INVESTIGATION OF THE IMPACT OF THE SOLAR POWER GENERATION FORECAST BY USING BIG DATA ANALYTICS ON THE LOCAL ELECTRICITY MARKET

BÜYÜK VERİ ANALİTİĞİ İLE YAPILAN GÜNEŞ ENERJİSİ ÜRETİM TAHMİNİNİN YEREL ELEKTRİK PİYASASI ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

OZAN OĞULCAN DEMİRTAŞ

Prof. Dr. AYŞEN SİVRİKAYA

Supervisor

Assoc. Prof. TUNCA DOĞAN

Co-Supervisor

Submitted to
Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfilment to the Requirements
for the Award of the Degree of Master of Science
in Renewable Energies.

2023

ABSTRACT

INVESTIGATION OF THE IMPACT OF THE SOLAR POWER GENERATION FORECAST BY USING BIG DATA ANALYTICS ON THE LOCAL ELECTRICITY MARKET

Ozan Oğulcan DEMİRTAŞ

Master of Science, Department of Renewable Energies

Supervisor: Prof. Dr. Ayşen SİVRİKAYA

Co- Supervisor: Assoc. Prof. Tunca DOĞAN

April 2023, 89 pages

Integration of Distributed Renewable Energy Sources (RES) into the existing energy system becomes more challenging as the number of RES increases due to their intermittent and variable nature. One way to address this issue is to use Local Electricity Markets (LEM) where consumers and producers can actively participate in trading locally produced electricity within their own Local Energy Communities (LEC). However, knowing the production value in advance (usually for a short period) is crucial for the formation of prices, evaluation of bids, and creating offers in local energy markets. Therefore, short-term load forecasting, which is an important parameter that helps electricity grid operators make decisions such as purchasing and selling electricity, load balancing, and maintenance planning, plays a significant role in system operations.

The aim of this study is to examine the possible role of a solar power plant whose short-term production value is estimated in advance through simulation in the local electricity day-ahead market and the effects it may have on electricity prices.

Additionally, this study aims to pre-shape pricing by obtaining bids before the day electricity will be supplied in this market.

In the first stage of the study, a high-capacity solar power plant was selected, and day-ahead electricity generation was estimated for this plant using past electricity production data and meteorological data from the plant's region. Due to the high variety and volume of the data, Big Data Analytics method was used in this analysis, and the analysis was carried out using machine learning techniques in Python programming. Three different models were examined, and the Light GBM model provided the best result for the day's electricity generation estimation In the second stage of the study, the forecasted electricity generation values for the modelled day were used in the local electricity market simulation model. Grid Singularity, an open-source and online software, was used to verify simulated scalable scenarios and evaluate LEMs economically. Firstly, a community was identified under Grid Singularity, and local market players were added for this community. Then, three different scenarios were developed to examine price formation, profitability, and the community's self-sufficiency thoroughly. In the first scenario, a solar power plant was not included in the community, and local market players were forced to meet all their electricity needs from the grid. In the second scenario, a solar power plant with high installed capacity was added to the system, and the simulation was run in this way. Finally, in the third scenario, two batteries with separate capacities of 10 kWh and 30 kWh were added to the system, unlike the second scenario, and the simulation was run again. In situations where solar energy could not be provided, local consumers purchased electricity from the battery, and it was observed that this increased the self-sufficiency of the community. When the results of all scenarios were evaluated, self-sufficiency rates were obtained as 0%, 65.0%, 69.0% & 77.0% (by depending on the battery power) respectively. The values indicates that the community can utilize the green electricity generated in the local market at the stated percentages. However, achieving these percentages fully is not possible due to the fact that solar energy is the primary renewable energy source in the community, and the production of the solar power plant is subject to fluctuations in meteorological values throughout the day. Moreover, it was achieved that penetration of substantial quantity of renewable energy into the system resulted

in a decrease of 26.7% and 30% in the average market price of electricity in the second and third scenarios, respectively, as compared to the first scenario. As a result, it has been observed that the integration of a high-capacity solar power plant into the local electricity market lowers market prices. Additionally, it has been emphasized that knowing the production that this plant will generate one day in advance allows market participants to take effective positions in the market.

Keywords: Big Data Analytics, Local Electricity Market, Machine Learning, Battery, Renewable Energy, Short Term Solar Power Generation Forecast Modelling

ÖZET

BÜYÜK VERİ ANALİTİĞİ İLE YAPILAN GÜNEŞ ENERJİSİ ÜRETİM TAHMİNİNİN YEREL ELEKTRİK PİYASASI ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Ozan Oğulcan DEMİRTAŞ

Yüksek Lisans, Temiz Tükenmez Enerjiler Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ayşen SİVRİKAYA

Eş Danışman: Doç. Dr. Tunca DOĞAN

Nisan 2023, 89 sayfa

Dağıtımlı Yenilenebilir Enerji Kaynaklarının (RES) mevcut enerji sistemini besleme entegrasyonu, kesintili ve değişken olduğu için RES sayısı arttıkça bu durum daha da zorlaşmaktadır. Bu sorunu ele almanın bir yolu, tüketicilerin ve üreticilerin kendi Yerel Enerji Toplulukları (LEC'ler) içinde yerel olarak üretilen elektriğin ticaretine aktif olarak katılabilecekleri Yerel Elektrik Piyasalarını (LEM'ler) kullanmaktır. Fakat, yerel enerji piyasalarında belirli bir süre öncesinden (genellikle kısa süreli) üretim değerinin bilinmesi, fiyatların oluşması, tekliflerin alınıp değerlendirilmesi açısından önem arz etmektedir. Bu nedenle elektrik şebekesi operatörlerinin elektrik enerjisi satın alma ve satma, yük değiştirme ve bakım planlaması gibi kararlar almasına yardımcı olan önemli bir parametre olan kısa süreli yük tahmini, sistem operasyonlarında önemli bir işlev görür.

Bu çalışmanın amacı önceden kısa süreli üretim değeri tahmin edilen bir güneş enerjisi santralinin simülasyon yoluyla elde edilecek bir yerel elektrik gün öncesi piyasasındaki olası rolünün ve elektrik fiyatları üzerinde doğuracağı sonuçların incelenmesidir. Ayrıca bu çalışmada, bu değere elektriğin temin edileceği günden

önce ulaşılması ile verilecek tekliflerin önceden alınarak piyasadaki fiyatlamaya önceden yön verilmesi hedeflenmiştir.

Çalışmanın ilk aşamasında, yüksek kapasiteli bir güneş enerjisi santrali seçilmiş ve bu seçilen tesisin geçmiş elektrik üretim verileri ile santralin bulunduğu bölgenin geçmiş meteorolojik verileri kullanılarak santral için gün öncesi elektrik üretim tahmini yapılmıştır. Verilerin çeşitliliği ve fazlalığı nedeniyle bu analizde Büyük Veri Analitiği yöntemi kullanılmış ve Python programlamada makine öğrenmesi tekniği ile analiz gerçekleştirilmiştir. 3 farklı model incelenmiş ve elektrik üretim tahmini yapılan gün için en iyi sonucu Light GBM modeli vermiştir.

Çalışmanın ikinci aşamasında, üretim tahmini yapılan gün için yapılan sözkonusu tahmini üretim değerleri yerel elektrik piyasası simülasyon modelinde kullanılmıştır. Simüle edilmiş ölçeklenebilir senaryoların doğrulanması ve LEM'lerin ekonomik açıdan değerlendirilmesi için açık kaynaklı ve çevrimiçi bir yazılım olan Grid Singularity kullanılmıştır. Öncelikle Grid Singularity altında bir topluluk belirlenmiş ve bu topluluk için yerel piyasa oyuncuları eklenmiştir. Daha sonra fiyat oluşumu, karlılık ve topluluğun öz yeterliliğini tam olarak incelemek için 3 farklı senaryo geliştirilmiştir. İlk senaryoda, güneş enerjisi santrali topluluğa dahil edilmemiş ve yerel piyasa oyuncuları tüm elektrik ihtiyaçlarını şebekeden karşılamak zorunda bırakılmıştır. İkinci senaryoda ise birinci senaryodan farklı olarak yüksek kurulu güce sahip bir güneş enerjisi santrali sisteme eklenmiştir. Son olarak 3. senaryoda 2. senaryodan farklı olarak sisteme ayrı ayrı 10 kWh ve 30 kWh kapasiteye sahip iki batarya eklenmiş ve simülasyon tekrar çalıştırılmıştır. Güneş enerjisinin sağlanamadığı durumlarda yerel tüketiciler bataryadan elektrik satın almış ve bu durumun toplumun kendi kendine yeterliliğini artırdığı görülmüştür. Tüm senaryolardaki sonuçlar değerlendirildiğinde, ilk senaryoda %0, ikinci senaryoda %65,0, üçüncü senaryoda ise batarya gücüne göre sırasıyla %69,0 ve %77,0 oranlarında topluluğun kendi kendine yeterlilik oranları elde edilmiştir. Bu değerler yerel piyasada üretilen temiz elektriğin belirtilen oranlarda topluluk tarafından kullanılabildiği anlamı taşımakta olup, bu oranların %100 olarak elde edilememesinin temel nedeni ise toplumdaki yenilenebilir enerji kaynağının güneş olması ve güneş santralinin üretiminin gün içinde değişen meteorolojik değerlere bağlı olmasıdır. Öte yandan, yüksek oranda yenilenebilir enerjinin

sisteme girişinin ortalama elektrik piyasa fiyatını birinci senaryodaki duruma göre, ikinci senaryoda %26,7 oranında, üçüncü senaryoda ise %30 oranında düşürdüğü gözlemlenmiştir. Sonuç olarak, yüksek kapasiteli bir güneş enerjisi santralinin yerel elektrik marketine entegrasyonun piyasa fiyatlarını düşürdüğü gözlemlenmiş olup, öte yandan bu santralin yapacağı üretimin bir gün önceden bilinmesinin piyasa katılımcılarının piyasada etkin pozisyonlar almasına olanak sağladığı vurgulanmıştır.

Anahtar Kelimeler: Büyük Veri Analitiği, Yerel Elektrik Piyasası, Makine Öğrenmesi, Batarya, Yenilenebilir Enerji, Kısa Süreli Güneş Enerjisi Üretim Tahmini Modellemesi

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

RES	Renewable Energy Sources
LEM	Local Electricity Market
LECs	Local Energy Communities
P2P	Peer to Peer
WEM	Wholesale Electricity Market
DERs	Distributed Energy Resources
ESCOs	Energy Services Companies
ICT	Information and Communication Technologies
NDCs	Nationally Determined Contributions
ANN	Artificial Neural Network
AL	Artificial Intelligence
ML	Machine Learning
DNOs	Distribution Network Operators
NRMSE	Normalized Root Mean Square Error
RMSE	Root Mean Square Error

1 INTRODUCTION

The integration of Distributed Renewable Energy Sources (RES) into the existing energy system is challenging due to the intermittent and variable nature of these sources, which becomes more complex as the number of RES increases. To address this problem, Local Electricity Markets (LEMs) can be utilized, allowing consumers and producers to actively participate in trading locally generated electricity within their Local Energy Communities (LECs). By using LEMs, the demand and supply of energy can be balanced at the local level, thereby reducing reliance on the wholesale market and minimizing the need for extensive electricity transmissions to or from the grid. The successful integration of Local Energy Community members into local energy markets with a high degree of automation requires interoperable information and communication technology such as blockchain. This technology can connect members through a market platform for peer-to-peer electricity trading, minimizing transaction costs.[1]

Short-term generation forecasting is crucial in local energy markets for price formation and evaluation of offers and bids. It helps electrical grid operators make important decisions such as purchasing and selling electrical energy, load switching, and maintenance planning. [2] However, forecasting short-term energy demand has become increasingly difficult because of the complexity of the power system due to the installation of smart grids and renewable energy sources such as wind and solar.[2] Despite this, it is possible to carry out accurate estimation analyses by using accurate historical data. Big Data Analytics is the most preferred analysis method in this regard, providing effective results for short-term forecasting, particularly for solar power plants. Short-term energy forecasting is especially vital for solar power plants, as it allows the system to be balanced day-ahead based on forecasts provided by the demand side to the day-ahead planning system.

The term "Big Data" refers to a novel concept that encompasses diverse digital content in varying volumes, which cannot be effectively processed using traditional database techniques. It comprises a mixture of structured, semi-structured, and unstructured data generated in large amounts and at high speeds.

A wide range of industries generate Big Data by either creating new data or digitizing existing ones. Knowledge in this field is a crucial factor for organizations seeking to gain a competitive edge. [3]

Using Big Data analytics, it is feasible to approximate the electricity generation value by analysing a vast amount of historical data from various renewable energy power plants. In order to verify the accuracy of the suggested technique for predicting short-term demand, actual data is employed to authenticate the model, and subsequently, the precision of the analysis is evaluated.[4] In the literature, there are numerous studies[5] that utilize data analytics methods based on historical data to forecast electricity generation from renewable energy plants. However, no study has explored the impact of the forecasted value on electricity prices in local electricity markets. Therefore, this study endeavours to explore this impact.

This thesis aims to explore how an increased penetration of renewable energy sources affects the prices of local electricity markets in the day-ahead market. To assess this impact, a simulation of the electricity market will be conducted. For the sake of accuracy, the simulation will focus on a local electricity market (LEM), which is better suited for the integration of many distributed renewable energy sources into the market than the national grid, resulting in efficient integration. Furthermore, in case of any potential electricity shortage or surplus, LEMs can interact with the wholesale electricity market.[6] In order to produce more accurate results, the generation data for the solar plant that will be included in the simulated local energy market will be anticipated one day in advance using historical data from the plant's location and previous generations of the chosen plant. Thus, with this study it will be emphasized the importance of having knowledge of the day-ahead price a day before. For market players and traders to plan their strategy in energy markets, day-ahead price forecasting is crucial.[7] Several meteorological conditions that can affect generation should be taken into account in order to assure accurate forecasting of solar power plant output. It is also important to assess how compatible these factors are with actual generation. To achieve more realistic results, the Big Data Analytics method will be employed, utilizing both historical meteorological data and previous generation data from the power plant. Big Data Analytics is a computer science technology that can be used to effectively process data from multiple sources, even those that may not be related, which is a challenge with traditional data analysis methods.[8]

Consequently, this research aims to contribute not only to the comprehension of how a high-capacity solar power plant influences market prices (known as the merit order effect) in the local electricity market, but also how the anticipation of electricity quantity value affects market variables such as bids, offers, self-sufficiency, profitability, and price formation. Additionally, this research could also encourage nations to switch from a single grid system to local energy markets. Due to the global expansion in the number of renewable power plants, this transformation may result in more local electricity markets.

2 THEORETICAL BACKGROUND AND LITERATURE

This section comprises three main parts. Firstly, the theoretical framework of electricity markets, including its models and components, is presented. Secondly, an overview of relevant studies in the literature exploring the relationship between renewable energy generation and electricity market prices is provided. Finally, the theoretical concept of local electricity markets and their components are explained.

2.1 General Characteristics of the Electricity Markets

The Electricity Market is a market consisting of generation, transmission, distribution, market operation, wholesale, retail sales, import and export activities of electrical energy and business & transactions related to these activities. Due to the fluctuation of electricity consumption based on months, days, and hours of the day, it is crucial to maintain a balance between the demand and supply of electrical energy every second. As a result, all operations in the electricity market are centered around achieving this balance.[9]

The primary aim of the electricity market is to offer affordable, reliable, and uninterrupted high-quality electrical energy. To achieve this objective, generation and trade activities have been shifted to a competitive structure to open the electricity market to competition and eliminate its natural monopoly characteristics.[10]

The two concepts that form the basis for the formation of more competitive electricity market are restructuring and deregulation. Existing companies can undergo restructuring, which may involve separating certain functions, merging others, or even creating entirely new companies. With the regulation, the prices of natural monopoly suppliers are determined and the entry of new players into the market is restricted. Regulation involves determining prices for natural monopoly suppliers and limiting the entry of new market players, while deregulation involves removing control over prices and allowing the entry of competitive suppliers into the market [11].

The first step in the liberalization of the electricity market was taken by Chile in the early 1980s and Chile was described as the pioneer of electricity reform.[12] On the other hand, deregulation of electricity market accelerated in the United States and the United Kingdom in the 1990s and has led to significantly increase grid efficiency as it allows for more competition.[13]

In a regulated market, the costs of energy, transmission, and distribution are set by regulatory and governmental entities. The market operates in a vertically integrated structure, and consumers are not permitted to choose their supplier. In contrast, in a deregulated market, prices are determined by the "invisible hand" of the market and market structure is horizontal. Different parts of the network are managed by different players, and competition is created among market participants.[13]

2.1.1 Components of Electricity Market

The main components of the electricity market are electricity generation companies, electricity transmission companies, electricity distribution companies, wholesale and retail companies and independent system operator.[14]

Electricity generation companies are companies that generates electricity. These companies are also responsible for the operation and maintenance of existing electricity generation plants.[15]

Electricity transmission companies consist of the transmission system, which is the most important element in the transmission of electricity from the power plants to the end users. The efficient and safe operation of the transmission system is essential for efficiency in the electricity markets.[15]

Electricity distribution companies transmit electricity to customers in certain geographic areas. A distribution company is a regulated electricity company that constructs and maintains transmission cables from the transmission network to the end users. Providing maintenance and voltage support as instant services are under their responsibility as well.[14]

Retail companies are nascent establishments in competitive electricity market industry. They have legal permits to sell electricity retail. The retailer purchases electricity and other necessary services in order to supply the required electricity to its customer and to sell electricity-related products and services in various packages. The retailer can also make an agreement with the end users indirectly through aggregators.[15]

On the other hand, wholesale refers to the sale of electrical energy and/or capacity for the resale. Trading takes place between power generation companies and intermediaries in markets. The intermediaries buy the electricity from wholesale companies and then resell it to retailers or other intermediaries. Transactions for buying and selling can be determined by agreements between companies or through organized wholesale markets. Power supply companies can vend to eligible consumers on both wholesale and retail bases.[15]

Providing the independent operational control of the grid, the independent system operator plays a role in managing transmission tariffs, ensuring system security, coordinating maintenance programs and long-term planning. The Independent System Operator functions autonomously and is not affiliated with any market participant, including transmission owners, generators, distribution companies, or users. Its responsibility is to ensure that all transmission system users are granted unrestricted access in a fair and impartial manner. The Independent System Operator holds the power to allocate and manage some or all resources, including implementing load reductions necessary to safeguard the system's security. This includes tasks such as eliminating transmission violations, stabilizing the balance between supply and demand, and maintaining the required network frequency. In addition, the Independent System Operator also sends appropriate economic signals that will motivate all market participants to invest in resources related to supporting efficient use and reducing constraints.[15]

There are typically two fundamental structures for independent system operators, which vary based on their goals and jurisdiction. The first structure is primarily focused on maintaining transmission security during power market operations. In this model, the independent system operator operates under a coordinated, multi-party commercial framework and does not have a role in the market itself. Its primary function is to provide security. As an example, The California independent systems operator has this kind of structure. Moreover, the independent system operator has no authority over forward energy markets and has very limited control over the planning of actual generation units.[15]

In the second structural model of the independent system operator, the power exchange is an integral part of the independent system operator operations. The power exchange system is an independent, non-government and non-profit organization that provides a competitive market through auctions for electricity trading. Power exchange calculates the market clearing price based on the highest bid. In some market structures, although the power exchange acts as an independent system operator, the independent system operator and the power exchange are separate entities.[15]

2.1.2 Electricity Market Models

There are three types of electricity market models: Pooling model, bilateral agreements model, and hybrid models.

The pooling model involves a centralized market that balances the market for buyers and sellers in the electricity market. Electric power buyers and sellers submit their price offers to the pool for the amount of electricity they wish to trade. In this market, sellers compete to supply power to the entire grid rather than to individual customers. If a market participant sets a high bid, they may not be able to sell their power, whereas buyers compete to purchase electricity, and if their offers are too low, they may not be able to acquire power. Producers with low production costs are incentivized in this market model.

In the pooling model, the independent system operator implements economic distribution and sets an independent (spot) price for electricity, giving clear signals to market participants for their consumption and investment decisions. The dynamics of the electricity market provide the motivation for spot market prices to equal the marginal costs of the most efficient power plants. In this market, winning bidders sell electricity at the spot market price offered by the winning bidder with the highest price.[16]

An alternative market model to the power pooling model is physical bilateral agreements model. According to this model, sellers and buyers set independent prices for the agreed-upon amount of electricity in their trade. Bilateral agreements model is characterized as a market-oriented model that allows for greater interaction between producers and buyers.[17] Sellers are power generation companies, and buyers are retail companies and eligible consumers. At the same time, the producers switch their positions to buyer here when they are unable to generate electricity due to power plant maintenance or breakdowns. Similarly, consumers can move to the position of seller. The terms and conditions of these contracts are determined independently of the system operator.[18]

The hybrid model is formed by combining various features of the bilateral agreements model and the pooling model. The hybrid model put forwards that purchase of electricity from pooling model is not mandatory and customers can make power supply contracts with suppliers or choose to purchase electricity at spot market prices. The pooling model provides services to all participants who choose not to sign bilateral agreements. Additionally, the system allows customers to negotiate power purchase contracts with suppliers, and suppliers can select the appropriate customers while offering customized pricing and services tailored to individual consumers.[16]

2.2 The Effects of Renewable Energy Generations on the Electricity Market Prices

Renewable Energy Sources (RES) are the most rapid growing sources for the electricity generation. Since 1990, the annual growth rates for photovoltaics and wind energy globally exceeded 30% and 20% respectively.[19] These outcomes have been reached thanks to supporting schemes and significant cost reductions for renewable energy systems. Moreover, the total installed capacities of PV and wind systems are expected to exceed the aggregated gas capacity in 2023 and total coal capacity in 2024. These developments in renewable energy sector are expected to contribute to mitigate the adverse effects of climate change crucially. As a result, the economy of power systems is changing all over the world due to lower implementation prices and benefits to the climate of PV and wind energy systems. In addition to that, lower implementation prices and environmental benefits of renewable energy may contribute to the tendency of consumers and producers to renewable sources as electricity generation source and thus to increase renewable electricity generation.[19] There are several advantages of PV, and wind power systems:

- Their installed capacities increase rapidly
- They are intermittent sources based on stochastic wind speed or solar irradiation. However, this can be managed by conducting supply analysis based on historical meteorological and generation data of these systems
- Their costs are almost stable
- Capacity additions on those systems are generally driven by support schemes like feed-in-tariffs[19]

As a result, the fast growth of these systems was started being seen as a major driver for decarbonisation globally. Another outcome of renewable energy sources is that penetration of these systems into wholesale market brings a decrease in the electricity prices in the wholesale market. This effect is called as "merit order effect of renewables" and it has been extensively discussed in the literature [19]

Increase in penetration of RES into wholesale electricity market has significant effect on electricity market prices since the marginal costs of RES are close to zero. Related cost for generation of the energy is attributed to construction of generator (Capex) totally. Thus, this contrasts with the generators of fossil-fuelled whose operational costs are dominative because of fuel and emissions.[20] Renewable Energy Generations play price taker role since the generation cost of them are always lower than market prices. Consequently, penetration of renewable energy generations to the grid, will reduce the cost of wholesale market prices since the conventional electricity generations with higher marginal cost are partly displaced with renewable generations. As a result of this, supply curve is shifted to the right and this effect is named as "Merit-Order Effect" in the literature.[21] Displacement of conventional generation also causes a reduction in the average price which renewable energy producers receive on the market. This term is referred to as "market cannibalisation".[22]

Merit order effect has effects on not only for RES suppliers but also for the energy policy and the prospects of decarbonisation of energy systems successfully. Therefore, it is essential to comprehend the scope and evolution of the merit-order effect in order to establish policies that effectively support both national and global decarbonization goals without unfairly disadvantage any stakeholders. This could have a variety of effects on how RES development will go in the future, such as how support policies will be created.[23]

In the literature, a lot of studies that have been carried out illustrated that RES generation decreases the wholesale electricity market prices. For instance, Ketterer[24], Paschen[25], Zipp[26], Benhmad & Percebois[27] & de Lagarde and Lantz[28] provided evidence about the merit order effect for Germany. Similar results have been achieved for other European countries, such as Denmark[29], Spain[30], Portugal[31], Ireland[32], Italy[33], Great Britain[34] and the Baltic states[35]. Outside of Europe, there are also studies from Australia[36], Canada[37], Massachusetts[38] and Texas[39] that show the merit order effect.

Based on the regression study (Halttunen et al.,2020) of electricity price on the variables of renewable energy (VRE) penetration, load, and seasonality, Figure 1 summarizes the merit-order effect on several countries' wholesale market. The change in price for a constant increase in VRE output (example: €/MWh per GW of output) is how the merit-order effect is commonly described in absolute terms.[23]

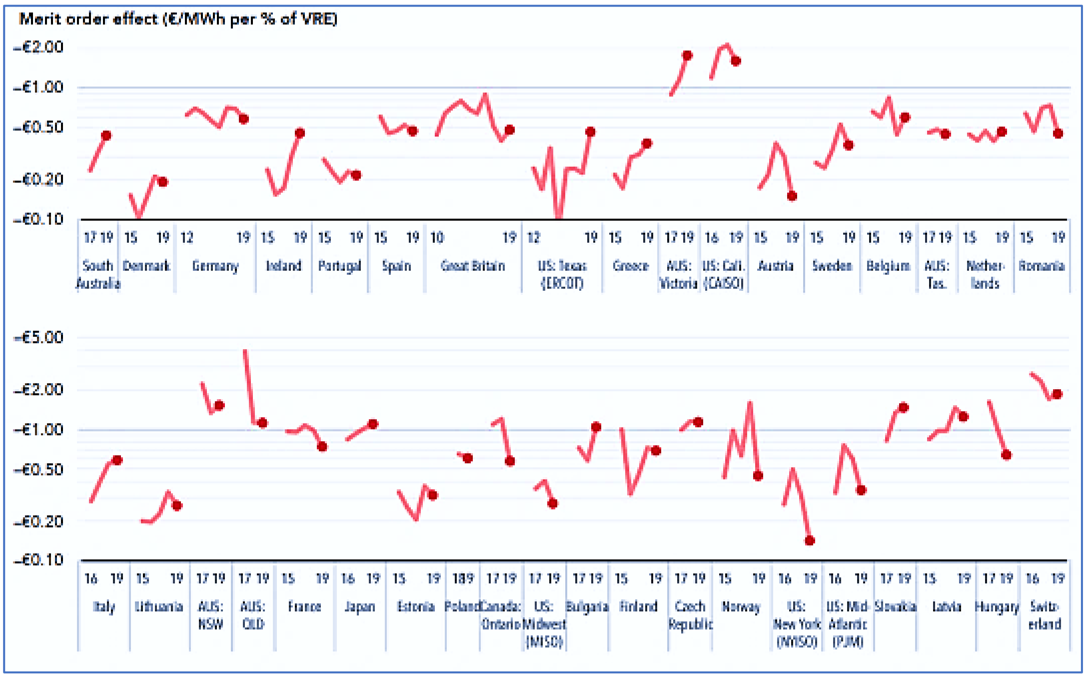


Figure 1. The annual average merit-order effect across 37 markets[23]

As might be expected, market prices were expected to fall further as the penetration of renewable energy increased. However, the results of the study of Halttunen et al. showed that merit-order effect gave weaker results with higher RES penetrations.[23] The relative merit-order effect ranged from €0.08 to €0.89 /MWh, with a mean of €0.41 ±0.21 /MWh (median €0.44 /MWh), focusing on the eight markets with RES penetration of above 20% (South Australia, Denmark, Germany, Ireland, Portugal, Spain, & Great Britain). The relative merit-order effect is larger in the other 29 markets, averaging €0.79 ±0.60/MWh (median: €0.61/MWh), where RES penetration is less than 20%. [23]

In addition, the merit-order effect has intensively been investigated in the literature and the studies have revealed that there are many factors that affect the size of merit order effect. These are technology type of renewables, market size, penetration rate of RES into grid, etc.[23]

In the following sections, the local electricity market and its components will be introduced first. Then, the big data analytics method will be employed to forecast the electricity generation amount of a selected solar power plant based on previous electricity generation and meteorological data. This forecasted value will be used in an online simulation software for local electricity market modelling to analyse the impact of penetration of high-capacity renewable electricity on market prices.

