72-hour lagged DAMPrices
-168	168-hour lagged DAMPrices
BM-24	24-hour lagged BMInfo
BM-48	48-hour lagged BMInfo
BM-72	72-hour lagged BMInfo
BM-168	168-hour lagged BMInfo
Wind-48	48-hour lagged WindForecast
Wind-72	72-hour lagged WindForecast
Wind-162	162-hour lagged WindForecast
Demand-24	24-hour lagged demand
Demand-48	48-hour lagged demand
Demand-72	72-hour lagged demand
Demand-162	162-hour lagged demand

Table 4.2: Overview of Delayed Factors

4.2 Stage 2: Data Analysis and Preparation

The provided dataset by company was very clean. The aim of this stage is to preprocess the data before passing it into learning algorithm. There are no missing values found. As all the attributes are numerical therefore not required to deal with the categorical data. Further, did some exploratory data analysis and scaled the data.

4.2.1 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) recognizes the data sets by summing up their key characteristics and by visually plotting them. This helps to explore the data and to understand the hidden features underlying within the data. For this thesis, performed EDA by visualisation technique on fundamental variables, described in table 4.1.

The provided data is time dependent and pandas library allows time-based indexing, re-sampling, and rolling windows such techniques that are used to gain insights on seasonality, trend and other interesting features of electricity consumption and production [48].

- Visualisation of Fundamental Data: The 4 fundamental variables in the data are visualised for the period 19/12/2018 to 13/05/2020. Figure 4.1 shows the demand and wind aggregated forecast while figure 4.2 shows the day-ahead electricity prices and balance market prices in Euro/MWH. The negative prices can be observed for balance market and the day-ahead market in figure 4.2.

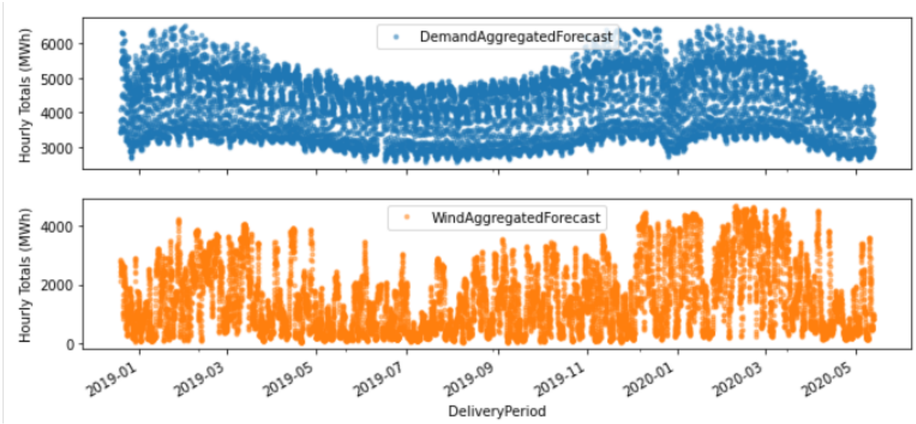

FIGURE 4.1: Demand Forecast & Wind Forecast for period 12/2018 to 05/2020

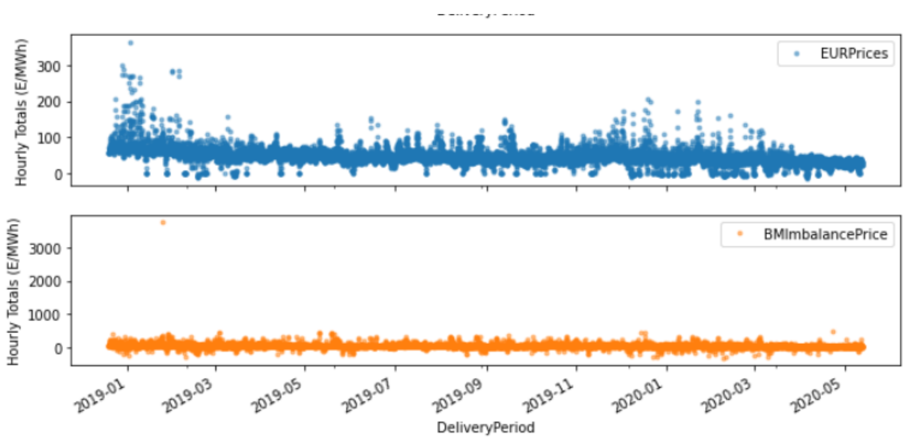

FIGURE 4.2: Day-ahead and balance market prices for period 12/2018 to 05/2020

- Visualisation of Demand Forecast for 24 Hours of the day: The figure 4.3 visualises the consumption of electricity on an hourly basis. The daily basis periodicity (day seasonality) is observed in demand for electricity by consumers. The demand for electricity increases from 6:00 am in the morning and reaches at peak in evening at 6:00 pm. Further, a huge drop in the demand is observed during night. The speculated reason for such a trend is that electricity is vigorously used in day time as businesses, households, and offices where electricity is more consumed are active.

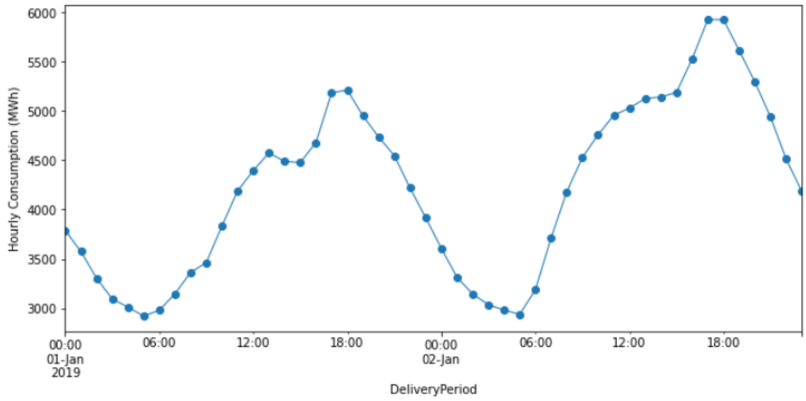

FIGURE 4.3: Demand Forecast for 24 Hours of the day

- Visualising Day-Ahead Market prices of 24 Hours of the day: The figure 4.4 visualises the day-ahead market electricity prices on an hourly basis. A similar pattern observed for demand forecast (figure 4.3) could be seen here. The electricity prices are high in the morning hours compared to night. From this get an intuition of a significant relationship between the electricity prices and the demand for electricity. Furthermore, it is observed that around 6:00 pm of the day, electricity is sold at the highest prices as compared to other hours of the day.