2.3 Local Electricity Market

Since LEMs have been grouped into several categories and used in a variety of use scenarios, there is no universal definition of them in the literature. LECs, microgrids, P2P trading, energy sharing, and energy exchanging are the other terms that are similar to LEMs, as noted in [40], which leads to confusion. For this reason, general terms of it which were introduced by European Commission are described in the below sections.

2.3.1 Local Electricity Market Definition

In 2016, European regulatory agencies proposed a novel electricity market model that is decentralized, intelligent, and networked to achieve energy policy objectives such as decarbonization, affordability, and supply security in the power sector[41]. This plan also suggests that consumers become prosumers so they can be more involved in their energy purchases and help to maintain the stability of the power system. In keeping with this goal, the European Commission coined the word "Local Electricity Communities (LECs)" and characterized it as an effective method of producing local energy at the community level and involving all consumers, prosumers, and producers in LEMs to enable them to trade localized energy.[1] Figure 2 represents the LEM schematically.

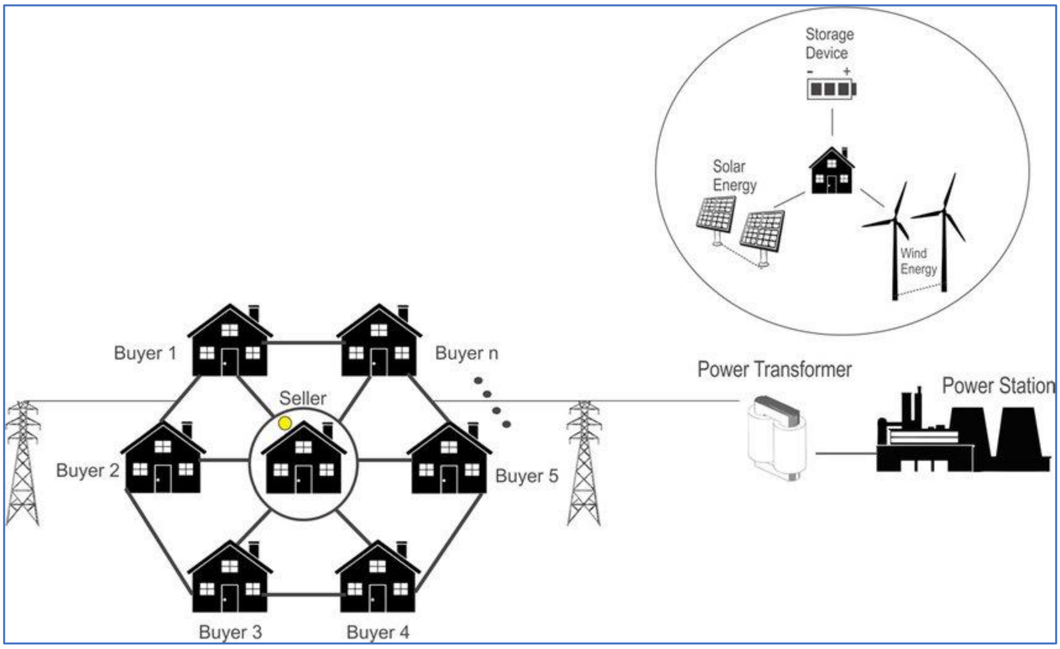


Figure 2 Schematic Representation of a LEM[42]

Local renewable energy production facilities compete on the same market at the wholesale level as centralized power plants with large capacities. There is a minimum amount of energy that power plants must produce in order to be able to sell their energy on the wholesale electricity market (WEM), as bids to supply energy on the WEM cannot be unreasonably low. The integration of small and decentralized power plants is hindered by their inability to generate the required amount of energy to participate in the wholesale electricity market.[43] The adoption of peer-to-peer markets at the local level removes the need for prosumers to compete with centralized power plants on the WEM. Due to the coexistence of LEMs and the WEM, as well as the fact that both markets utilize the same electrical grid, the idea of local electricity markets should be taken into consideration when planning to implement the wholesale electricity market. This suggests that new market-regulatory frameworks developed for the integration of local peer-to-peer(P2P) marketplaces into the WEM should ensure the coexistence of various market types.[44]

As it was indicated in the previous paragraph, market mechanisms are the most important parts of LEMs. In LEMs, market mechanisms can be implemented as P2P or Order Book Mechanism. Fundamentally, the order book market mechanism involves a solitary price, referred to as the market clearing price, which is paid by consumers and received by prosumers/producers. Conversely, in the P2P market mechanism, prices can vary for each individual trade since they are one-to-one transactions. [1]

By setting their own prices, LEMs seek to maximize the use of distributed RES inside LECs while also promoting the active participation of LEC members in the local energy trading platform[45]. Another objective is for LEMs to aid distribution system operators and their operations in managing grid bottlenecks and addressing electricity balance issues at the local level [6]. To establish an appropriate market for the exchange of local energy, it is necessary to establish a set of rules that define how the market participants interact with one another and engage in trade.

Consequently, a decentralized market mechanism and its physical parameters are the basis for the design of LEM. In the context of market matching and pricing procedures, electricity is sold and purchased on LEM trading platforms. Moreover, LEMs facilitate information flow among all market participants, whereas electrical flows occur on the grid infrastructure. Furthermore, LEMs encompass renewable energy producers in addition to residential and business users and prosumers, who may engage in the trading of local electricity via LEMs.[1]

2.3.2 Advantages of Local Electricity Market

There are four main advantages that are offered by LEMs. These advantages are described below.

- **Customer Level Advantage:** In Local Electricity Markets, market participants that can generate their own electricity, by optimizing the use of local energy sources (including RES and Energy storage) they can consume their own generated electricity and at the same time they can sell the excess of the electricity in the local market. This customer type is called as “prosumer” in LEM. When compared to the traditional electricity "price-taker consumers," this gives them a greater degree of energy independence and control, which are crucial driving force for them to participate in LEMs.[46][47]

- **Network Operators Level Advantage:** The customer-owned distributed energy sources generation found in LEMs will have a variety of effects on how distribution and transmission networks run on a daily basis. The addition of Distributed Energy Resources (DER) capacity, as well as enhanced flexibility and more effective overall network operations, can lessen the need for network operators to undertake additional investments and reinforcements in the distribution grid.[47]
- **Advantages for Providers of Energy, Technology and Service Level:** The use of contemporary technologies and ongoing product innovation will be crucial for LEMs. Opportunities will arise for market participants in this situation to realign their strategy and create new goods and/or services in response to changing consumer needs. As its models and governance mechanisms change, new market participants will appear. Most of the time, these actors will act as third-party brokers between consumers and network and market operators. For instance, by offering their core services, aggregators benefit both operators and customers while making money at the same time. Another illustration is energy services companies (ESCOs), which can provide clients with significant value by lowering their energy demand. Moreover, a significant move towards increased market transparency can be achieved by aligning ESCOs' revenue with the savings achieved for their clients. In the area of proactive information exchange between stakeholders, a crucial component of LEMs, new business opportunities will present themselves. This interchange will be driven by information and communication technologies (ICTs), which are crucial for making the distribution network more effective and adaptable. The expansion of new participants into current LEMs will be prompted by the existence of sizable commercial prospects to create and offer relevant flexibility and energy management services.[47]

- **National Level Advantages:** A significant project with numerous advantageous societal effects is the transition from a conventional energy model to a new paradigm grounded in decentralized and customer-centric energy production and distribution. LEM initiatives not only acknowledge but also extend the paradigm of sustainability by pledging improved market openness, more just power allocation, and a more even distribution of systemic expenses and incentives, all of which are fundamental to societal development. By facilitating the increase of clean energy production, particularly from renewable energy sources (RES), LEMs make a meaningful contribution to reducing local and global greenhouse gas emissions. This is in harmony with the goals of the Paris Climate Agreement and the Nationally Determined Contribution (NDC) of participating nations. [47]

3 METHOD: BIG DATA ANALYTICS

3.1 Big Data Analytics

The data volume and data diversity in the world are increasing rapidly. With the introduction of internet technologies and social media into every stage of our lives and even our mobile phones, people can make the data available even in their daily activities. Manually operated vehicle requirements of yesterday are referred to as smart devices today and almost all of them generate data with sensors. The increasing generation of such dense and diverse data from different sources has created a new concept which is “Big Data”. Big Data is a set of structural, semi-structural and unstructured data produced in high volume, speed, and variety. Since Big Data is heterogeneous data in different volumes and consists of various digital contents, it cannot be processed using traditional database techniques.[48] Figure 3 summarizes the advantages of using Big Data analytics. As it can be understood from the figure, Big Data has several advantages such as, cost reduction, effective decision-making, creates new products & services.

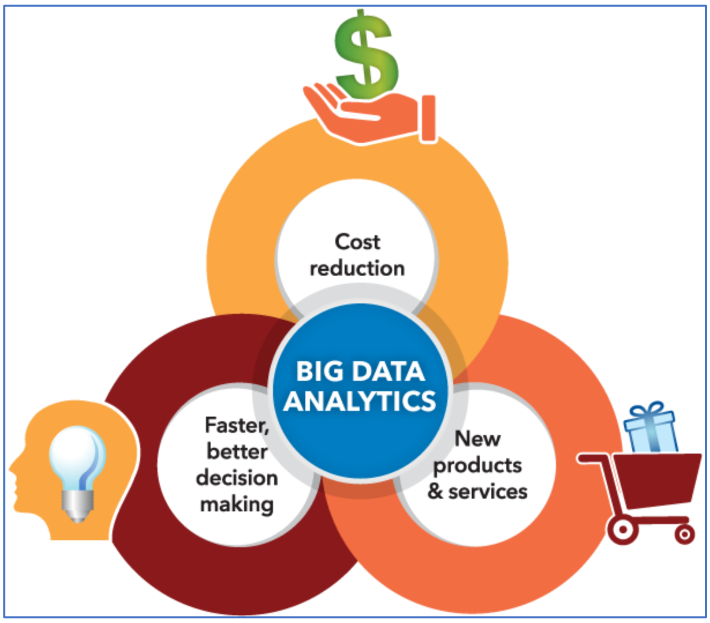


Figure 3. Advantages of Using Big Data Analytics[49]

There are 3 main components (3V) that characterize Big Data analytics which are variety, velocity, and volume. In addition to 3V, which defines Big Data Analytics, in some sources, Veracity and Value components are also included as main components and makes that 5V instead of 3V. In Figure 4 below, components of Big Data are represented [48].

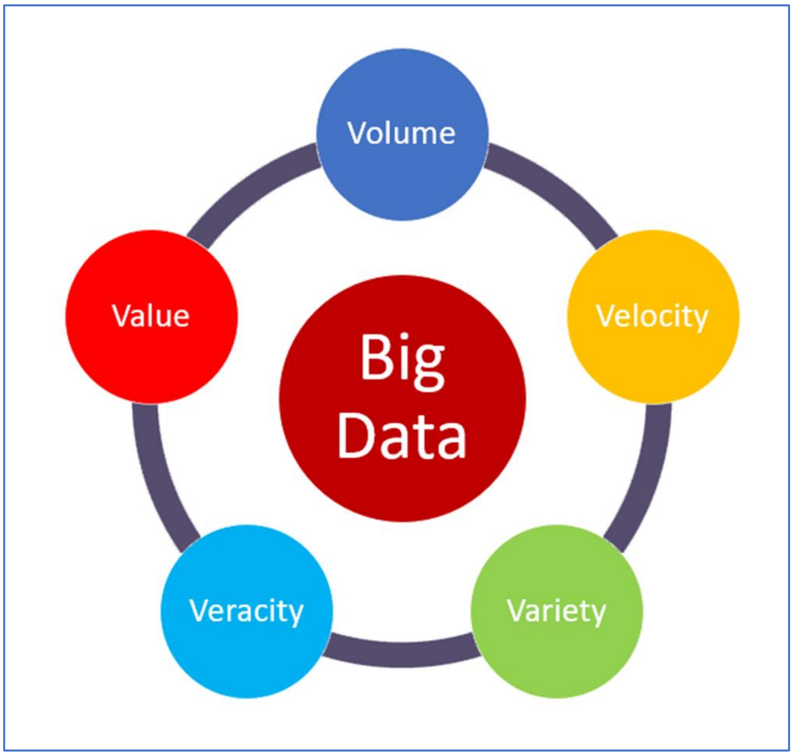


Figure 4. 5V Components of Big Data[50]

Definition of these components are explained below separately.

Volume: The size of the data, which is expressed in measures such as Gigabytes, Terabytes, Petabytes, can also be expressed as volume. Large amounts of collected data constitute the volume dimension of Big Data.

Velocity: The data obtained for processing can be produced at variable speeds. While large volumes of static data can be a problem, fast generated data also creates processes that need to be managed.

Veracity: Veracity refers to the accuracy and reliability of Big Data. The data must be dependable enough to be utilized in making business decisions. The high diversity of Big Data complicates the process of ensuring the quality and reliability of the analysed data. In such cases, the data must be pre-processed to obtain the correct data.

Variety: Different data sources cause data diversity. The data obtained can be structural, semi-structural and unstructured. The diversity dimension of Big Data is provided because the types of used data can differ.

Value: Data that does not turn into information and does not produce value is useless. With classical data analysis methods, it is difficult to analyse large amounts of data and the desired values are not produced. It is possible to produce important and valuable outputs for systems with the analysis of Big Data.[51]

3.1.1 Applications of Big Data Analytics

Big Data analytics provides many benefits and convenience and facilitates effective analysis, so it is used in many areas such as, banking sector, communication, media and entertainment sectors, healthcare sectors, education sector, production sector, government services, insurance sector, retail and commerce sector, transportation sector and lastly in energy sector. [48]

3.1.1.1 Application of Big Data Analytics in Energy Sector

Big Data Analytics has been widely used in Energy Sector. First of all, with the better and qualified sources, it provides better workforce management, hereby identification of operational problems before the system failures. In this way, it enables operators to review the problem easily and quickly. For instance, using smart meters that collect information every 15 minutes is a more effective way of controlling customers' consumption information and energy infrastructure compared to traditional meters that gather data only once a day. Energy Big Data encompasses not just smart meter readings but also a vast quantity of data from additional sources such as weather data and geographic information systems. One instance of integrating data from energy generation and consumption, geographic information systems, and weather - including temperature, atmospheric pressure, humidity, cloud cover, wind speed, and wind direction - is to assist in locating new renewable energy generation devices, ultimately enhancing power generation and energy efficiency[48].

3.1.1.2 Application of Big Data Analytics in Renewable Energy Sector

There are lots of main advantages of using Big Data Analytics in renewable energy industry. The main advantage is that by using historical meteorology data, future weather conditions can be forecasted. Similarly, as mentioned in the previous section, historical operational data can also facilitate the effective organization of operation and maintenance processes. Also, use of historical operational, meteorological, and technical data can reduce renewable energy generations costs, making renewable energy projects more reliable and creditable. As it is indicated previously renewable energy projects are counted as risky projects since generation of energy depends on natural resources and the environmental conditions. However, by using Big Data analytics, historical meteorological data belong to the project site can be analysed and this enables financers to know whether this project will generate expected energy amount in the future or not. In addition, Big Data allows market players to forecast the number of future generations of related plants, which in turn facilitates the identification of the market value of energy in advance [52].

3.1.2 Techniques Used in Big Data Analytics

Since many areas started to utilize Big Data, scientists have created a wide range of methods and tools for collecting, organizing, analysing, and visualizing big data. They still fall short of fulfilling a range of needs. To get a high value from Big Data Analytics, new techniques are needed to be developed. To extract the important information from Big Data, multidisciplinary approaches are required. Numerous fields are involved in Big Data techniques, such as statistics, data mining, machine learning, neural networks, social network analysis, signal processing, pattern recognition, optimization strategies, and visualization techniques. These disciplines have a wide variety of specialized methodologies, which constantly overlap [53]. The used techniques are illustrated in Figure 5.

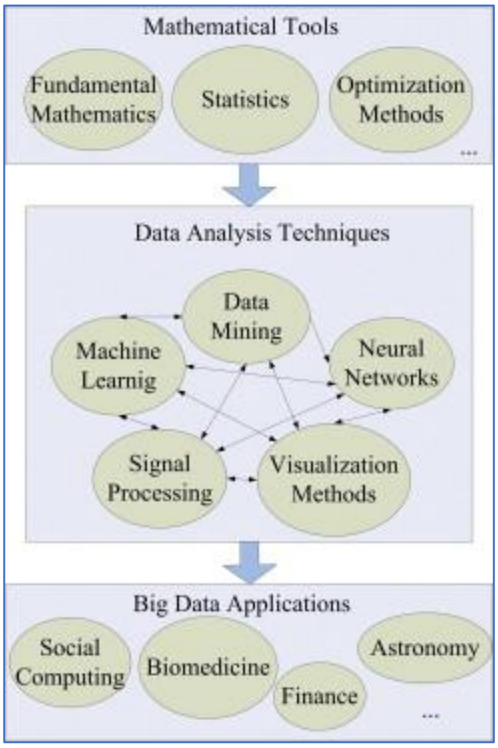


Figure 5. Techniques in Big Data Analytics[53]

Data mining: Clustering analysis, classification, regression, and association rule learning are just a few of the methods used in data mining, which is a collection of methods for identifying valuable information (patterns) in data. Comparing typical data mining techniques to Big Data mining presents greater challenges and to ease these challenges machine learning and statistical methodologies are used.

Neural networks: The artificial neural network (ANN) is an established method with numerous applications. Pattern recognition, image analysis, adaptive control, and other fields all make use of it successfully. The majority of the ANNs for artificial intelligence that are currently in use are based on control theory, classification optimization, and statistical estimations.

Visualization methods: These techniques are used to produce tables, pictures, diagrams, and other types of understandable data displays. Big Data visualization is more challenging than visualizing standard, relatively small data sets due to the complexity involved in managing the 3Vs or 4Vs (i.e., volume, velocity, variety, and veracity).

Machine learning: A significant branch of artificial intelligence called machine learning aims to create algorithms that let computers develop new actions based on actual data. Machine learning is characterized by its capacity for intelligent decision-making by gathering autonomous information. To handle Big Data, machine learning techniques are needed to scale up, including supervised learning and unsupervised learning.

Traditional data analysis methods are inadequate for fully harnessing the benefits of Big Data. The sheer volume of information available for analysis makes it impractical, if not impossible, to verify every assumption made about the data. Additionally, the complexity of Big Data means that potential linkages and relationships between data can be easily overlooked. Due to its high throughput and excellent performance with massive data sets, machine learning is a great tool for uncovering hidden correlations or links between data. Because the machine learning algorithm "learns" from the current data and applies the discovered rules on new entries, the more data we have, the more valuable it becomes.[54]

This thesis focuses on forecasting the electricity generation of a solar power plant by using different types of meteorological and the realized past electricity generations data, it applies the machine learning technique.

3.1.2.1 Machine Learning

A subdomain of computer science called "machine learning" is used to evaluate data and automate the creation of analytical models. Algorithms for machine learning are designed to learn from the data that is already available. When models are introduced to fresh data sets, they adjust on their own, which is a trait that results from machine learning's iterative feature, and this is a crucial part of machine learning. For the purpose of generating certain and repeatable decisions and results, these models are learning from earlier calculations. New machine learning approaches have been developed in recent years by researchers that work compatible with Big Data.[54] Google’s autonomous car and the recommendation systems of Amazon or Netflix can be given as examples to these developments. In Google’s autonomous car example, the vehicle is outfitted with software that processes and analyses all data received in order to provide safe navigation. The concept of machine learning is exemplified by this application example.[55] On the other hand, in 2014, a constrained Bolzman machine and a form of matrix factorization served as the foundation for the Netflix recommendation engine. These are illustrations of real-world applications of machine learning.[54]

Because machine learning is effective with vast and varied amounts of data, interest in it has grown. Additionally, computing processing is more efficient and powerful. As a result, models are created quickly and automatically for the analysis of big and complicated data as well as for faster delivery and more accurate results. The usage of these models produced extremely accurate forecasts that allowed for the taking of smarter decisions and intelligent actions in real time without the involvement of humans.

Supervised learning and unsupervised learning are the two different categories of machine learning approaches. Supervised learning, which is based on a defined classification model from which the computer should learn, is frequently employed in challenges in which the data should be classified. On the other hand, unsupervised learning is used when the computer must learn how to do a task without any instructions. This kind of learning generalizes it close to the real world. Even though supervised learning is more effective than unsupervised learning, in reality unsupervised learning is more frequently used.[54]

3.1.2.1.1 Machine Learning Applications for Solar Energy Forecasting

Today, the globe is moving toward redefining the energy balance and utilizing renewable energy sources to generate electricity. Sen et. al suggests that more environmentally friendly and long-lasting electrical system can only be created if bigger proportions of renewable electricity can be added to the energy mix.[56] But the energy context is characterized by fluctuating prices, shifting demand, and unstable renewable energy production. In this situation, since it is the cleanest and most abundant renewable energy source available, the most promising renewable option for generating electricity is solar energy. However, when it comes to solar energy, instability in the energy generation is influenced by the materials used to make photovoltaic (PV) cells as well as other elements like non-stationary meteorological variables. Therefore, it is crucial to employ a suitable model for predicting the production of renewable energy that can learn from past meteorological data and assist in the energy sector's operational optimization and decision-making. In other words, in order to make solar power plants more competitive in the energy market and less dependent on fossil fuels for economic and social growth, solar energy forecasting is a crucial component.[57]

PV forecast models are divided into two groups in literature, namely indirect and direct forecast models. The former utilizes various methods to predict solar radiation on different time scales, which is then transformed into power using panel characteristics. On the other hand, direct forecasts are made directly from the generated power of the plant.[58],[59] In addition, there are four distinct categories into which methods for predicting solar energy can be classified: statistical, physical, artificial intelligence (AI), and hybrid.[57] Nevertheless, Liu et al. state that AI, and in particular, Machine Learning (ML) techniques, have been widely employed in this area due to their significant learning and regression capabilities. [60]

3.1.3 Applications of Big Data Analytics to Local Electricity Markets

There are two different electricity market types which are centralized and decentralized. In the centralized market traditional single grid connection system, which is also called vertical system has been adopted. Under the Vertical System, local energy producers and consumers have limited access to the electricity market, but they may participate in central energy or frequency markets if they gather to a sufficient scale. However, from the consumer's perspective, the current vertical business model may not be beneficial for local energy producers, especially those who use renewable technology, as it provides limited options for utilizing their output. In particular, the grid only recycles a very small amount of the extra intermittent energy from prosumer renewable resources in their raw form.

In decentralized market type, which is Local Electricity Market, adopted system is Horizontal System. The Horizontal System allows for the horizontal sharing of renewables among flexible users and energy producers, which encourages customers to share responsibility for maintaining network security and reducing the costs associated with the exclusive central control of Distribution Network Operators (DNOs) when integrating low-carbon technologies into the system. It would also enable customer flexibility to resist the local supply uncertainty.[61] Schematic representation of Vertical and Horizontal Transactions are given in Figure 6. As it can be understood from the figure, in vertical market type, excess amount of electricity generated by producers is purchased by grid at a relatively low price (£0.03/kWh) and it is recycled by the grid and sold to consumers at a higher price (£0.17/kWh). However, in horizontal market type, producers can sell their renewable electricity at a better price (£0.10/kWh) to the consumers directly by using a trading platform which can be accessed by producers and consumers.

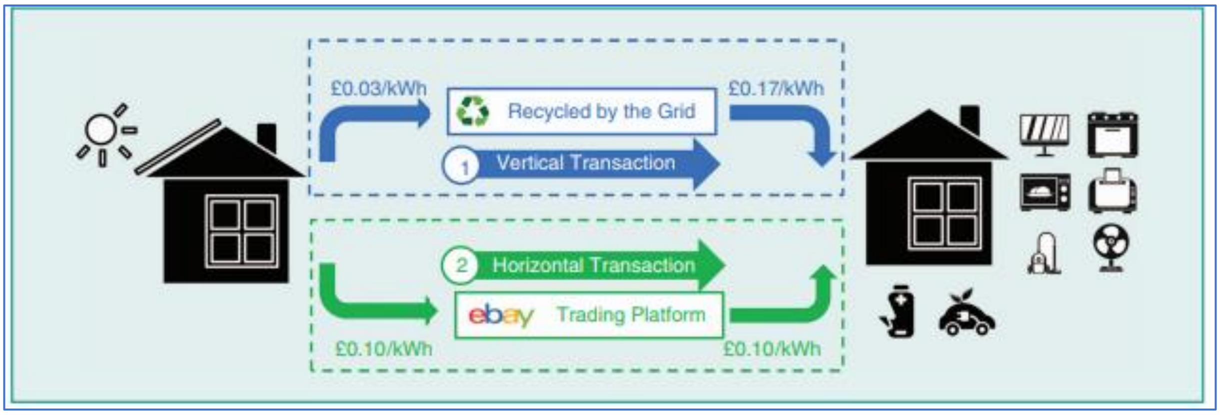


Figure 6. Vertical and Horizontal Systems[61]

In local electricity market, prosumers are thought to be able to trade without using the central system by cooperating horizontally. Such regional market activities would send out signals and encourage local consumers to alter their patterns of demand and follow the output of regional producers, absorbing the uncertainty locally. The prevalent half-hourly energy trading system is made for conventional, centralized, large-scale generating. Distributed renewable energy producers differ greatly from one another in many ways. Pricing and balancing intermittent electricity generation with demand flexibility depends on determining the amount of generated electricity that can be sold over time and space. The amount of energy that atypical solar energy producer can provide to a central power market change significantly over time. Figure following displays an example PV generator's daily output together with its average diurnal value.

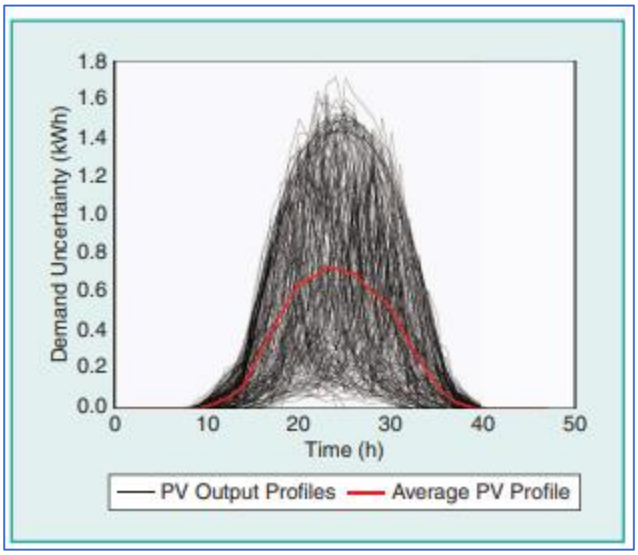


Figure 7. Daily generations of a PV system over a year (The red curve shows the average diurnal value)[61]

The uncertainty in electricity generation from renewable sources can make it challenging for markets to balance supply and demand within the local community. Energy might be traded in multiple time blocks (from seconds and minutes to hours), in various quantities, and, most importantly, with various supply reliability on a P2P local energy market. To provide an energy trading system that is multi-time, multiscale, and multi-reliability and is optimized for tracking supply/demand levels and their variability, Big Data analytics will be essential. In this regard, Big Data Analytics can support the creation of Local Energy Markets (LEMs) in the following ways[61]:

-**Resource characterisation** is necessary to comprehend the biggest and smallest trade units, the biggest and the smallest time blocks, and the biggest and the least reliable supply sources.[61]

-**Probability forecasting:** This will be utilized to predict the probable amount of energy to be produced and consumed during the bidding periods, which range from seconds to hours. To evaluate the uncertainty associated with the prediction, a probability distribution of each supply and demand level will be estimated for each bidding period.[61]

-**Ultrafast settlement:** The related surpluses or shortages would need to be resolved because there may be substantial uncertainty between forecast energy and real-time trade. Big Data analytics may enable quick trading and settlement, depending on the P2P market's size and time frame.[61]

Due to the interdependence of real-time forecasting, pricing, and matching, very effective data processing is needed to handle a potentially very large number of offerings. To ensure that the market accurately represents the availability and condition of the offerings, forecasting must be done quickly and accurately. In local electricity markets, quality of matching is determined by forecasting of generated electricity amount and pricing.

Big Data analytics will be a crucial enabler because they can inform consumers and market players, set prices, and balance demand flexibility with intermittent generation through in-depth analysis of metered data along with socioeconomic and weather data. Moreover, market operation must be more effective than the central half-hourly market for local markets to compete, and this effectiveness can be provided by Big Data analytics.

As previously mentioned in earlier sections, this study will highlight the significance of knowing the day-ahead electricity generation quantity and price a day in advance. This is because day-ahead forecasting of generation and price is crucial for local electricity market participants and traders to reduce supply risk in the market and to effectively plan their trading strategies[7]. For this reason, Big Data analytics is used to forecast day-ahead electricity generation of chosen solar power plant of simulated local electricity market in this thesis.

4 SOLAR POWER PRODUCTION FORECAST WITH BIG DATA ANALYTICS

This chapter gives the details of the technical and spatial characteristics of the solar power plant to be used for the thesis, the characteristics, and types of data to be used for the forecasting analysis, and finally the results of the forecasting analysis.

4.1 Specifications of Chosen Solar Power Plant

The main focus of this thesis is on a simulated local electricity market that comprises a solar power plant. The plant consists of 99,708 photovoltaic panels and covers an area of 519,220 square meters, with a total installed capacity of 26 MW. Its annual electricity production averages at 36,000 megawatt-hours, which is sufficient to meet the daily energy needs of 10,000 people, including housing, industry, metro transportation, government offices, and environmental lighting.[62] When considering only household electricity demand, the solar plant is capable of powering 12,000 homes.[63] Additionally, the latest Turkey National Electricity Grid Emission Factor for solar power generation plants has been calculated as a combined margin emission factor of 0.6488 tCO_2/MWh by the Ministry of Energy and Natural Resources of the Republic of Turkey. Based on this value, it can be said that the power plant has approximately reduced 23,356 tons of CO2 emissions for the generation of 36,000 megawatt-hours. [64][63]

4.1.1 Location of the Solar Power Plant

The power plant is located in the city of Niğde in the Central Anatolian Region of Türkiye. The location of the city is illustrated on the map of Türkiye given in Figure 8.

Figure 8. Location of City Niğde on the map of Türkiye[65]

4.1.2 Climate Conditions of the Location

Niğde is situated in the northern hemisphere and generally has a warm and hot climate. The city receives more rainfall during winter than summer and has an average annual temperature of 9.1°C, with average annual precipitation measured at 513 mm.[66] Furthermore, the graph presented with Figure 9 illustrates the average daily sunshine hours per month in Niğde. As evident from the graph, July has the highest number of sunny hours per day in Niğde, with an average of 12.35 hours of sunshine and a total of 382.95 hours of sunshine throughout the month [66]. Conversely, Niğde experiences the least number of daily sunny hours in January, with an average of 6.36 hours per day and a total of 197.19 hours of sunshine over the course of the month. A total of 3347.52 hours of sunshine are recorded annually in Niğde, with an average of 109.91 hours per month. [66]

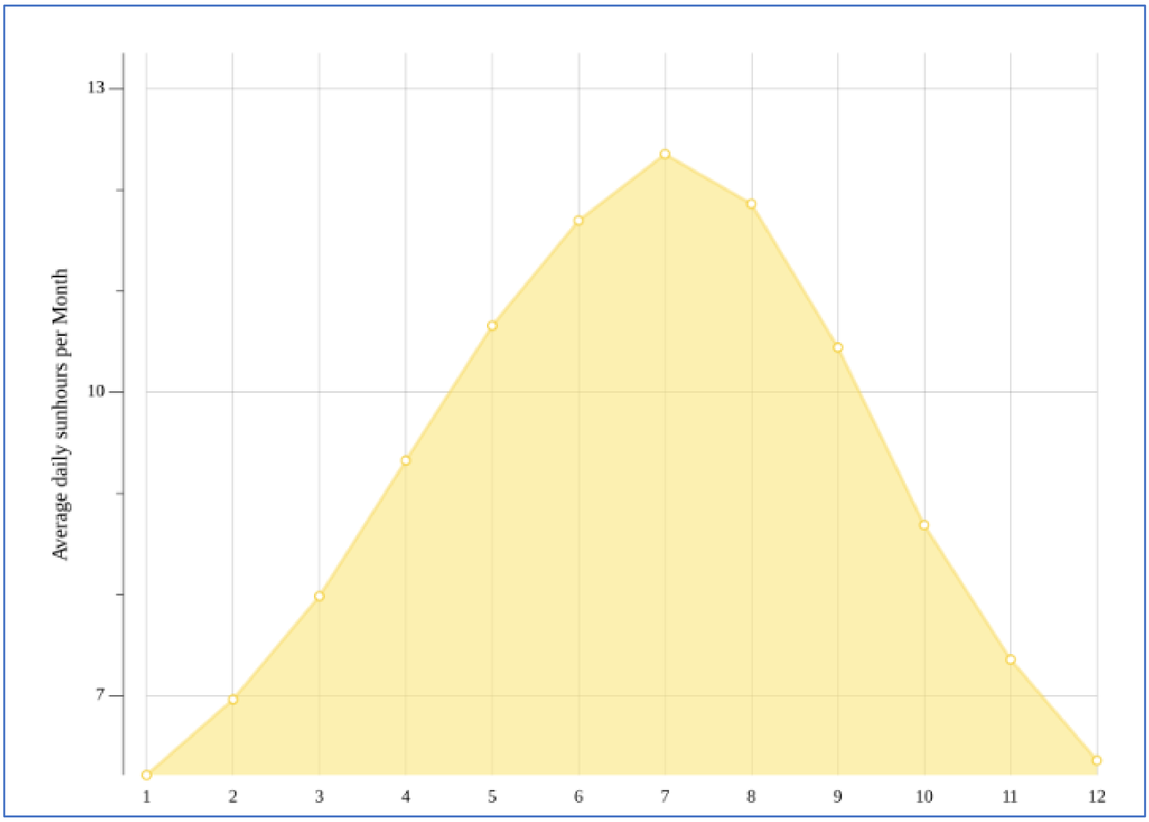


Figure 9. Average Daily Sun hours per Month in Niğde, Türkiye[66]

On the other hand, when analysing Türkiye's irradiation map, it can be observed that Niğde falls under the area that receives a total solar radiation of 1650-1700 KWh/m2-year, as depicted in the Figure 10[66].

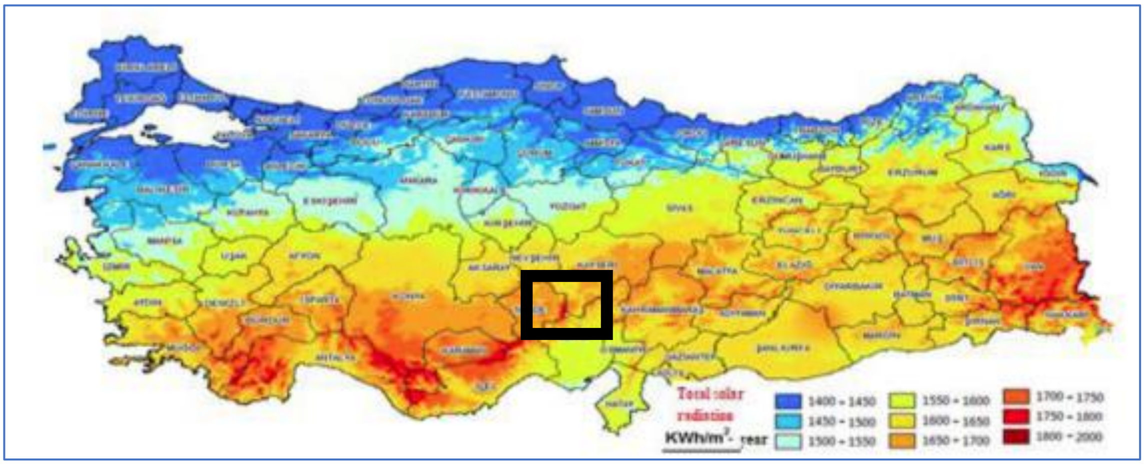


Figure 10. Solar Irradiation Map of Türkiye.[67]

4.2 Forecasting analysis

This section outlines the procedures for the forecasting analysis and provides key information about the data utilized in the analysis.

4.2.1 Input Variables and Forecast Variable

Many studies in the literature have utilized big data analytics methods to forecast data in solar power plants by compiling and analysing historical data [68]. Table 1 is a summary of some of the studies conducted in this field in the literature.

Table 1. Examples of solar power plant generation forecasting studies from the literature

Study, Year	Forecasting period	Input Variables	Forecasted output	Data Period
Zhang et al., 2015[69]	24-h ahead	Site 1 = Wind speed, Temperature of ambient, temperature of cell, Site 2 = cloud cover, GHI, fog, wind speed, temperature of ambient Site 3 = Altitude of Solar, GHI, cloud cover, DNI, Temperature of ambient	Solar power	Site 1 = 1st July 2010– 31st December 2011 Site 2 = 1st January – 31st December 2006 Site 3 = 1st January – 31st December 2011
Lima et al.,2016[70]	24-h ahead	Weather data by model of WRF , Observational ground data	solar irradiance	2009 &2011 including seasons rainy & dry
Wang et al.,2017[71]	15-min ahead 30 min ahead 1h ahead 2h ahead	Retroactive Solar generations	Solar Generations	Jan-Dec (2015)
Hossain et al.,(2017)[72]	Day ahead 1-h ahead	Avg. solar irradiance Temperature Module temperature wind speed Retroactive PV output	Solar Generations	1st January 2015–30th September 2016 (12months)
Mohammadi et al.,2015[73]	Daily and monthly	Solar radiation. Duration of sunshine Min, avg., max, temperatures Relative humidity Water vapor pressure	Solar Radiation (Horizontal)	From Jan 1992 to Dec 2005

Ekici, 2014[4]	Next-day	Avg. &max ambient temperatures Duration of sunshine Solar insolation (Day before)	Solar insolation (Daily)	2000–2003

In a review study which was carried out for solar photovoltaic generation forecasting method by Sobri & Koohi-Kamali & Rahim[5], the most important and used parameters in the previous forecasting analyses were examined, such as, solar irradiation, past PV generations, temperature and wind speed. Then, the meteorological parameters specified in the following section were narrowed accordingly.

4.2.2 Forecast Horizon (Dataset preparation)

This thesis utilizes the data for the selected solar power plant's retroactive generations, which is retrieved from Transparency Platform of EPİAŞ. Transparency Platform provides reliable, fair, and transparent data to Türkiye's energy markets. EPİAŞ, in compliance with the Board Decision, gathers information from relevant institutions, organizations, and businesses to publish data on its Transparency Platform, including information on the marketplaces where it conducts trades.[1] The data span covers the period between 9th October 2019 and 18th September 2021 for the chosen solar power plant for this study. Since the plant started its operation the oldest data that could be reached was 09.10.2019.

In addition to retroactive generation data, this study also employs historical meteorological data at the power plant's location for the forecast modeling. These data were obtained from Solcast DATA API, a platform developed by Solcast that provides forecast, live, and historical solar irradiance, PV power, and weather data.[2] The meteorological data for the period between October 9th, 2019 and September 18th, 2021 were retrieved using the coordinates of the region via the Solcast DATA API toolkit.

[1] 6282-4 decision of the Energy Market Regulatory Board dated 13/05/2016 (https://seffaflik.epias.com.tr/transparency/)
[2] https://toolkit.solcast.com.au/legacy-live-forecast

The types of meteorological data have been listed below separately.

- **Air Temperature:** the temperature of the air two meters above ground.
- **Solar Azimuth Angle:** the angular distance between the sun's horizontal rays and true north, with positive values in the west and negative values in the east which varies between -180 and 180. When the value is -90, the sun is in the east, 0 is in the north, and 90 is in the west.
- **Cloud Opacity:** the cloud-induced attenuation of incoming sunlight ranges between 0% (no clouds) and 100% (full attenuation of incoming sunlight).
- **Dewpoint Temperature**: the temperature at 2 meters above sea level where dewpoint is present.
- **Diffuse Horizontal Irradiance (DHI):** a horizontal surface's exposure to diffuse radiation which is called diffuse sky radiation as well. Irradiance that is scattered by the atmosphere makes up the diffuse component.
- **Direct Normal Irradiance (DNI):** Irradiance coming from the sun's direction (10th percentile clearness). Additionally known as beam radiation.
- **Clear-sky Global Horizontal Irradiance (GHI):** the diffuse illumination that a horizontal surface receives (if there are no clouds). It is known as diffuse sky radiation. In the case of a clear sky, the diffuse component is irradiance that is scattered by the atmosphere (i.e., no water or ice clouds in the sky).
- **Precipitable Water:** water that falls as rain across the entire air column.
- **Relative Humidity:** relative humidity at a height of two meters. The percentage of water vapor needed for saturation at the same temperature that makes up relative humidity. 50% indicates that the air is 50% saturated.
- **Snow Depth Water Equivalent: the liquid-water equivalent of snow depth.**
- **Surface Pressure:** the air pressure at the ground level.
- **Wind Direction (10m):** Wind direction at 10 meters above the ground. Zero represents true north. It varies from 0 and 360. When the value is 270, the wind's direction is west.
- **Wind Speed (10m):** the speed of the wind at 10m above the ground level.

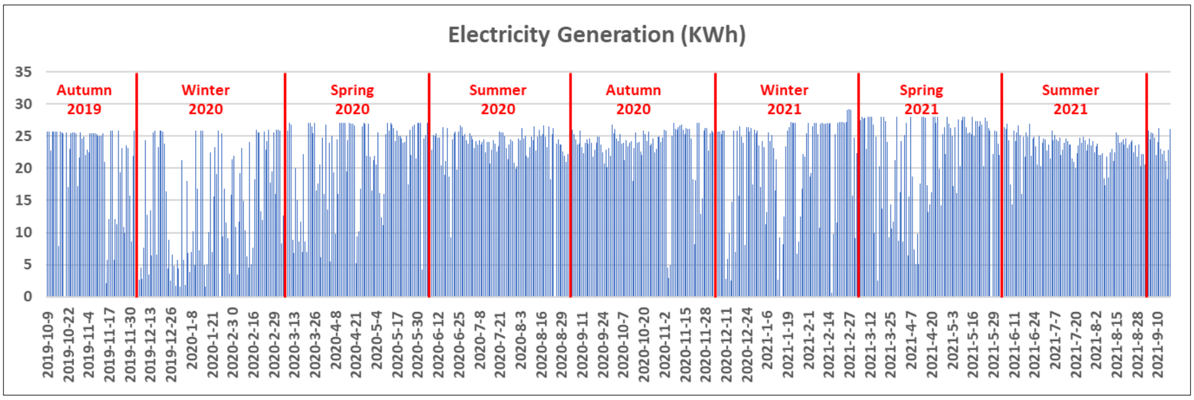
- **Solar Zenith Angle:** the angle formed by the zenith and the sun's direction (directly overhead). The zenith angle is 90 degrees at sunrise and sunset and 0 degrees when the sun is directly overhead.
- **Albedo Daily:** Visible light's average surface reflectance during the day, represented as a proportion between 0 and 1. The number 0 stands for total absorption. The number 1 stands for total reflection. This figure is an interpolated daily average and does not account for reflectivity's diurnal angular dependence.[74]

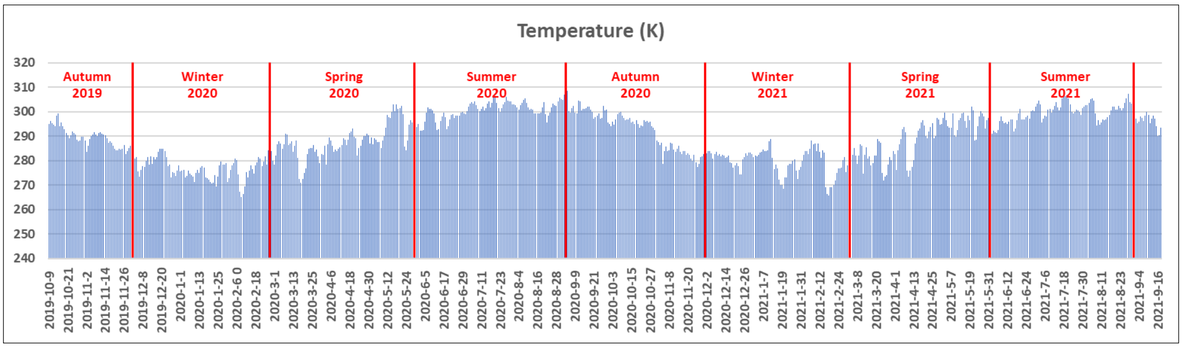
As previously noted, the literature on forecasting was scrutinized to identify the parameters that were most important and commonly employed. Subsequently, the meteorological parameters that had been previously listed were refined accordingly. In summary, in the forecasting analysis in this thesis, besides retroactive electricity generation data, the variables used are cloud opacity, precipitation, direct normal irradiance, temperature and wind speed.

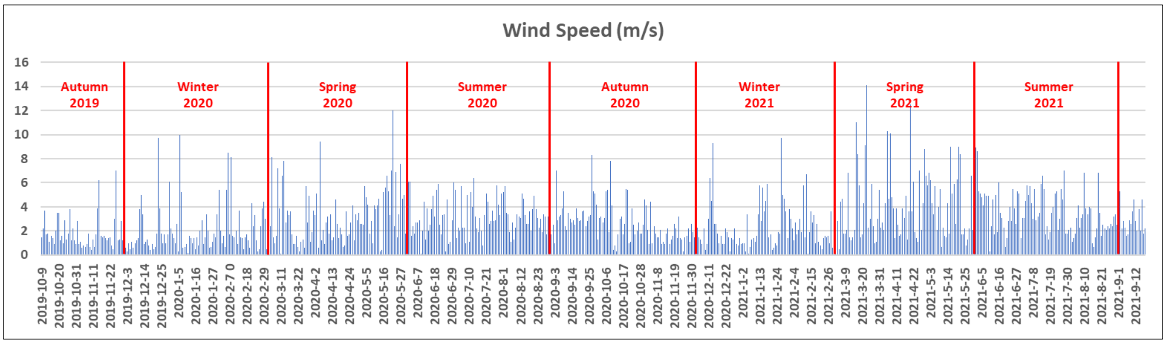
4.2.3 Examination of Input Variables Used in the Analysis

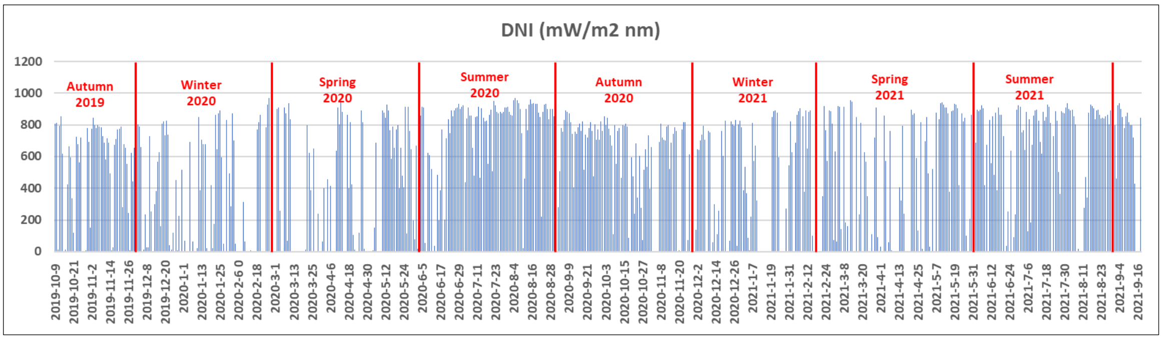
In this analysis, both for the retroactive electricity generations and the meteorological data, the data span covers the period between 9th October 2019 and 18th September 2021. Distribution of all input variables that were used in the analysis are illustrated in the given graphs below. The data used in this study is hourly and covers a period of two years. To clearly illustrate the distribution of the data, representative meteorological and past PV generation input variables were filtered for each day of the two-year data, specifically for daytime (12:00) and night-time (00:00) periods.