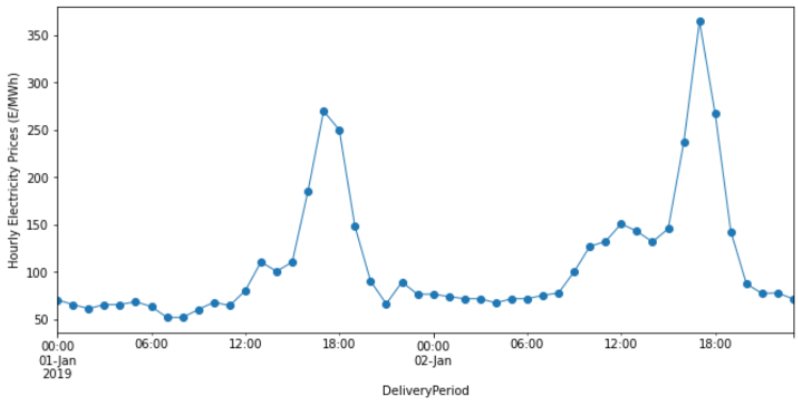

FIGURE 4.4: Day-Ahead Market prices of 24 Hours of the day

- Visualising Balance Market prices of 24 hours of the day: The figure 4.5 visualises the balance market electricity prices on an hourly basis. From the figure, not such significant pattern is observed within the day.

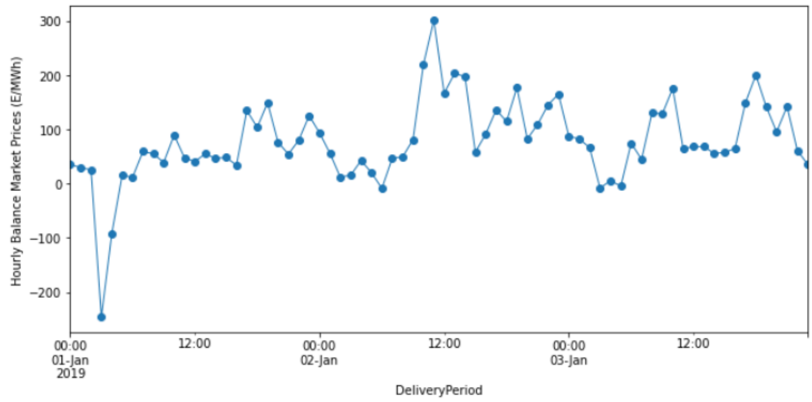

FIGURE 4.5: Balance Market prices of 24 hours of the day

- Visualising Wind Forecast of 24 Hours of the day: The figure 4.6 visualises the wind forecast for the production of electricity on an hourly basis. From the figure, not such significant pattern is observed within the day as the magnitude of wind changes each day and is volatile in nature.

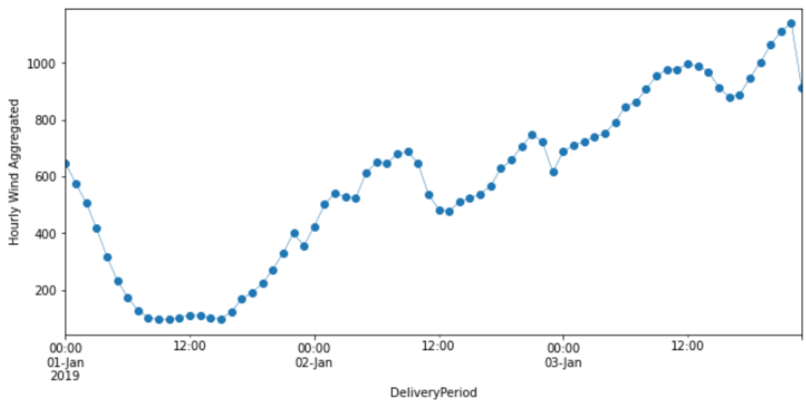

FIGURE 4.6: Wind Forecast of 24 Hours of the day

4.2.2 Scaling Data

The purpose of this step is to adjust numerical column values in the data set to a standard scale, without distorting variations in values ranges. This step is required as the provided data features have different ranges. Normalization and Standardisation are the two popular methods to scale the data.

In standardization scaling method, each data point is subtracted from the mean of the data and divided by its standard deviation. The standardized data has mean of 0 and standard deviation of variable is 1. Hence, the data ranges from -1 to 1.

In Equation 4.1, X is the data point, $\bar{X}$ is the mean of the data, and S_x is the standard deviation of the data. This equation is used to scale the data in standardisation method.

$$X_{std} = \frac{X - \bar{X}}{S_x}, \tag{4.1}$$

In normalization scaling method, a variable is scaled in a range of 0 to 1. This method is also referred as min-max scaling [49]. Equation 4.2 is used to scale the data in normalisation method. Here, X is the data point, min(X) is minimum value of X, and max(X) is maximum value of X attribute.

$$X_{norm} = \frac{X - min(X)}{max(X) - min(X)}, \tag{4.2}$$

In [1], [49], normalization technique is preferred relatively more for the application of neural network. Hence, for this thesis normalization technique is used to bring the values of all features on a same scale. MinMaxScaler() function of scikit-learn is used to normalize the data in python.

4.3 Stage 3: Training

The scaled data is split into train-validation set. The competing issues while splitting the data into train and test set is that neither of the set should have high variance. If the training data split ratio is less, than the performance statistics might have greater variance. High variance occurs when the model learns too much from the training data and this results into the problem of "Over-fitting". However, when the model learns not much from the training data (high bias), it results into the problem of "Under-fitting"[50]. Hence, 80:20 ratio is generally more considered and used the same ratio to split for this thesis.

4.3.1 Feature Selection

The mutual information and correlation feature selection techniques are used in this thesis to select the relevant features.

Figure 4.7 shows the proposed flow of feature selection block in the thesis. The processed data from preprocessor is passed through the feature selection block. Further, the predictions are made on the selected candidates in forecast engine.

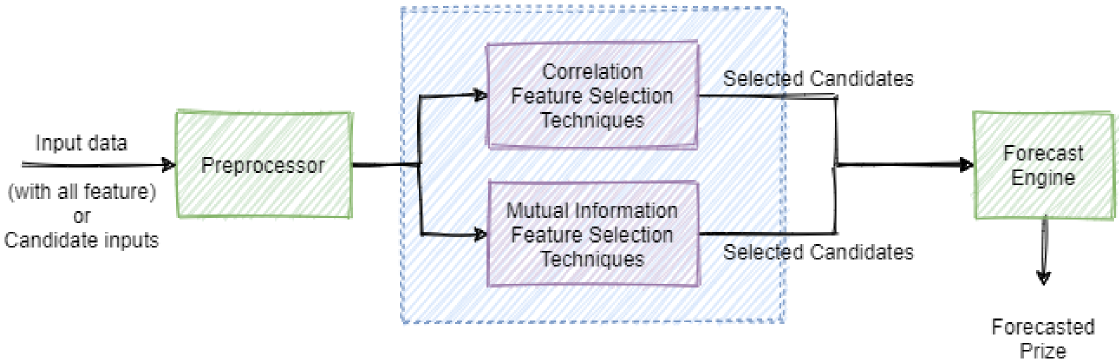

FIGURE 4.7: Flow of Feature Selection in the Solution Architecture Design

4.3.1.1 Correlation Feature Selection Method

This technique is the classic approach to quantify the relation between the two variables. The term "correlation" means the strength of relationship between the two factors. This relationship can be measured by Peasrson's correlation coefficient and is calculated by using equation 4.3 where Covariance(X,Y) is the covariance between X and Y, S_x is the standard deviation of X, and S_y is the standard deviation of Y.

$$Pearson'sCorrelationCoefficient = \frac{Covariance(X,Y)}{S_x * S_y}, \tag{4.3}$$

The result from Pearson's correlation coefficient ranges from -1 to 1. The result can be interpreted as a value above 0.5 or below -0.5 shows strong correlation and a value close to 0 means weak correlation between two variables. Furthermore, the positive correlation means any change in the one variable affects the change in the other variable in the same direction while if the variables changes in different direction it is called negative correlation. However, the 0 correlation coefficient means no relationship between the variables and known as neutral correlation [51].