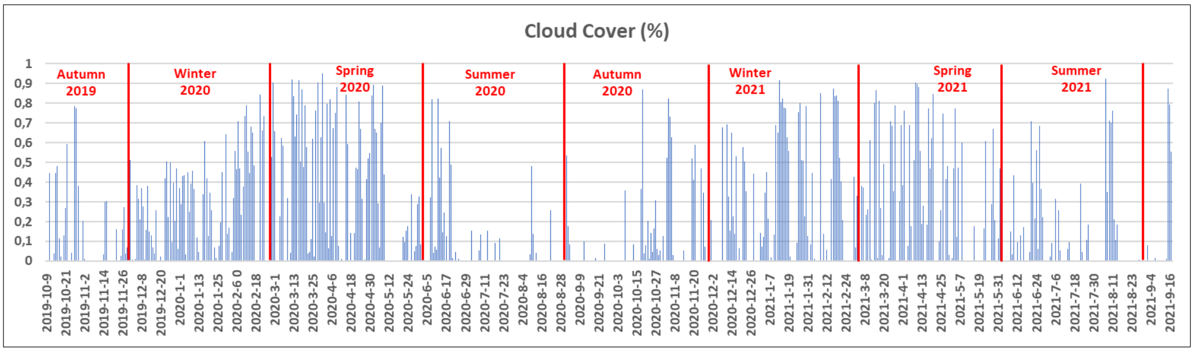
a) Day-time input variables (12:00)

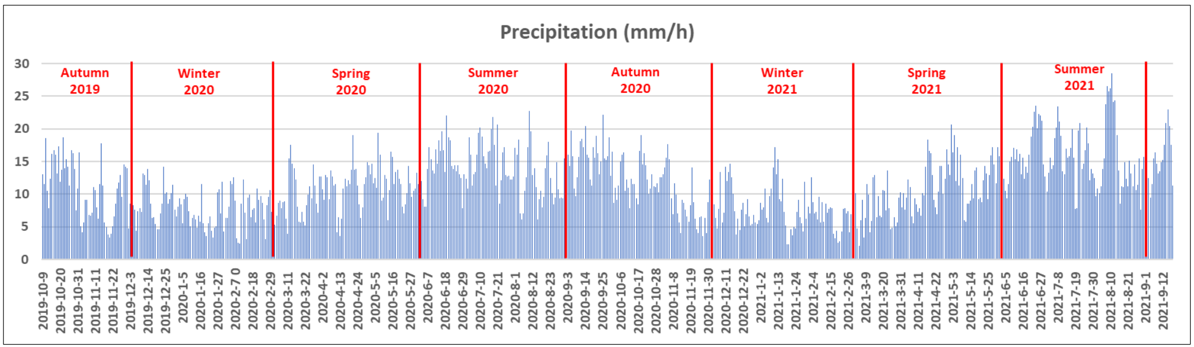




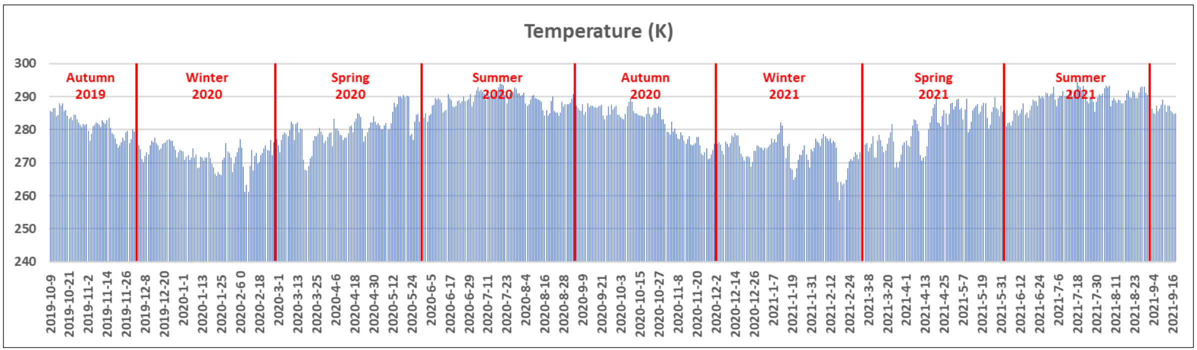


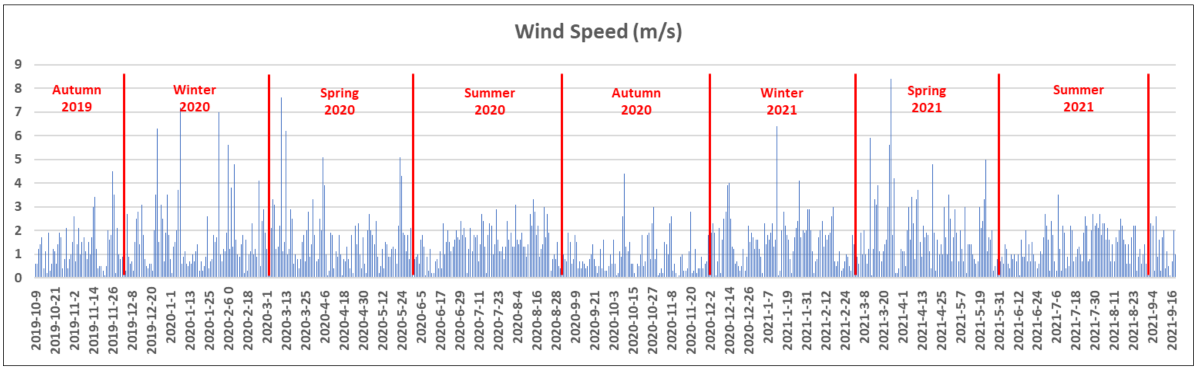


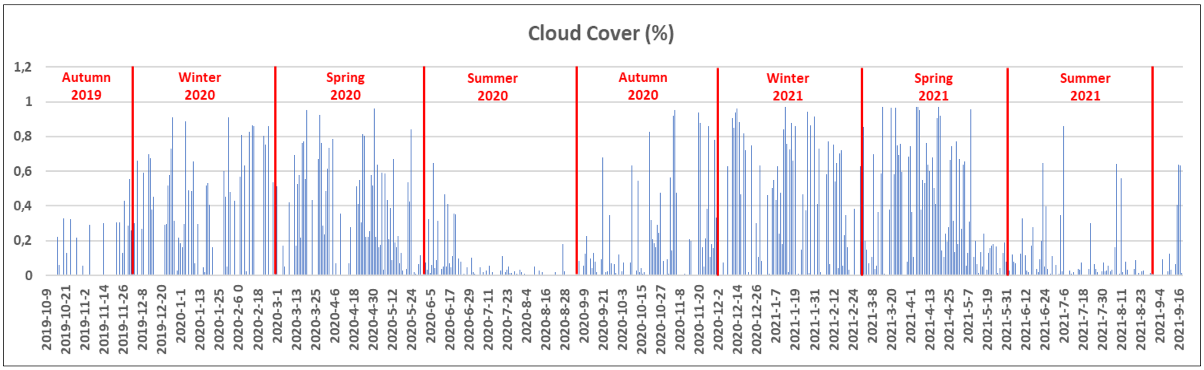


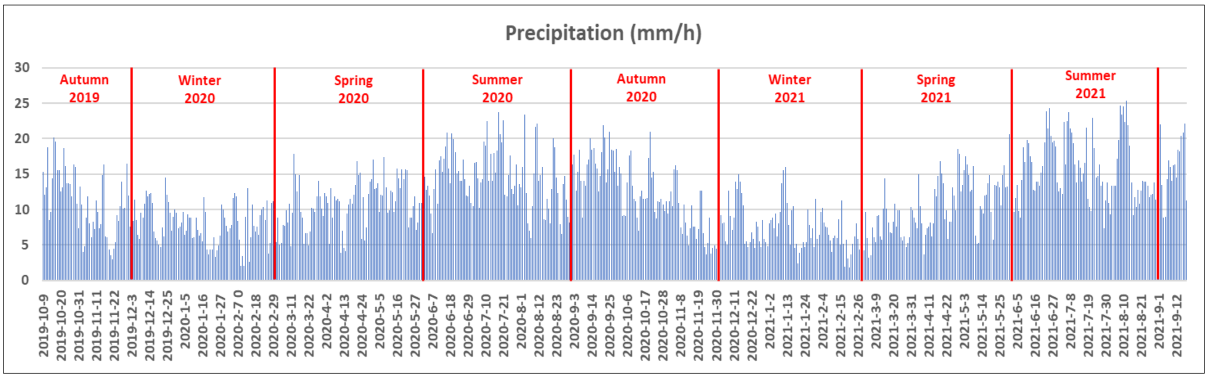


b) Night-time input variables (00:00)









As there is no electricity generation or DNI data during the night-time, graphs for these variables are not included. The graphs that show the distribution of input variables demonstrate that the selected location has a high potential for DNI since it receives consistent irradiation throughout all seasons. Additionally, the distribution of other variables such as temperature, cloud cover, precipitation, and wind speed are consistent with seasonal norms.

4.2.4 Applied Forecast Method

The steps followed before starting the forecasting analysis are summarized below:

- Defining the Problem and Determining the Data Requirement
- Data Discovery
- Data Processing, Cleaning
- Model Selection
- Standardization
- Creating the Training and Test Set
- Parameter Selections
- Visualization of Output and Metrics

This thesis predicts the next day's generation quantities of Solar Power Plants in hourly frequency for the day-ahead local electricity market. Therefore, while the dependent variable is the generation data, the independent variables affecting the generation of solar power plants are weather events and breakdown-maintenance conditions. This is a regression problem as the dependent variable is the generated electricity values. The difference between estimated generation and actual generation is expected to be minimized. To measure this, an error metric such as mean absolute error, root mean squared error or normalized root mean squared error, which are used in the literature to solve the optimization problem, can be used. Generally, the closer the error metrics mentioned are to zero, the more successful the analysis will be.

In this forecast analysis Python was used since it enables to use of machine learning thanks to its libraries. Python is a general-purpose programming language that mainly supports object-oriented programming and functional programming to a certain extent. The procedure of the forecasting started by opening necessary libraries for machine learning in Python. Afterwards, historical generation data and meteorological data was introduced to the software. Then the data was checked for missing values and the distribution of the areas in the data was examined. For the details of the weather data and generation data, firstly, the processes of obtaining the number of rows and columns and obtaining the statistical outputs were applied. Results of these processes are given in Table 2 and Table 3 separately.

Table 2. Statistical Outcomes of Generation Data

For Generation Data	
Number of Rows &Columns of the Input Generation Data	17064 & 2
Mean	6.656
Std	9.298
Min	0
25%	0
50%	0.07
75%	13.45
Max	30

Table 3. Statistical Outcomes of Weather Data

For Weather Data					
	Meteorological Parameters that are used in the Model				
	cloud_cover	precipitation	Irradiation	temperature	wind_speed
Number of Rows & Columns of the Input Weather Data	17064 & 6	17064 & 6	17064 & 6	17064 & 6	17064 & 6
Mean	0.2	11.4	228.1	284.9	2.0
Std	0.3	4.8	340.7	9.2	1.6
Min	0.0	1.6	0.0	257.7	0.0
25%	0.0	7.6	0.0	277.6	0.9
50%	0.1	11.0	0.0	285.0	1.6
75%	0.4	14.6	498.0	292.0	2.6
Max	1.0	29.0	1038.0	308.9	14.1

In the next step, the data was processed and cleaned by removing duplicate values and filling in any missing values. The drop duplicates method was used to remove any duplicated values, and all data intended as input for the model was checked for completeness. Although there were no null values in the dataset, some rows may have been expected but had no records due to the dataset's structure. The appropriate number of rows and columns were determined to ensure that the model period matched the total number of hours between the start and end dates. The resulting number of rows and columns matched the time period of the model, indicating that the input data contained information for all hours.

After cleaning and processing, the categorical data was transformed into numerical data to make it suitable for the model. This process was applied to all generation and weather data used in the model, and calendar features were added to the dataset. The correlation function was used to examine the linear relationship between variables, and the plot of the results of this analysis is presented in Appendix-1.

The accuracy of the model was improved in the following stage by applying a delay to the radiation data using the lag creator function. It is necessary to create "lagged" copies of the time series in order to investigate any potential serial dependencies (such as cycles) in the data. Lag means to advance the times in an index by one or more steps, or to advance the times in the values of a time series by one or more stages. In both scenarios, the observations from lagged series will seem to have occurred at a later time. By introducing a delay to a time series, past values can be made to seem current with the values that are being predicted (in the same row, in other words). So, when describing serial dependence, lag series might be a useful characteristic.[75]

Then, by using scikit-learn library and importing minmax scaler into the script, normalization process was applied to reduce the model's sensitivity to the scale of the features by bringing each feature between 0 and 1.

To measure the performance of a specific component, data was divided by date rather than using the train-test split function when creating the training and test sets. Since the primary objective of this thesis is to highlight the significance of forecasting the market price for the following day from the viewpoint of a market maker, and subsequently, taking a position in the market. Thus, to achieve this, the final day of dataset was used for testing, and the rest of the data was utilized as training dataset. In other words, the data from 9 October 2019 to 17 September 2021 was set aside as the training set and the data from 18 September 2021 was designated as the test set as a result. However, even if the analysis is performed for just one day, to demonstrate that the approach used is robust, the model will be run again by changing the training and test datasets at regular intervals. In addition, to observe the seasonality effect of the model, a daily forecasting modelling study will be conducted for each month within the last year’s data.

In the follow-up, the error metric was decided upon. As the solar production values during the night are 0, the mean absolute percentage inaccuracy is not a reliable metric. In the literature, normalized root mean square error (NRMSE) is most frequently employed.[5]. However, NRMSE is explicitly defined in the model as an external error metric because this is not a part of the Scikitlearn[3] library. On the other hand, besides NRMSE, Root mean square error (RMSE) and R^2 metrics were also determined to examine the error of the analysis. [5] Mathematical formulas of RMSE and NRMSE are given in Equation (1).

[3] Scikit-learn is one of the most widely used Python packages for data science and machine learning.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(x_i^m - x_i^s\right)^2}{\mathrm{n}}}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{x_{max}^m - x_{min}^m}$$

(1)

where max and min reflect the variable's maximum and minimum values for the time period under consideration, and m and s stand for "measured" and "simulated," respectively.[76]

Consequently, for NRMSE, a number between 0 and 1 is generated, with numbers closer to 0 denoting models that fit data better.[77]. For RMSE, better results are achieved with a lower RMSE. Lastly, if the R-squared value is close to 1, it indicates that the regression model is a good one as it explains a significant portion of the variation in the response variable values.

Random Forest and Hyperparameter Selection

It is possible to specify every possible configuration combination to test. GridSearchCV, a method that examines all defined combinations, can be used to do this. This tool is applied to improve the accuracy of Random Forest model[78]. Random Forest algorithm is a type of machine learning model that randomly produces various models and creates classification by training each decision tree on a different observation sample over multiple decision trees.[79] At this part of the model, by using GridSearch tool, the parameter set that gives the highest accuracy was found.

The random forest model was first tested without the GridSearchCV enhancement, and then it was tested again using the same test set but with the parameters that produced the smallest error. Without the GridSearchCV enhancement, the Random Forest model's NRMSE value was determined to be

0.244; with the upgrade, it was found to be 0.144. GridSearchCV with enhancement produced more accurate results because its NRMSE value is closer to zero.

Modelling with Lightgbm (LGBM)

In 2017, as a component of the Microsoft DMTK project focused on distributed machine learning, the boosting algorithm called LightGBM was developed. [80] In contrast to other boosting algorithms, LightGBM offers several benefits including speedy processing, efficient handling of large datasets, low usage of system resources such as RAM, accurate prediction capabilities, support for parallel learning, and compatibility with GPU (Graphics Processing Unit) learning. According to the article "LightGBM: A Highly Efficient Gradient Boosting Decision Tree" in which the model was introduced, compared to other models, LightGBM operates at a speed that is 20 times faster.[80] For this reason, the NRMSE value was recorded as 0.044 with the use of the LGBM model. When compared to GridSearchCV models, the LGBM model gave the most accurate result because NRMSE value is the one that is closest to zero.

4.3 Results of the Analysis

Among the applied models, the Light GBM model gave the best performance. After LGBM, the algorithm with high performance has been obtained by the random forest with improvement of GridSearchCV. Table 4 summarizes the error metric values of the applied models for the forecasting analysis.

Table 4. Error metric values of applied Models

Error Metrics	NRMSE	RMSE	R^2
Random Forest without improvement of GridSearch	0.2436	6.5409	0.6242
Random Forest with improvement of GridSearch	0.1437	3.7479	0.8766
Light GBM	0.0435	1.1358	0.9887

Figure 11 is displaying the correlation between the estimated and real value based on the LGBM model's findings. The test set only covers the date of September 18, 2021. Below result represents what has been forecasted and what was actually achieved on 18.09.2021. The x-axis indicates production in MWh, while the y-axis indicates time in hours.

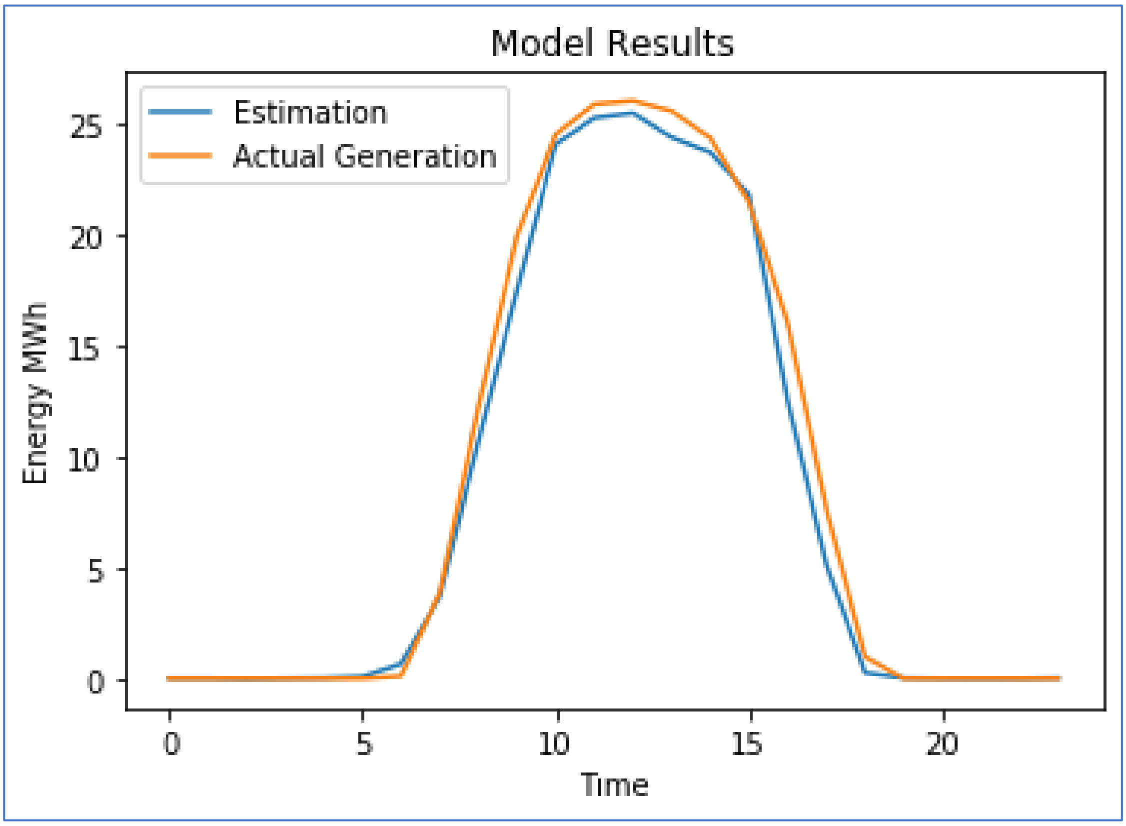


Figure 11. Model Results (LGBM)

4.3.1 Evaluation of the Model

For the accuracy of the applied and adopted model (LGBM), the same analysis was also carried out by using different training and test periods among input variables. First of all, 25% of the input variables was used as a training set and the remaining data was evaluated as test set. In other words, the data from 9 October 2019 to 3 April 2020 was set aside as the training set and the data from 4 April 2020 to 18 September 2021 was designated as the test set as a result. According to this arrangement, achieved result is given in Figure 12. The x-axis indicates production in MWh, while the y-axis indicates time in hours. As can be seen from the graph, there is a significant overlap between the estimation and actual generation values.

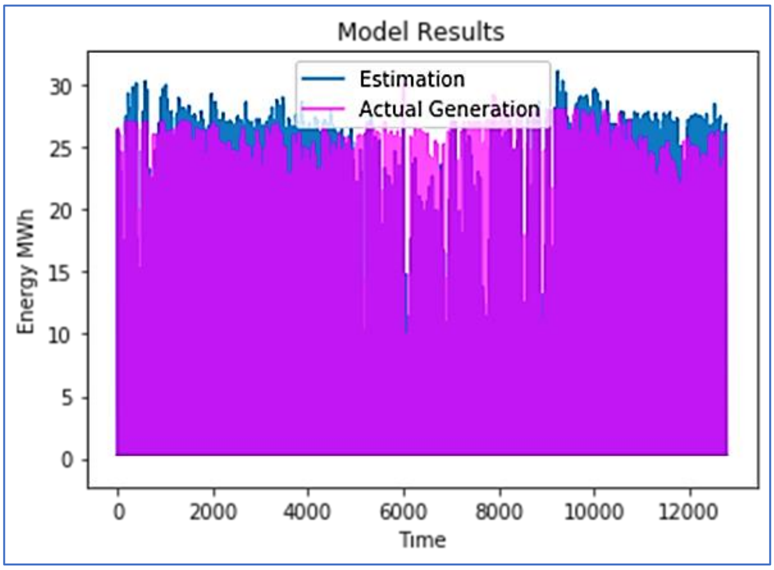


Figure 12. Model Results (LGBM) (25% training data set & 75% test data set)

Secondly, the same model was run again by using 50% of the data as a training set, and 50% of the data as test set. In this scenario, the data from 9 October 2019 to 28 September 2020 was set aside as the training set and the data from 29 September 2020 to 18 September 2021 was designated as the test set as a result. It can be seen that in this scenario, model is able to see varying of the input variables throughout a year including all seasons. Result of this scenario is represented with Figure 13 below. The graph illustrates that obtained result for this scenario gave more successful results compared to scenario 1.

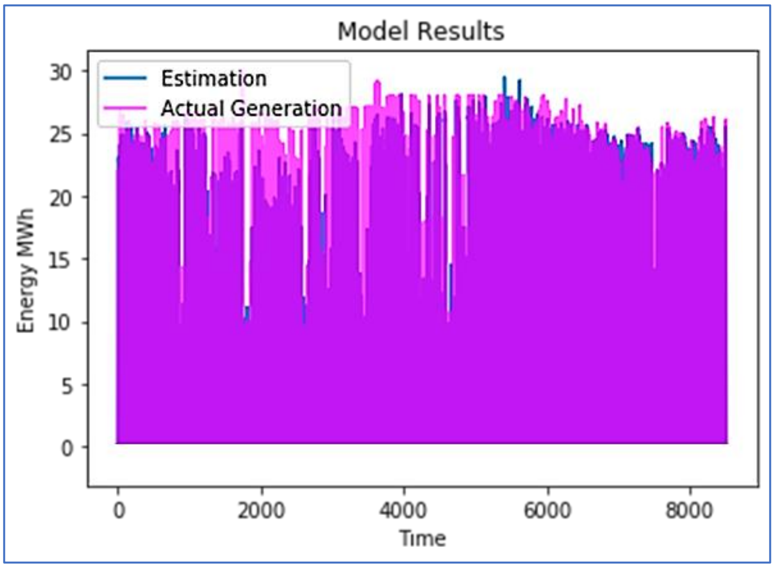


Figure 13. Model Results (LGBM) (50% training data set & 50% test data set)

Lastly, the model was run again by changing the periods of training and test sets as 75% training data input, 25% test data input. In this one, the data from 9 October 2019 to 25 March 2021 was set aside as the training set and the data from 26 March 2021 to 18 September 2021 was designated as the test set. This data set arrangement was resulted like in the graph which is provided in Figure 14. Among the three scenarios, the most successful model results have been achieved with this scenario.

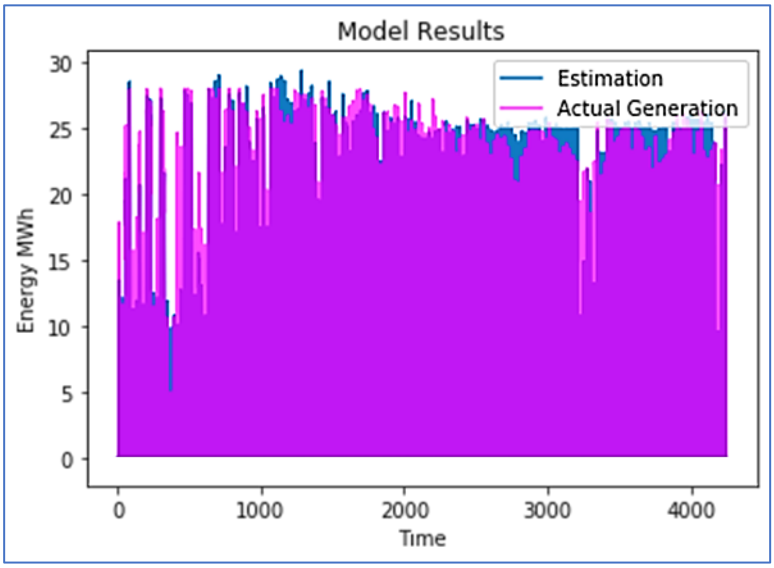


Figure 14. Model Results (LGBM) (75% training data set & 25% test data set)

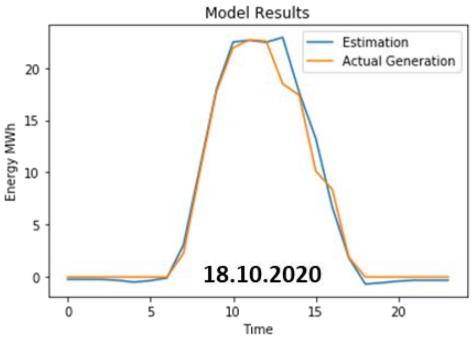
According to different scenarios adopted in the forecasting model above, model performances were checked accordingly. Thus NRMSE, RMSE and R^2 error metrics were calculated for each scenario separately. Outcomes of the analysis are provided in Table 5.

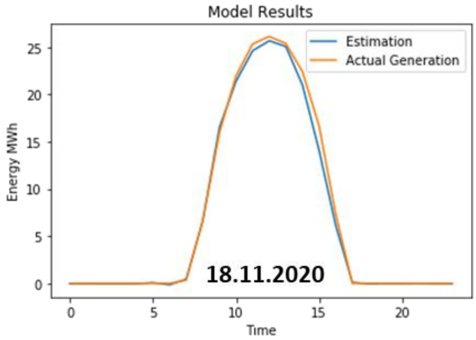
Table 5. Error metric values of LGBM for different training-test periods

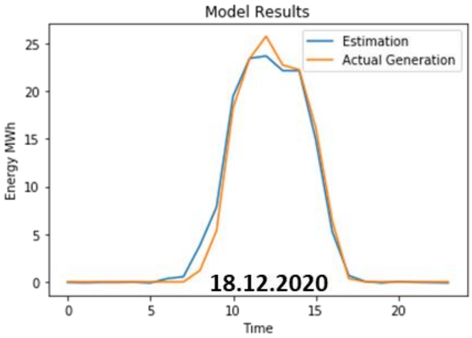
	Error Metrics		
Training-Test Percentages	**NRMSE**	**RMSE**	**R^2**
25%-75%	0.1000	3.0023	0.9024
50%-50%	0.0783	2.3487	0,9385
75%-25%	0.0702	1.9691	0.9589

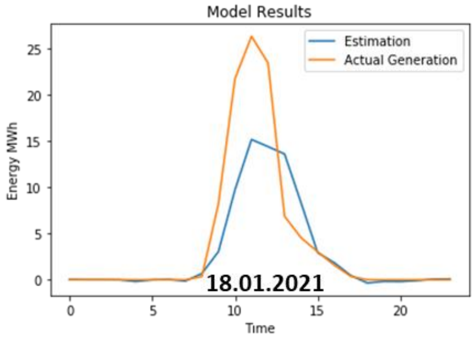
The results from analysing the model's errors indicate that there is little difference between using a 50% or 75% training set. The model covers most of the potential situations it may encounter within a year, and this is used as the training data set, accounting for 50% of all data. Therefore, using one year's worth of data as training is sufficient, as the 75% scenario did not differ significantly from the 50% scenario, meaning extreme values are not excluded from the training set. The objective of the analysis was to obtain electricity generation data for a specific day (September 18, 2021), and almost all of the data set was utilized as training set, which led to a higher accuracy in the results obtained.