The correlation statistic having only positive values (calculated by converting correlation coefficient) is used for selecting best correlated features. Larger the positive value, say score tends to have strong relation and more likely to be used for modelling[52].

For implementation in python, f_regression() function of scikit-learn and SelectKBest class is used. The f_regression() function calculates the correlation statistics and SelectKBest class is used to select the number of k features on which f_regression() function is passed [52].

The figure 4.9 shows 10 to 12 features have larger scores, relatively more important compared to other features and further could be selected for modelling.

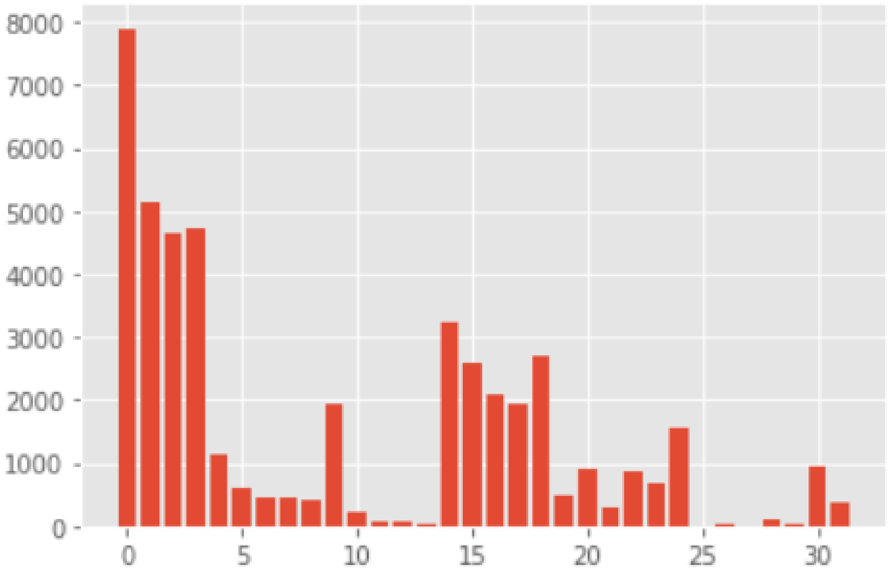

FIGURE 4.8: Correlation Feature Importance

Hence, depending on the score selected 15 features that are used for modeling. The list of selected features are shown in Figure 4.9.

```
# Selected top 15 features using Correlation Feature Selection Method
features_corr.columns

Index(['-24', '-48', '-168', '-72', 'DemandAggregatedForecast', 'Demand-168',
       'Demand-24', 'Demand-48', 'WindAggregatedForecast', 'Demand-72',
       'PUMP_STORAGE_-168', 'BMImbalancePrice', 'InterconnectorNetTotal_-24',
       'GAS_-168', 'PEAT_-168'],
      dtype='object')
```

FIGURE 4.9: List of selected features using Correlation Feature Selection Method

4.3.1.2 Mutual Information Feature Selection Method

This feature selection method is based on the Mutual Information concept that is mainly used while constructing the decision trees. Mutual Information calculates the dependency between two random variables, and entropy measures the amount of information within the variable[52].

Mutual Information between two random variables say X & Y can be calculated by using equation 4.4 , where H(Y) is the entropy of Y and H(Y|X) is the conditional entropy of Y given the variable X

$$I(X;Y) = H(Y) - H(Y|X), \quad (4.4)$$

The Entropy(H) can be calculated by equation 4.5, where p is the probability of success and q is the probability of failure.

$$H(Y) = -p * log_2 p - q * log_2 q, \quad (4.5)$$

The equation 4.4 could be used for feature selection criteria. Lets assume Y as target variable and X as subset of features. Hence, the features having high mutual information with target variables are selected as they reduces the uncertainty on the values taken by the output. The advantage of the mutual feature selection method over correlation feature selection method is that it is capable of capturing non-linear relationship between variables [53].

For implementation in python, mutual_info_regression() function of scikit-learn and selectKBest class is used. It results the score for each variable and variables with larger score are selected.

The figure 4.10 clearly shows around 15 features have relatively more mutual information with target variable compared to other variables. Hence the subset of these 15 features are selected for modelling.

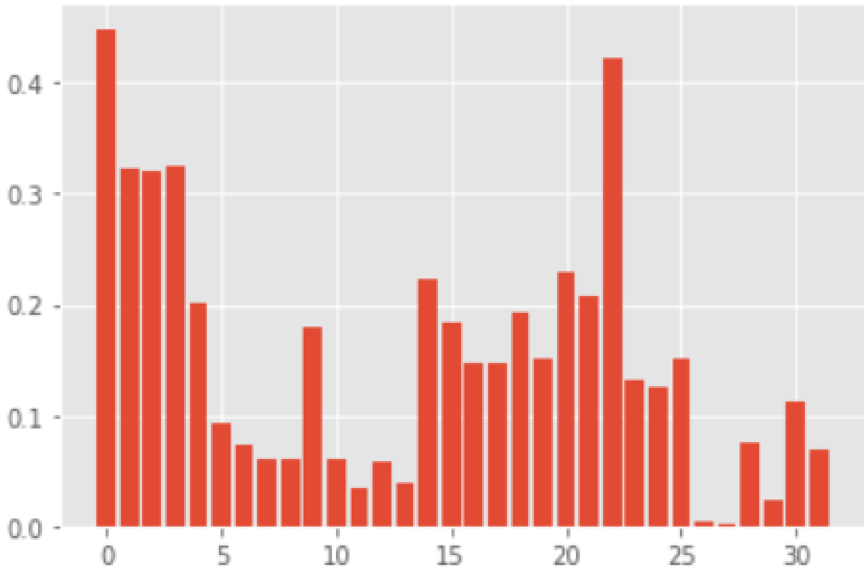

FIGURE 4.10: Mutual Information Feature Importance

Hence, the list of selected features are shown in Figure 4.11.

```
[144] # Selected top 15 features using Mutual Information Method
      features_MI.columns

Index(['-24', 'PEAT_-168', '-168', '-48', '-72', 'GAS_-168',
       'DemandAggregatedForecast', 'HYDRO_-168', 'BMImbalancePrice',
       'Demand-168', 'Demand-24', 'WindAggregatedForecast', 'MULTI_FUEL_-168',
       'BIOMASS_-168', 'Demand-72'],
      dtype='object')
```

FIGURE 4.11: List of selected features using Mutual Information Feature Selection Method

4.3.2 Supervised Learning Algorithms

For this thesis applied linear regression, random forest, support vector machine, and extra-gradient boosting such machine learning algorithms on all the available features and features selected from the feature selection techniques.