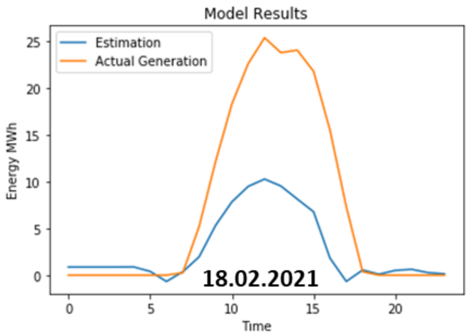
To examine the seasonal validity of the applied and adopted model (LGBM), the same forecasting analysis was run by testing a chosen single day from each month within last 1 year of the data set. To be able to be compatible with targeted estimation day for the analysis, which is 18 September 2021, from 18 October 2020 to 18 September 2021, the same model was applied for the same day of each month within the last year of the dataset. In other words, to investigate the impact of seasonality on the model, a single day from each month of the preceding year was selected and utilized as the test dataset. In this analysis, the dataset preceding the selected test date was used as the training dataset. Obtained results for each day of these months are given in Figure 18 separately.

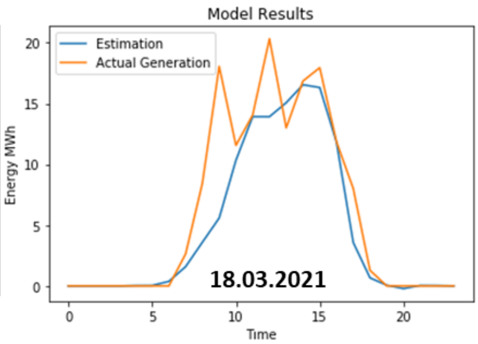

Model Results
Estimation
Actual Generation
Energy MWh
Time
18.10.2020

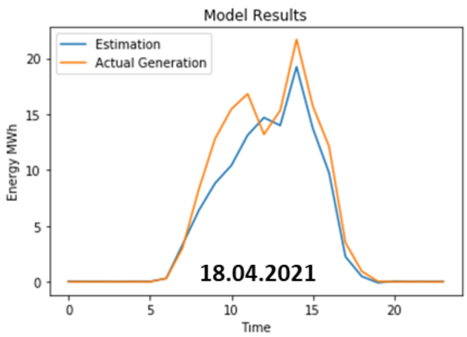

Model Results
Estimation
Actual Generation
Energy MWh
Time
18.11.2020

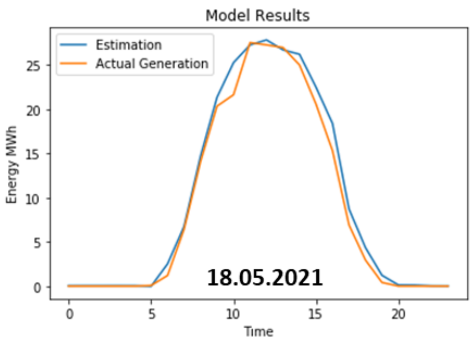

Model Results
Estimation
Actual Generation
Energy MWh
Time
18.12.2020

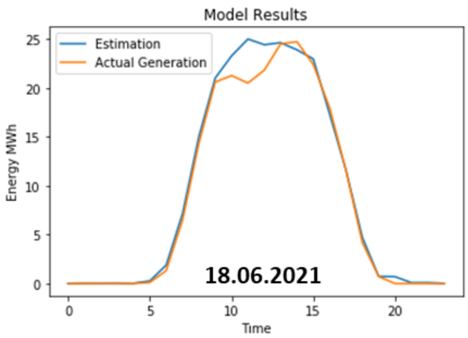

Model Results
Estimation
Actual Generation
Energy MWh
Time
18.01.2021

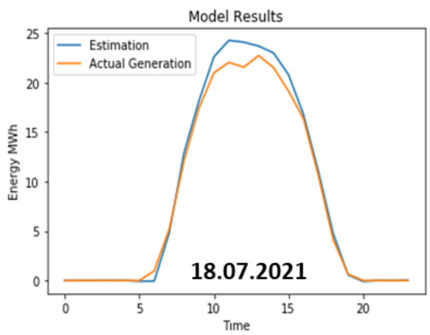

Model Results
Estimation
Actual Generation
Energy MWh
Time
18.02.2021


Model Results
Estimation
Actual Generation
Energy MWh
Time
18.03.2021


Model Results
Estimation
Actual Generation
Energy MWh
Time
18.04.2021


Model Results
Estimation
Actual Generation
Energy MWh
Time
18.05.2021


Model Results
Estimation
Actual Generation
Energy MWh
Time
18.06.2021


Model Results
Estimation
Actual Generation
Energy MWh
Time
18.07.2021

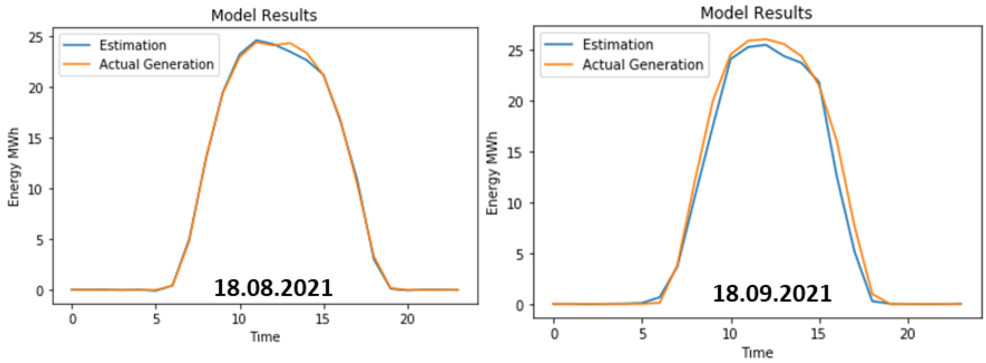


Figure 15. Model Results (LGBM) for single day from 18 October 2020 to 18 September 2021

The performance of the forecasting model was evaluated based on several selected scenarios. As a result, NRMSE, RMSE, and R^2 error measures were determined individually for each scenario. Table 6 provides the analysis' results.

Table 6. Error metric values of LGBM for single day generation estimation from 18 October 2020 to 18 September 2021

Date	**LGBM Error Metrics**		
	NRMSE	**RMSE**	**R^2**
18.10.2020	0.0535	1.21240	0.9800
18.11.2020	0.0266	0.6965	0.9952
18.12.2020	0.0378	0.9752	0.9889
18.01.2021	0.1622	4.26070	0.7033
18.02.2021	0.3124	7.90080	0.3389
18.03.2021	0.1589	3.22970	0.8023
18.04.2021	0.0842	1.82650	0.9356
18.05.2021	0.0453	1.24710	0.9863
18.06.2021	0.0487	1.20470	0.9851
18.07.2021	0.0431	0.9782	0.9886
18.08.2021	0.0110	0.26933	0.9993
18.09.2021	0.0435	1.13580	0.9890

As can be interpreted from Table 6, efficient, and similar results were obtained from the model except for 1-2 days. As an example, when examining the meteorological data for February 18th, 2021, which was the day with the least efficient model result, it was observed that the irradiation values were much lower compared to other days. It can be said that the main reason for the model's inefficient result for this day is the low level of irradiation, which resulted in a low production amount. The seasonal effect analysis began with the training set comprising 50% of the entire model set and gradually increased to 99.9%. This suggests that to achieve accurate results, it is adequate to use a data set that covers all seasonal events within a one-year period.

In conclusion, the model has gained more explanatory power as a result of repeating the analysis using different training and test datasets. As a result, it has helped us to obtain noteworthy outcomes that show the model's reliability.

4.3.2 Comparison of the Model with other Forecasting Studies in the Literature

To be able to evaluate the results of the applied and approved model, error measurements of the forecasts were also compared with the similar forecasting studies that have been carried out in the literature. Input variables, horizon of the forecasts and the data period of the studies have been considered to evaluate the forecast analysis of this study. Within this scope, different studies from the literature have been analysed and detailed in Table 7.

Table 7. Solar Power Forecasting Studies in the Literature

Number	Author and year	Location	Method of Forecast	Horizon of Forecast	Error measurement for Forecast	Input Variables	Forecast Variable	Data Period
1	Kardakos et al. (2013)[81]	Greece	ANN	Day ahead	NRMSE = 11.26-11.42%	Solar radiation	PV power	01.01.2011-31.12.2012
2	Mellit et al. (2013)[82]	Türkiye	ANN	Hourly	RMSE = less than 0.2%	Solar Irradiance, Temperature of Air	PV power	01.01.2011-24.02.2012
3	Almonacid et al. (2014)[83]	Spain	Dynamic ANN	1-h ahead	R^2= close to 1 & RMSE =3.38%	Solar Irradiance, Temperature of Air, Past PV power generations	PV power	2011-2012
4	Giorgi et al. (2013)[84]	Italy	ANN	1, 3, 6, 12, 24-h ahead	NRMSE= 1h (10.91%), 3h (15.61%), 6h (18.89%), 12h (18.80%), 24h (23.99%)	Past PV Power, Temperature of the Module, Air Temperature, Solar Irradiance	PV power	05.03.2012-5.03.2013
5	Teo et al. (2015)[69], [85]	N/A	ANN	N/A	RMSE = 3.8574%	Air Temperature, Module Temperature, Total Daily Energy, Solar Irradiance, PV power	PV power	08.06.2014-6.07.2014
6	Zhang et al. (2015)[69]	USA, Denmark, Italy	kNN, WkNN	Day ahead	NRMSE= USA (9.82%, WkNN), Denmark (8.72%, kNN), Italy (10.37% WkNN)	USA = Air Temperature, cell temperature, and wind speed Denmark = cloud cover, GHI, fog, wind speed, temperature of ambient Italy = Altitude of Solar, GHI, cloud cover, DNI, Temperature of ambient	PV power	USA = 01.07.2010-31.12.2011, Denmark =01.01.2006-31.12.2006, Italy= 01.01.2011-31.12.2011
7	Graditi et al. (2016)[86]	Italy	Sandia, Regression, ANN	N/A	Sandia model: NRMSE = 10.09% R^2 = 0.978 Regression model: NRMSE= 7.01 R^2 = 0.980 ANN model: NRMSE= 6.66% R^2= 0.982	Global Radiation, Air Temperature, Wind Speed, Module Temperature	PV power	2006-2012

8	Hossain et al. (2017)[72]	Malaysia	Extreme Learning Machine	Day ahead	RMSE =17.89–35.39%	Solar Irradiance, Air Temperature, Temperature of the Module, Wind Speed, PV Generations	PV power	01.01.2015-30.09.2016
9	Shi et al. (2012)[87]	China	Support Vector Machine	Day ahead	RMSE= 2.10	Historical Generations, Next day's weather report	PV power	13.01.2010-29.10.2010
10	Wang et al. (2016)[88]	Coloane island of Macau	Regularized PFLRM	Day ahead	RMSE= 59.1704	FPC score, wind speed, intercept, pressure, air temperature, insulation, humidity	PV power	01.01.2011-30.06.2012
11	Li et al. (2016)[89], [90]	Coloane island of Macau	MARS	Day ahead	RMSE=119.0	Air Temperature, wind speed, dew temperature, air pressure, precipitation, insolation duration, humidity	PV power	01.01.2011-20.06.2016
12	Massidda et al. (2017)[90]	Germany	MARS	Day ahead	RMSE= 177.8	NWP data, historical PV generations	PV power	2014
13	Fernandez et al. (2012)[91]	Spain	ANN-MLP	1-39h ahead	Average RMSE= 11.79%	Historical PV generations, Weather parameters	PV power	02.06.2007-27.05.2008
14	Yang et al. (2014)[92]	Taiwan	SVR	Day ahead	RMSE= 350.2	Temperature, precipitation probability	PV power	01.05.2012-30.04.2013
15	Dolara et al. (2015)[93]	Italy	PHANN	Day ahead	NRMSE= 20.6% (60 days training), 15.0% (90 days training), 13.4% (120 days training)	Weather forecasts, clear sky model	PV power	240 days (from 1st of Jan 2012)
16	Antonanzas et al. (2017)[94]	Spain	Blended Model (SVR & DNN & XGB & RF)	Day ahead	NRMSE = 22.49%	GHI, ambient temperature, humidity, speed of wind, position of the sun, extra-terrestrial irradiance	PV power	2009-2010

As it was described earlier, in the forecasting model conducted in this thesis study, 5 different meteorological data (cloud cover, precipitation, wind speed, air temperature, irradiation) and past electricity generations for approximately 2 years were used as input variables and consequently, the best results were obtained with LGBM. Accordingly, the NRMSE, RMSE and R^2 error values were calculated as 4.35%, 1.14, and 0.9887 respectively.

For example, in study number 1(Kardakos et al., 2013), it is observed that the length of the data period used is similar to the model studied in this thesis, which is 2 years. However, while 5 different meteorological data and past production values are used as input variables in the model in this thesis, it is seen that only solar radiation is used as the input variable in study number 1. The NRMSE value obtained in this study using the ANN model varies between 11.26% and 11.42%. This value is approximately 7% higher than the model used in this thesis, which was conducted using the machine learning method. On the other hand, in study number 4 (Giorgi et al.,2015), as similar to this thesis' model, solar irradiance, air temperature, module temperature and past electricity generations were used as input variables. However, length of the data period for these variables was taken for 1 year only. As a result, while NRMSE value has been calculated as 4.35% in the model of this thesis, it has been calculated as 23.99% for day ahead forecasting by using ANN method in study number 4. As can also be understood from the previous studies conducted in the literature, the success rate of the models increases as the number and period of input variables used in the forecasting models increase, and the horizon of forecast becomes shorter.

Taking all of this into consideration and with the comparison of the results obtained and similar studies in the literature given in Table 7 above, it can be said that the success of the forecasting analysis does not depend only on the model used, but also the great variety and big volume of the data used. The more successful results can be obtained by increasing the volume and variety of the used data in Big Data Analysis.

5 DESIGNING LOCAL ELECTRICITY MARKET SIMULATION

The open-source and online software of Grid Singularity[95] has been used for the validation of simulated scalable scenarios and LEMs evaluation from an economic perspective. Grid Singularity is a recognized open-source energy technology business that co-founded the Energy Web Foundation and created the Grid Singularity Exchange to put people and the environment at the center of the energy market (EWF). Grid Singularity models and runs networked, grid-aware energy markets, giving every market player the greatest degree of trading flexibility.[95]

5.1 Types of Spot Markets

Bids and offers are matched in a local energy market (LEM) in accordance with the chosen clearing mechanism. Studies reveal that various clearing procedures have varying advantages and disadvantages in terms of market efficiency, fairness, and user options. On the exchange run by Grid Singularity (formerly D3A), there are three types of spot markets.[95]

One-Sided Pay-as-Offer Market

In the One-Sided Pay-as-Offer market, agents for energy producers, including prosumers (sellers), post offers with energy prices based on the trading strategy for the assets.

Agents acting on behalf of buyers have the ability to review available offers in their local market, eliminate unaffordable options, and select the most suitable offer. The price of the offer is established through an agreed-upon energy rate between the buyer and seller, known as pay-as-offer. As a result, the trading rate may fluctuate for transactions completed within the same timeframe. As the auction is continuous, offers can be accepted at any point, even prior to the conclusion of each market slot once they have been submitted.[95]

Double Sided Pay-as-Bid Market

Buyers have the ability to submit their own offers alongside those of sellers in the Two-Sided Pay-as-Bid market. A Market Agent is created and managed for each market, which gathers and matches bids and offers submitted by trade agents and distributes them to other markets in the region. The Market Agent's role is to transmit bids and offers to be linked marketplaces. As the auction is ongoing, bids and offers can be immediately matched once submitted, even before the market slot expires. Trade agents also have the option to reject bids and offers. The matching process illustrated in Figure 16 demonstrates that the market continually matches bids and offers.[95] As it can be seen from the figure, in this type of market, the consumer bided 0.12 Euros for 4 kWh. The most compatible offer from the producers for this offer is 0.11 Euro for 6 kWh, and consequently a trade was made between these two participants. On the other hand, another consumer bided 0.22 Euros for 8 kWh, and this bid was matched with another producer's offer which is 0.14 Euro for 8 kWh, and the trade was realized between them. Unmatched offers are waiting in the market as unmatched until the next slot.

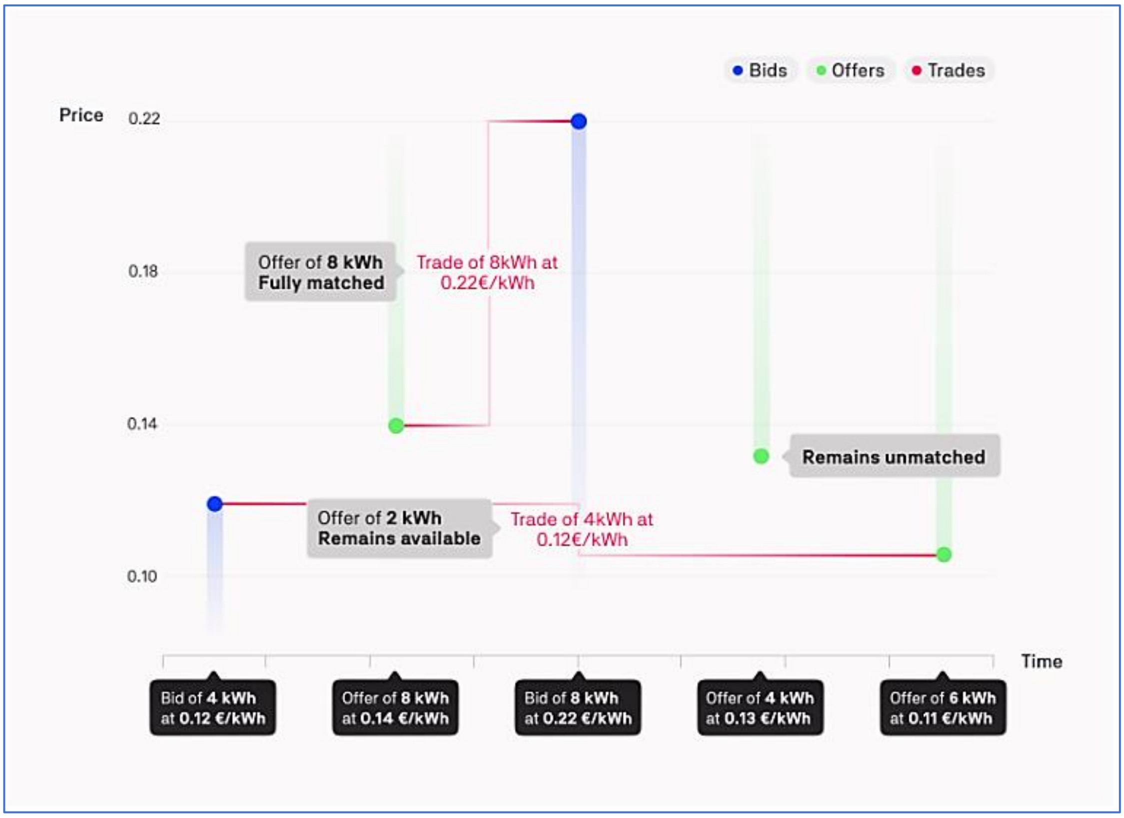


Figure 16. Double Sided Pay-as-Bid Market mechanism.[95]

Double Sided Pay-as-Clear Market

Buyers are able to submit bids alongside sellers' offers in the Two-Sided Pay-as-Clear market. The market receives, matches, and distributes bids and offers made by trading agents to other markets via the Market Agent (MA). Trade agents have the option to reject bids and offers. Market Agents are established and managed for each market (region) to transmit bids and offers to be connected marketplaces.[95]

A merit-order-effect method is currently in place for matching bids and offers, where bids and offers are combined and cleared within a predetermined clearing time. At the conclusion of each interval, offers and bids are organized in ascending and descending order, respectively, and the equilibrium quantity of energy and price is determined. The point at which the agreed-upon bid curve for the buyers falls below the agreed-upon offer curve for the sellers is the clearing point, which determines the amount of energy that is accepted trading volume for a particular energy rate clearing price. The lowest offers are paired with the highest bids. The clearing price (cents/kWh) is the matching rate. The leftover offers and bids, which are shown in the plot below to the right of the clearing point, are not cleared at this clearing point and are instead stay in the market for subsequent matching.[95]

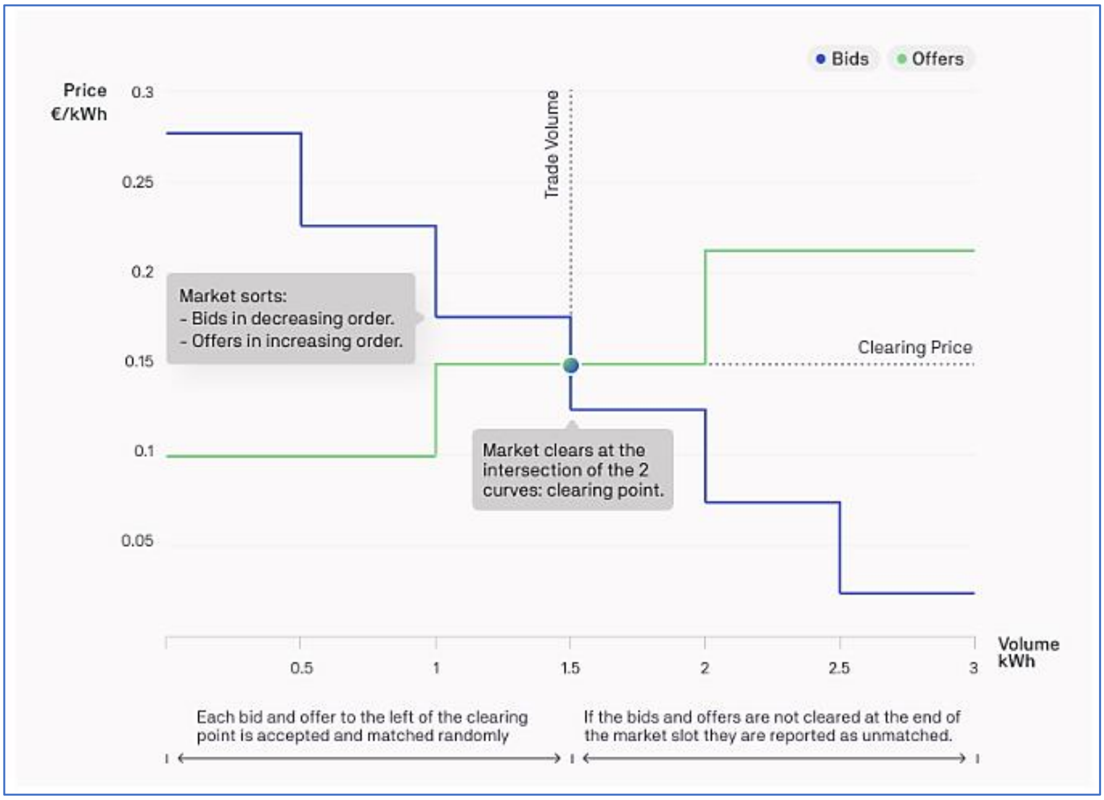


Figure 17. Double Sided Pay-as-Clear Market mechanism[95]

In this thesis, in all simulated scenarios, as a spot market type, “Double Sided Pay-as-Clear Market” has been adopted.

5.2 Market Slots and Market Ticks

Market Slots

The energy spot market's default setting for time slots is 15 simulated minutes, resulting in 96 market slots for a one-day simulation. However, the length can be adjusted during simulations, allowing it to be increased up to 60 minutes.

Bids and offers are matched either within each slot or at the end, depending on the market type. Assets may face penalties for any physical energy they produce or consume that is not exchanged if bids and offers remain open at the conclusion of a market slot.[95]

Market Ticks

Ticks are generated by splitting each slot into smaller segments. In other words, ticks are further division of each slot. The default configuration for a tick is 15 seconds of simulated time, but this setting can be changed. Simulated time is the time unit within a simulation, as opposed to real-time, which is the amount of time that the simulation actually takes. For example, seven days of trade can be simulated in a simulation in a matter of minutes or hours, and the setup can be altered. There are 60 ticks in the standard 15-minute trading period since there are 4 ticks for a 1-minute slot by default.

For instance, in a pay-as-bid market, the market is cleared at the conclusion of each tick. A Market Agent propagates an unmatched order to all nearby marketplaces after two ticks.[95]

In the simulated scenarios, length of the spot market and tick length was taken as 60 minutes and 60 seconds, respectively.

5.3 Description of Scenarios and Simulations

Three different integration scenarios of different PV systems to the grid and market are explored in this study. These scenarios include different community structures, consumers, producers, and different load profiles. Each LEC member is connected to the electricity grid to generate or consume electricity and participate in the local electricity trade.

The study in Chapter 4 produced a forecasted value for the solar power plant for September 18, 2021. Since the main scope of this thesis is to emphasize the importance of predicting the market price for the next day from the perspective of a market maker and then using this information to take a position in the market, generation forecast has been made only for one day as 18.09.2021. The machine learning models used learn through a probabilistic distribution and attempt to predict real values, so there is a small margin of error in all predictions. This situation is valid for both daytime and night-time generation. Therefore, instead of predicting zero for night-time generation, it can be forecast values that are close to zero. Since it is known that night-time generation should be zero, when evaluating the results through scenarios, night-time production was considered negligible and ignored. In a real-world usage scenario, this is likewise the proper course of action. Hence, estimated generations between 0:00–05:00 and 19:00–23:00 have been assumed to be zero and the estimated PV generations have been taken accordingly for use in simulation studies in order to produce more accurate and conservative results in local electricity market simulations. This estimated generation value will be utilized as the generation data for the solar power plant in the simulations conducted as part of the scenarios. Moreover, the date of 18.09.2021 is taken as the basis for the execution date in all scenarios. In other words, the working date of the local electricity market has been taken as 18.09.2021 in simulated scenarios. Additionally, the capital costs of the included solar power plant and batteries have not been taken into account in the scenarios. The scenarios are explained in detail in the next sections.

5.3.1 Scenario 1: Local Community without Solar Power Generation

In this scenario, the market actors' consumption load profiles have first been defined. This scenario assumes that there are 10 families in this community, each with a unique consumption load profile, and that there are 6 different commercials.

Figure 18. Representation of Scenario-1 from GridSingularity Simulation Software

Details of the local market players are given in below.

Households

Household 1 (HH1): Family with 3 children & job

Household 2 (HH2): Young couple with job

Household 3 (HH3): Family with 2 children

Household 4 (HH4): Family with 2 children & job

Household 5 (HH5): Family with 3 children

Household 6 (HH6): Flat-sharing students

Household 7 (HH7): Family with 1 child & job

Household 8 (HH8): Middle-aged couple with job

Household 9 (HH9): Retired Couple

Household 10 (HH10): Single with 1 child

Commercials

- School

- GYM
- Wastewater Treatment Plant (WWTP)
- Hairdresser
- Bakery
- Dining Restaurant

While average daily consumption load profiles of households were obtained from the library of Grid Singularity[95], average daily consumption load profiles of commercials were obtained from the literature. [96] In Scenario-1, there is no renewable energy penetration into the market. Accordingly, all consumers purchased electricity from the grid for their electricity consumption at a constant price. The market price for the electricity used from the grid has been taken as 0.30 EUR/kWh by default in the simulation.

5.3.2 Scenario 2: Local Community with Solar Power Generation

In this scenario, the local market participants behaved similarly to the first scenario, with the exception of the addition of a 26 MWe solar power plant to the local market.

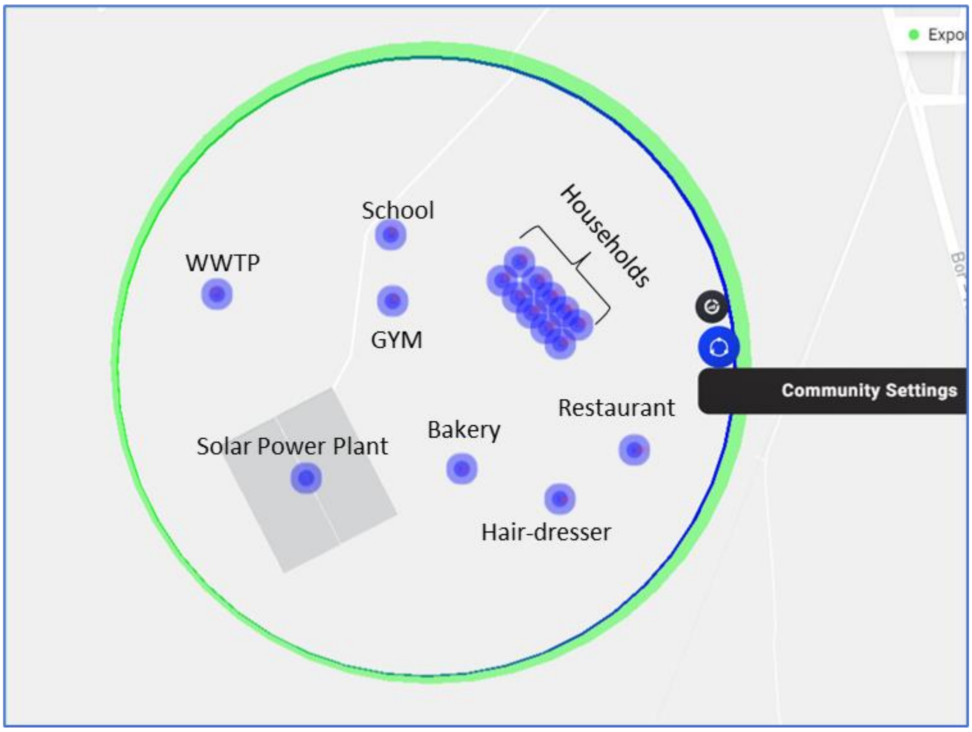


Figure 19. Representation of Scenario-2 from GridSingularity Simulation Software

This scenario features a consumer with an initial buying rate of 0 cents/kWh and a final buying rate of 30 cents/kWh, which increases by 0.51 cents/kWh with each market update. On the other hand, the solar energy supplier's initial selling rate is set at 30 cents/kWh and decreases by 0.51 cents/kWh to reach a final selling rate of 0 cents/kWh. The market price for grid electricity usage remains constant at 0.30 EUR/kWh in this scenario.

5.3.3 Scenario 3: Local Community with Solar Power Generation and Installation of Battery

The situation involves introducing a 10-kWh battery to the system along with the second scenario. The battery is assumed to have an initial capacity or state of charge of 50%. The purpose is to examine how installing batteries affects market prices, particularly in cases where the solar plant isn't generating power and people are buying clean energy from the battery instead of the grid. The investigation seeks to comprehend the impact on market prices. Representation of the scenario is given in Figure 20 below.

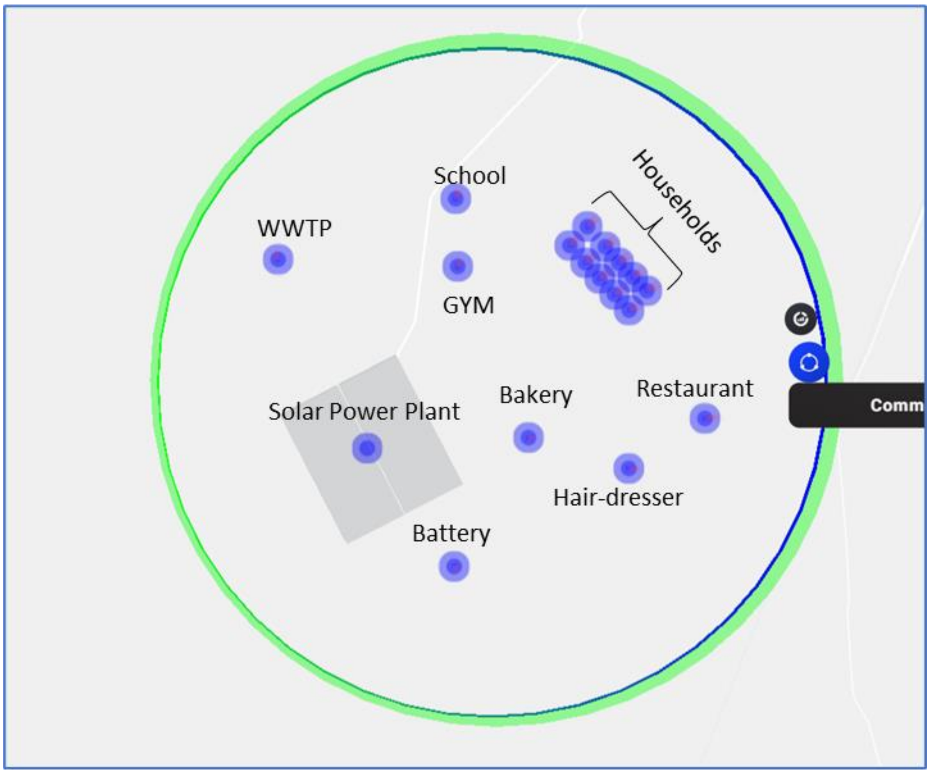


Figure 20. Representation of Scenario-3 from GridSingularity Simulation Software

In this scenario, the initial and final buying-selling rates for consumers and suppliers are the same as in Scenario 2. However, for the battery, an initial selling rate of 27.5 cents/kWh and a final selling rate of 25.1 cents/kWh have been set with a decrease rate of 0.04 cents/kWh per market update. Additionally, again for the battery, an initial buying rate of 0 cents/kWh and a final buying rate of 25 cents/kWh have been established with an increasing rate of 0.25 cents/kWh per market update.

5.4 Results of the Simulations

The 3 scenarios defined above were run separately on GridSingularity, and the simulation results are explained separately for each scenario in the following sections. In all scenarios, grid fee is assumed as zero.

5.4.1 Results of Scenario 1

After scenario 1 was run, the market price remained constant as expected, and all electricity consumers had to purchase electricity from the grid at a fixed price, which is 30 cents/kWh. Therefore, the share of community electricity consumption that is supplied by its own renewable energy assets, which is called self-sufficiency of the community, has been obtained as 0.0%. On the other hand, details of the energy bills and the traded net energy data according to Scenario 1 is provided in Table 8.

Table 8. Energy bills and the traded net energy in the community (Scenario 1)

Asset	Bought		Sold		Total Balance		
	Energy (kWh)	Paid (€)	Energy (kWh)	Revenue (€)	Energy (kWh)	Percentage of Energy Demand in Total Demand	Total (€)
HH1	12.89	3.87	0.00	0.00	12.89	6.98%	**3.87 (Paid)**
HH2	4.92	1.48	0.00	0.00	4.92	2.66%	**1.48 (Paid)**
HH3	19.17	5.75	0.00	0.00	19.17	10.38%	**5.75 (Paid)**
HH4	15.12	4.54	0.00	0.00	15.12	8.18%	**4.54 (Paid)**
HH5	13.99	4.20	0.00	0.00	13.99	7.57%	**4.2 (Paid)**
HH6	7.08	2.12	0.00	0.00	7.08	3.83%	**2.12 (Paid)**
HH7	10.23	3.07	0.00	0.00	10.23	5.54%	**3.07 (Paid)**
HH8	7.23	2.17	0.00	0.00	7.23	3.91%	**2.17 (Paid)**

HH9	9.08	2.72	0.00	0.00	9.08	4.91%	**2.72 (Paid)**
HH10	7.12	2.13	0.00	0.00	7.12	3.85%	**2.13 (Paid)**
GYM	12.84	3.85	0.00	0.00	12.84	6.95%	**3.85 (Paid)**
WWTP	11.68	3.50	0.00	0.00	11.68	6.32%	**3.5 (Paid)**
Bakery	12.48	3.74	0.00	0.00	12.48	6.75%	**3.74 (Paid)**
School	17.43	5.23	0.00	0.00	17.43	9.43%	**5.23 (Paid)**
Restaurant	10.75	3.22	0.00	0.00	10.75	5.82%	**3.22 (Paid)**
Hairdresser	12.75	3.83	0.00	0.00	12.75	6.90%	**3.83 (Paid)**
Grid Market	0.00	0.00	184.76	55.43	184.76	100.00%	**55.43 (Revenue)**
Totals	184.76	55.43	184.76	55.43	0.00	-	**0 (Neutral)**

As it can be understood from Table 8, the community members demanded 184.76 kWh electricity on the simulated day and had to purchase electricity from the grid at a fixed price which is 30 cents/kWh. Trading profile and the variation of the electricity demand of the community during the day is given by Figure 21, Figure 22, Figure 23 separately.

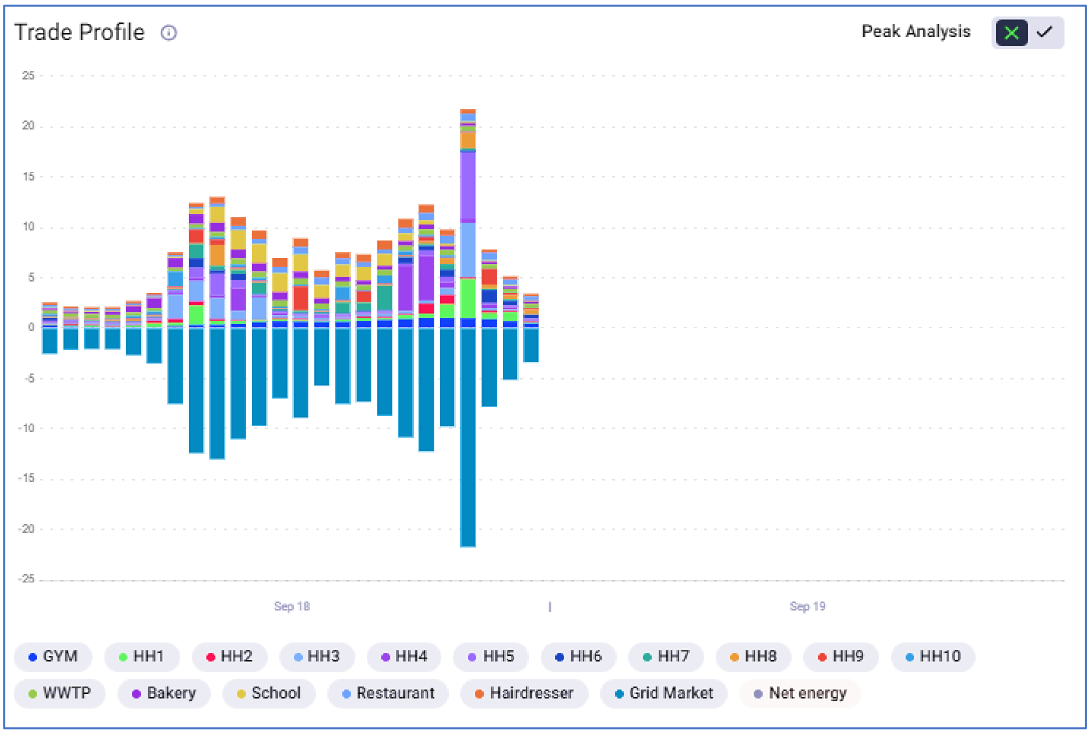


Figure 21. Trade Profile of the community (Scenario 1)

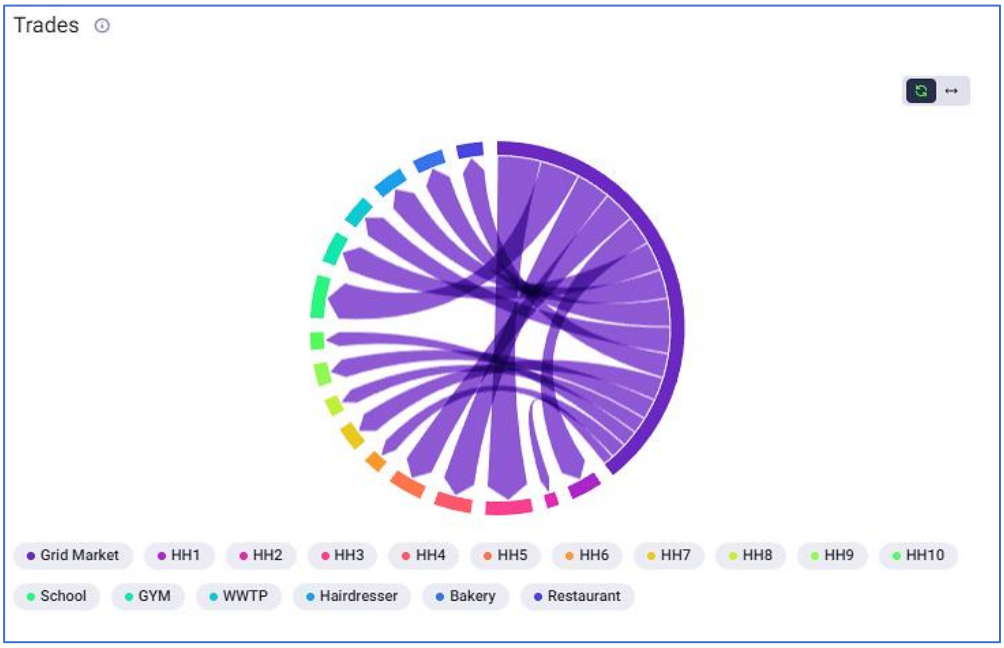


Figure 22. Schematic representation of trading profile (Scenario 1)

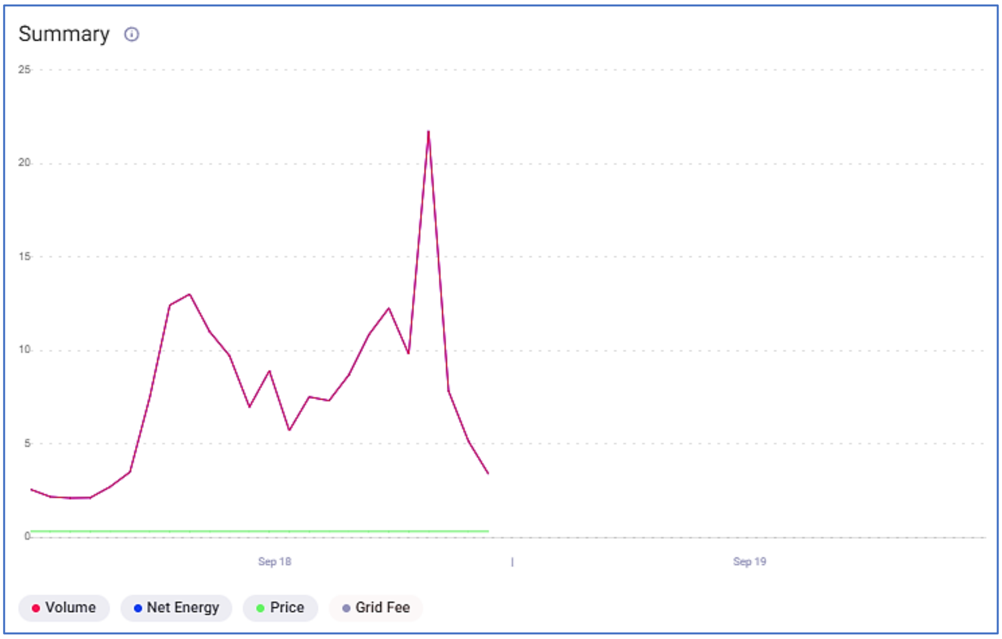


Figure 23. Variation of the electricity demand of the community during the day (Scenario 1)

The graph in Figure 23 shows that the trend for volume and net energy was the same. Line has been depicted with the colour purple as a result.

Moreover, price variation curve is provided with Figure 24. As it was mentioned earlier, since the only electricity supplier is grid, electricity prices remained constant during the simulated day.

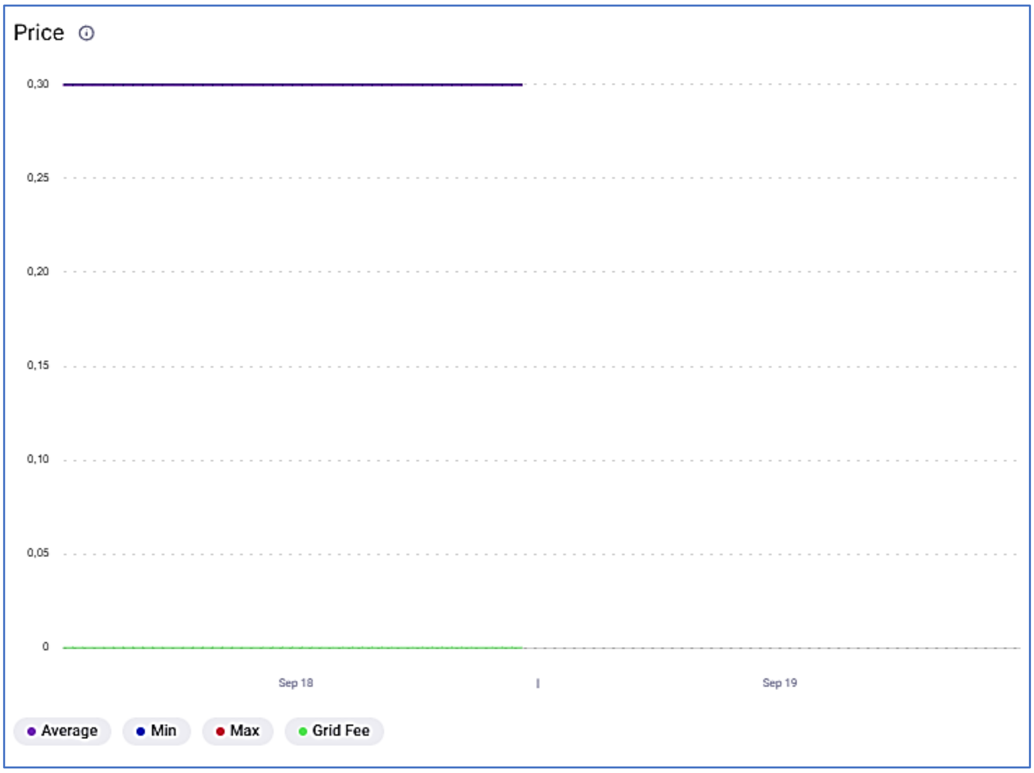


Figure 24. Electricity price variation curve on the simulated day (Scenario 1)

As a result of this scenario, community stayed neutral with 0 € savings.

5.4.2 Results of Scenario 2

As the result of Scenario 2, since there is an asset that produces renewable energy in the community and can sell electricity freely in the local market, the average market price has decreased to 22 cents/kWh. Self-sufficiency of the community reduced reliance on the electricity grid and was obtained as 65.0%. This means that the community can use the clean electricity produced in the community at a rate of 65.0%. The reason why this rate could not reach to 100% is that the renewable energy source in the community is the sun and the production of the solar plant depends on the meteorological values that change during the day. Details of the energy bills and the traded net energy data according to Scenario 2 is provided in Table 9.

Table 9. Energy bills and the traded net energy in the community (Scenario 2)

Asset	Bought		Sold		Total Balance	
	Energy (kWh)	Paid (€)	Energy (kWh)	Revenue (€)	Energy (kWh)	Total (€)
HH1	12.89	3.12	0.00	0.00	12.89	**3.12 (Paid)**
HH2	4.92	1.07	0.00	0.00	4.92	**1.07 (Paid)**
HH3	19.17	4.03	0.00	0.00	19.17	**4.03 (Paid)**
HH4	15.12	2.64	0.00	0.00	15.12	**2.64 (Paid)**
HH5	13.99	3.3	0.00	0.00	13.99	**3.3 (Paid)**
HH6	7.08	1.6	0.00	0.00	7.08	**1.6 (Paid)**
HH7	10.23	1.78	0.00	0.00	10.23	**1.78 (Paid)**
HH8	7.23	1.69	0.00	0.00	1.69	**1.24 (Paid)**
HH9	9.08	1.75	0.00	0.00	9.08	**1.75 (Paid)**
HH10	7.12	1.28	0.00	0.00	7.12	**1.28 (Paid)**
GYM	12.84	2.69	0.00	0.00	12.84	**2.69 (Paid)**
WWTP	11.68	2.45	0.00	0.00	11.68	**2.45 (Paid)**
Bakery	12.48	2.47	0.00	0.00	12.48	**2.47 (Paid)**
School	17.43	2.87	0.00	0.00	17.43	**2.87 (Paid)**
Restaurant	10.75	2.27	0.00	0.00	10.75	**2.27 (Paid)**
Hairdresser	12.75	2.44	0.00	0.00	12.75	**2.44 (Paid)**
Solar Power Plant	0.00	0.00	195474.92	15646.83	195474.92	**15646.83 (Revenue)**
Grid Market	195353.1	15628.25	62.94	18.88	195290.16	**15609.37 (Paid)**
Totals	195537.86	15665.71	195537.86	15665.71	0.00	**0 (Neutral)**

As it can be understood from Table 9, in total 184.76 kWh electricity was demanded by the community members on simulated day, while 121.82 kWh of it is supplied by solar power plant, 62.94 kWh had to be supplied by grid. As a result, overall average market price has been obtained as 22 cents/kWh. Trading profile and the variation of the electricity demand of the community during the day are given by the Figure 25 and Figure 26 separately. Since the trade between SPP and Grid Market has been carried out in high volumes compared to local market players, traded volumes of local market players are not able to be seen in the graph clearly. Therefore, as an example, only the trade details that has been taken place at 12 pm on the simulated day was shown in Figure 25.

Figure 25. Trade Profile of the community (Scenario 2)

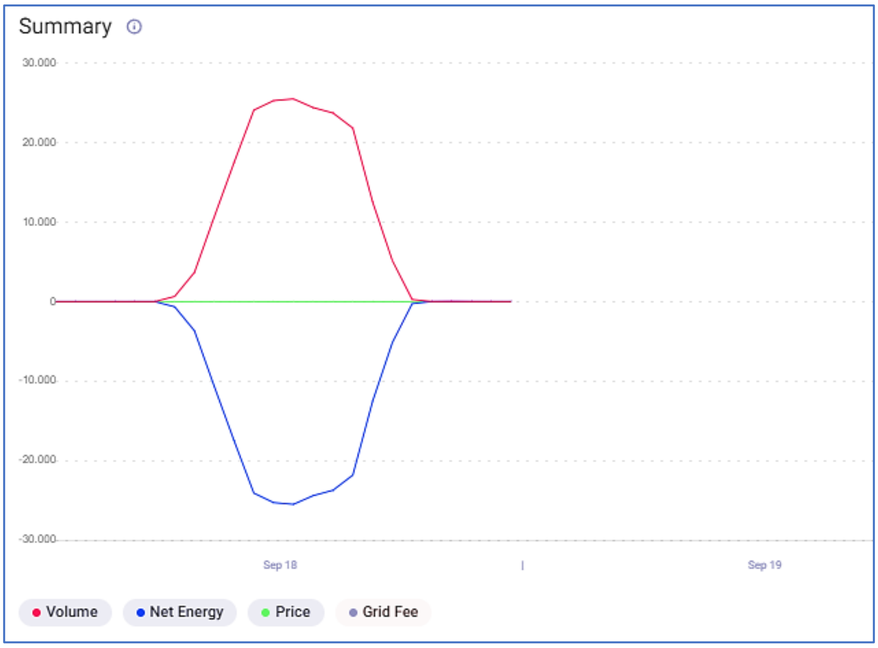


Figure 26. The graph that represents volume traded and net energy (Scenario 2)

With Figure 26 above, total volume of the generated electricity and the net energy profiles are illustrated. Here the net energy represents the excess amount of electricity that is sold to the grid. Moreover, price variation curve is provided with the Figure 27. It is seen that the market price is 30 cents/kWh from 00:00 to 05:00

and from 19:00 to 23:00. As it was mentioned earlier, since there is an asset that produces renewable energy in the community and can sell electricity freely in the local market, electricity consumers did not have to purchase all their demanded electricity from the grid at a constant 30 cents/kWh, consequently the average market price has decreased to 22 cents/kWh.

Figure 27. Electricity price variation curve on the simulated day (Scenario 2)

On the other hand, after the electricity produced in the solar power plant met the 121.82 kWh electricity needs of the community, the excess amount of electricity was sold to the grid as it is illustrated in the Figure 28 below.