4.3.2.1 Linear Regression (LR)

Linear Regression is a supervised learning algorithm that was developed in the fields of statistics. It is both statistical and machine learning algorithm that is used for the regression data. It predicts the target value by analysing the linear relationship between independent features and the target feature. Consider X as an input feature and Y as target feature, equation 4.6 is used to predict the output where 'm' is the coefficient of 'X' and 'c' is the intercept.

$$Y = c + m(X), \tag{4.6}$$

The model is trained by fitting the best regression line that is determined by finding the best values of both intercept (c) and coefficient of features (m) to predict the target values. Implemented this model as it is a popular model for evaluating feature selection methods, because it could work better if the model excludes irrelevant features [52].

4.3.2.2 Random Forest (RF)

Random Forest is a supervised learning algorithm based on the bagging approach (growing trees parallel from sub-samples) and is developed by aggregating trees. The several random subsets of the training dataset is used to build the various decision trees with each subset. Thus, aggregates each decision tree prediction or say a vote to get the final and better result. Random Forest avoids overfitting, easy to deal with a large number of features, and helps with feature selection based on importance [54].

4.3.2.3 Support Vector Regression (SVR)

Support Vector Machine (SVM) is a generalised learning algorithm used for both classification and regression problem. When used for regression data than it is referred as support vector regression (SVR) algorithm. The main objective of the algorithm is to recognize the optimal hyper-plane in N-dimension, or an appropriate line that fits best to the data within the specified margin of error. The advantage of the algorithm compared to other regression algorithms is that the dimension of input features does not affects its computational complexity [55]. In [56], SVR gives better results with mutual information feature selection technique and this is the reason to implement this modelling technique.

4.3.2.4 Extreme-Gradient Boosting (XGB)

Extreme Gradient Boosting (XGBoost) is a tree-based ensemble learning algorithm that is based on the concept of gradient tree boosting. Boosting means to convert the weak learners into strong learners. In gradient boosting the forest of regression trees are used in a sequential learning process known as weak learners and predicts the outcome after summing up the scores from each tree and minimizes the loss function using gradient descent. It has high computational speed compared to models like AdaBoost and random forests [54].

4.3.3 Artificial Neural Network (ANN)

4.3.3.1 Architecture

As explained in section 2.2.3, ANN consist of 3 types of layer and each layer has certain number of neurons and this forms the core structure of neural network, called as its architecture. For implementation of this project below is the architecture designed (shown in Figure 4.12).

- **Input Layer:** The number of neurons in this layer is same as number of features, and as mentioned in section 4.3.1, models will run on all the features, selected features from section 4.3.1.1 and selected featured from section 4.3.1.2. Hence, each ANN model will have different number of input features, yet from the code implementation they are known to be 32, 15 and 15 input neurons for each of the model.

- **Hidden Layers:** The number of hidden layers and neurons in each layer are arbitrary, can be experimental. For this project it is decided to be kept as 3 hidden layers and each having 70 neurons for every model.

- **Output Layer:** The number of neuron in the output layer depends on the problem statement, as the project is dealing with a regression problem, it is expected to give just a single value. Hence, there will be just 1 neuron in output layer for every model.

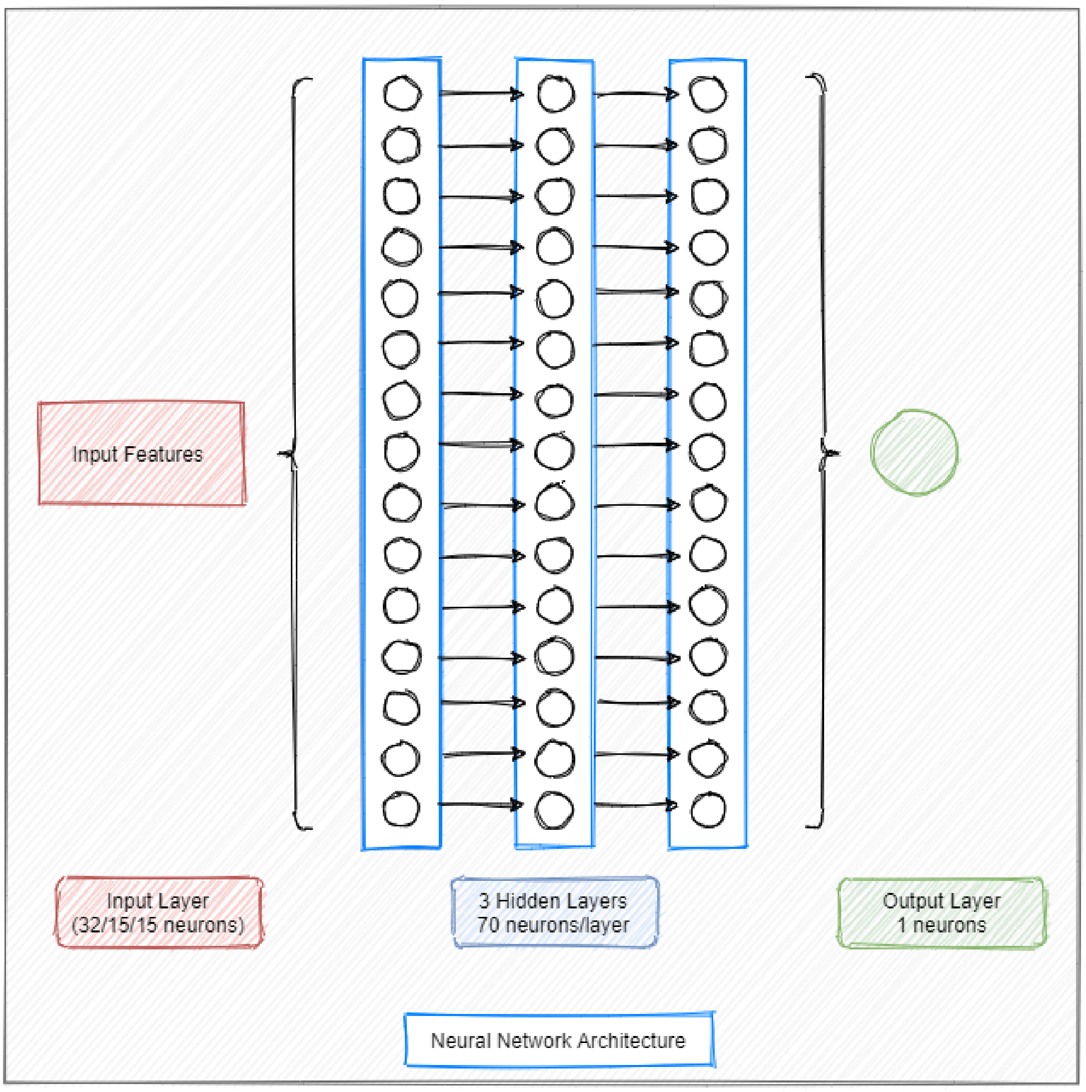

FIGURE 4.12: ANN Architecture Designed

4.3.3.2 Model Specifications

1. **Model with no feature selection technique**

    - **Neurons - Input/Hidden/Output Layers:** 32/(70/70/70)/1

- **Epochs:** 500, Overfitting - 200, Best - 150th
- **Optimiser:** Adam
- **Activation Function:** RelU
- **Loss:** Mean Absolute Error
- **Training - Validation Loss curve:** Figure 4.13

FIGURE 4.13: All features train validation-loss Curve

2. **Model with Correlation feature selection technique**

   - **Neurons - Input/Hidden/Output Layers:** 15/(70/70/70)/1
   - **Epochs:** 1000, Overfitting - 800, Best - 500th
   - **Optimiser:** Adam
   - **Activation Function:** RelU
   - **Loss:** Mean Absolute Error
   - **Training-Validation Loss:** Figure 4.14

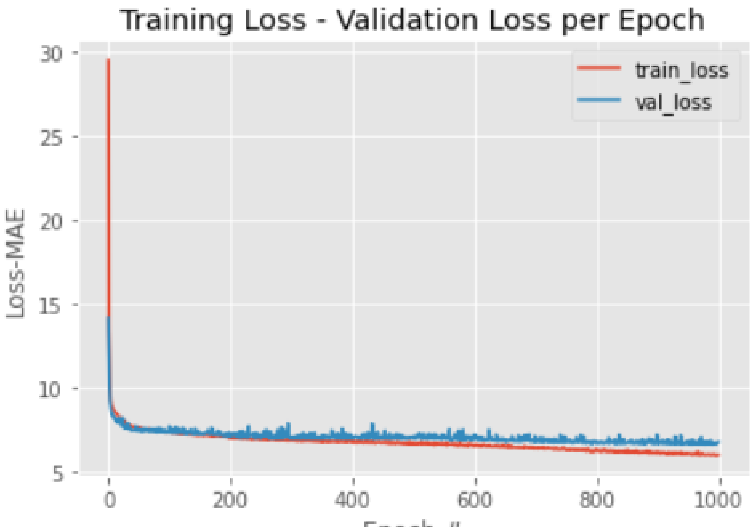

FIGURE 4.14: CFS train validation-loss Curve

3. **Model with Mutual information feature selection technique**

   - **Neurons - Input/Hidden/Output Layers:** 15/(70/70/70)/1
   - **Epochs:** 1000, Overfitting - 800, Best - 561th
   - **Optimiser:** Adam
   - **Activation Function:** ReLU
   - **Loss:** Mean Absolute Error
   - **Train-Validation Loss Curve:** Figure 4.15

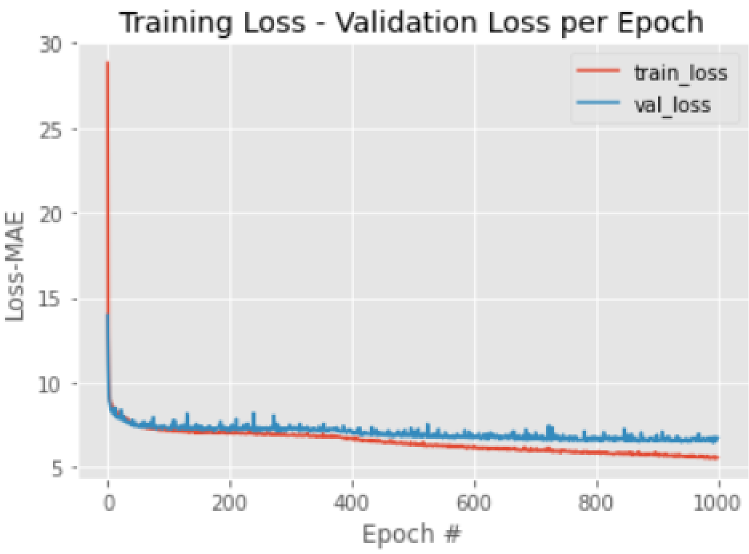

FIGURE 4.15: MI train validation-loss Curve

4.4 Stage 4: Testing

The models built in stage 3 are tested on test set having 20 percent share of the data. There are certain metrics like mean absolute error (MAE), mean absolute percent error (MAPE), and root mean square error (RMSE) to evaluate the performance of the forecasting models [57].

4.4.1 Evaluation Metrics

- **Mean Absolute Error(MAE)**

  This measures the average of the absolute difference between the actual and the predicted data points. This can be calculated by equation 4.7.

$$MAE = (\frac{1}{n})\sum_{t=1}^{n}\left|A_t - F_t\right|, \tag{4.7}$$

- **Root Mean Squared Error (RMSE)**

  It measures the standard deviation of the predicted errors. This can be calculated by equation 4.8

$$RMSE = \sqrt{(\frac{1}{n})\sum_{t=1}^{n}(A_t - F_t)^2}, \tag{4.8}$$

- **Mean Absolute Percentage Error (MAPE)**

  It measures the percentage error of the prediction. Lower the MAPE (the percentage error is less), more accurate the model is. It can be calculated by equation 4.9

$$MAPE = (\frac{1}{n})\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_(t)}\right| * 100\%, \tag{4.9}$$

For this thesis mean absolute error (MAE) is used for comparing the models performance. As mean absolute percentage error (MAPE) would not provide interpretable results for zero and negative values. Also, there is not much difference in root mean square error and mean absolute error values as both depends on absolute values [2].

Chapter 5

Testing and Evaluation

The various machine learning algorithms and artificial neural network on all features, and features selected using correlation feature selection method and mutual information feature selection method has carried out in the previous chapter. The models has been trained and tested on test data. This chapter's goal is to evaluate the final system by comparing the performance of models on test dataset.

5.1 Visualising the performance of ANN & ML models

This section is subdivided into three cases that visualises the comparison of both machine learning models and artificial neural network model performances.

5.1.1 Case 1: All features selected

This case represents the comparison of each model built on all the available features in the dataset. It represents the each models mean absolute error (MAE) value after testing on unseen data (test set).