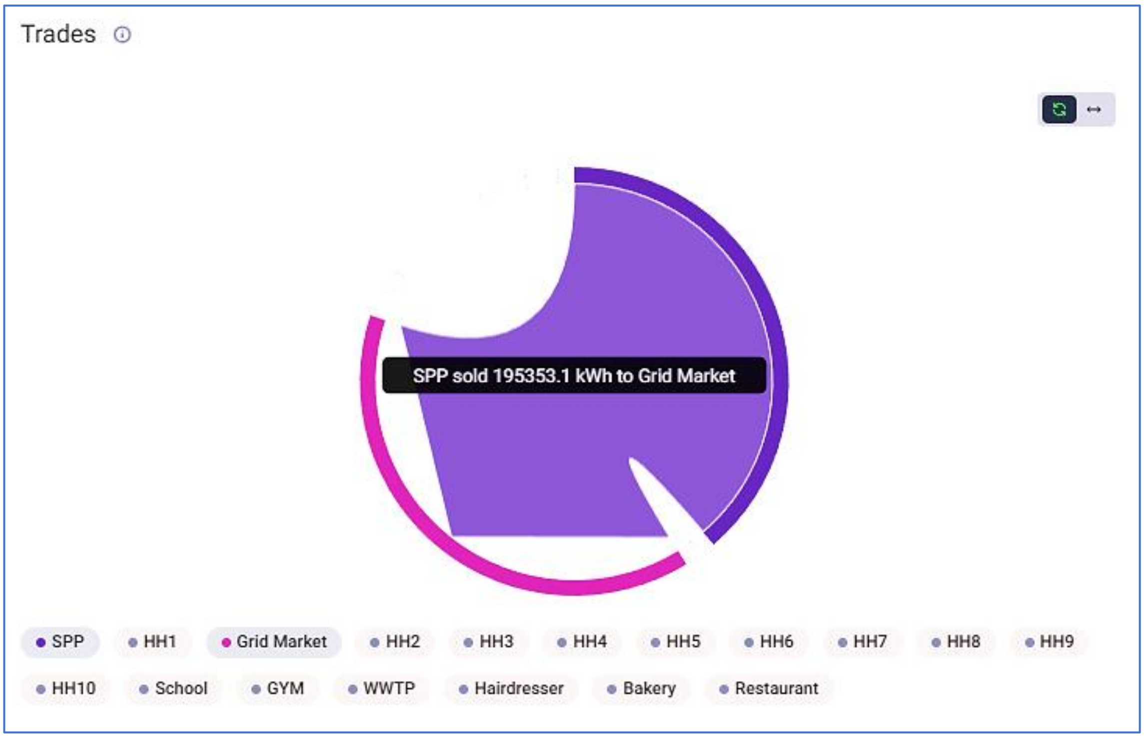


Figure 28. Schematic representation of trading between SPP and Grid Market (Scenario 2)

As a result of this scenario, community had 26.80 € profit thanks to integration of solar power plant into local electricity market.

5.4.3 Results of Scenario 3

As a result of Scenario 3, the self-sufficiency of the community increased to 69.0 %, in addition to the solar power plant, thanks to the 10-kWh capacity battery installed in the community. The battery purchases electricity from the solar power plant and stores it to sell to the local community consumers when renewable energy is not generated. This means that the community can use the clean electricity produced in the community at a rate of 69.0 % without buying electricity from the grid. However, as mentioned earlier, the community's only renewable energy source is the sun, and the production of the solar plant depends on changing meteorological conditions throughout the day, so the self-sufficiency rate could not reach 100%. Additionally, the insufficient installed capacity of the battery was another reason why the rate was not 100%. Table 10 provides details of the energy bills and the traded net energy data according to Scenario 3.

Table 10. Energy bills and the traded net energy in the community (Scenario 3)

Asset	Bought		Sold		Total Balance	
	Energy (kWh)	Paid (€)	Energy (kWh)	Revenue (€)	Energy (kWh)	Total (€)
HH1	12.89	3.06	0.00	0.00	12.89	**3.06 (Paid)**
HH2	4.92	1.03	0.00	0.00	4.92	**1.03 (Paid)**
HH3	19.17	3.99	0.00	0.00	19.17	**3.99 (Paid)**
HH4	15.12	2.61	0.00	0.00	15.12	**2.61 (Paid)**
HH5	13.99	3.28	0.00	0.00	13.99	**3.28 (Paid)**
HH6	7.08	1.57	0.00	0.00	7.08	**1.57 (Paid)**
HH7	10.23	1.78	0.00	0.00	10.23	**1.78 (Paid)**
HH8	7.23	1.68	0.00	0.00	7.23	**1.68 (Paid)**
HH9	9.08	1.75	0.00	0.00	9.08	**1.75 (Paid)**
HH10	7.12	1.27	0.00	0.00	7.12	**1.27 (Paid)**
GYM	12.84	2.64	0.00	0.00	12.84	**2.64 (Paid)**
WWTP	11.68	2.44	0.00	0.00	11.68	**2.44 (Paid)**
Bakery	12.48	2.44	0.00	0.00	12.48	**2.44 (Paid)**
School	17.43	2.86	0.00	0.00	17.43	**2.86 (Paid)**
Restaurant	10.75	2.23	0.00	0.00	10.75	**2.23 (Paid)**
Hairdresser	12.75	2.43	0.00	0.00	12.75	**2.43 (Paid)**
Battery	7.5	0.75	10.00	2.59	2.5	**1.84 (Revenue)**
Solar Power Plant	0.00	0.00	195474.92	15646.98	195474.92	**15646.98 (Revenue)**
Grid Market	195345.6	15627.65	52.94	15.88	195292.66	**15611.77 (Paid)**
Totals	195537.86	15665.46	195537.86	15665.46	0.00	**0 (Neutral)**

As it can be understood from the table, in this scenario with the battery addition, in total 192.26 kWh (184.76 kWh of it from consumers, 7.5 kWh of it from battery) electricity was demanded within the local community on the simulated day. While 121.82 kWh of 184.76 kWh is directly supplied by solar power plant, 10 kWh of it is indirectly supplied by solar power plant thanks to installation of battery. Remaining 52.94 kWh had to be purchased from grid. As a result, overall average market price has slightly changed. Trading profile and the summary of the electricity interaction of the community during the day are given by Figure 29 and Figure 30 below.

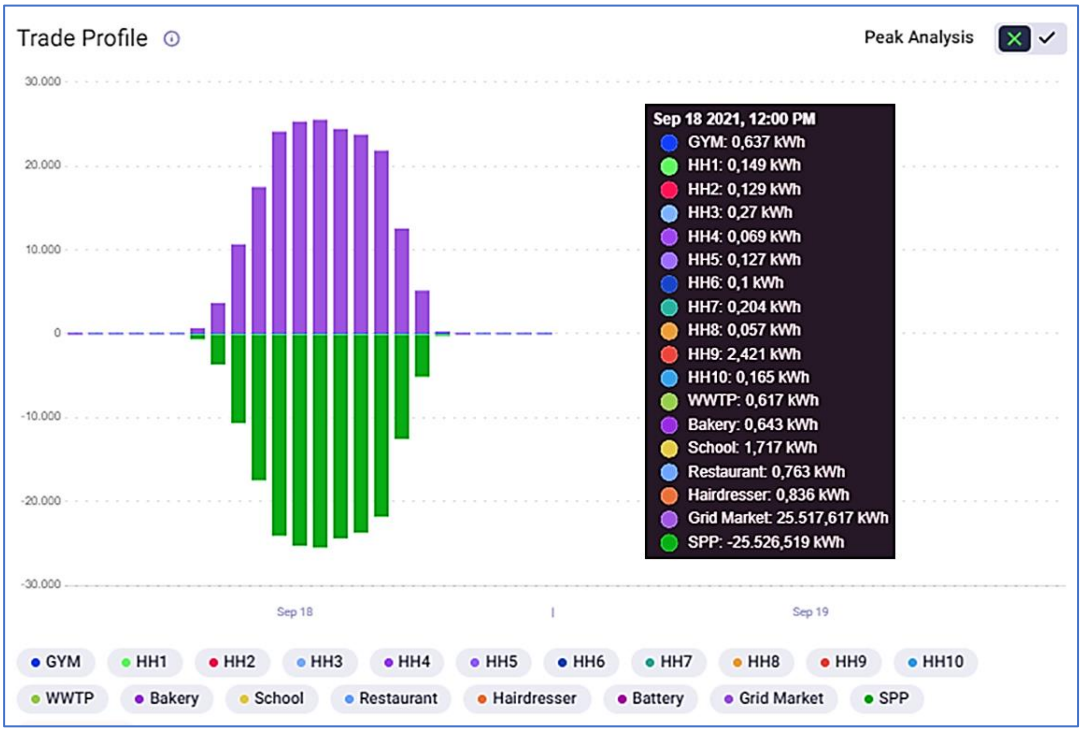


Figure 29.Trade Profile of the community (Scenario 3)

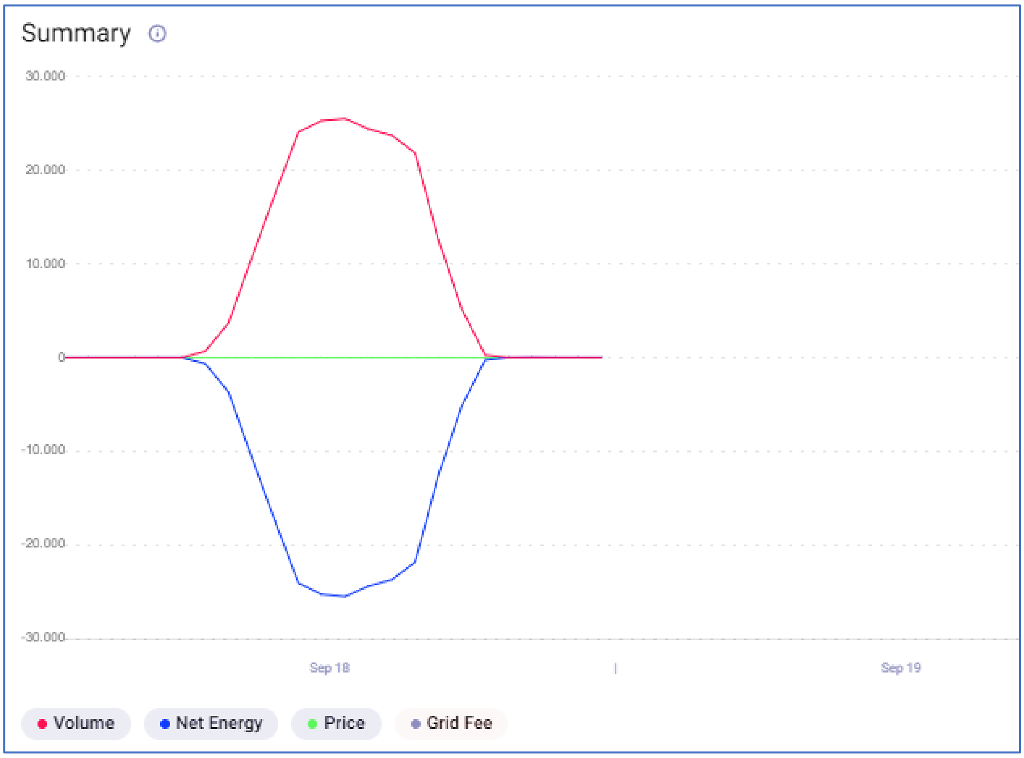


Figure 30. The graph that represents volume traded and net energy (Scenario 3)

With Figure 30 above, total volume of the generated electricity and the net energy profiles are illustrated. Here the net energy represents the excess amount of electricity that is sold to the grid. Moreover, price variation curve is provided with the Figure 31. As it was mentioned earlier, addition of battery with a 10-kWh capacity into local electricity market change the average market price slightly. Therefore, average market price has decreased to 21 cents/kWh.

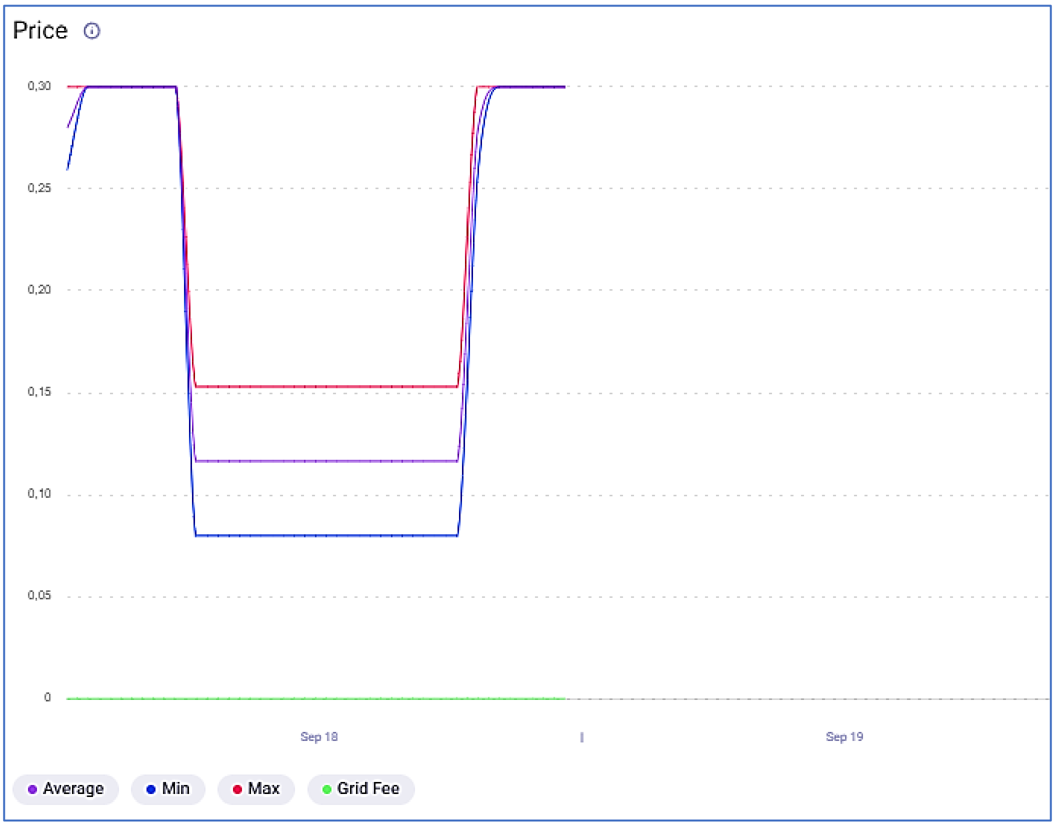


Figure 31. Electricity price variation curve on the simulated day (Scenario 3)

On the other hand, after the electricity produced in the solar power plant met the 121.82 kWh electricity needs of the community, the excess amount of electricity from solar power plant was sold to the grid as it is illustrated in Figure 32 below.

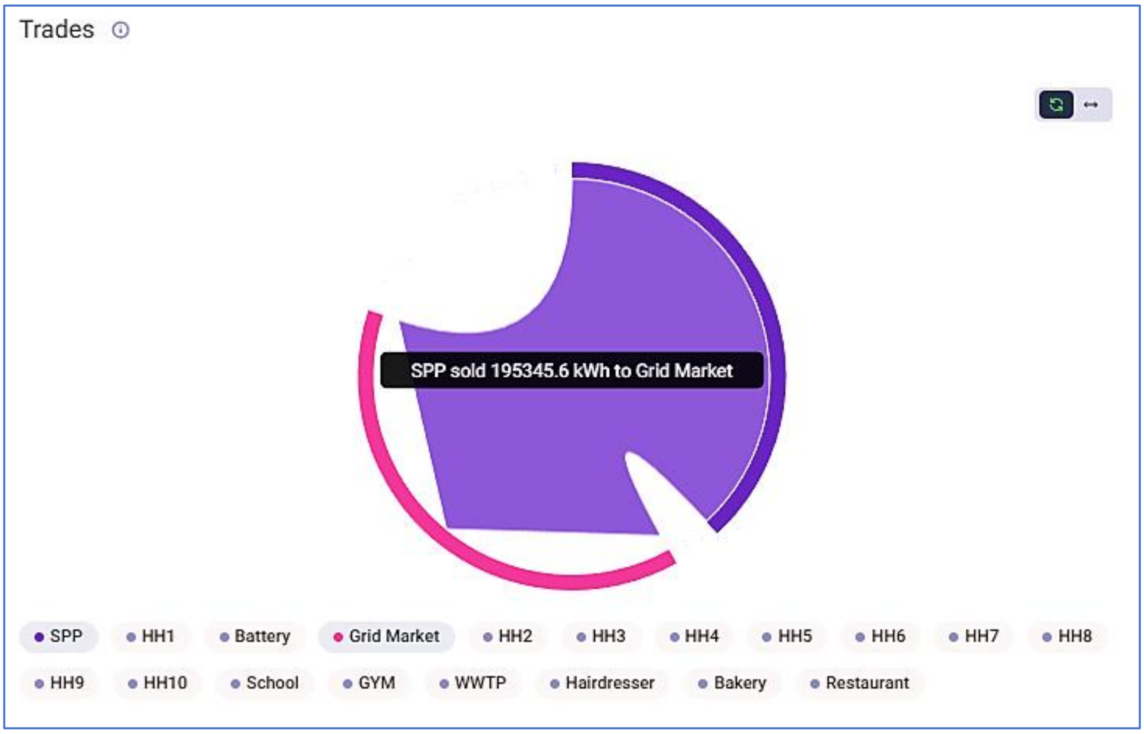


Figure 32. Schematic representation of trading between SPP and Grid Market (Scenario 3)

The representation of the trade that is carried out between solar power plant and battery is illustrated with Figure 33 below. On the other hand, while dark purple arrow represents the electricity sale from solar power plant to battery, light purple arrow represents the electricity sales from battery to local community members.

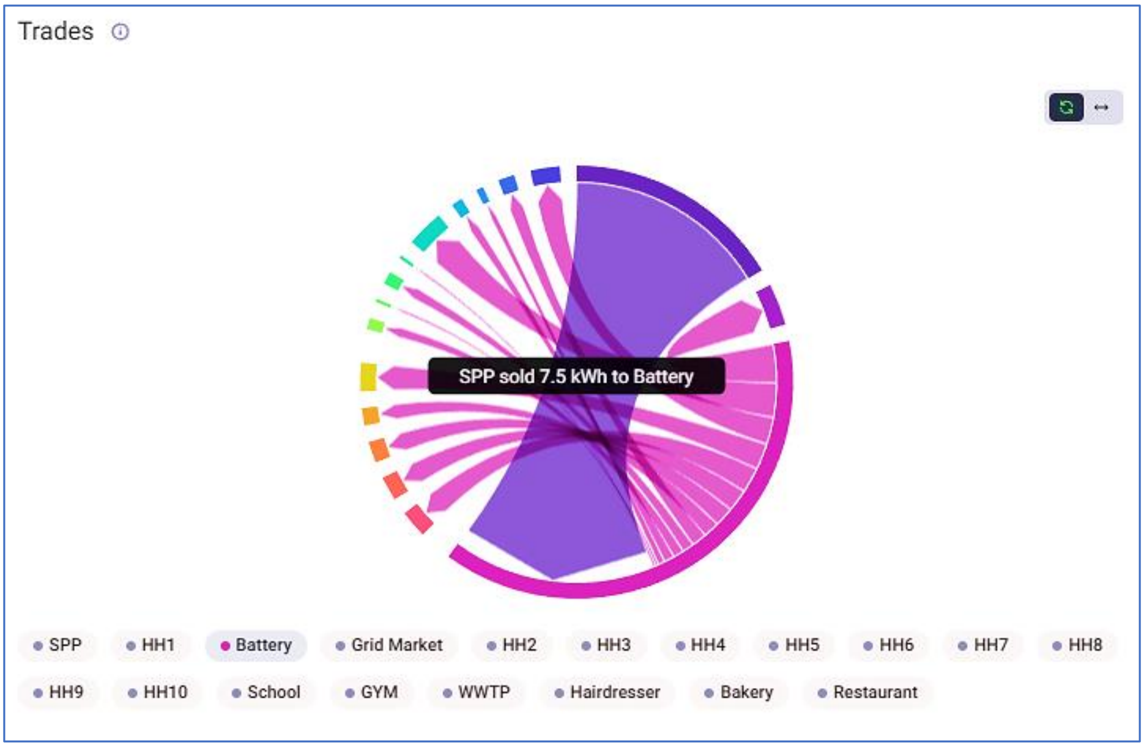


Figure 33. Schematic representation of trading between SPP and Battery (Scenario 3)

As a result of this scenario, even if the average market price of the electricity has not changed significantly, the overall profit of the local community has increased. Therefore, community had 30.65 € profit thanks to implementation of 10 kWh battery besides of solar power plant into local electricity market.

In addition to this, Scenario 3 has been re-executed by taking the battery capacity of 30 kWh so that the effect of the battery on the local market can be observed more clearly. In this situation, solar power plant sold 22.5 kWh electricity to battery as it is illustrated in Figure 34. In this figure again, while dark purple arrow represents the electricity sale from solar power plant to battery, light purple arrow represents the electricity sales from battery to local community members.

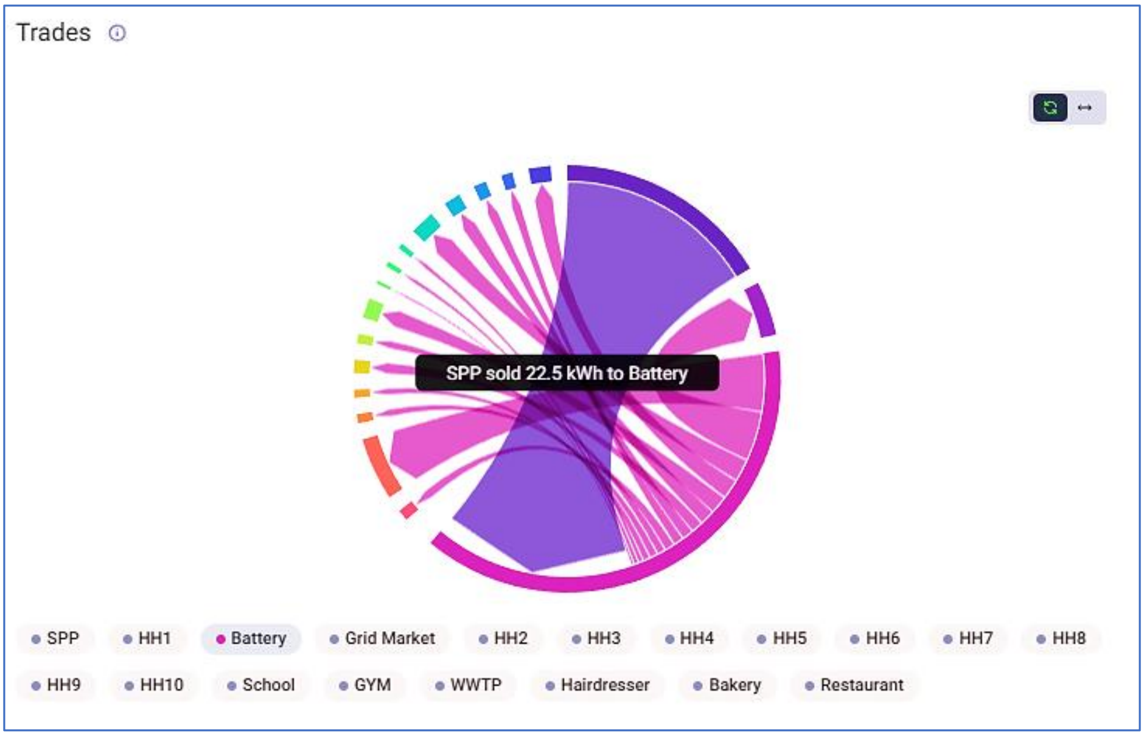


Figure 34. Schematic representation of trading between SPP and Battery with an 30 kWh capacity (Scenario 3)

As a result of this capacity addition on the battery, self-sufficiency of the community has increased from 69.0% to 77.0%. In other words, the community can use the clean electricity produced in the community at a rate of 77.0%. However, average market price has not changed, and it remained at 21 cents/kWh. The overall profit of the local community has increased, and community had 38.35 € profit thanks to increase in the capacity on battery from 10 kWh to 30 kWh.

6 CONCLUSION AND RECOMMENDATIONS

A country's economic progress and degree of social welfare can be significantly influenced by its energy policy. The need for energy skyrocketed following the industrial revolution when new discoveries started to be applied widely in business. Energy is now regarded as one of the most fundamental inputs in the manufacturing process for the realization of economic and social progress. Nevertheless, rising energy consumption also results in rising carbon emissions. Increased greenhouse gas (GHG) emissions are unavoidable as a result of urbanization and the resulting energy demand. In order to achieve the emission reduction, it is required to look at the elements related to energy use that influence changes in GHG emissions. According to the International Energy Agency (IEA), renewable energy will account for the fastest increase in global energy consumption. Because renewable energy is a carbon-free energy source that may be able to help with climate change issues, there has also been an increase in interest in implementing green renewable energy sources in the energy system.

The cost of energy produced from renewable sources may now compete with the traditional method of producing electricity from fossil fuels thanks to advancements in technology. The necessity for sustainable energy sources that will endure as long as the planet spins without relying on any outside source has also been exposed by the decrease in fuel consumed by traditional energy generation sources, economic considerations, and environmental problems. As a result, the integration of renewable energy sources into the traditional energy system is beginning extremely swiftly. The incorporation of Distributed Renewable Energy Sources (RES) into the current energy system is complicated because it is intermittent and fluctuates. This becomes even more difficult as the number of RES grows. To deal with this problem, Local Electricity Markets (LEMs) offer a solution. LEMs allow individuals and producers to engage in the trading of locally generated electricity within their Local Energy Communities. Local Energy Markets assist in balancing local energy demand and supply, as well as reducing dependency on the wholesale market and the need for lengthy electrical transmissions to or from the grid. The anticipation that higher penetration of renewable energy will lower electricity prices, known as the merit order effect, is another advantageous outcome of the integration of renewable energy into the system.

Keeping the cost of power generation to a minimum requires a balanced approach to managing supply and demand. All market players share the responsibility for maintaining this balance. This involves predicting the demand at any given time and planning the power plants that will meet that demand in real-time. To do this accurately, production planning and demand forecasting are conducted one day before electricity distribution.

This thesis explores the use of Big Data Analytics to predict the short-term electricity generation of a solar power plant with a large installed capacity. It then evaluates how this data could affect the prices of the local day-ahead electricity market using simulation software. Specifically, the study investigates the impact of a high installed capacity solar power plant on local energy market prices. It also examines how anticipating the value of power generation can affect other factors in the local energy market, such as self-sufficiency, profitability, and price formation.

Initially, a solar power plant with a high capacity was selected, and its day-ahead energy output was predicted by analysing retroactive electricity generation data for the plant and meteorological data for the area where it is located. Due to the large volume and variety of data, Big Data Analytics was employed, and Python programming was utilized to execute the machine learning technique. Out of the three models tested, the Light GBM model provided the most accurate forecast for the electricity production on the specific day (18.09.2021) for which the forecast was generated.