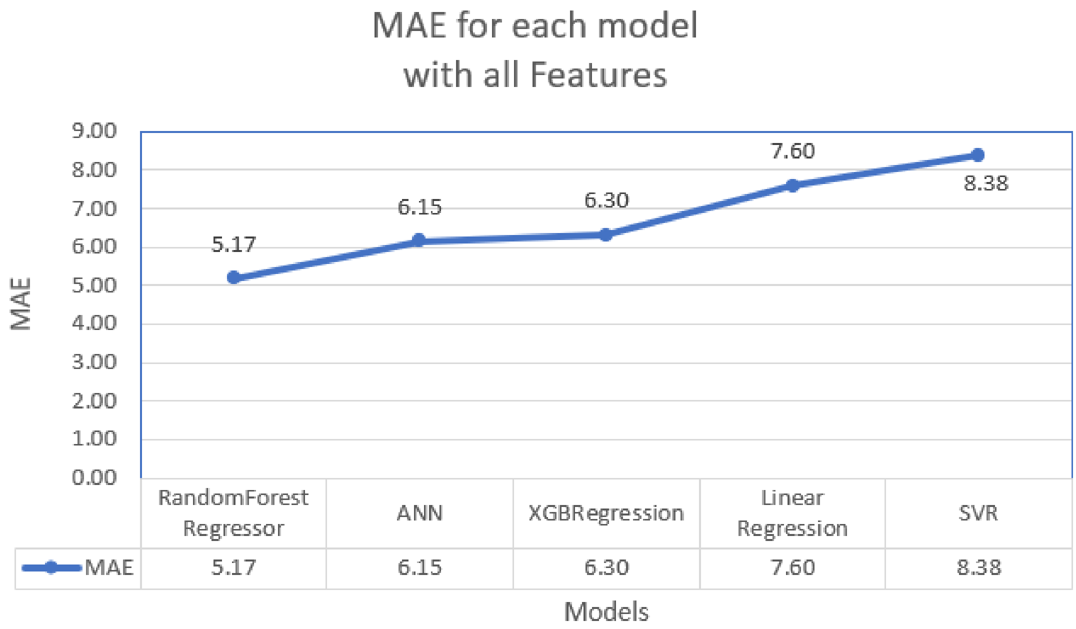

FIGURE 5.1: Comparison of models with all available features

Figure 5.1 shows that the model built on random forest algorithm has the minimum mean absolute error of 5.17 compared to other models including artificial neural network model having mean absolute error of 6.15. There is not much difference found in the performance of both ANN model and XGBregression model. However, Linear Regression model and support vector regression (SVR) model did not performed well having MAE of 7.60 and 8.38 respectively.

5.1.2 Case 2: Features selected by using Correlation Feature Selection Method

This case represents the comparison of each model built on top 15 features selected by using Correlation feature selection method (section 4.3.1.1). It represents the each models mean absolute error (MAE) value after testing on unseen data (test set).

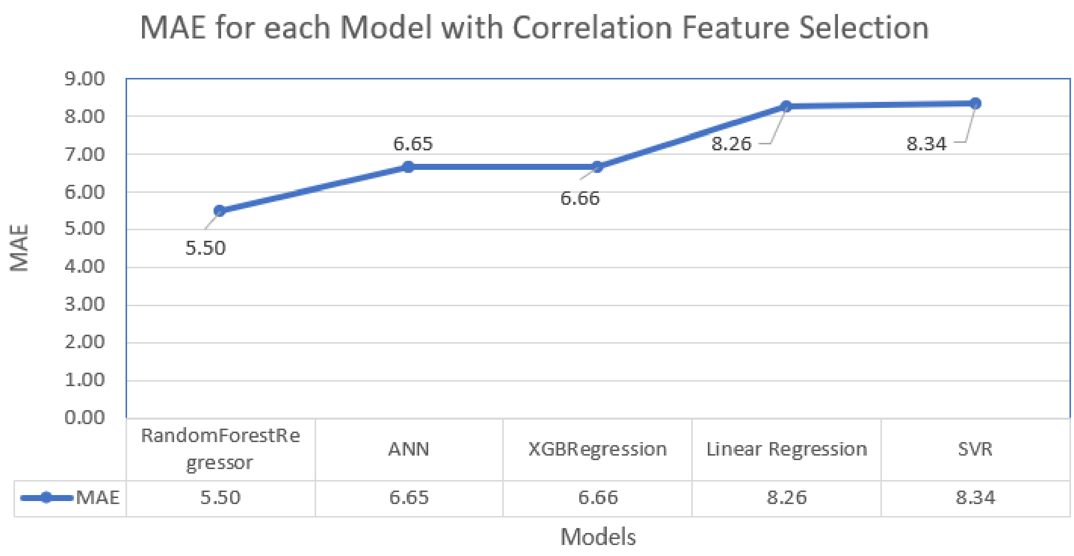

FIGURE 5.2: Comparison of models using Correlation features selection method

Figure 5.2 shows that the random forest regression model has the minimum mean absolute error of 5.50 compared to other models including artificial neural network model having mean absolute error of 6.65. XGBregression model and ANN model performance is similar as the difference in both the models performance is negligible. However, both linear regression model and support vector regression (SVR) model did not performed well having MAE of 8.26 and 8.34 respectively.

5.1.3 Case 3: Features selected by using Mutual Information Feature Selection Method

This case represents the comparison of each model built on top 15 features selected by using Mutual Information feature selection method (section 4.3.1.2). It represents the each models mean absolute error (MAE) value after testing on unseen data (test set).

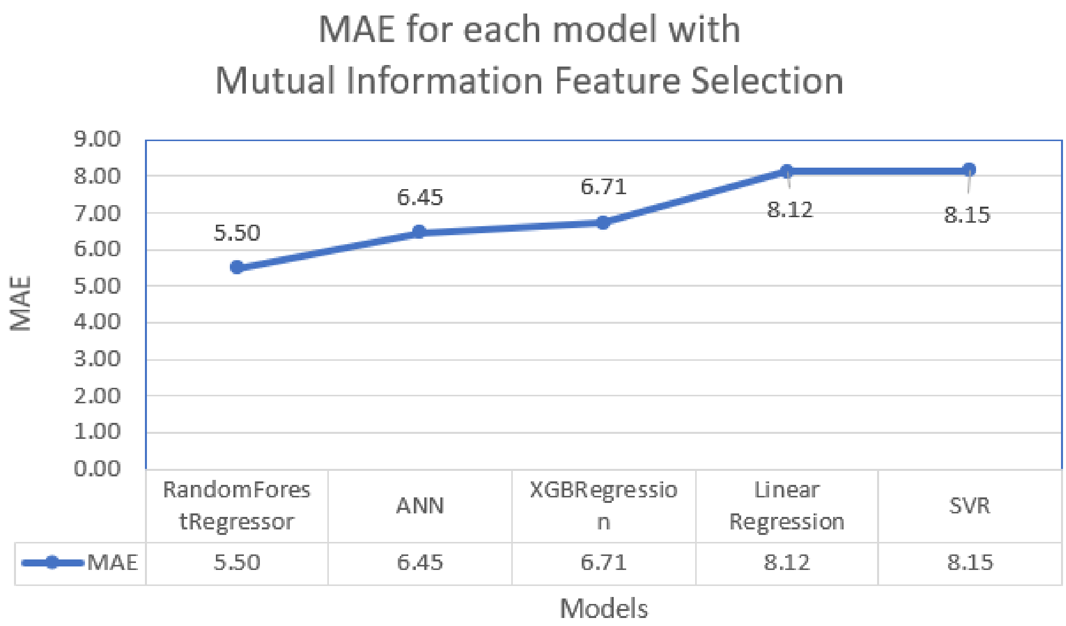

FIGURE 5.3: Comparison of models using Mutual Information feature selection method

Figure 5.3 shows that the performance of random forest regression model is better compared to other ml and ANN models having the minimum mean absolute error (MAE) of 5.50. However, ANN model having mean absolute error of 6.45 performs better than the other remaining ml models i.e. XGBRegressor, linear regression model, and support vector regression (SVR) model. Furthermore, it could be observed that linear regression model and SVR model did not provided much satisfactory results.

5.2 Evaluating the performance of models with & without feature selection

For the comprehensive evaluation of all the models with and without selected features, create a matrix of feature selection to the models that display the comparable metric of mean absolute error (MEA) for each combination of models and features used (Figure 5.4).

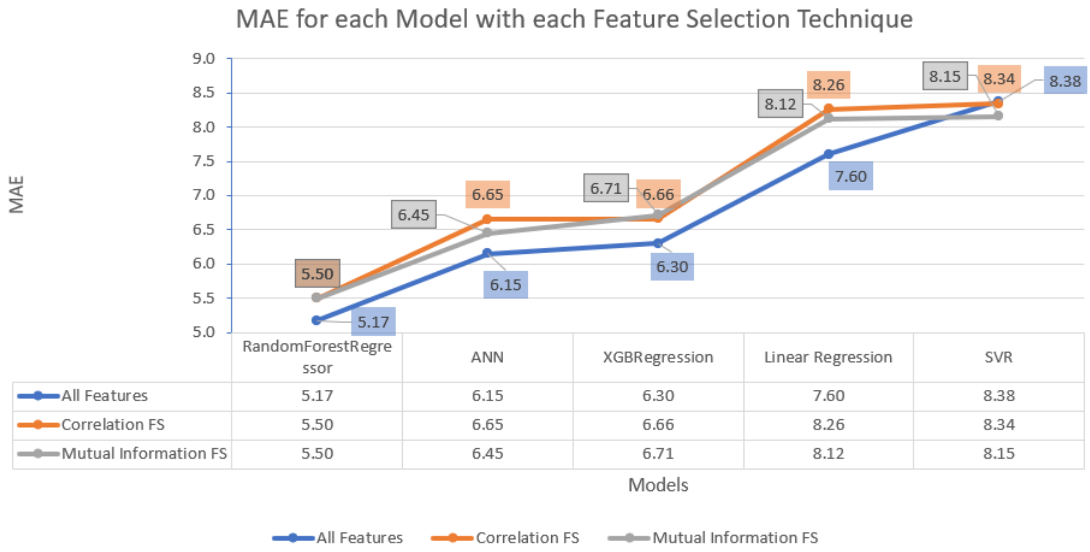

FIGURE 5.4: Comparison of models with and without feature selection technique

Figure 5.4 shows the performance of each model with all the features selected, features selected using the correlation feature selection method, and features selected using the mutual information selection method.

Observations:

- Random Forest model built on all the features of the dataset provided the best results having lowest mean absolute error of 5.17.
- The performance of models based on both correlation feature and mutual information feature selection methods does not show much significant difference.
- Support Vector Regression model works relatively better with mutual information feature selection method compare to SVR model with all features and correlation technique selected features.
- For artificial neural networks, model having all features performs better compared to the models using feature selection techniques.
- Random Forest model based on machine learning supervised algorithm outperformed the artficial neural network model having MAE of 5.17 and 6.15 respectively.

Chapter 6

Discussion and Conclusions

6.1 Discussion

The primary motivation of this thesis is to investigate the application of machine learning and neural network architecture in predicting the day-ahead market electricity prices. The forecasting of prices is achieved by implementing various supervised machine learning algorithms and artificial neural network depending on the features selected. The dataset provided by CapSpire is a multivariate time series data recorded hourly for period 19/12/2018 to 13/05/2020. The dataset was split into train and test set with the ratio of 80:20. The feature selection techniques like correlation feature selection and mutual information feature selection method were applied on training data to find the best interest of explanatory variables. The machine learning algorithms such as random forest (RF), linear regression (LR), support vector regression (SVR), and extreme-gradient boosting (XGBoost) and artificial neural network (ANN, having one input layer, three hidden layer, and one output layer) were trained on all the features available and the features selected using feature selection techniques. Furthermore, these models were tested on the test data set, and to evaluate the performance of the models efficiently used mean absolute error (MAE) as an evaluation metric.

6.2 Conclusion

The proposed solution to forecast the electricity prices is incorporation of building the both machine learning and artificial neural network models based on the features selected. After analysing the results from testing and evaluation mentioned in Chapter 5, the random forest model with all the features significantly outperform the other models (including ANN) with and without selected features having minimum absolute error of

5.50 Euro/MWh. Also, artificial neural network model works significantly better than other models except random forest model. However, support vector regression model performed relatively better with features selected using mutual information selection method compared to model with all the features. Moreover, models build using feature selection technique did not provide much satisfactory results. Hence, the random forest model with all the features is the best model to predict the electricity prices based on this thesis analysis.

6.3 Future Work

Forecasting electricity prices is an ongoing and well-researched topic. This project can be enhanced further by deep digging down to the work done for this project. The performance of models can be improved further by hyper-parameter tuning and optimizing the model. The parameters of ANN depends on the structure and learning algorithm, has a significant impact on the models performance. The only way to tune these parameters is trial-and-error method which is time consuming and computationally expensive. Hence, there is a potential to investigate more complex neural network model for forecasting, which was limited in this work due to the lack of high computational efficiency.

Bibliography

[1] D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner, “Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks,” *Applied Energy*, vol. 162, no. C, pp. 218–230, 2016.

[2] U. Ugurlu, I. Oksuz, and O. taş, “Electricity price forecasting using recurrent neural networks,” *Energies*, vol. 11, p. 1255, 05 2018.

[3] E. plc. (2017) ”industry guide to the i-sem”. Accessed on 04-July-2020. [Online]. Available: https://www.sem-o.com/documents/general-publications/I-SEM-Industry-Guide.pdf

[4] A. Ebrill. (2016) ”cer factsheet on the single electricity market”. Accessed on 20-July-2020. [Online]. Available: https://www.cru.ie/wp-content/uploads/2011/07/cer11075.pdf

[5] O. Abedinia, N. Amjady, and H. Zareipour, “A new feature selection technique for load and price forecast of electrical power systems,” *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62–74, 2017.

[6] C. Management. About capspire. Accessed on 1-July-2020. [Online]. Available: https://www.capspire.com/about-capspire-global-consulting-solutions/

[7] J. Lago, F. De Ridder, P. Vrancx, and B. De Schutter, “Forecasting day-ahead electricity prices in europe: The importance of considering market integration,” *Applied Energy*, vol. 211, pp. 890 – 903, 2018.

[8] T. I. W. E. Association. (2019, 09) ”blog: How ireland’s electricity is traded”. Accessed on 30-July-2020. [Online]. Available: https://www.iwea.com/latest-news/2652-blog-how-ireland-s-electricity-is-traded

[9] E. plc. (2016) ”quick guide to the integrated single electricity market”. Accessed on 14-July-2020. [Online]. Available: http://www.eirgrid.ie/site-files/library/EirGrid/EirGrid-Group-I-SEM-Quick-Guide.pdf

[10] W. Söderström. Trading platform. Accessed on 20-June-2020. [Online]. Available: https://www.cropex.hr/images/radionica/CROPEX_TRADING_CLEARING_MEMBERSHIP.pdf

[11] D. W. Bunn, "Forecasting loads and prices in competitive power markets," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 163–169, 2000.

[12] N. Amjady and M. Hemmati, "Energy price forecasting - problems and proposals for such predictions," *IEEE Power and Energy Magazine*, vol. 4, no. 2, pp. 20–29, 2006.

[13] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030 – 1081, 2014.