The main objective here is to draw attention to qualitative changes by observing the impact of the integration of a high-capacity renewable energy system into a local electricity market on market prices. By doing this, it is aimed to highlight the significance of forecasting the market price for the following day from the viewpoint of a market maker and subsequently, utilizing this information to take a position in the market. In the local electricity market simulation that was run in this study, the community's electricity demand was assumed to be constant. As the renewable energy facility that penetrated the system has a high capacity, it will be able to meet the community's electricity needs under all weather conditions. To ensure this, forecasting analysis has been repeated on a monthly basis to evaluate all months of the year for observing the effect of seasonal changes. As a result, it has provided an additional explanatory power to the approach that the electricity to be produced will be sufficient for the demand in the community under all circumstances. Therefore, since the price elasticity of demand was zero and the market was balanced only by demand, the simulation was run for only one specific day. Subsequently, the values obtained from the generation forecasting for September 18th, 2021 were incorporated into the local electricity market simulation model. Grid Singularity, an open-source and online software, was used to validate the simulated scalable scenarios and evaluate LEMs from an economic perspective. The first step involved using Grid Singularity to identify a community and then adding local market players to this group. The daily electricity load characteristics of these new local market participants were gathered from the literature. Next, three separate scenarios were developed to thoroughly examine price formation, profitability, and community self-sufficiency.

In the first scenario, the community did not have a solar power plant, and thus all electricity was obtained from the grid, resulting in no change in market pricing and a self-sufficiency rate of 0%. In the second scenario, a high-capacity solar power plant was added to the system, resulting in a 26.7% reduction in the average power market price according to simulation results. The community's self-sufficiency was found to be 65.0%, indicating that the community could use 65.0% of the locally generated clean electricity. However, this rate could not reach 100% due to the fact that the community relied solely on the sun as a renewable energy source, and the solar power plant's output was dependent on changing meteorological conditions throughout the day. Lastly, in contrast to the second scenario, the third scenario involved the addition of a 10-kWh battery to the system, which allowed local consumers to purchase electricity from it when solar energy was unavailable. This led to an increase in the community's self-sufficiency. The inclusion of the 10-kWh battery improved the self-sufficiency rate from 65.0% in Scenario 2 to 69.0% in Scenario 3. Another simulation was carried out with a battery capacity of 30 kWh, which further increased the community's level of independence from 69.0% to 77.0%. Additionally, this scenario demonstrated a 30% reduction in the average market price compared to Scenario 1.

The intermittent feature stands out as the most important obstacle to the widespread use of renewable energy. Lithium-ion batteries will be able to provide better and more widespread use of solar energy in terms of their use in energy storage systems.[97] With the result of third scenario, this statement has been confirmed. However, even in scenarios where batteries are not used, the penetration of a high-capacity solar power plant into the local electricity market has increased self-sufficiency from 0% to 65% and led to an 8 cent/kWh decrease in prices. On the other hand, in scenarios where the batteries with different capacities are included, self-sufficiency has reached up to 77%, but only a 1 cent/kWh decrease in market prices has been observed compared to the scenario without the battery but with solar energy penetration. Thus, it has been observed that the use of batteries indirectly increases renewable energy use and self-sufficiency within the community, while it does not create major changes in market prices.

On the other hand, the pre-establishment of electricity prices in this market, prior to the day of supply, showed that pricing could be determined beforehand by considering the bids and offers from market players. Moreover, by knowing the amount of electricity that will be generated one day ahead, market participants take positions one day in advance, thereby reducing the supply risk posed by the intermittent nature of renewable energy.

As a result, while Halttunen et al.'s findings showed that the effect of increased adoption of renewable energy sources in the electricity market has less impact on the merit-order effect, this study emphasizes that incorporating solar power plants with high capacity into local electricity markets can lead to reduced market prices, improved self- sufficiency, and greater profitability for the community.

Countries can achieve more cost-effective renewable energy prices for both producers and consumers by developing local electricity markets instead of relying solely on a national grid and increasing the renewable energy share in these markets. Furthermore, creating sustainable communities in this way can help countries meet their obligations under the Paris Agreement to combat climate change and promote a cleaner world. Governments can promote the use of renewable energy sources in these areas by implementing incentive programs.

In addition, the installation of renewable energy power plants in these communities can be certified under a voluntary carbon standard or renewable energy certification program, which will allow for the sale of certificates or carbon credits to generate additional revenue and boost the community's profitability rate. Moreover, these certificates could be sold to communities where renewable energy integration is not possible, allowing them to offset their scope-2 emissions by purchasing these environmental commodities, and ultimately helping them become carbon neutral communities.

In conclusion, in this thesis, simulations were run for only one day to assess the effect of a high solar energy capacity on local electricity market prices. This was done since the overall local market's electricity consumption was assumed to be completely constant. The power plant's capacity allows it to meet market electricity demand even at minimum generation levels. In other words, the outcomes from choosing a different day would not create difference as the demand would not change regardless of the electricity generation. As a result, attention has been drawn to qualitative changes rather than quantitative changes in the market. As a recommendation, to obtain quantitative outcomes on the market price, this study could be expanded with other studies also by considering demand changes varying seasonal conditions. Furthermore, it is advised to carry out optimization evaluations to ascertain the appropriate capacity for integrating battery storage systems and renewable energy plants for maximal self-sufficiency before developing sustainable communities. This can facilitate the creation of efficient carbon-neutral communities.

7 References

[1] Sanem Yagmur Kement, “Designing a local electricity market using blockchain technology,” University of Freiburg, Freiburg, 2020.

[2] X. Zhang, M. Cheng, Y. Liu, D. H. Li, and R. M. Wu, “Short-Term Load Forecasting Based on Big Data Technologies,” *Applied Mechanics and Materials*, vol. 687–691, pp. 1186–1192, Nov. 2014, doi: 10.4028/www.scientific.net/AMM.687-691.1186.

[3] Y. Riahi and S. Riahi, “Big Data and Big Data Analytics: concepts, types and technologies,” *International Journal of Research and Engineering*, vol. 5, no. 9, pp. 524–528, Nov. 2018, doi: 10.21276/ijre.2018.5.9.5.

[4] B. B. Ekici, “A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems,” *Measurement*, vol. 50, pp. 255–262, Apr. 2014, doi: 10.1016/j.measurement.2014.01.010.

[5] S. Sobri, S. Koohi-Kamali, and N. Abd. Rahim, “Solar photovoltaic generation forecasting methods: A review,” *Energy Convers Manag*, vol. 156, pp. 459–497, Jan. 2018, doi: 10.1016/j.enconman.2017.11.019.

[6] Í. Munné-Collado, E. Bullich-Massagué, M. Aragüés-Peñalba, and P. Olivella-Rosell, *Local and Micro Power Markets*. UK: John Wiley & Sons, Ltd Chichester, 2019.

[7] A. Gianfreda, F. Ravazzolo, and L. Rossini, “Comparing the forecasting performances of linear models for electricity prices with high RES penetration,” *Int J Forecast*, vol. 36, no. 3, pp. 974–986, Jul. 2020, doi: 10.1016/j.ijforecast.2019.11.002.

[8] N. Mostafa, H. S. M. Ramadan, and O. Elfarouk, “Renewable energy management in smart grids by using big data analytics and machine learning,” *Machine Learning with Applications*, vol. 9, p. 100363, Sep. 2022, doi: 10.1016/j.mlwa.2022.100363.

[9] H. Çetintaş, İ. Murat, and ∗∗∗ B., “Elektrik Piyasalarında Yeniden Yapılanma ve Türkiye Elektrik Piyasasında Yapısal Dönüşüm 1,” 2015. [Online]. Available: http://dergipark.ulakbim.gov.tr/usakoeyb/

[10] E. Tüketiminin *et al.*, “Türkiye’de Elektrik Piyasasının Rekabete Açılması ve Rekabetçi Elektrik Piyasasında The Opening of Electricity Market to Competition in Turkey and Determinants of Electricity Consumption in Competitive Electricity Market: ARDL Bounds Test,” *JOURNAL OF EMERGING ECONOMIES AND POLICY 2021*, vol. 6, no. 1, pp. 148–164, 2021, [Online]. Available: http://dergipark.org.tr/joeep

[11] S. Hunt, “Making Competition Work in electricity.”

[12] A. Galetovic and C. M. Muñoz, “Regulated electricity retailing in Chile,” *Energy Policy*, vol. 39, no. 10, pp. 6453–6465, 2011, doi: 10.1016/j.enpol.2011.07.047.

[13] Q. Zhang and I. E. Grossmann, “Enterprise-wide optimization for industrial demand side management: Fundamentals, advances, and perspectives,” *Chemical Engineering Research and Design*, vol. 116, pp. 114–131, Dec. 2016, doi: 10.1016/j.cherd.2016.10.006.

[14] Hacı Yusuf Çınar, *Elektrik Piyasasında Dağıtım Faaliyetine İlişkin Lisanslama Rejimi ve Denetim*. Astana Yayınları, 2020.

[15] T. Danışmanı, S. Onaygil, E. Kayahan, E. Bilim, T. A. Dalı, and T. Programı, “İSTANBUL TEKNİK ÜNİVERSİTESİ ↔ ENERJİ ENSTİTÜSÜ YÜKSEK LİSANS TEZİ ELEKTRİK ENERJİSİ PİYASALARI ve ÇİMENTO SEKTÖRÜNDE ELEKTRİK ENERJİSİ TÜKETİM TAHMİNİNİN ÖNEMİ,” 2019.

[16] M. , Y. H. , L. Z. Shahidehpour, *Market Operations in Electric Power Systems: forecasting, scheduling, and risk management*. 2002.

[17] E. Onaiwu, “HOW DOES BILATERAL TRADING DIFFER FROM ELECTRICITY POOLING?”

[18] L. Augusto Barroso PSR, M. de Energia, and T. H. Cavalcanti, “CLASSIFICATION OF ELECTRICITY MARKET MODELS WORLDWIDE On behalf of CIGRE Task Force C5.2.1.”

[19] W. Antweiler and F. Muesgens, “On the long-term merit order effect of renewable energies,” *Energy Econ*, vol. 99, Jul. 2021, doi: 10.1016/j.eneco.2021.105275.

[20] H. , and S. S. T. Klinge Jacobsen, “Curtailment of renewable generation: Economic optimality and incentives,” in *Energy Policy*, 2012.

[21] F. , R. M. , and G. M. Sensfuß, “The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany.,” in *Energy Policy*, 2008.

[22] J. López Prol, K. W. , Steininger, and D. Zilberman, “The cannibalization effect of wind and solar in the California wholesale electricity market.,” in *Energy Econ.*, 2020.

[23] K. Halttunen, I. Staffell, R. Slade, R. Green, Y.-M. Saint-Drenan, and M. Jansen, “Global assessment of the merit-order effect and revenue cannibalisation for variable renewable energy.”

[24] J. C. Ketterer, “The impact of wind power generation on the electricity price in Germany,” in *Energy Econ.*, 2014, pp. 270–280.

[25] M. Paschen, “Dynamic analysis of the German day-ahead electricity spot market,” in *Energy Econ.*, 2016, pp. 118–128.

[26] A. Zipp, “The marketability of variable renewable energy in liberalized electricity markets – An empirical analysis,” in *Renew. Energy*, 2017, pp. 1111–1121.

[27] F. Benhmad and J. Percebois, “Photovoltaic and wind power feed-in impact on electricity prices: The case of Germany.,” in *Energy Policy*, 2018, pp. 317–326.

[28] C. Martin de Lagarde and F. Lantz, “How renewable production depresses electricity prices: Evidence from the German market.,” in *Energy Policy*, 2018, pp. 263–277.

[29] W. , C. Z. Hu and B. Bak-Jensen, “The Relationship Between Electricity Price and Wind Power Generation in Danish Electricity Markets.,” in *In 2010 Asia-Pacific Power and Energy Engineering Conference*, 2010.

[30] L. Gelabert, P. Linares, and X. Labandeira, “An ex-post analysis of the effect of renewables and cogeneration on Spanish electricity prices,” in *Energy Econ.*, 2011, pp. S59–S65.

[31] R. Prata, P. M. S. Carvalho, and I. L. Azevedo, “ Distributional costs of wind energy production in Portugal under the liberalized Iberian market regime.,” in *Energy Policy*, 2018, pp. 500–512.

[32] V. di Cosmo and L. Malaguzzi Valeri, “ Wind, storage, interconnection and the cost of electricity generation,” in *Energy Econ.*, 2018, pp. 1–18.

[33] S. Clò, A. Cataldi, and P. Zoppoli, “The merit-order effect in the Italian power market: The impact of solar and wind generation on national wholesale electricity prices,” in *Energy Policy*, 2015, pp. 79–88.

[34] I. Staffell, “Measuring the progress and impacts of decarbonising British electricity.,” in *Energy Policy*, 2017, pp. 463–475.

[35] V. Bobinaite and I. Priedite, *Assessment of Impacts of Wind Electricity Generation Sector Development: Latvian Case. Procedia - Soc. Behav. Sci.*, vol. 213. 2015.

[36] S. Forrest and I. MacGill, “Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian National Electricity Market.,” in *Energy Policy*, 2013, pp. 120–132.

[37] J. MacCormack, A. Hollis, H. Zareipour, and W. Rosehart, “The large-scale integration of wind generation: Impacts on price, reliability and dispatchable conventional suppliers.,” in *Energy Policy*, 2010, pp. 18–28.

[38] R. K. Kaufmann and D. Vaid, “ Lower electricity prices and greenhouse gas emissions due to rooftop solar: Empirical results for Massachusetts.,” in *Energy Policy*, 2016, pp. 345–352.

[39] C. K. Woo, I. Horowitz, J. Moore, and A. Pacheco, “The impact of wind generation on the electricity spot-market price level and variance: The Texas experience,” in *Energy Policy*, 2011, pp. 3939–3944.

[40] C. Weinhardt, E. Mengelkamp, and J. Diesing, “Tracing Local Energy Markets: A Literature Review,” 2019, doi: 10.13140/RG.2.2.17644.21128.

[41] European Commission, “EU Directive, “DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on common rules for the internal market in electricity (recast),” 2016. https://eur-lex.europa.eu/resource.html?uri=cellar:c7e47f46-faa4-11e6-8a35-01aa75ed71a1.0014.02/DOC_1&format=PDF (accessed Nov. 21, 2022).

[42] J. S. PankiRaj, A. Yassine, and S. Choudhury, “An auction mechanism for profit maximization of peer-to-peer energy trading in smart grids,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 361–368. doi: 10.1016/j.procs.2019.04.050.

[43] B. Biegel, L. H. Hansen, J. Stoustrup, P. Andersen, and S. Harbo, “Value of flexible consumption in the electricity markets,” *Energy*, vol. 66, pp. 354–362, Mar. 2014, doi: 10.1016/j.energy.2013.12.041.

[44] M. Maldet *et al.*, “Trends in local electricity market design: Regulatory barriers and the role of grid tariffs,” *J Clean Prod*, vol. 358, Jul. 2022, doi: 10.1016/j.jclepro.2022.131805.

[45] E. Mengelkamp and C. Weinhardt, “Decentralizing Energy Systems Through Local Energy Markets: The LAMP-Project,” 2018. [Online]. Available: https://www.sonnen-batterie.com/de/home

[46] W. H. Timmerman, G. B. Huitema, H. G. Sol, and Ipskamp, *Facilitating the Growth of Local Energy Communities*.

[47] G. Mendes, J. Nylund, S. Annala, S. Honkapuro, O. Kilkki, and J. Segerstam, “LOCAL ENERGY MARKETS: OPPORTUNITIES, BENEFITS, AND BARRIERS,” 2018.

[48] E. Aktan, B. Yönetimi, H. Makaleler, and M. Bilgisi, “Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu Big Data: Application Areas, Analytics and Security Dimension Article Info.” [Online]. Available: http://dergipark.gov.tr/by

[49] SAS, “Big Data Analytics - What it is and why it matters.”

[50] Surya Gutta, “Data Science: The 5 V’s of Big Data,” *Analytics Vidhya*, May 04, 2020. https://medium.com/analytics-vidhya/the-5-vs-of-big-data-2758bfcc51d (accessed Mar. 23, 2023).

[51] M. Akin and S. Sagiroglu, “Büyük Veri Perspektifinden Trafik Tahmini Traffic Prediction Based on Big Data Perspective,” in *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, IEEE, Nov. 2019, pp. 1–6. doi: 10.1109/UBMYK48245.2019.8965545.

[52] C. Watts, “Managing the Risk in Renewable Energy. Department of Agricultural & Resource ,” 2011.

[53] A. Ali, J. Qadir, R. ur Rasool, A. Sathiaseelan, A. Zwitter, and J. Crowcroft, “Big data for development: applications and techniques,” *Big Data Anal*, vol. 1, no. 1, Dec. 2016, doi: 10.1186/s41044-016-0002-4.

[54] S. L. NITA, L. DUMITRU, and A. BETERINGHE, “MACHINE LEARNING TECHNIQUES USED IN BIG DATA,” *Scientific Bulletin of Naval Academy*, vol. 19, no. 1, pp. 466–471, Jun. 2016, doi: 10.21279/1454-864x-16-i1-078.

[55] S. L. Poczter and L. M. Jankovic, “The Google Car: Driving Toward A Better Future?,” *Journal of Business Case Studies (JBCS)*, vol. 10, no. 1, p. 7, Dec. 2013, doi: 10.19030/jbcs.v10i1.8324.

[56] S. Sen and S. Ganguly, “Opportunities, barriers and issues with renewable energy development – A discussion,” *Renewable and Sustainable Energy Reviews*, vol. 69. Elsevier Ltd, pp. 1170–1181, 2017. doi: 10.1016/j.rser.2016.09.137.

[57] I. Jebli, F. Z. Belouadha, M. I. Kabbaj, and A. Tilioua, “Prediction of solar energy guided by pearson correlation using machine learning,” *Energy*, vol. 224, Jun. 2021, doi: 10.1016/j.energy.2021.120109.

[58] U. K. Das *et al.*, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81. Elsevier Ltd, pp. 912–928, Jan. 01, 2018. doi: 10.1016/j.rser.2017.08.017.

[59] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Solar Energy*, vol. 136. Elsevier Ltd, pp. 78–111, Oct. 15, 2016. doi: 10.1016/j.solener.2016.06.069.

[60] D. Liu and K. Sun, “Random forest solar power forecast based on classification optimization,” *Energy*, vol. 187, Nov. 2019, doi: 10.1016/j.energy.2019.115940.

[61] F. Li, R. Li, Z. Zhang, M. Dale, D. Tolley, and P. Ahokangas, “Big Data Analytics for Flexible Energy Sharing: Accelerating a Low-Carbon Future,” *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 35–42, May 2018, doi: 10.1109/MPE.2018.2796358.

[62] Enerji Atlası, “Güneş Enerji Santralleri .” https://www.enerjiatlasi.com/gunes/cingilli-ges.html (accessed Mar. 23, 2023).

[63] “GESDERGISI- BUYUK OLCEKLI GUNES PROJELERI,” *MONETA Tanıtım*, Nov. 04, 2019. https://www.gesdergisi.com/cingilli-ges-uretime-basladi/ (accessed Mar. 23, 2023).

[64] Enerji Verimliliği ve Çevre Dairesi Başkanlığı, “TÜRKİYE ULUSAL ELEKTRİK ŞEBEKESİ EMİSYON FAKTÖRÜ BİLGİ FORMU,” Sep. 2022.

[65] “Niğde(il) Wikipedia,” *Niğde (il)*. https://tr.wikipedia.org/wiki/Ni%C4%9Fde_%28il%29 (accessed Mar. 23, 2023).

[66] Climate Data, “İklim Niğde.” https://tr.climate-data.org/asya/tuerkiye/nigde/nigde-254/ (accessed Nov. 25, 2022).

[67] “Güneş Enerjisi Potansiyelimiz,” Mar. 01, 2023. https://enerji.gov.tr/bilgi-merkezi-enerji-gunes (accessed Mar. 15, 2023).

[68] J. F. Torres, A. Troncoso, I. Koprinska, Z. Wang, and F. Martínez-Álvarez, “Big data solar power forecasting based on deep learning and multiple data sources,” *Expert Syst*, vol. 36, no. 4, Aug. 2019, doi: 10.1111/exsy.12394.

[69] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, “Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators,” *IEEE Trans Smart Grid*, vol. 6, no. 5, pp. 2253–2262, Sep. 2015, doi: 10.1109/TSG.2015.2397003.

[70] F. J. L. Lima, F. R. Martins, E. B. Pereira, E. Lorenz, and D. Heinemann, “Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks,” *Renew Energy*, vol. 87, pp. 807–818, Mar. 2016, doi: 10.1016/j.renene.2015.11.005.

[71] H. Wang *et al.*, “Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network,” *Energy Convers Manag*, vol. 153, pp. 409–422, Dec. 2017, doi: 10.1016/j.enconman.2017.10.008.

[72] M. Hossain, S. Mekhilef, M. Danesh, L. Olatomiwa, and S. Shamshirband, “Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems,” *J Clean Prod*, vol. 167, pp. 395–405, Nov. 2017, doi: 10.1016/j.jclepro.2017.08.081.

[73] K. Mohammadi, S. Shamshirband, C. W. Tong, M. Arif, D. Petković, and S. Ch, “A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation,” *Energy Convers Manag*, vol. 92, pp. 162–171, Mar. 2015, doi: 10.1016/j.enconman.2014.12.050.

[74] “Solcast Data API.” https://docs.solcast.com.au/ (accessed Nov. 26, 2022).

[75] Ryan Holbrook and Alexis Cook, “Time Series as Features,” *What is Serial Dependence?* https://www.kaggle.com/code/ryanholbrook/time-series-as-features (accessed Nov. 27, 2022).

[76] C. Cornaro, M. Pierro, V. Puggioni, and D. Roncarati, “Outdoor Characterization of Phase Change Materials and Assessment of Their Energy Saving Potential to Reach NZEB,” *Buildings*, vol. 7, no. 4, p. 55, Jun. 2017, doi: 10.3390/buildings7030055.

[77] Zach, “What is Considered a Good RMSE Value?,” May 10, 2021. https://www.statology.org/what-is-a-good-rmse/ (accessed Dec. 18, 2022).

[78] Will Koehrsen, “Hyperparameter Tuning the Random Forest in Python,” *Towards Data Science*, Jan. 10, 2018. https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74 (accessed Nov. 27, 2022).

[79] Mirac Özturk, “Python ile Sınıflandırma Analizleri – Rastgele Orman (Random Forest) Algoritması,” *Python, Veri Bilimi*, Apr. 13, 2022. https://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi/ (accessed Nov. 27, 2022).

[80] G. Ke *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree.” [Online]. Available: https://github.com/Microsoft/LightGBM.

[81] E. G. Kardakos, M. C. Alexiadis, S. I. Vagropoulos, C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis, “Application of time series and artificial neural network models in short-term forecasting of PV power generation,” in *2013 48th International Universities’ Power Engineering Conference (UPEC)*, IEEE, Sep. 2013, pp. 1–6. doi: 10.1109/UPEC.2013.6714975.

[82] A. Mellit, S. Sağlam, and S. A. Kalogirou, “Artificial neural network-based model for estimating the produced power of a photovoltaic module,” *Renew Energy*, vol. 60, pp. 71–78, Dec. 2013, doi: 10.1016/j.renene.2013.04.011.

[83] F. Almonacid, P. J. Pérez-Higueras, E. F. Fernández, and L. Hontoria, “A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator,” *Energy Convers Manag*, vol. 85, pp. 389–398, Sep. 2014, doi: 10.1016/j.enconman.2014.05.090.

[84] M. G. De Giorgi, P. M. Congedo, and M. Malvoni, “Photovoltaic power forecasting using statistical methods: impact of weather data,” *IET Science, Measurement & Technology*, vol. 8, no. 3, pp. 90–97, May 2014, doi: 10.1049/iet-smt.2013.0135.

[85] T. T. Teo, T. Logenthiran, and W. L. Woo, “Forecasting of photovoltaic power using extreme learning machine,” in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, IEEE, Nov. 2015, pp. 1–6. doi: 10.1109/ISGT-Asia.2015.7387113.

[86] G. Graditi, S. Ferlito, and G. Adinolfi, “Comparison of Photovoltaic plant power production prediction methods using a large measured dataset,” *Renew Energy*, vol. 90, pp. 513–519, May 2016, doi: 10.1016/j.renene.2016.01.027.

[87] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines," *IEEE Trans Ind Appl*, vol. 48, no. 3, pp. 1064–1069, May 2012, doi: 10.1109/TIA.2012.2190816.

[88] G. Wang, Y. Su, and L. Shu, "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renew Energy*, vol. 96, pp. 469–478, Oct. 2016, doi: 10.1016/j.renene.2016.04.089.

[89] Y. Li, Y. He, Y. Su, and L. Shu, "Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines," *Appl Energy*, vol. 180, pp. 392–401, Oct. 2016, doi: 10.1016/j.apenergy.2016.07.052.

[90] L. Massidda and M. Marrocu, "Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany," *Solar Energy*, vol. 146, pp. 141–149, Apr. 2017, doi: 10.1016/j.solener.2017.02.007.

[91] L. A. Fernandez-Jimenez *et al.*, "Short-term power forecasting system for photovoltaic plants," *Renew Energy*, vol. 44, pp. 311–317, Aug. 2012, doi: 10.1016/j.renene.2012.01.108.

[92] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output," *IEEE Trans Sustain Energy*, vol. 5, no. 3, pp. 917–926, Jul. 2014, doi: 10.1109/TSTE.2014.2313600.

[93] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "A physical hybrid artificial neural network for short term forecasting of PV plant power output," *Energies (Basel)*, vol. 8, no. 2, pp. 1138–1153, 2015, doi: 10.3390/en8021138.

[94] J. Antonanzas, R. Urraca, A. Pernía-Espinoza, A. Aldama, L. A. Fernández-Jiménez, and F. J. Martínez-de-Pisón, "Single and Blended Models for Day-Ahead Photovoltaic Power Forecasting," 2017, pp. 427–434. doi: 10.1007/978-3-319-59650-1_36.

[95] "Grid Singularity." Accessed: Nov. 27, 2022. [Online]. Available: https://gridsingularity.com/singularity-map

[96] "Standardlastprofile Strom," *bdew*, Jan. 01, 2017. https://www.bdew.de/energie/standardlastprofile-strom/ (accessed Nov. 29, 2022).

[97] B. Avcı *et al.*, "BATARYA VE YENİLİKÇİ ENERJİ TEKNOLOJİLERİ ÇALIŞMA GRUBU RAPORU," 2020. Accessed: Dec. 21, 2022. [Online]. Available: https://tusside.tubitak.gov.tr/sites/images/Tusside/batarya_ve_yenilikci_enerji_teknolojileri_calisma_grubu_raporu.pdf

APPENDICES

APPENDIX 1. THE PLOT OF THE RESULT OF CORRELATION ANALYSIS