[14] M. Ventosa, Álvaro Baıllo, A. Ramos, and M. Rivier, "Electricity market modeling trends," *Energy Policy*, vol. 33, no. 7, pp. 897 – 913, 2005.

[15] I. Vehviläinen and T. Pyykkönen, "Stochastic factor model for electricity spot price—the case of the nordic market," *Energy Economics*, vol. 27, no. 2, pp. 351 – 367, 2005, special Issue on Electricity Markets.

[16] R. Aid, L. Campi, and N. Langrené, "A structural risk-neutral model for pricing and hedging power derivatives," Oct. 2010, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00525800

[17] N. V. Karakatsani and D. W. Bunn, "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients," *International Journal of Forecasting*, vol. 24, no. 4, pp. 764–785, 2008.

[18] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 744 – 763, 2008, energy Forecasting.

[19] H. Geman and A. Roncoroni, "Understanding the fine structure of electricity prices," *The Journal of Business*, vol. 79, no. 3, pp. 1225–1262, 2006.

[20] S. L. K. CHRISTENSEN, Timothy; HUM, "It never rains but it pours : Modeling the persistence of spikes in electricity prices," *The Energy journal (Cambridge, MA)*, 2009.

[21] J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *European Journal of Operational Research*, vol. 204, no. 1, pp. 139 – 152, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037722170900705X

[22] F. Ziel and R. Weron, “Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks,” *Energy Economics*, vol. 70, no. C, pp. 396–420, 2018.

[23] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, “Forecasting next-day electricity prices by time series models,” *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 342–348, 2002.

[24] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, “Day-ahead electricity price forecasting using the wavelet transform and arima models,” *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[25] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, “The effect of wind generation and weekday on spanish electricity spot price forecasting,” *Electric Power Systems Research*, vol. 81, no. 10, pp. 1924–1935, 2011.

[26] J. C. Bezdek, “What is computational intelligence?” 12 1994.

[27] J. Zhang and Chuntian Cheng, “Day-ahead electricity price forecasting using artificial intelligence,” in *2008 IEEE Canada Electric Power Conference*, 2008, pp. 1–5.

[28] P. Mandal, T. Senjyu, and T. Funabashi, “Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market,” *Energy Conversion and Management*, vol. 47, no. 15, pp. 2128 – 2142, 2006.

[29] J. Catalão, S. Mariano, V. Mendes, and L. Ferreira, “Short-term electricity prices forecasting in a competitive market: A neural network approach,” *Electric Power Systems Research*, vol. 77, no. 10, pp. 1297 – 1304, 2007.

[30] I. P. Panapakidis and A. S. Dagoumas, “Day-ahead electricity price forecasting via the application of artificial neural network based models,” *Applied Energy*, vol. 172, no. C, pp. 132–151, 2016.

[31] H. Heidenreich. What are the types of machine learning? Accessed on 19-July-2020. [Online]. Available: https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f

[32] Unraveling artificial intelligence for the non-geeks the non-technical insight. Accessed on 17-July-2020. [Online]. Available: https://mc.ai/unraveling-artificial-intelligence-for-the-non-geeks%E2%80%8A-%E2%80%8Athe-non-technical-insight/

[33] Brian. (2018, February) ”scaling data for deep learning”. Accessed on 28-July-2020. [Online]. Available: http://www.feeny.org/scaling-data-deep-learning/

[34] J. Brownlee. (2014, 10) "an introduction to feature selection". Accessed on 18-July-2020. [Online]. Available: https://machinelearningmastery.com/an-introduction-to-feature-selection/

[35] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique," *Energy Conversion and Management*, vol. 50, no. 12, pp. 2976 – 2982, 2009.

[36] H.-T. Pao, "Forecasting electricity market pricing using artificial neural networks," *Energy Conversion and Management*, vol. 48, no. 3, pp. 907 – 912, 2007.

[37] J. A. Carta, P. Cabrera, J. M. Matías, and F. Castellano, "Comparison of feature selection methods using anns in mcp-wind speed methods. a case study," *Applied energy*, vol. 158, pp. 490–507, 2015.

[38] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, "The effect of wind generation and weekday on spanish electricity spot price forecasting," *Electric Power Systems Research*, vol. 81, no. 10, pp. 1924 – 1935, 2011.

[39] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the ontario competitive power system market," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 366–374, 2004.

[40] Y.-Y. Hong and C.-P. Wu, "Day-ahead electricity price forecasting using a hybrid principal component analysis network," *Energies*, vol. 5, no. 11, p. 4711–4725, Nov 2012.

[41] G. Van Rossum and F. L. Drake Jr, *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[42] T. E. Oliphant, *A guide to NumPy.* Trelgol Publishing USA, 2006, vol. 1.

[43] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445. Austin, TX, 2010, pp. 51–56.

[44] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[46] F. Chollet *et al.* (2015) Keras. Accessed on 21-July-2020. [Online]. Available: https://github.com/fchollet/keras

[47] J. Brownlee. (2019, 21) How to convert a time series to a supervised learning problem in python. Accessed on 14-July-2020. [Online]. Available: https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/

[48] J. Walker. (2019, 01) Tutorial: Time series analysis with pandas. Accessed on 29-June-2020. [Online]. Available: https://www.dataquest.io/blog/tutorial-time-series-analysis-with-pandas/

[49] Jacob. (2017, 08) "data science: Scaling of data in python". Accessed on 14-July-2020. [Online]. Available: https://medium.com/@stallonejacob/data-science-scaling-of-data-in-python-ec7ad220b339

[50] G. F. Volpi. (2019, 08) 6 amateur mistakes i've made working with train-test splits. Accessed on 10-August-2020. [Online]. Available: https://towardsdatascience.com/6-amateur-mistakes-ive-made-working-with-train-test-splits-916fabb421bb

[51] J. Brownlee. (2020, 06) "how to calculate correlation between variables in python". Accessed on 14-July-2020. [Online]. Available: https://machinelearningmastery.com/how-to-use-correlation-to-understand-the-relationship-between-variables/

[52] B. Jason. (2020, 06) How to perform feature selection for regression data. Accessed on 13-July-2020. [Online]. Available: https://machinelearningmastery.com/feature-selection-for-regression-data/

[53] B. Frénay, G. Doquire, and M. Verleysen, "Is mutual information adequate for feature selection in regression?" *Neural Networks*, vol. 48, pp. 1 – 7, 2013.

[54] J. Nikulski. (2020, 03) The ultimate guide to adaboost, random forests and xgboost. Accessed on 13-July-2020. [Online]. Available: https://towardsdatascience.com/the-ultimate-guide-to-adaboost-random-forests-and-xgboost-7f9327061c4f

[55] T. Sharp. (2020, 05) An introduction to support vector regression (svr). Accessed on 19-July-2020. [Online]. Available: https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

[56] Z.-W. Qiu, "Mutivariable mutual information based feature selection for electricity price forecasting," in *2012 International Conference on Machine Learning and Cybernetics*, vol. 1. IEEE, 2012, pp. 168–173.

[57] V. Roman. (2019, 01) Supervised learning: Basics of linear regression. Accessed on 05-August-2020. [Online]. Available: https://towardsdatascience.com/supervised-learning-basics-of-linear-regression/